



ISSN(Print):2328-3491, ISSN(Online): 2328-3580, ISSN(CD-ROM): 2328-3629

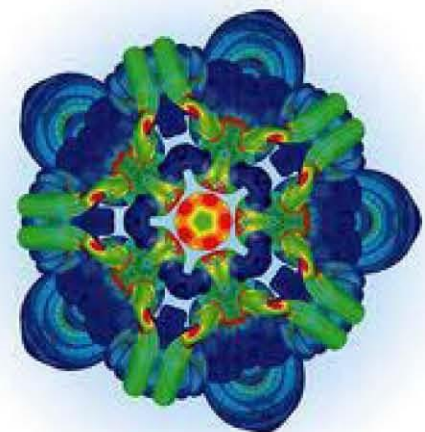


National Conference on Information Technology,
Electronics and Management (NCITEM-2018),
July 26-27, 2018

Organised by:
Centre for Development of Advanced Computing (CDAC),
Noida, Uttar Pradesh, INDIA

**A
I
J
R
S
T
E
M**

American International Journal of Research in Science, Technology, Engineering & Mathematics



International Association of Scientific Innovation and Research (IASIR)

(An Association Unifying the Sciences, Engineering, and Applied Research)

STEM International Scientific Online Media and Publishing House

Head Office: 148, Summit Drive, Byron, Georgia-31008, United States.

Offices Overseas: Germany, Australia, India, Netherlands, Canada.

Website: www.iasir.net, E-mail (s): iasir.journals@iasir.net, iasir.journals@gmail.com, aijrstem@gmail.com

**Editorial Board Members of National Conference on Information Technology,
Electronics and Management (NCITEM-2018), July 26-27, 2018**

Chief Editor

Mr. V. K. Sharma

Patron, NCITEM-2018 &
Director,
Centre for Development of Advanced Computing (C-DAC),
B 30 & C-56/1, Institutional Area, Sector 62, Noida,
Uttar Pradesh Pin-201307, INDIA

Editors:

Dr. Arti Noor

Convener, NCITEM-2018 &
Associate Director,
Centre for Development of Advanced Computing (C-DAC),
B 30 & C-56/1, Institutional Area, Sector 62, Noida,
Uttar Pradesh Pin-201307, INDIA

Ms. Priti Razdan

Convener, NCITEM-2018 &
Associate Director,
Centre for Development of Advanced Computing (C-DAC),
B 30 & C-56/1, Institutional Area, Sector 62, Noida,
Uttar Pradesh Pin-201307, INDIA

Ms. Kriti Saroha

Co-Convener, NCITEM-2018 &
Joint Director,
Centre for Development of Advanced Computing (C-DAC),
B 30 & C-56/1, Institutional Area, Sector 62, Noida,
Uttar Pradesh Pin-201307, INDIA

Mr. Sanjay Ojha

Co-Convener, NCITEM-2018 &
Assistant Professor,
Centre for Development of Advanced Computing (C-DAC),
B 30 & C-56/1, Institutional Area, Sector 62, Noida,
Uttar Pradesh Pin-201307, INDIA

Organizing Committee Members of the National Conference on Information Technology, Electronics and Management (NCITEM-2018), July 26-27, 2018

- 1 Bhawana Tyagi
- 2 Isha Gupta
- 3 Kiran Walia
- 4 Manish Kumar
- 5 Mary Jacintha
- 6 Munish Kumar
- 7 Neha Bajpai
- 8 Rahul Mishra
- 9 Ravi Shankar Rai
- 10 Rekha Saraswat
- 11 Rosy Verma
- 12 Shashi Maurya
- 13 Shipra Sharma
- 14 Sunita Prasad
- 15 Tushar Patnaik
- 16 Vivek Arya

TABLE OF CONTENTS

(Special Issue of National Conference on Information Technology, Electronics and Management
(NCITEM-2018))

Paper Code	Paper Title	Page No.
NCITEM 18-101	Reconfigurable Controller for Power Electronics Designs Ajeesh A, Nevin Samuel and Nisha R N	01-06
NCITEM 18-103	Area-Speed Modular Architecture for GF (2^m) Multiplier using Vedic Mathematics Shravan Kumar Sehgal , Dr. Sunita Prasad	07-11
NCITEM 18-104	Design and Implementation of Polar Code Folded Encoder Architecture and Successive Cancellation Decoder on FPGA Vineet Jha, Ravi Payal	12-17
NCITEM 18-105	A General parallel -Serial Architecture for Montgomery Algorithm using Vedic Mathematics (Montgomery modular multiplication) Mangal Deep Gupta, Dr. Sunita Prasad	18-22
NCITEM 18-106	Synthesis of Network on Chip Topologies for Area Efficiency Ana Kumar, Dr. C. K. Jha	23-27
NCITEM 18-107	Design of a Low Power Pulse Triggered D flip flop using Modified Signal Feed Through Scheme Asiyah Zainab Naqvi	28-35
NCITEM 18-109	Look ahead clock gating with probability driven MBFF integration for dynamic power optimization Mohd.Adnan, Dr. Sunita Prasad	36-42
NCITEM 18-110	Design and Implementation of Power Efficient Dynamic Circuit for High Fan-In Circuits Ditipriya Banerjee , Dr. Arti Noor	43-50
NCITEM 18-112	Gesture control using RADAR Ambuj Maurya, Isha Gupta	51-55
NCITEM 18-114	Interfacing of Multiple FPGA using I2C protocol Harshita Varshney, Ankur Vats, Isha Gupta	56-62
NCITEM 18-117	Designing and Verification of AMBA AHB LITE Protocol Manjeet Kumar, Dr. Sunita Prasad	63-67
NCITEM 18-118	Piezo-electric Energy Harvesters: A Battery less Solution for Advance Technology Garima Bandhawakar Wakhle	68-72
NCITEM 18-202	Face Recognition by Removing Illumination Effect Shivani Rawat, Tushar Patnaik	73-78
NCITEM 18-203	Efficient Security in Cloud Storage for Dropbox Preeti Chhikara and Sanjay Ojha	79-82
NCITEM 18-204	Customer Review Rating Using Opinion Mining in Cloud Computing Madhu and Sanjay Ojha	83-87
NCITEM 18-205	Efficient resource utilization using load balancing in Cloud computing Supriya and Sanjay Ojha	88-91
NCITEM 18-210	Palmprint Recognition using Local Binary Pattern Arshi Husain, Tushar Patnaik	92-96
NCITEM 18-211	Noisy and Low resolution document analysis Gunjan Sharma, Tushar Patnaik	97-101
NCITEM 18-212	Age and Gender Prediction for Face book Profile Ankita Sharma, Kalpana Johari	102-104

NCITEM 18-213	Offline Signature Verification Mrignainy Kansal,Tushar Patnaik	105-109
NCITEM 18-214	Analysis of Households Electricity Consumption using Data Mining Techniques Nitin Singh Bisht, Kriti Saroha	110-114
NCITEM 18-215	Flight Crash Investigation Using Data Mining Techniques Vishali , Kriti Saroha	115-118
NCITEM 18-216	Analysis of Asthma Disease with Data Mining Techniques Manmeet Singh , Kriti Saroha	119-121
NCITEM 18-219	Privacy preserving techniques-Outline of Security in Big data Shipra Varshney, Ekta Jain	122-126
NCITEM 18-220	Implementation of Held–Karp Algorithm in Home Delivery System Rakesh Ranjan, Anoop Kumar Yadav	127-129
NCITEM 18-221	Comparison and Analysis of routing protocols in Simulated and Real Time Environment based on the performance of DTN (Delay/ Disruption Tolerant Networks) Jatin Nagpal	130-135
NCITEM 18-222	Cloud Computing Security and Integrity Sri Nath Dwivedi, Shivendra Verma, Rajat Awasthi, Tushar Sharma	136-140
NCITEM 18-223	Prediction of Flight Delays using Data Mining Techniques Neema Chaudhary , Kriti Saroha	141-144
NCITEM 18-224	Punjabi Unicode Inscript Keyboard for MS Word Ashish Batra, Dr. Suneet Madan	145-147
NCITEM 18-226	Prediction of Survival in Thyroid Cancer using data mining techniques Jasvinder Kaur and Kriti Saroha	148-151
NCITEM 18-227	Data Security and Data Privacy Issues Based On Internet of Things Dr. Prabhat Mathur, Manisha Mathur	152-157
NCITEM 18-230	Data Analysis on Cyber Crime Motives Anmol Mittal, Neha Bajpai	158-162
NCITEM 18-231	Brief Outline & Comparison of Various Energy Efficient Routing Protocols in MANET Mona Malik and Rekha Saraswat	163-165
NCITEM 18-301	Importance Of SEO Dupen	166-169
NCITEM 18-302	Training Need Assessment: Creating Skill Inventory At C-DAC, Noida: An Empirical Study Ms. Pooja Atul Kokanthankar, Mrs. Priti Razdan, Dr. M. Mary Jacintha, Mrs. Shipra Sharma	170-176
NCITEM 18-303	Impact of E-Commerce Marketplaces on Business Performance in reference to Retail Industry Aditya Kumar	177-181
NCITEM 18-305	Analysis of Market Strategy of Smartphone Providers Shreya Chawla, Dr.M. Mary Jacintha, Vivek Arya, Shipra Sharma	182-190



Reconfigurable Controller for Power Electronics Designs

Ajeesh A¹, Nevin Samuel² and Nisha R N³

Power Electronics Group

Centre for Development of Advanced Computing

Vellayambalam, Thiruvananthapuram

INDIA

Abstract: This paper describes the reconfiguration competence of an FPGA based high-speed controller for Power Electronic application. The controller is conceived to replace the conventional microcontroller/Digital Signal Processor based controller design, for real-time control and monitoring of Power Electronic systems. The basic control elements for the system are integrated with hardware and software into an FPGA to form flexible controller architecture. The design avoids obsolescence risk and increased cost of the conventional controller. The flexibility of the controller is illustrated by reconfiguring it to meet the requirements of two applications; voltage mode control of boost converter and V/f control of 3-phase ac induction motor. The results obtained have been analyzed and surmised that the FPGA based controller can be easily configured for different PE application, which provides better results within a minimal amount of time.

Keywords: HSRPEC; FPGA; reconfigurable; Power electronics; controller; obsolescence; Boost converter; induction motor

I. Introduction

Lifespan of power electronic systems vary from 10 to 20 years, whereas DSPs or microcontrollers used in its controllers face obsolescence risk after 2 years [1][2]. Conventional limitations of these off-the-shelf devices such as resource limitations are overcome with the help of other onboard peripherals and PLDs [4][5]. This method is lopsided by increased power requirement, failure rate, component density and cost. Further there is the risk of obsolescence of the components. ASIC Controllers were considered as an alternative to the conventional processors. ASIC designs are however neither flexible nor reusable in nature and hence provides no scope for the designer for improvements and changes in field.

High Speed Re-configurable Power Electronic Controller (HSRPEC) is a SOPC based cost-effective reconfigurable controller which offer flexibility to modify hardware at field for various Power Electronics Application[3][4]. This design explores the idea of using control and communication elements as hardware IP modules inside an FPGA. Thus it is faster and reusable than its equivalent software implementation. In this paper, HSRPEC is configured for DC-DC Boost converter application by using various Off the shelf IPs (NIOS, On-chip memory, JTAG UART, System ID, Flash etc.) and custom IP cores (PWM, SPI ADC, PI Controller). Later the controller design is updated for V/f control of 3-phase AC Induction motor by adding extra custom IPs (timer, PWM 3-phase, alpha-beta to abc).

II. Design Methodology

In SOPC scheme, the user can implement a system for a specific application with minimum efforts than required in conventional processor hardware and its software development. The SOPC methodology mainly has two stages namely hardware design and software design [7].

A. Hardware Design

In Hardware section, the first step is to analyse the application and identify all components required for the design. All the unavailable components in off the Shelf IP library is designed in HDL as custom IPs and added into the user IP library as custom IP cores. This IPs must be synthesizable and verified using standard verification methods (OVM/UVM).

Depending on the soft-core processor, only the top-level entity of HDL code needs to be modified for bus interfaces and embodied to SOPC library. For accessing each custom IPs, simple easily understandable custom instructions/functions are designed. After hardware development a hardware description file is generated by the tool. This file is a complete system description and is used by the downstream tool for the software development.

B. Software Design

In software section, the controller software for soft CPU is developed using a software development tool. The selection of application software tool is based on soft CPU used in the design. All software development tasks can be accomplished by using IDE including editing, building, and debugging programs. Off-the-shelf component modules have built-in driver files for the software interface. For the custom developed IP cores, the onus is on the designer to create these driver files.

The following steps are executed for creating software for the design:

- Create the application project
- Create the Board specific package for the system.
- Iterate through one or both of these steps until the design is complete.

The optimized C/assembly code is written with simple understandable functions and ported to the FPGA using boot loaders.

C. Simulation

ModelSim or any other tool can be used for evaluating the functionality of the hardware developed in SOPC methodology along with the application software developed in the IDE.

III. HSRPEC Controller Specifications

HSRPE Controller is designed based on SOPC methodology. HSRPEC card stacked with the Peripheral Interface card is connected to the Power Electronics systems. The specifications of HSRPEC and Peripheral Interface cards are shown in Table I.

Figure 1 HSRPEC card



Figure 2 SOPC--Peripheral Interface card

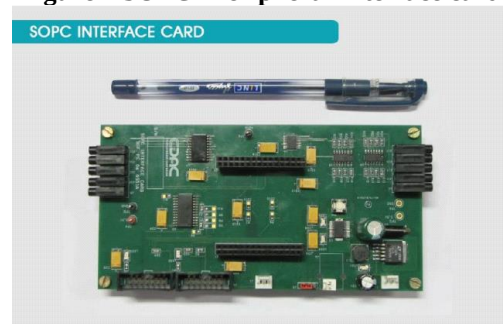


Table I Specifications of HSRPEC and Peripheral Interface Cards

HSRPEC Board		Peripheral Interface Card	
Features	Description	Features	Description
FPGA	Altera EP3C25E144C8	ADC	12bit+ 1 sign bit , 8 channels
Flash	EPCS16N serial flash for FPGA configuration	DAC	12bit+ 1 sign bit 8 channels
Digital I/O	25; 3.3 V LVTTTL Configurable for I/O or for communication (UART, SPI etc.)	GPIO	12 channel PWM, 2 Digital Input, +5V
Host interface	JTAG	Supply	12 Volt
Supply	3.3 V		

IV. Boost Converter

The control structure of DC-DC boost converter is shown in fig 3[8]. After analysing the system IPs required for the design is identified. The output voltage is sampled and fed back to the input to form error signal which is fed to the PI controller IP (PI_PEG_V1). It will generate the reference signal for the PWM generator IP (PWM_PEG_V1) based on the error value. The PWM generates PWM pulses by comparing the reference value obtained from the PI Controller with an internally generated carrier. The PWM pulses from PWM generator IP act as switching pulses for the IGBT. Depending on the error value the duty ratio of PWM pulses also varies so as to minimize the error. The voltage feedback from the converter is digitized using SPI compatible serial ADC (ADC_PEG_V1) in SOPC PEC peripheral Interface card. For debugging purpose, Serial DAC Controller IP (DAC_PEG_V1) is also configured. The IPs used for voltage mode control of DC-DC Boost Converter is listed in Table II. The block schematic of DC-DC Boost Converter with HSRPEC Controller is shown in Fig 4. After hardware development, the software for the controller is designed in IDE software development tool. In this way, HSRPEC Controller card is configured as the controller for voltage mode control of DC-DC Boost Converter.

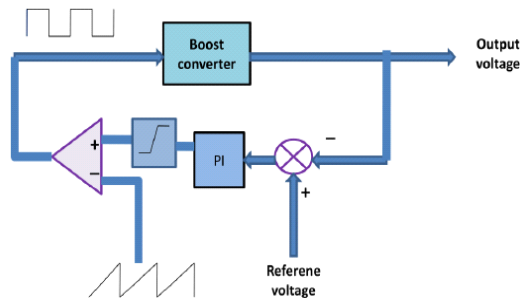


Figure 3 Control structure

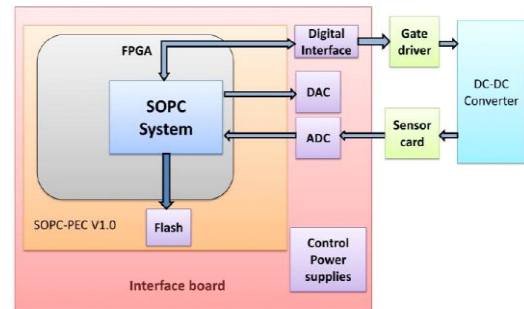


Figure 4 Block view - HSRPEC with dc-dc converter

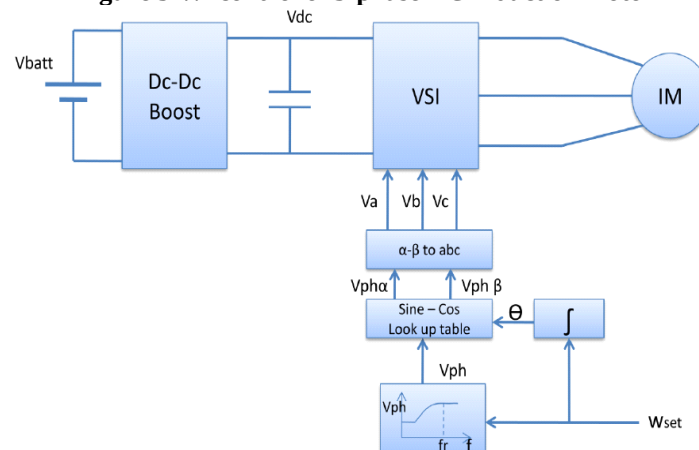
Table III IPs used in DC-DC Boost Converter

IP	Details	Vendors
Off the shelf IPs		
CPU core	NIOS/f CPU (Soft CPU)	Altera
Memory	On-chip RAM	Altera
Interface	JTAG-UART	Altera
Memory Interface	CFI Flash Interface controller, EPCS device controller	Altera
System Identifier (Software protection)	System ID	Altera
On-Chip memory connection support	Tri state bridge	Altera
Custom IPs		
PWM Generator		Custom Development
PI Controller		Custom Development
SPI-ADC Controller		Custom Development
SPI-DAC Controller		Custom Development

V. V/F Control of 3 - Phase Ac Induction Motor

The voltage and frequency drive control topology gives a variable output voltage and frequency to control the speed of AC Induction motor [9]. The V/f drive control can be design based on Boost Converter. The boost circuit is integrated within the design to generate the required ac voltage at the desired frequency. The Reconfigurable architecture of HSRPEC allows us to update boost converter controller hardware for the AC induction motor. The Controller for AC Induction motor is realized by including additional IPs (alpha beta to abc, PWM 3 phase, Timer) to the SOPC controller hardware of DC-DC Boost Converter. The block schematic of V/F control of 3-phase AC induction motor is shown in Fig 5.

Figure 5 V/f control of 3-phase AC induction motor



In the V/f control, the speed of induction motor is controlled by the adjustable magnitude of stator voltages and frequency in such a way that the air gap flux is always maintained at the desired value at the steady state [9]. The IPs used for Hardware development of the application is listed in Table III. In this application, serial ADC IP (ADC_PEG_V1) is used to input the speed of the motor which is integrated to get an angular position, θ .

$$\theta = \int \omega_{set} dt$$

θ based look-up tables for $\sin(\theta)$ and $\cos(\theta)$ for a cycle are evaluated. Based on the θ evaluated from speed, values of $\sin(\theta)$ and $\cos(\theta)$ are obtained from the look-up tables which correspond to alpha and beta. The phase voltage is calculated using the equation.

$$V_{ph} = V_o + k \cdot \omega_{set} \quad ; \text{ where } V_o = \text{stator drop, } k = V/\omega, \omega_{set} = \text{electrical rotor speed}$$

V_{ph} , alpha and beta values are given as input to alpha_beta_to_abc IP which generates PWM references corresponding to V_a , V_b and V_c voltages to be applied to the inverter bridge. The PWM pulses vary the frequency of the supply and the phase voltage is varied accordingly to keep the flux constant. After hardware development, the controller software is designed in IDE tool using custom instruction.

Table III IPs used in V/f control of 3-phase AC induction motor

IP	Details	Vendors
Off the shelf IPs		
CPU core	NIOS/f CPU (Soft CPU)	Altera
Memory	On-chip RAM	Altera
Interface	JTAG-UART	Altera
Memory Interface	CFI Flash Interface controller, EPCS device controller	Altera
System Identifier (Software protection)	System ID	Altera
On-Chip memory connection support	Tri state bridge	Altera
Custom IPs		
PWM Generator		Custom Development
PI Controller		Custom Development
SPI-ADC Controller		Custom Development
SPI-DAC Controller		Custom Development
Timer		Custom Development
PWM 3phase generator		Custom Development
Alpha beta to abc		Custom Development

VI. Application Evaluation and Results

The HSRPE controller along with a custom designed interface card for power stage is used in a 60V-150V DC-DC boost converter. It is found that the HSRPEC operating at 100MHz took 800n sec. for the control loop execution.

Figure 6 Boost converter with HSRPEC

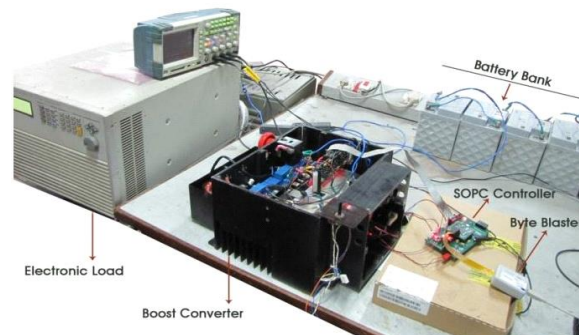
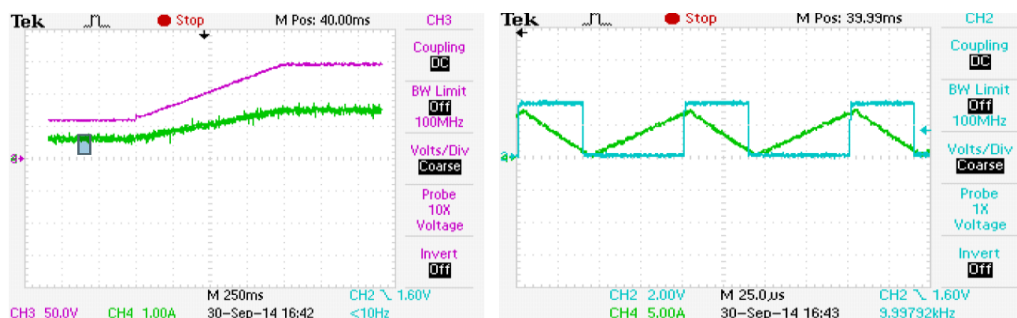


Figure 7 Load voltage and current; inductor current and PWM signal



The Reconfigured HSRPEC along with peripheral interface card is used in V/F control of 3 phase AC Induction motor. The result shows alpha_beta_to_abc IP took only 500ns to perform the conversion. V_a , V_b and V_c values from Nios II through alpha_beta_abc IP is also evaluated.

Figure 8 Test setup of three phase AC Drive hardware assembly

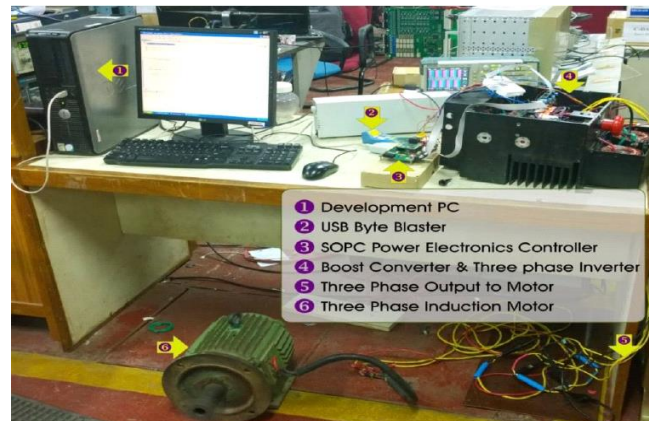


Figure 9 Line voltage (pink) & current (green)

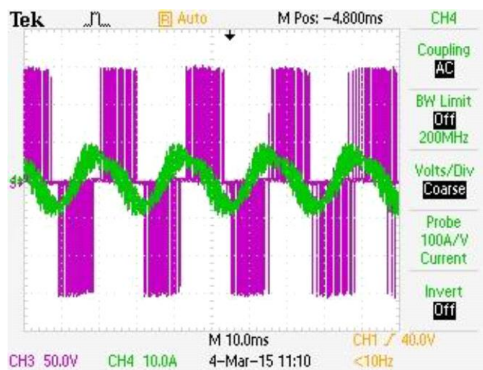


Fig10. Line voltages V_{ab} (pink) & V_{bc} (blue)

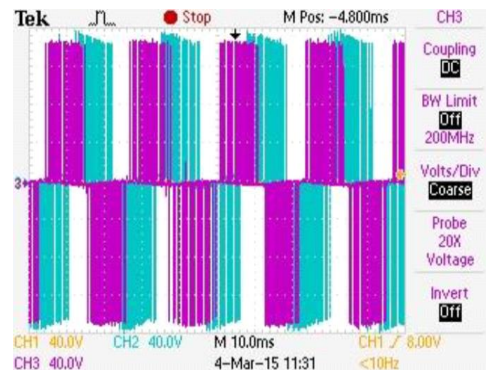


Fig 11 V_{ab} , V_{bc} & line current at 1300 rpm ω_{set}

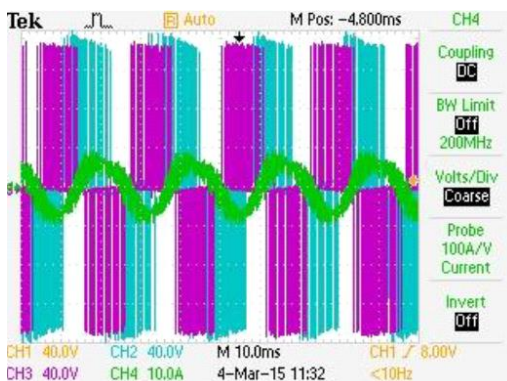
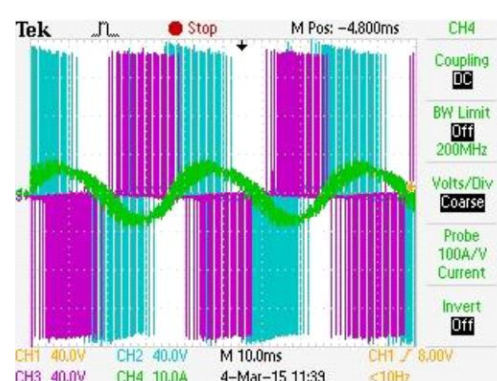


Fig12 V_{ab} , V_{bc} & line current at 758 rpm ω_{set}



VII. Conclusion

Power Electronics systems in railway, military etc. have a life-span of the order of 10-20 years. It is essential for the manufacturing industry to provide support for such systems over its entire service life. The SOPC methodology has the potential to provide a reconfigurable design and good solution to processor obsolescence and related overheads. HSRPEC is a SOPC based cost-effective reconfigurable controller for Power Electronic applications. In this paper, we discussed the reconfigurable competence of HSRPEC by using two applications; voltage mode control of boost converter and V/f control of 3 phase ac induction motor. The results obtained have been analyzed and surmised that HSRPEC developed along with power electronics specific IP is both cost-effective, ensures long term support for power electronic systems and provides better performance over

conventional systems.

References

- [1] P. Sandbord, "Trapped on Technology's Trailing Edge", IEEE Spectrum,
- [2] R. C. Stogdill, "Dealing with obsolete parts" IEEE Design & Test of Computers, Aug 2002, pp. 17-25
- [3] L. Anghel, R. Velazco, S. Saleh, S. Deswaertes, A. El Moucary, "Preliminary Validation of an Approach Dealing with Processor Obsolescence", 18th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, p. 493, 2003
- [4] Armando Astarloa, Unai Bidarte, Jesus Lazaro, Jon Andreu, and Jose Luis Martin, Configurable-System-on-Programmable-Chip for Power Electronics Control Applications, *International Conference on Reconfigurable Computing and FPGAs*, pp. 169 – 174, 2008
- [5] E. Monmasson, L. Idkhajine, I. Bahri, M-W- Naouar and L. Charaabi, Design methodology and FPGA-based controllers for Power Electronics and drive applications, *Pacific-Asia Conference on Circuits, Communications and System*, pp. 360-363, 2009
- [6] K. Sugahara, S. Oida, T. Yokoyama, High performance FPGA controller for digital control of power electronics applications, *IEEE 6th International Power Electronics and Motion Control Conference*, pp. 1425 - 1429, 2009
- [7] Jason G. Tong, Ian D. L. Anderson and Mohammed A. S. Khalid, Soft-Core Processors for Embedded Systems, *International Conference on Microelectronics*, 2006
- [8] Robert W. Erickson, Dragan Maksimovic, Fundamentals of Power Electronics, *Second Edition*, 2001
- [9] L. Umanand, Power Electronics: Essentials & Applications, 2010



Area-Speed Modular Architecture for GF (2^m) Multiplier using Vedic Mathematics

¹Shravan Kumar Sehgal, ²Dr. Sunita Prasad (Principal Technical Officer)

^{1,2}School of Electronics, C-DAC Noida
India

Abstract: Arithmetic operations over GF (2^m) is used in many application and the major application are found in the field of cryptography. It is a modern technique used for modular multiplication. This paper proposes a modified or adjusted approach for AES algorithm which is speed and area efficient technique. This paper can be operated on large bits defined in cryptographic standard which prompts increment of the general zone speed productivity of the design. Multiplication process of large bits is a time and area consuming process and both these parameters are inversely proportional to each other. For improving the overall efficiency the paper combines the bit-parallel with the bit-serial and for bit parallel Vedic Mathematics is used for the processing of the inputs. This paper proposes a modular architecture it doesn't need to upgrade the multiplier in the event that we choose to change the field measure utilized. Proposed architecture is evaluated for Spartan 6 FPGA.

Keywords: FPGA, GALOIS FIELD, VEDIC MATHEMATICS, AES ALGO, CRYPTOGRAPHY.

I. INTRODUCTION

Effective hardware executions of finite field GF (2^m) is widely used in several applications signal processing, coding theory, commutative algebra, advance encryption standard, but mainly in cryptography like elliptic curve cryptography, pairing cryptography. Finite field arithmetic has following operations addition, division, multiplication, squaring, inversion, and exponentiation. The main focus of this paper is in AES algorithm application and solution dedicated to reconfigurable hardware like Field Programmable Gate Array (FPGA) [1]. The multiplication in basically in two steps the first one is polynomial multiplication and the second is modular reduction. In this we implement parallel multiplier using Vedic Mathematics.

A modular architecture for multiplication is discussed in this paper which can be redesigned for different size of bits than proposed by NIST. Paper proposes a method of building a smaller multiplier with 12 bits for multiplication. The target hardware on which this method is implemented is Spartan 6. The polynomial basis multiplier is widely used because of its simplicity in designing also it offers scalability in higher orders which is used in real time applications.

II. MULTIPLICATION FOR FINITE FIELD

Multiplication operation in the finite fields is very difficult process. Few other operations which are complicated include inversion, squaring, exponentiation and division. Repeated multiplications can be used to carry out these operations. Hence efficiently carrying out finite field multiplication for serving the target applications is critical step.

Multiplication basically requires a Multiplier, Multiplicand and a Modulus. For the cryptographic application it require application on large number of bits. The performance efficiency of the multiplier is dependent on the field element terms and numbers of non-zero irreducible polynomial.

The implementation of finite field multiplication is heavily dependent by the choice basis which is used for representing the field' elements for example, The "Polynomial Basis" (PB) and the "Dual Basis" (DB) and "Normal Basis" (NB)[3].

Proposed approach considered polynomial basis (PB), which is very commonly used, and irreducible trinomials for designing of bit-parallel finite field multipliers. For the bit-parallel multiplier design Vedic Multiplication has been used.

The polynomial basis can be depicted as binary polynomial of the form:

$$p = p(x) = (x + a)^n = \sum_{i=0}^{m-1} p_i x^i = p_0 + p_1 x + \dots + p_k x^k + \dots + p_{m-1} x^{m-1} \text{ where } p_i \in GF(2)$$

The designing is focused to FPGA's. For the task of simulation, Xilinx device SPARTAN 6 is used. The structural design is made in Verilog and synthesized, placed, routed and then implemented by using Xilinx ISE

14.7 environment and their tools. Complete simulation is carried out under ISE 14.7 or ModelSim simulator (Mentor Graphics).

Outlining a productive equipment calculation for Multiplication in the GF (2^m) requires a exhaustive examination of existing techniques furthermore, arrangements. The investigation regarding the matter is extremely cutting-edge. Numerous designs were proposed for GF (2^m) multipliers. The current arrangements can be named bit-parallel, bit-serial, digit-serial/parallel multipliers or systolic designs. Each kind of arrangement has distinctive properties: bit-parallel arrangements are quick however substantial, bit-serial are moderate however slow. The Digit and additionally systolic structures offer generally some trade off amongst speed and territory of the arrangement.

Models of the multipliers in the GF (2^m) can be characterized into bit serial, bit parallel, digit serial, and composite models. Bit-parallel multipliers simultaneously produce all outcome bits which are synchronized in single clock cycle. Multipliers which are Bit-serial, yield a single bit in each clock and have lower equipment cost[2]. Digit-serial multipliers give d ($1 \leq d \leq m$) bits in the result in single clock cycle. Combined multipliers can acquire sub quadratic space many-sided quality outlines. Bit-parallel multipliers have a tendency to be quicker so that they are appealing for some applications.

GF (2^m) multiplication process consists of following two steps: Multiplication and reduction[1]. Two numbers are multiplied one is taken as Multiplier and other as Multiplicand. The multiplier is fixed as it is and multiplicand is divided in smaller bits. The result obtained is reduced by using Galois Field reduction. It performs XOR operation hence no carry is being propagated. In the multiplicand case we use Vedic Mathematics.

III. ARCHITECTURE

On observing Eq. 2 it was found that it became time inefficient when considered for large polynomials and performed computation in bit serial fashion. But on performing parallel bit operation, area inefficiency was found to be an issue. In order to tackle the above discussed issues a method in which the vector polynomial q was processed bit by bit in parallel with 12-bit words[2]. According to research 12-bit is the appropriate size and it gives good results. For the field size m , we have to perform $\lceil m/4 \rceil$ operations.

$$s(x) = p(x)q(x) = p(x) \sum_{i=0}^{m-1} q_i x^i + p(x) \sum_{i=4(\lceil \frac{m}{4} \rceil - 1)}^{m-1} q_i x^i$$

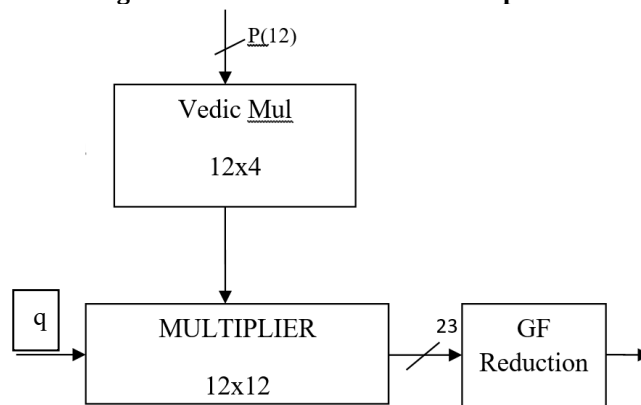
$$s = pq^{[0]} + pq^1 + \dots + pq^{\lceil \frac{m}{4} \rceil - 1}$$

Partitions are to be made of the multiplicand q into 4-bit words. By doing this $\lceil m/4 \rceil$ $pq^{[i]}$ operation is performed, after that the final product is obtained. Architecture has two blocks: 12x4 bits Vedic multiplier block and multiplier performing 12x12 multiplications. For performing $\lceil m/4 \rceil$ it requires $\lceil m/4 \rceil \times T_{mul16x4}$ cycles, and $T_{mul16x4}$ is the delay by multiplier block. The multiplier block gives 23 bit result since carry is not being propagated in the multiplication operation. This result is then reduced by GF reduction module to obtain 12 bit results. First multiplication is performed:

$$s(x) = p(x)q(x) = p(x) \sum_{i=0}^{m-1} q_i x^i = \sum_{i=0}^{m-1} q_i p(x) x^i = q_0 p(x) + q_1 p(x)x + q_2 p(x)x^2 + \dots + q_{m-1} p(x)x^{m-1}$$

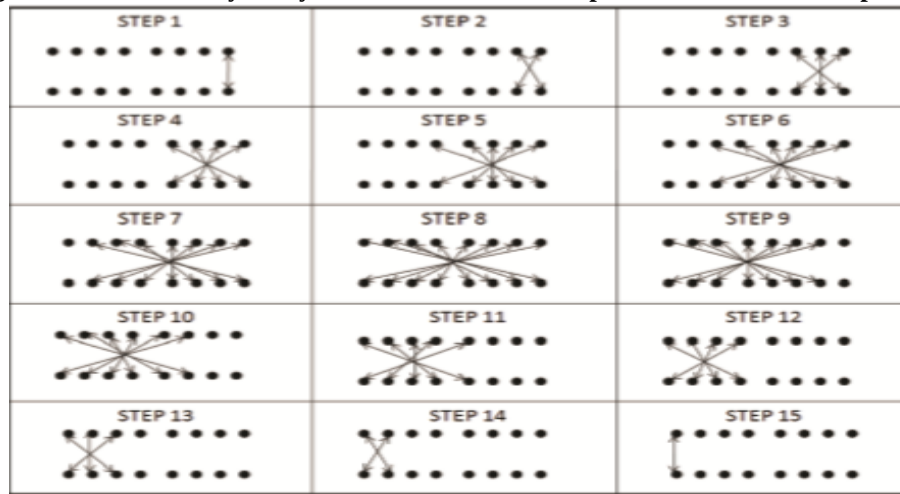
The multiplication is performed by using Vedic Mathematics and it is first implemented for smaller bits but as this is a modular design it can also be implemented for large number of bits. This design is simple and efficient.

Figure. 1. Architecture of GF Multiplier



Designing of 12x4 bit sub-multiplier block which performs operation is the most important task. For larger number of bits to perform multiplication operation it is done by parallel method and the parallel architecture uses the Vedic Multiplication concept. In Vedic multiplication large number of bits is multiplied in a single step. After performing the task of 12x4 multiplications, the task of 12x12 multiplications is carried out. For the multiplication process of two numbers p and q first p is taken which is the multiplier and kept as it is and then q is taken which is the multiplicand and divided in to 4 bits to make it of form [m/4]. The multiplication of 12x4 is done in parallel by the use of Vedic mathematics. In Vedic arithmetic procedure whatever be the issue there is one general system material to all cases and moreover various uncommon example issues. It is mostly in light of 16 principle sutras and 13 sub-sutras. Vedic science is a champion among the most wonderful areas of the twentieth century logical history. It has great informative regard in light of the way that the sutras contain systems for playing out some essential numerical exercises in fundamental ways and results are obtained quickly. Thus Vedic science is easy to learn, speedier to utilize and less obligated to error than traditional technique.

Figure. 2. Urdhwa Tiryakbhyam Sutra of Vedic Multiplier FOR 8*8 bit Multiplication.



The ancient Vedic Mathematics sutra known as Urdhwa Tiryakbhyam Sutra, by its definition, means vertical and crosswise. This further implies the multiplication that occurs between multiplier and multiplicand bits which are present at their extremes. The significant improvement of the proposed technique is that the product of the two numbers can be obtained in just one step. Further-more the process of multiplying two individual single bits reduces to just an AND operation, the VLSI implementation of the approach proves to be design efficient both in terms of area as well as speed. In this the addition operation is not performed instead XOR operation hence there is no carry generated. Hence delay in terms of time is very less.

Each 12x12 multiplier is implemented by the following equations:

$$s[0] = (p[0] \& q[0]),$$

$$s[1] = (p[1] \& q[0]) \sim (p[0] \& q[1]),$$

$$s[2] = (p[2] \& q[0]) \sim (p[0] \& q[2]) \sim (p[1] \& q[1]),$$

$$s[3] = (p[3] \& q[0]) \sim (p[0] \& q[3]) \sim (p[2] \& q[1]) \sim (p[1] \& q[2]),$$

$$s[4] = (p[4] \& q[0]) \sim (p[3] \& q[1]) \sim (p[1] \& q[3]) \sim (p[2] \& q[2]),$$

$$s[5] = (p[5] \& q[0]) \sim (p[4] \& q[1]) \sim (p[3] \& q[2]) \sim (p[2] \& q[3]),$$

$$s[6] = (p[6] \& q[0]) \sim (p[5] \& q[1]) \sim (p[4] \& q[2]) \sim (p[3] \& q[3]),$$

$$s[7] = (p[7] \& q[0]) \sim (p[6] \& q[1]) \sim (p[5] \& q[2]) \sim (p[4] \& q[3]),$$

$$s[8] = (p[8] \& q[0]) \sim (p[7] \& q[1]) \sim (p[6] \& q[2]) \sim (p[5] \& q[3]),$$

$$s[9] = (p[9] \& q[0]) \sim (p[8] \& q[1]) \sim (p[7] \& q[2]) \sim (p[6] \& q[3]),$$

$$s[10] = (p[10] \& q[0]) \wedge (p[9] \& q[1]) \wedge (p[8] \& q[2]) \wedge (p[7] \& q[3]),$$

$$s[11] = (p[11] \& q[0]) \wedge (p[10] \& q[1]) \wedge (p[9] \& q[2]) \wedge (p[8] \& q[3]),$$

$$s[12] = (p[11] \& q[2]) \wedge (p[10] \& q[3]),$$

$$s[13] = (p[11] \& q[2]) \wedge (p[10] \& q[3]),$$

$$s[14] = (p[11] \& q[3]);$$

IV. RESULTS

For the purpose of comparing the Vedic Mathematics architecture with the existing architecture both the architectures have been implemented. Both the architecture has been implemented on SPARTAN 6 FPGA as the results are as shown in the Table:

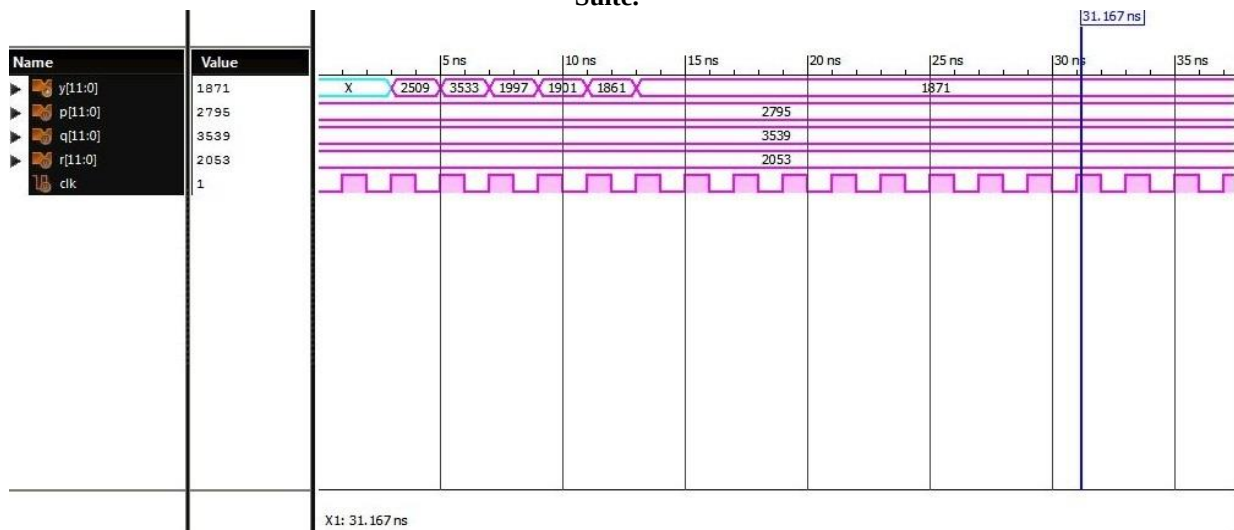
TABLE I MODULAR FINITE FIELD MULTIPLIER

	12_4 Multiplier using array multiplication	12_4 Multiplier Vedic Mathematics
No. of 4 input LUTs	33	35
No. of occupied slices	17	18
Average fan-out	2.55	2.57
No. of IOBs	31	31
Timing Constraint :Delay	7.334nsec	7.334nsec

TABLE II IMPLEMENTATION RESULT FOR TWO TYPES OF MULTIPLIERS

	Multiplier using Vedic Mathematics	Multiplier using Array Multiplication
Number of Slice LUTs	91	39
Average Fan-out	3.53	3.9
Number of IOBs	13	49
Timing Constraint :Delay	5.611nsec	8.656nsec
Number of 4 input LUTs	169	359
Number of XOR gates	65	65
Number of Registers	8	36

Figure.3: Simulation Result of GALOIS FIELD Multiplication of 12 bit data performed on ISE Design Suite.



The above figure depicts the Simulation Result in which two 12 bits numbers are multiplied and then reduced by using GALOIS Field reduction by dividing it with 12 bit prime number. So as to reduce the result of multiplication which is obtained as 23bits after the multiplication back to 12bits.

V. CONCLUSIONS

On comparing our answer with existing ones isn't simple [2]. Numerous references introduce simply hypothetical depiction, i.e. assessment of anticipated quantity of gates utilized and delay. A correlation of answers has been played out with the ones, exhibited in the paper, which appear basing on the comparative thought as the proposed one which is in the paper, the multiplier has been implemented on SPARTAN 6 FPGA. The bit size for the computation is taken to be $m=12$. The answers are planned also, streamlined to work for substantial field sizes, the solutions have been represented in tables I & II.

Delay in case of GALOIS Field Multiplier using Vedic Mathematics has been reduced significantly. Also area in terms of number of registers and number of LUTs used has also been reduced by a significant amount. Here the objective of reducing the overall time-area product has been achieved successfully. The Vedic Multipliers reduces the time complexity of parallel multiplication and these Multipliers also reduce the carry dependency. The computational time required of the proposed Vedic multiplier is less as compared to Array multipliers. This design is implemented on SPARTAN 6 FPGA for modulus value 4.

VI REFERENCES

- [1]. Huang Q, Liu L, Huang H. and Wei S, 2017, May. "Area-efficient polynomial modular multiplication over $GF(2^n)$ and application to AES". *The 2017 IEEE 9th International Conference on (pp. 1128-1132). IEEE. In Communication Software and Networks (ICCSN)*,
- [2]. Pamula D. and Hryniewicz E, 2013, April. "Area-speed efficient modular architecture for $GF(2^m)$ multipliers dedicated for cryptographic applications". Design and Diagnostics of Electronic Circuits & Systems (DDECS), 2013 IEEE 16th International Symposium on 8-10 April 2013 INSPEC Accession Number: 13614452 IEEE 2013 Karlovy Vary, Czech Republic.
- [3]. Anto Bennet M, Manimaraboopathy M, Maragathavalli, P. and Dinesh Kumar T.R, 2014. "Low Complexity Multiplier For $Gf(2^m)$ Based All One Polynomial". IEEE Transactions on Very Large Scale Integration (VLSI) Systems on 12 January 2012 INSPEC Accession Number: 13188360 IEEE Circuits and Systems Society IEEE Computer Society IEEE Solid-State Circuits Society.
- [4]. Xie J, jun He J. and Meher P.K, 2013. "Low latency systolic montgomery multiplier for finite field $GF(2^m)$ based on pentanomials". School of Information Science and Engineering, Central South University, Changsha 410083, China. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 21(2), pp.385-389.
- [5]. Pan J.S, Azarderakhsh R, Kermani M.M, Lee C.Y, Lee W.Y, Chiou C.W. and Lin J.M., 2014. "Low-Latency Digit-Serial Systolic Double Basis Multiplier over $GF(2^m)$ Using Subquadratic Toeplitz Matrix-Vector Product Approach". *IEEE Transactions on Computers* 02 October 2012 INSPEC Accession Number: 14281628 IEEE Computer Society.
- [6]. Cho Y.I., Chang, N.S, Kim, C.H Park, Y.H. and Hong S, 2012. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 20(10), "New bit parallel multiplier with low space complexity for all irreducible trinomials over $GF(2^n)$ " pp.1903-1908.
- [7]. Xu JianBo, Dai Zibin, Xuan Yan, Su Yang, 2011. IEEE. "Research on Reconfigurable Multiplier Unit Based on $GF(2^8)$ Field of Symmetric Cryptography". 2011 9th IEEE International Conference on ASIC.

VII. ACKNOWLEDGEMENT

I take this opportunity to acknowledge all those who have guided me in this paper. I express my earnest gratitude towards P.T.O. Dr. Sunita Prasad, my mentor for her valuable encouragement and his guidance. I am grateful for their most cooperative attitude and suggestions, without which I could not have been able to do this work and My colleagues their motivation and suggestions were invaluable in the project work. I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Design and Implementation of Polar Code Folded Encoder Architecture and Successive Cancellation Decoder on FPGA

¹Vineet Jha, ²Ravi Payal (Principal Technical Officer)
School of Electronics
CDAC, NOIDA, Uttar Pradesh
INDIA

Abstract: Among all channel encoding technique recently proposed polar code is a very strong candidate for very high speed capacity achieving wireless communication. Arikan proposed polar code as capacity achieving class of linear block codes, having logarithmic encoding and decoding architecture. The fully parallel polar code encoder is not very suited for large block length code due to its obvious disadvantage of high hardware complexity. Hence in this brief we discuss the new folded polar code encoder architecture for long polar codes and implementation is done on XILINX Spartan6 FPGA xc6slx45 as target device. This architecture shows a very promising result for low hardware complexity polar code encoder architecture.

Keywords: Data Flow Graph(DFG), Polar codes, Folding, Successive cancellation decoder(SCD), Field programmable gate array

I. Introduction

The compression of redundant data, error free transmission of data through physical channel and at the same time achieving the Shannon capacity are the important consideration in the scope of the information theory[1]. The information theorist always strive for designing of the coding theory with high performing error correcting and detecting code and the Shannon limit. Arikan[8] proposed a superior BER performance channel encoding technique known as polar codes. Polar codes belong to the class of linear systematic block codes with capacity achieving performance. Synthesizing the polar codes it is required to assume that inputs X have uniform distribution and are transmitted at maximum mutual information $I(X;Y)$. [7][8] presented the systematic polarization of channel in a recursive fashion and dividing the channel into set of extreme channels with very high capacity channel or high mutual information and very low capacity channel or virtually zero mutual information. High capacity channels are good channel through which data can be transmitted without any noisy perturbation in data and at bad channel frozen bit are transmitted. Through this systematic splitting of channel into good and bad channel encodes the data into polar codes. Polar codes shows very high performance across discrete memory less Binary symmetric channel and binary erasure channel. [8] proposes SC decoder a greedy iterative tree search algorithm for decoding of polar codes. The encoding and decoding complexity of polar codes is $O(N \log N)$. The phenomenon of polar codes achieving capacity prominently happens at asymptotically large block length for superior error correcting performance. So to achieve capacity the code block length should be at least 2^{20} bits or more. But with this increase in block length the hardware complexity of encoder and decoder increases steeply. The storage requirement for data bit during encoding and decoding increases enormously with large block length. The parallel encoder suffers from this large hardware complexity asymptotic code length.

This paper discusses the polar code encoding process in the view point of ASIC and FPGA implementation and analyzes the folding transform architecture [6] in the context of polar code encoder. The proposed polar code architecture is highly suitable for asymptotic long polar codes further this paper also discusses about the successive cancellation decoder. Sec-II talks about the basics of polar codes, encoding and decoding. Sec-III discusses the proposed architecture. Sec-IV presents the FPGA result and comparisons and finally the conclusion at the end.

II. Polar Encoding

The channel polarization [8] splits the channel into perfect maximum capacity channel and completely noisy channel as the block length N asymptotically approaches infinity. As the knowledge of the set of K reliable sub channels and $N-K$ noisy channels are known a priori. The data is transmitted through K reliable channels and noisy channels are given frozen bit which construct a (N,K) linear polar code. The channel polarization is simply a method of synthesizing the artificial channels, it is a rule that dictates particular technique for channel splitting, although the random permutation can be seen as a fast converging polarizer, but not very efficient in terms of hardware implementation. Then [8] suggested a systematic technique to convert a mediocre channel

into extremes channels. Arikan kernel $F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ is used to obtain the generator matrix G_N as follows $G_N = F^{*n}$, G_N is obtained as kronecker product on kernel F by n times as $n = \log N$. The extremes of original BEC channel in form of artificial channel with either very small erasure probability or very high erasure probability is obtained by recursively growing the encoding tree at the depth $n = \log N$. This is mathematically described by triangular generator matrix. Polar codes belong to the class of linear block code, hence the encoding is described as mapping of input vector U to output code word X by G_N generator matrix according to the encoding equation $X = G_N * U$. In this paper we assume that input vector U is arranged in normal order, while codeword X in bit reversed order for the sake of simpler explanation of encoding.

A. Parallel encoding

The fully parallel straight forward architecture as presented in [7] having encoding complexity $O(N \log N)$ for N block length. Where $\log N$ is the number of stages of encoding graph. A $(16, k)$ block code fully parallel architecture has four stages and contain 32 XOR gates in total with 8 XOR gates at each stages. As it can be seen in fig. 1 that fully parallel architecture does not require any sequential element, hence code word is obtained in first cycle with maximum throughput.

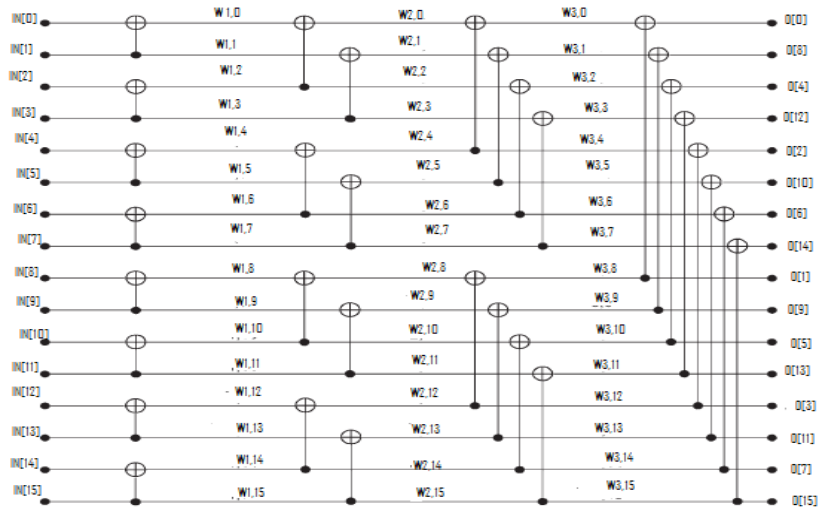


Fig 1. (16,k) parallel encoder architecture

B. SC decoding

[8] first proposed the successive cancellation decoder, this is one of the fundamental decoder for polar codes. It uses two way recursive algorithm on N received likelihood in N iterations. SCD is traditionally a greedy tree node traversal algorithm. In which we intend to find the tree path having 2^k leaves at the end. The SC decoder uses following equations.

$$\begin{bmatrix} l1 \\ l2 \end{bmatrix} \xrightarrow{\text{Transformation}} \begin{bmatrix} f(l1, l2) \\ g(l1, l2) \end{bmatrix}$$

$$\begin{bmatrix} f(l1, l2) \\ g(l1, l2) \end{bmatrix} = \begin{bmatrix} \frac{l1 \cdot l2 + 1}{l1 + l2} \\ l2 \cdot l1 \text{ or } l2/l1 \end{bmatrix}$$

For circuit implementation above equation further approximated as

$$\begin{bmatrix} f(l1, l2) \\ g(l1, l2) \end{bmatrix} = \begin{bmatrix} \text{sgn}(l1) \cdot \text{sgn}(l2) \min(|l1|, |l2|) \\ l2 + (-1)^u \cdot l1 \end{bmatrix}$$

Here $l1, l2$ are likelihood at the input of decoder and $f(l1, l2)$ $g(l1, l2)$ are the likelihood function which is evaluated and u is the hard estimated bit at the decoder output as shown in above equation. The SCD mainly is the network of memory elements and XOR gates. Memory element stored the result of intermediate estimated value. Hence in SCD decoder the latency is the main concern due to its sequential nature.

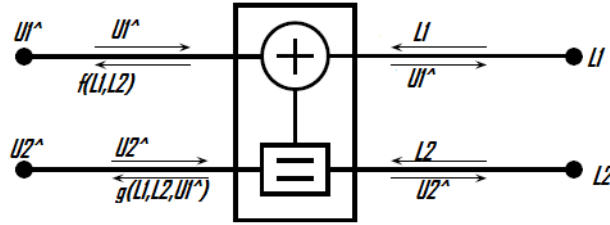


Fig.2 SC decoder data flow architectures

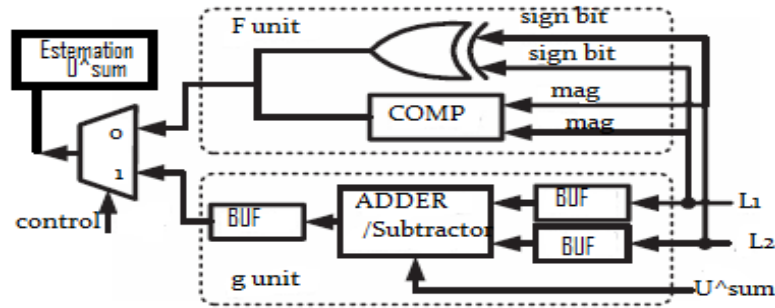


Fig.3 SC decoder processing element

III. Proposed Encoder Architecture

The implementation of fully parallel encoder architecture although intuitive but such a design becomes very hardware extensive as block length approaches millions bits. [6] Discusses the concept of hardware reusability using time multiplexed design for FFT architecture and further [5] uses the same concept for polar encoder architecture. The result is of course the tradeoff between speed and hardware complexity. Here this paper proposes a folded architecture for (16,k) polar code encoder. The fully parallel architecture is first changed into folded form [5], further lifetime chart and register minimization and allocation is introduced in folded design.

A. Folding transformation

The folding transformation is a time multiplexed technique which introduces the hardware reusability by performing multiple operation on single functional unit in the multiplexed fashion. It is considered a effective technique to save extensive hardware usage in design. In folding transform with the use of register set ,multiplexer and well defined controlled logic, the flow of data from one state to next state and scheduling of operation in operational unit is accomplished. Folding transform introduces hardware pipelining. In figure1. data flow graph(DFG) of (16,k) polar code is shown, in which node on graph represent kernel matrix operation F and $w(i,j)$ represent the j^{th} edge of corresponding i^{th} stage. It is very much apparent that the DFG of polar code is similar to FFT [6] [8] ,except the fact that polar code DFG uses kernel matrix F unlike FFT butterfly unit. In proposed 16 bit folded architecture the folding factor is 4, where 4 bit processing happens at a time. Two functional unit per stage and each unit processes two bit and accomplishes the task. The most crucial design factor is to determine the folding set. The folding set schedule the operation to be executed in particular order by the functional unit. First operations are divided into separate folding sets. Each folding set either have only odd or even scheduled operations. At the first stage four parallel inputs are in natural order has two folding set ($S_{10}, S_{12}, S_{14}, S_{16}$) and ($S_{11}, S_{13}, S_{15}, S_{17}$) , each set has four functional unit elements and their position represents the order of operation in which functional unit executes. In the first stage S_{10} and S_{11} executes operation simultaneously at the starting and S_{12} , S_{13} in the next time unit and so on. In second stage the folding set is in the same order as previous stage ($S_{20}, S_{22}, S_{24}, S_{26}$) and ($S_{21}, S_{23}, S_{25}, S_{27}$). Further determination of folding set of stage n , a property is used which says that functional unit operates on the inputs whose indices are differing by 2^{n-1} .For the case of stage three data inputs with index differing by four are processed , for example w_{20}, w_{24} from S_{20} , S_{22} are the inputs for S_{30} . But for operation both input should be valid for processing in functional unit , the stage three operation aligned to the delayed input data. A delay element is introduced to align the inputs to the stage operational unit. Stage three folding set are ($S_{36}, S_{30}, S_{32}, S_{34}$) and ($S_{37}, S_{31}, S_{33}, S_{35}$). Similarly the folding set for stage four is obtained by cyclic shifting by two unit right of previous stage folding sets. For stage four folding set are ($S_{42}, S_{44}, S_{46}, S_{40}$) and ($S_{43}, S_{45}, S_{47}, S_{41}$). In general for a stage index n which is less than equal to $logp$, p is parallelism index has folding set as interleaving operations by one unit in order, for n greater than $logp$, the set is determined by cyclic shift right of the folding set of previous stage by $n - logp$.

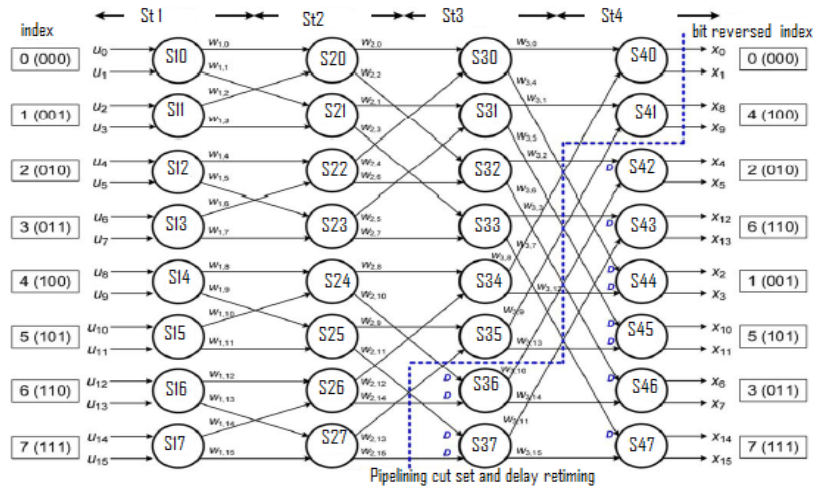


Fig 4. DFG for folded architecture

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$D(W_{1j})$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$D(W_{2j})$	1	1	2	2	0	0	1	1	1	1	-2	-2	0	0	-3	-3
$D(W_{3j})$	2	2	-2	-2	-0	-0	-0	-0	0	0	0	0	-2	-2	2	2
$D'(W_{1j})$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$D'(W_{2j})$	1	1	2	2	0	0	1	1	1	1	2	2	0	0	1	1
$D'(W_{3j})$	2	2	2	2	4	4	4	4	0	0	0	0	2	2	2	2

Table 1. Original delay requirement $D(w(i,j))$ and $D'(W(i,j))$

Now the delay required between the operation from one stage to other stage is determined by [5] $D(W(i,j)) = Fd + v - u$, this delay is between functional units from set U to V after folding where F is the folding factor, d is the intrinsic delay before folding, u and v are the indices from the set U and V respectively. The $D(w(i,j))$ is tabulated in table. For folding architecture to be feasible we need to get rid of negative delay, as delay should be greater than equal to zero. To do this pipelining technique is applied to the edges having negative delays in which delay element are added called delay retiming. Further the delays $D'(w(i,j))$ are calculated for restructured nodes. We get 48 internal delay elements to implement (16,k) four parallel architecture. Further to minimize the delay elements linear lifetime analysis and register allocation [6] used which reduces the number of delay to 12 from 48 by reusing the register set in time multiplexed fashion. Finally the 4-parallel 4-folded pipelined architecture for (16,k) polar code encoder consist of four stages, in which first two stages are purely combinational and contain in total 4 XOR gates in four functional unit ,two in each stage. In stage three and four the combination of register set, multiplexer and XOR gates are used. In stage three there are two XOR gates, four register and two multiplexer are used. Similarly for stage four eight register two XOR gates and two multiplexer are used. The output is obtained at the starting of third cycle. The architecture take input in consecutive index order and at the output bits are in reversed fashion.

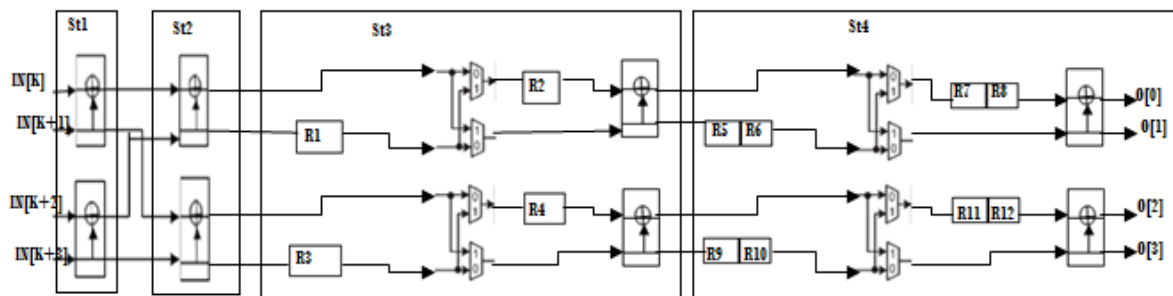


Fig.5 Proposed folded architecture for (16,k) polar code

IV. Implementation Results and Comparisons

Device utilization	Fully parallel encoder(16,k)			[5]Partially parallel encoder architecture(16,k)			Proposed folded encoder architecture(16,k)		
	Used	Available	%	Used	available	%	Used	available	%
Number of slice register	0	54576	0%	4	30064	0.013%	17	54576	0.0311%
Number of slice LUTs	21	27288	0.076%	25	15032	0.17%	19	27288	0.069%
Number of IOBs	32	218	14%	131	190	68.95%	9	218	4%
Number of BUFG	0	16	0%	1	16	6.25%	1	16	6.25%

Table 2. Device utilization data for Xilinx Spartan 6

Device utilization	Successive cancellation decoder (8,k)		
	Used	Available	%
Number of slice register	7	54576	0.012%
Number of slice LUTs	210	27288	0.76%
Number of IOBs	49	218	22.4%
Number of BUFG	1	16	6.25%

Table 3. Device utilization data for SC decoder(8,k) on Spartan 6

All the design is synthesized and simulated on Xilinx ISE design suite 14.7 and our target device is Xilinx Spartan 6 FPGA for implementation. First (16,k) fully parallel encoder architecture is implemented and design summary is obtained and tabulated then proposed architecture is implemented and the synthesis report is compared with fully parallel design and with [5] architecture. The proposed encoder architecture shows improvement over fully parallel design as well as [5] architecture in terms of slice LUTs, IOBs etc. The timing information is also compared among different architecture for fully parallel architecture the maximum combinational delay is 6.88ns and for proposed folded architecture the maximum combinational delay is 6.87ns furthermore the time period and maximum frequency of operation is 2.190ns and 456.66Mhz. The timing details for the SC decoder is maximum combinational delay is 17 ns and frequency of operation is 474.72 Mhz.

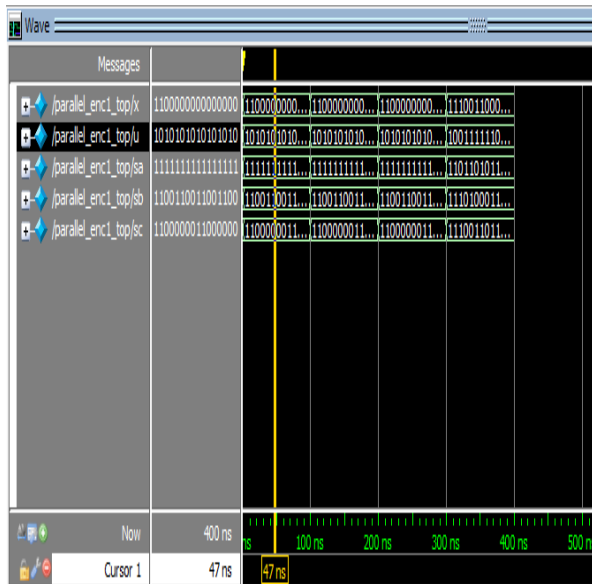


Fig 6. Fully parallel conventional encoder (16,k)

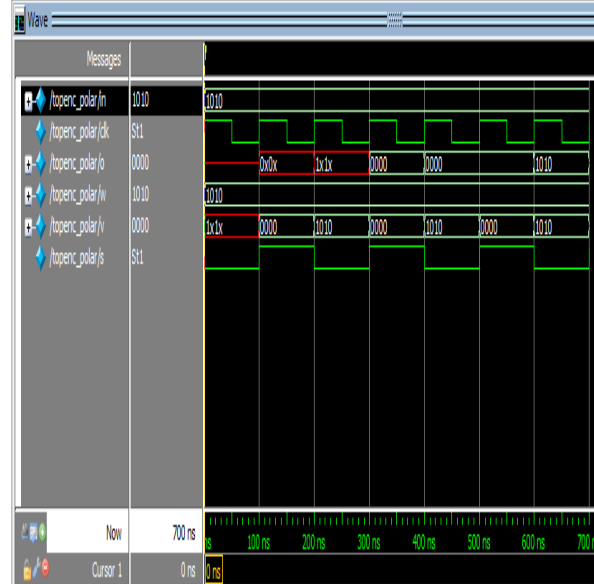


Fig 7. (16,k) folded polar code encoder

V. Conclusion

In this brief paper polar code encoder architecture is designed and implemented for (16,k) polar code and this folded architecture shows substantially promising result and in can be concluded that this architecture can be employed for large block length code and it is capable for saving large silicon area for hardware resource constraint applications specially can be use full in FPGA based design. Although the architecture save area and increase resource reusability but the trade off with speed is there is further scope of improvement in design in terms of latency of both encoder and SC decoder.

References

- [1] Le Gal, B., Leroux, C. and Jegou, C., 2017, October. Successive cancellation decoder for very long polar codes. In Signal Processing Systems (SiPS), 2017 IEEE International Workshop on (pp. 1-6). IEEE.
- [2] Yuan, B. and Parhi, K.K., 2017. LLR-based successive-cancellation list decoder for polar codes with multibit decision. IEEE Transactions on Circuits and Systems II: Express Briefs, 64(1), pp.21-25.
- [3] Arpure, A. and Gugulothu, S., 2016, April. FPGA implementation of polar code based encoder architecture. In Communication and Signal Processing (ICCSP), 2016 International Conference on (pp. 0691-0695). IEEE.
- [4] Vangala, H., Hong, Y. and Viterbo, E., 2016. Efficient algorithms for systematic polar encoding. IEEE communications letters, 20(1), pp.17-20.
- [5] Yoo, H. and Park, I.C., 2015. Partially parallel encoder architecture for long polar codes. IEEE Transactions on Circuits and Systems II: Express Briefs, 62(3), pp.306-310.
- [6] Ayinala, M., Brown, M. and Parhi, K.K., 2012. Pipelined parallel FFT architectures via folding transformation. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 20(6), pp.1068-1081.
- [7] Arikan, E., 2011. Systematic polar coding. IEEE communications letters, 15(8), pp.860-862.
- [8] Arikan, E., 2009. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Transactions on Information Theory, 55(7), pp.3051-3073.



A General parallel -Serial Architecture for Montgomery Algorithm using Vedic Mathematics (Montgomery modular multiplication)

¹Mangal Deep Gupta, ²Dr. Sunita Prasad
Centre For Development Of Advanced Computing (CDAC), Noida
INDIA

Abstract — Montgomery algorithm is modern technique for modular multiplication, rigorously used in applications pertaining to Elliptical curve cryptography (ECC) [6]. This paper tests parallel -serial design of Montgomery modular multiplication with 12 bit integer data. An effective parallel –serial architecture of Montgomery Algorithm using Vedic Mathematics is presented and its performance in terms of complexity is evaluated. The methodology discussed in the paper illustrates an architecture in which Vedic Multipliers are used to reducing the time complexity of Montgomery Modular Multiplications. In multiplication of two large integers ,computational time and required area both are inversely proportional. So in our proposed architecture using both parallel –serial multipliers partially. Finally, evaluate the proposed architecture on SPARTAN 6 FPGAs.

Index Terms- Vedic Multipliers, Montgomery modular multiplication, Elliptical curve cryptography (ECC).

I. Introduction

In cryptography role of multiplications is much important because for encrypt or de-encrypt of any data we have to apply multiplications with key numbers but after multiplication result bit length increase ,so for reducing data ,applied Modular reduction ,that is called Modular Multiplication. Modular multiplication of large bits is the primary operation in majority of cryptography such as Rivest–Shamir–Adleman (RSA) [1], [3] and elliptic curve cryptography (ECC) [6],[7]. Standard modular multiplication is one of the costliest operations because it requires division, which in itself is complex. The Montgomery modular multiplication method [4] is an effective technique in which complex division is replaced with the multiplication and binary shifting. Montgomery Algorithm makes uses of a pre-computation method based on modulus m to gain this advantage. Montgomery Algorithm can be implemented in following two distinctive ways, where modulus is denoted by 'm'.

- 1) The entire multiplier is multiplied with the multiplicand. At that point, the subsequent item is compress by utilizing the Pre-computation of modular inverse: $\epsilon = -m^{-1} \bmod 2^\delta$, where n represents an integer at least equal to the bit length of multiplier.
- 2) In second way for modular multiplication ,multiplication and reduction both are combined .the multiplier is multiplied with δ bits of the multiplicand in each iteration .in each iteration reduction operation performed parallel. In process of reduction, the Pre-computation of modular inverse: $\epsilon = -m^{-1} \bmod 2^\delta$ is used. In multiplication of two large integers ,computational time and required area both are inversely proportional for example in serial multiplication required area is less but computational time is more and in parallel multiplication required area is more but computational time is less. So in my proposed architecture using both parallel –serial multipliers partially.

In paper [5], carry save adders are used to implement the architecture of digit-serial Montgomery modular multiplier but in the proposed architecture to implement Montgomery algorithm, the use of Vedic Multipliers are illustrated , to reduce time complexity of Modular Multiplication. In this proposed architecture for Montgomery Modular Multiplication, 12 bit length of multiplicand, multiplier, modulus and $\delta=4$ are considered for Parallel multiplication.

In Montgomery multiplication, x , y , and m are the residual representations of few integers $\tilde{x}$ (multiplicand), $\tilde{y}$ (multiplier) and $\tilde{\eta}$ (modulus).

$$x = \tilde{x} 2^\delta \bmod m, \quad y = \tilde{y} 2^\delta \bmod m, \quad \eta = \tilde{\eta} 2^\delta \bmod m.$$

This representation is commonly known Montgomery domain. We observe that following equation is satisfied by Montgomery multiplication:

$$\eta = x y 2^{-\delta} \bmod m$$

$$\tilde{\eta}.2^{\delta} \bmod m = (\tilde{x} 2^{\delta} \bmod m)(\tilde{y} 2^{\delta} \bmod m) 2^{-\delta} \bmod m$$

$$\tilde{\eta} = \tilde{x} \tilde{y} \bmod m$$

As observed, the Montgomery form representation of the Montgomery modular multiplication $\tilde{\eta} = \tilde{x} \tilde{y} \bmod m$ is represented by $\eta = xy 2^{-\delta} \bmod m$.

The Montgomery modular multiplication $\eta = xy 2^{-\delta} \bmod m$ can be divided into two multiplication steps (As discussed earlier), $s = x * y$ and the reduction step $r = s 2^{-\delta} \bmod m$.

Without use of division, computation of Montgomery modular reduction $\eta = s 2^{-\delta} \bmod m$ can be follows:

$$\eta = s 2^{-\delta} \bmod m \equiv \tilde{\eta} = (s + \theta q) / 2^{\delta}$$

where $q = s \epsilon \bmod 2^{\delta}$ is a parameter depends on the pre-computation $\epsilon = -m^{-1} \bmod 2^{\delta}$.

II. Montgomery Algorithm

Figure 1: Proposed architecture of Montgomery Algorithm

Inputs: n bit integers x, y, m .
 output: $\eta = (xy2^{-n}) \bmod m$
 Pre-computation: $\epsilon = -m^{-1} \bmod 2^{\delta}$
 A.) $s[n:0] = 0$
 B.) for $i = 0$ to $\frac{n}{\delta} - 1$
 B1. $\alpha = y[\delta(i+1) - 1 : \delta i]$
 $v[n+\delta:0] = x * \alpha + s[n:0]$
 B2. $q = \epsilon * v[\delta - 1 : 0] \bmod 2^{\delta}$
 $s[n:0] = (m * q + v[n+\delta:0]) / 2^{\delta}$
 C.) if $n = s[n:0] - m$ else $n = s[n-1:0]$

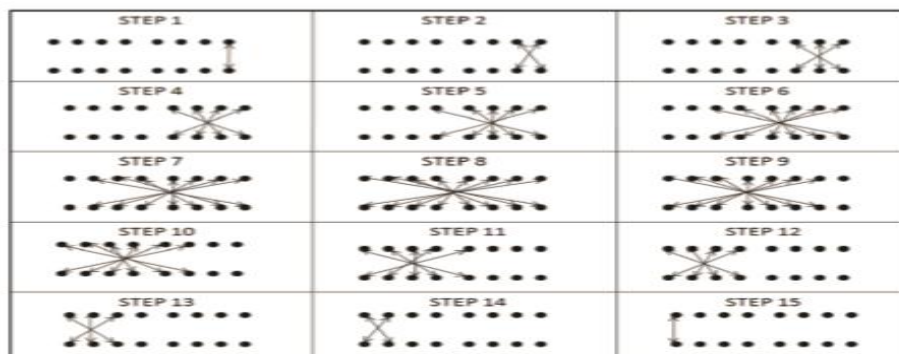
Fig 1. Show that block diagram of proposed architecture of Montgomery modular multiplication using Vedic mathematics. In this architecture different blocks are connected to each other and these block are called modules. In this architecture uses two accumulator for addition operation $(x * \alpha + s[n:0])$ and $(m * q + v[n+\delta:0])$. required three register for hold the data of s , modulus m and $(\epsilon = -m^{-1} \bmod 2^{\delta})$. In this algorithm used three different types of multiplier first one is $12 * 4$ bit for multiplication of $(x * \alpha)$, second is $12 * 4$ bit for multiplication of $(m * q)$ and last is $4 * 4$ bit for multiplication of $(\epsilon * v[\delta - 1 : 0])$. Hence final output from algorithm is in the Montgomery form. so we have to required convert result from Montgomery form to normal form.

III. VEDIC MATHEMATICS

Vedic science is amongst the most vital system of the twentieth century logical history. It has awesome instructive esteem in light of the fact that the sutras contain simple strategies for playing out some basic numerical activities like Multiplication in straightforward ways and results are acquired fastly. Subsequently Vedic Mathematics [8] is anything but difficult to learn, quicker to utilize and less at risk to misinterpretation than complex system.

The Urdhwa Tiryakbhyam Sutra is one of the largest sutras in age-old Vedic Maths. As per the definition, the meaning of Urdhwa Tiryakbhyam is "vertically crosswise". The suggested increase happens between extraordinary bits of the multiplier and multiplicand. The main utility of this methodology is that the product result is obtained in one step. This method is area and speed efficient.

Figure 2: Urdhwa Tiryakbhyam Sutra of Vedic Multiplier of 8*8 bit



For eg. Consider A and B, which are two numbers whose product is to be found and are 12 bit and 4 bit long respectively. This parameter is according to our proposed architecture of Multipliers. Suppose P is product term obtained when multiplying A and B. Subscripts under both alphabets A and B represent their respective bit places. Subscript denoted by '0' represents the LSB and the higher subscript is used to denote the MSB.

$$\begin{aligned}
 P_0 &= A_0 \& B_0 \\
 [1:0]P_1 &= (A_1 \& B_0) + (A_0 \& B_1), \\
 [2:0]P_2 &= (A_2 \& B_0) + (A_0 \& B_2) + (A_1 \& B_1) + P_1[1], \\
 [2:0]P_3 &= (A_3 \& B_0) + (A_0 \& B_3) + (A_2 \& B_1) + (A_1 \& B_2) + P_2[1], \\
 [2:0]P_4 &= (A_4 \& B_0) + (A_3 \& B_1) + (A_1 \& B_3) + (A_2 \& B_2) + P_2[2] + P_3[1], \\
 [2:0]P_5 &= (A_5 \& B_0) + (A_4 \& B_1) + (A_3 \& B_2) + (A_2 \& B_3) + P_3[2] + P_4[1], \\
 [2:0]P_6 &= (A_6 \& B_0) + (A_5 \& B_1) + (A_4 \& B_2) + (A_3 \& B_3) + P_4[2] + P_5[1], \\
 [2:0]P_7 &= (A_7 \& B_0) + (A_6 \& B_1) + (A_5 \& B_2) + (A_4 \& B_3) + P_5[2] + P_6[1], \\
 [2:0]P_8 &= (A_8 \& B_0) + (A_7 \& B_1) + (A_6 \& B_2) + (A_5 \& B_3) + P_6[2] + P_7[1], \\
 [2:0]P_9 &= (A_9 \& B_0) + (A_8 \& B_1) + (A_7 \& B_2) + (A_6 \& B_3) + P_7[2] + P_8[1], \\
 [2:0]P_{10} &= (A_{10} \& B_0) + (A_9 \& B_1) + (A_8 \& B_2) + (A_7 \& B_3) + P_8[2] + P_9[1], \\
 [2:0]P_{11} &= (A_{11} \& B_0) + (A_{10} \& B_1) + (A_9 \& B_2) + (A_8 \& B_3) + P_9[2] + P_{10}[1], \\
 [2:0]P_{12} &= (A_{11} \& B_1) + (A_{10} \& B_2) + (A_9 \& B_3) + P_{10}[2] + P_{11}[1], \\
 [2:0]P_{13} &= (A_{11} \& B_2) + (A_{10} \& B_3) + P_{11}[2] + P_{12}[1], \\
 [1:0]P_{14} &= (A_{11} \& B_3) + P_{12}[2] + P_{13}[1], \\
 [1:0]P_{15} &= P_{13}[2] + P_{14}[1];
 \end{aligned}$$

Where suffix with p represent the number of bit for hold the data of addition operation in each equations. for example [2:0]P₁₀ is a three bit data represented P₁₀. Final result of Multiplication using Urdhwa Tiryakbhyam Sutra of Vedic Multiplier is given by:

$$P = \{P_{15} + P_{14}[0] + P_{13}[0] + P_{11}[0] + P_{10}[0] + P_9[0] + P_8[0] + P_7[0] + P_6[2] + P_5[0] + P_4[0] + P_3[0] + P_2[0] + P_1[0] + P_0\}$$

IV. Area and time complexity of different implementations

Table 1: Area and time complexity of Vedic multiplier and array multiplier

	12_4 Multiplier using Vedic Mathematics:	12_4 Array Multiplier :
No. of 4 input LUTs	117	102
No. of occupied slices	62	52
Average fan-out	3.36	3.39
No. of IOBs	32	32
Adder/ Subtractor	58	68
Timing Constraint :Delay	23.708ns	38.838 ns

Table 2: Area and time complexity of Montgomery Modular Multiplication using Vedic Multiplier and Array Multiplier

	Montgomery Algorithm using Vedic Mathematics:	Montgomery Algorithm using Array Multiplier:
No. of 4 input LUTs	1219	1158
No. of 4 input LUTs	674	620
Average fan-out	3.61	3.61
No. of IOBs	49	49
Adder/ Subtractor	424	484
Register	5	5
Latches	27	27
Comparator	10	10
XORs	3	3
Timing Constraint :Delay	34.849 ns	50.292 ns
Area-Time Product	42.48093μs	58.238136 μs

V. PERFORMANCE COMPARISONS

Table 1 represent the comparison of time and area complexity parameters of two different multiplier that is Vedic multiplier and Array multiplier, Which are show in detail. and observed that Maximum path delay of Vedic multiplier is much better than Array multiplier.

Table 2 represent the comparison of time and area complexity parameters of Montgomery modular multiplication using two different approach .in Montgomery algorithm required different bit length multiplier so in proposed architecture used 12×4 bit Vedic multiplier and 4×4 bit Vedic multiplier .in first approach design Montgomery algorithm using Vedic multiplier and in second approach design Montgomery algorithm using Array multiplier. Hence in terms of time complexity first one approach is better than second approach. Finally compare the area ,time product which is also minimum in first approach.

Figure 1: Parallel-Serial Architecture of Montgomery Algorithm using Vedic Multipliers

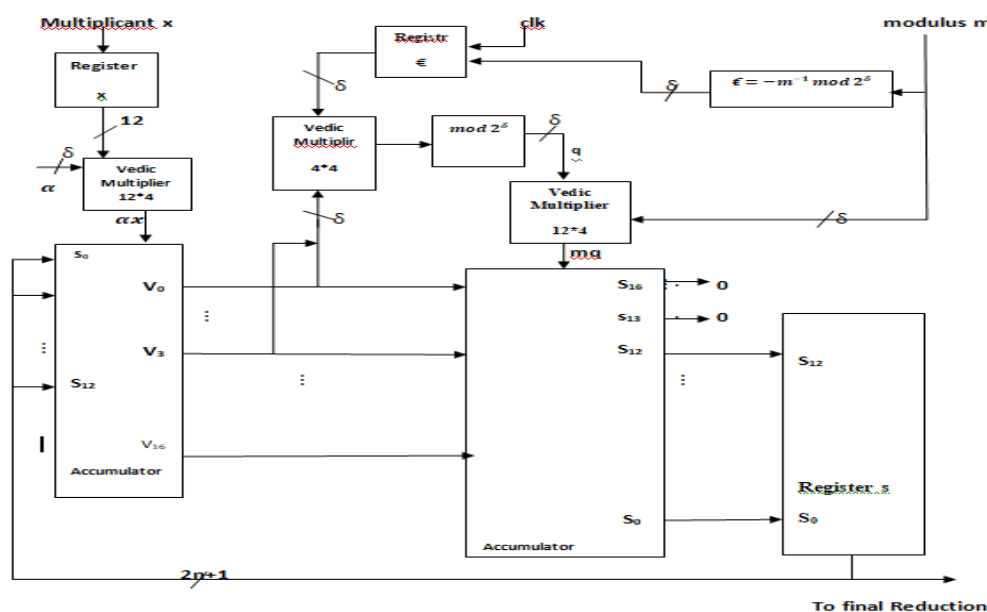
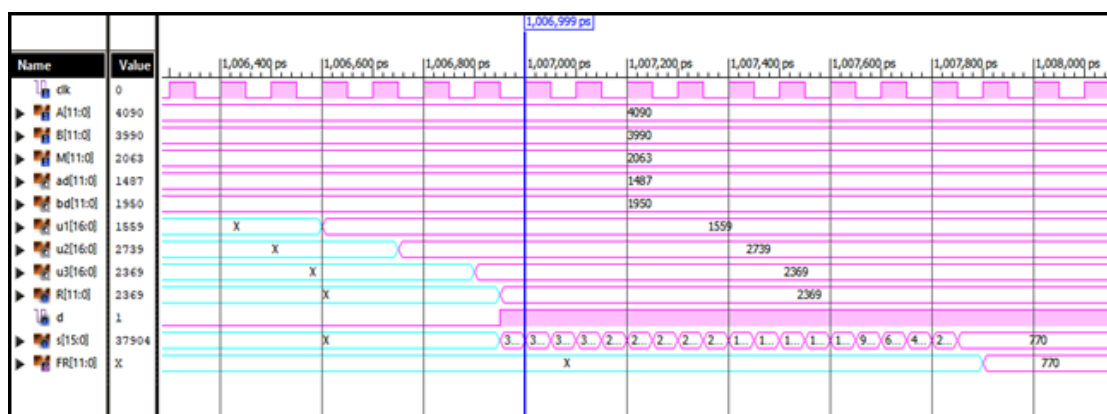


Figure 3: Simulation Result of Montgomery Modular Multiplication of 12 bit data from ISE Design Suite 14.7.



VI.CONCLUSION

This paper presents a application of Vedic Multipliers to implementation of Montgomery algorithm for Modular multiplication and also design of parallel- serial architecture of Montgomery Modular Multiplication . In this paper also design Montgomery Modular Multiplication using Array Multiplier for comparison with proposed architecture that is Montgomery Modular Multiplication using Vedic Multipliers .In this Paper ,design a more general serial parallel architecture for Montgomery Modular Multiplier ,to Implementation of any Modulus m and four bit digit size(δ).For reduction of critical Carry propagation path , using Vedic Multipliers . In this describe in detail how these Vedic Multipliers reduce the time complexity in parallel multiplication and

these Multipliers also reduce the carry dependency . The overall computational time required by the proposed Vedic multiplier is more faster than Array multipliers. that design benefits arrives at the mere cost of design area complexity . Proposed Vedic multiplier utilize to implementation of Montgomery Modular Multiplication, which gives too much improvement in time complexity. this proposed design also implemented on SPARTAN 6 FPGA. After implementation on FPGA, which getting the result that shows the minimum time and area product is obtain with digit size ($\delta = 4$) when compared with implementation of Montgomery Modular Multiplication using array multipliers.

VII. REFERENCES

- [1] Yang CC, Chang TS, Jen CW. A new RSA cryptosystem hardware design based on Montgomery's algorithm. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing. 1998 Jul;45(7):908-13.
- [2] Shieh MD, Lin WC. Word-based Montgomery modular multiplication algorithm for low-latency scalable architectures. IEEE transactions on computers. 2010 Aug;59(8):1145-51.
- [3] Miyamoto A, Homma N, Aoki T, Satoh A. Systematic design of RSA processors based on high-radix Montgomery multipliers. IEEE Transactions on very large scale integration (VLSI) Systems. 2011 Jul;19(7):1136-46.
- [4] Kuang SR, Wu KY, Lu RY. Low-cost high-performance VLSI architecture for Montgomery modular multiplication. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2016 Feb;24(2):434-43.
- [5] Erdem SS, Yanık T, Çelebi A. A General Digit-Serial Architecture for Montgomery Modular Multiplication. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2017 May;25(5):1658-68.
- [6] Koblitz N. Elliptic curve cryptosystems. Mathematics of computation. 1987;48(177):203-9.
- [7] Miller VS. Use of elliptic curves in cryptography. In Conference on the theory and application of cryptographic techniques 1985 Aug 18 (pp. 417-426). Springer, Berlin, Heidelberg.
- [8] Bansal Y, Madhu C, Kaur P. High speed vedic multiplier designs-A review. In Engineering and Computational Sciences (RAECS), 2014 Recent Advances in 2014 Mar 6 (pp. 1-6). IEEE.



Synthesis of Network on Chip Topologies for Area Efficiency

Ana Kumar¹, Dr. C. K. Jha²

¹Department of Electronics, Banasthali Vidyapith, Rajasthan, India

²Department of Computer Science, Banasthali Vidyapith, Rajasthan, India
Banasthali Vidyapith, Rajasthan
India (New Delhi)

Abstract: Execution of a SoC (system-on-chip) emphatically relies upon its hidden design and related routing strategies. An auxiliary adaptable interconnection design might altogether expand the execution of a NoC as far as versatility, particularity, transport latency and parallelism in communication [1]. With the constant increment of the quantity of on-chip segments, issues, like correspondence delay, framework throughput, power utilization and expansive pass on area begin to develop in conventional two-dimensional (2D) integrated circuits (ICs). Notwithstanding the higher adaptability and parallelism displayed by 2D-Network-on-Chip (NoC) it is still not an extreme answer for future vast scale Systems-on-Chip (SoCs). NoC reconciliation in three dimensions (3D-Network-on-Chip) is proposed as a potential arrangement offering higher pace, low latency, lower dynamic power consumption, and high parallelism propelled incorporation advances are making plausible the expansion of topology blend of on-chip systems from 2 dimensions to 3 dimensions [2]. Further, to design 3 dimensions NoC the topology is arranged utilizing vitality effective topology switches taking into account of physical circuit-exchanging as found in FPGAs. This paper will contain investigation of different network on chip topologies, synthesis of topologies, and analysis on the basis of area resources. The work will also contain evaluation for most area efficient NoC Topology.

Keywords: topologies; network on chip; system on chip; FPGA, 3-D; 2-D;

I. Introduction

In this era, every chip contains more than 100 million transistors. The gate length of transistors is measured in terms of nanometers. Along with this, Moore's Law states that the number of transistors on a chip doubles in every 18 months. Hence System-on-chip solves the problem of connecting to components on the printed circuit board, by integrating them into a single chip. It enables the designer to put system-on-chip that moves everything from the board onto chip eventually. System on the chip is a high- performance microprocessor, which can be programmed and provide instruction to μp to do the task. SoC is also an effort to integrate different types of silicon IP on the same chip, like memory, μp , random logic and analog circuitry. The SoC includes an embedded processor, ASIC logic, analog circuitry, and embedded memory. The System on Chip software includes an Operating system, compiler, simulator, firmware, driver, protocol stack integrated development environment (debugger, linker) and an application interface (C++/C, assembly)[3].

Network on Chip(NoC) topology can be viewed as a great tool for improvising the overall performance of an Integrated Circuit (IC). NoC outlines guarantee to convey quick, reliable, energy saving communication between on-chip segments. Packet switched network is better for NoC for application traffic [4]. The NoC design idea is acquired from layered based PC network protocol and packet switched network which comprises of on-chip switches to exchange the packets, links to impart with the neighboring router and Network Interfaces (NIs) which act as a correspondence medium amongst router and handling components on a definite topology [5]. NoC offers high adaptability and the consistency of a network structure, supporting less difficult interconnect models and more prominent adaptation to non-critical failure. The accepted interconnect spine of the network joined with network protocols to upgrade the adaptability of such systems. NoC platforms are differentiated from each other with the routing algorithm used and topology. These two features affect the computation capability of NoC [2]. NoC overcomes drawbacks of System on Chip by replacing dedicated signal wires.

II. Performance Criterion

When a network is formed between, these are the following criteria from which the overall performance is measured for a topology.

- A. **Degree:** Number of edges connected to the node in a topology is termed as node degree of a topology. It plays an important role in network cost. Topology also varies from the location of a node in a network.

Like, for a 2D mesh, nodes situated in the corner have degree 3 and intermediate tiles are connected with 4 adjacent tiles and a core [6].

- B. Diameter:** It is the length of the maximum shortest path between any two nodes in hopes of a network. For some topologies, the diameter of the network also depends upon the number of nodes. The smaller the diameter, the less time will be taken to send a message from node to node [7].
- C. Average Distance:** The average distance between any two nodes. Like, for uniform dimension mesh it is given by $D \times \sqrt{n} / 3$. For torus, it will be $n/2$ [7].
- D. Network Cost:** The product diameter X node degree is good for measuring the relationship between cost and performance of a multiprocessor system. Inter-node distance, message traffic density, and fault tolerant depend on diameter and degree [8].
- E. Bisection width:** It is the minimum number of the link which can be removed to divide the network into two halves.

III. Basic Network on Chip Topologies

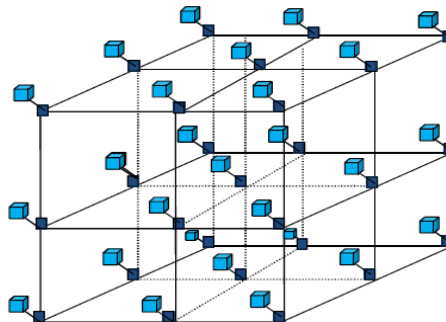
The evaluation is done on the basis of following basic network on chip topologies, listed below:

A. 3-D Mesh Architecture

Mesh architecture is one of the simplest architecture of NoC design. In mesh architecture, the connection is present between one node and four neighborhood node. These nodes are connected for transfer of data. Packet transfer is done by various routing algorithms. Dimension order routing is generally used for routing in 3-D Mesh. Throughput, latency, fault tolerance and link utilization help in determining which topology is more appropriate. Every node is linked to 2N of its nearby nodes in N-Dimensional Mesh as presented in the figure. Altering the size of the network doesn't alter connections per node. But the increase in diameter corrupts the performance.[9] Here are some fundamental assessment parameters, when there are n-nodes:

- Degree: 6
- Diameter: $N((n^{1/N})-1)$
- Network Cost: $6 N((n^{1/N})-1)$

Figure 1 3-D Mesh Topological Structure

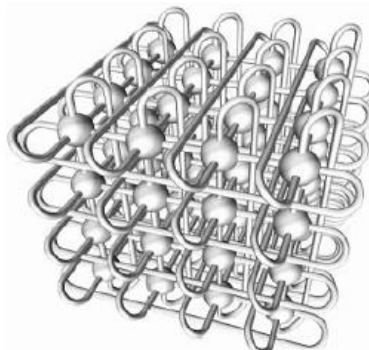


B. 3-D Torus Architecture

It is like an advanced and modified version of Mesh. This contains an additional set of wrap around edges covering the terminal nodes by connecting them. This introduced wrap around edge not only reduces the path while traversing but also reduces overall cost and performance of a network [10][11]. Some basic properties of this architecture are as follows:

- Degree: 6
- Diameter: $(N/2)(n^{1/N})$
- Network cost: $6 (N/2)(n^{1/N})$

Figure 2 Interconnection scheme in a 3-D Torus



C. Hypertorus Architecture

This topology is designed using Hypertorus as a basic module. It uses Cartesian product of Hypercube and Hypertorus. This design has a constant degree of 4 which proves it to be beneficial when it is compared to other topologies[12]. Here, is a list of essential topological parameters for this topology:

- Degree: 4
- Diameter: $0.35 \sqrt{N} + 4$
- Network cost: $1.4\sqrt{N} + 16$

IV. Synthesis Result Comparison and Area Efficiency

To retrieve novel features, a number of experiments on different algorithms with different parameters are performed to determine the topology of best network performance. Synthesis is a process in which a designed behaviour that is modelled using an HDL is translated into an implementation consisting of logic gates. The experimental work includes design and synthesis of the network on chip topologies on Xilinx ISE Design Suite 14.7, the device is chosen for synthesis is Spartan 3E, and hence the area efficiency is calculated. The implementation is done in Verilog. This area efficiency is calculated in terms of number of slices, number of slice flip-flops, number of 4I/P LUTs, number of bonded IOBs, and number of GCLKs. The implementation is done for $N=3$ and 64 number of nodes. Here, are the results obtained after synthesis of various basic Network on Chip Topologies:

Figure 3 Technology Mapping Diagram of 3-D Mesh

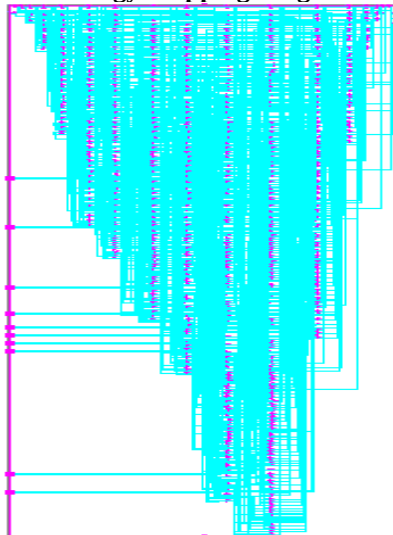


Figure 4 Technology Mapping Diagram of Torus Topology

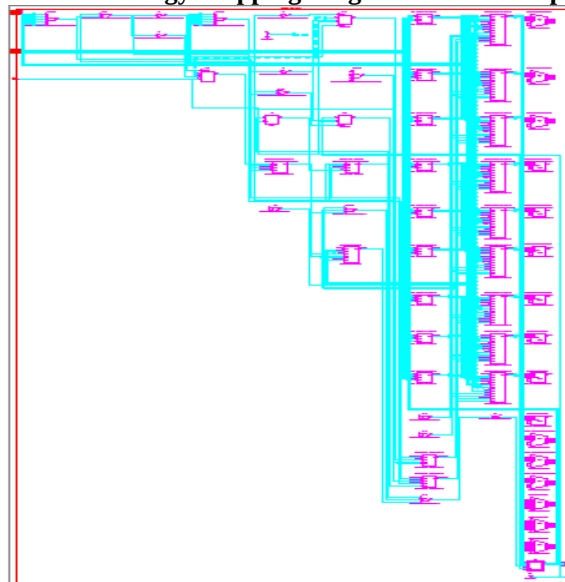


Figure 5 Technology Mapping Diagram of Hyper-torus Topology



Table I Shows the Design Summary Comparison of various NoC Topologies (Spartan 3E)

Parameter for Device Utilization Summary	3-D Mesh Topology	3-D Torus Topology	Hyper-Torus Topology
Number of Slices(out of 960)	260	40	25
Number of Slice Flip Flops(out of 1920)	17	13	44
Number of 4 I/P LUTs(out of 1920)	487	79	36
Number of Bonded IOBs(out of 66)	69	30	301
Number of GCLKs(out of 24)	1	1	1

Spartan 3E is one of the most efficient FPGA board, but here for basic NoC topologies, as represented in Table I, the area utilized by the topologies are exceeding than the area available, i.e. number of bonded IOBs for 3D Mesh topology is exceeding the available number of bonded IOBs with a utilization summary of 104%. Similarly, for Hyper-Torus topology numbers of Bonded IOBs required are exceeding the approachable Bonded IOBs, here the device utilization summary is of 456%. So, due to this demand of area utilization parameters, the work has been implemented for Spartan 3A DSP.

Table II Shows the Design Summary Comparison of various NoC Topologies (Spartan 3A DSP)

The parameter for Device Utilization Summary	3-D Mesh Topology	3-D Torus Topology	Hyper-Torus Topology
Number of Slices (out of 16640)	270	40	25
Number of Slice Flip Flops(out of 33280)	17	13	44
Number of 4 I/P LUTs(out of 33280)	507	79	36
Number of Bonded IOBs(out of 309)	69	30	301
Number of GCLKs(out of 24)	1	1	1

Spartan 3A DSP is an extended and enhanced version of Spartan 3A family. It is suitable and preferred for high volume applications. When the implementation is done on this Spartan device it has been found that the exceeding problem of device utilization for the topologies has been resolved as it provides more area and better performance. This is a graphical representation of a comparison of utilized area parameters for NoC topologies.

Figure 6 Graph representing comparison in device utilization parameters for NoC Topologies

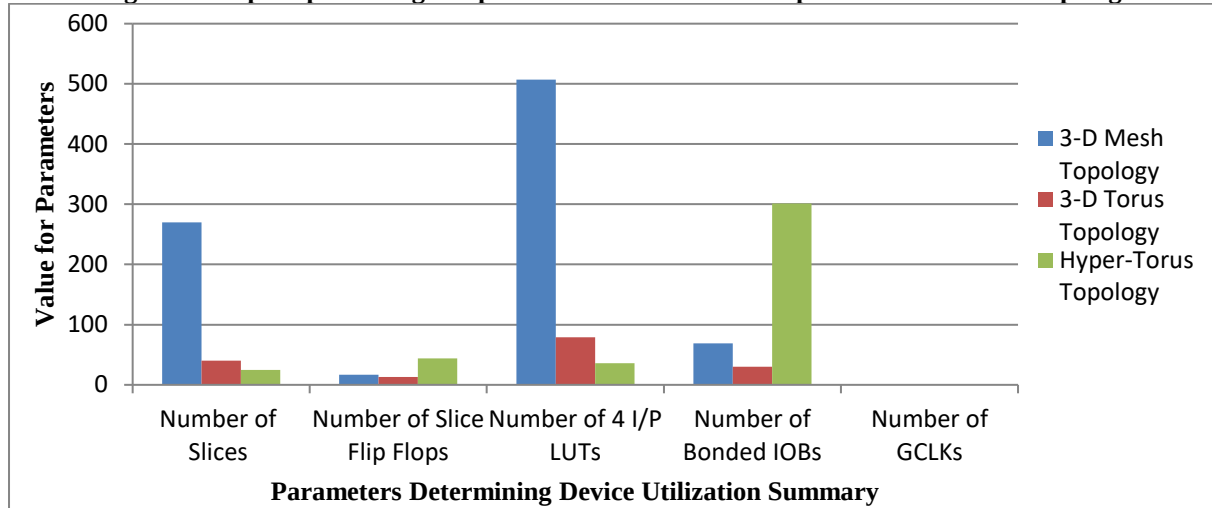


Table III Shows the comparison of Performance Metrics of Network on Chip Topologies

Topology	Degree	Diameter	Network Cost
3-D Mesh	6	9	54
Torus	6	6	36
Hyper-Torus	4	7	28

V. Conclusion

The network cost of Hypertorus is lesser than the 3-D Mesh and Torus, making it best topology only when Network Cost is taken into consideration as represented in Table III.

Area efficiency is evaluated from a number of occupied slices and numbers of 4 I/P LUTs. From the graph, the comparison of NoC Topologies is being depicted. From the Table II and from the graph of Figure 6, it has been represented that 3-D Mesh consumes a maximum number of slices, and number of 4 I/P LUTs. The number of bonded IOBs for 3-D Mesh is more than that of Torus but lesser than Hypertorus. On, the other hand when Torus and HyperTorus are compared, it has been found that Torus utilizes a number of 4 I/P LUTs and more number of Flip Flops than Hyper-torus. But Hypertorus consumes a maximum number of bonded IOBs than Torus and 3-D Mesh. So, when the requirement of implementing combinational logic and sequential logic is more than the number of pins available Torus is preferred. And when number of pins are not an issue, Hypertorus is preferred. So, it has been evaluated that area efficiency of Hypertorus proves to be better than Torus and 3-D Mesh network on chip topologies.

VI. References

- [1] Prasad, N., Chattopadhyay, S., & Chakrabarti, "ZMesh: An Energy-Efficient Network-on-Chip Topology for Constant-Geometry Algorithms." In Nanoelectronic and Information Systems (iNIS), 2015 IEEE International Symposium IEEE, Dec 2015, pp. 146-151.
- [2] Kumar, M. Pawan, et al. "A method for integrating network-on-chip topologies with 3D ICs." VLSI (ISVLSI), 2011 IEEE Computer Society Annual Symposium on. IEEE, 2011 pp.60-65
- [3] Kumar, A., Tyagi, S., & Jha, C. K. (2017). Performance analysis of network-on-chip topologies. Journal of Information and Optimization Sciences, vol. 38 issue6, pp. 989-997.
- [4] Kumar, A., Tyagi, S., & Jha, C. K. "Operational Evaluation of Network on Chip Topologies" IAETSD Journal for Advanced Research in Applied Sciences 5,1 January 2018 pp. 16-22
- [5] Bhople, Gaikwad "A Comparative Study of Different Topologies for Network on Chip Architecture", International Journal of Computer Applications, in Recent Trends in Engineering Technology, Mar 2013, pp.1-3
- [6] M.M. Hafizur and Xiaohong Jiang et. al "Network Performance of Pruned Hierarchical Torus Network" Sixth IFIP International Conference on Network and Parallel Computing, IEEE, Oct 2009, pp. 9-15
- [7] Woo-see Ki and Hyeong-Ok Lee et. al "The new Torus Design Based on 3-Dimensional Hypercube" In Advanced Communication Technology, Feb 2009, Vol.1, pp. 615-620
- [8] Faiz Faisal and M.M. Hazifur "Symmetric Tori Connected Torus Network" In Computers and Information Technology ICCIT'09 12th International Conference on IEEE, Dec 2009, pp. 174-179
- [9] Alakesh Kalita and Kaushik Ray et. al. "A topology for Network on Chip" In Information Communication and Embedded Systems (ICICES) International Conference IEEE, Feb 2016, pp. 1-7
- [10] N. Golpalakrishna and M. Sathish et. al. "A Torus Embedded Hypercube Scalable Interconnection Network for Parallel Architecture" In IEEE International Advance Computing Conference (IACC), Mar 2009, pp. 858-861
- [11] M.M. Hazifur and Susumu Horiguchi "High Performance Hierarchical Torus Network Under Matrix Transpose Traffic Patterns " In Parallel Architectures, Algorithms and Networks, 2004. Proceedings 7th International Symposium on IEEE, May 2004, pp. 111-116
- [12] Mohammad Khan and Abdul Ansari " A Review of Hypertorus Based Topologies for Network on Chip" Proceedings of International Conference Emerging Trends in Computing (ICETC-2011), March 2011, pp. 116-121



Design of a Low Power Pulse Triggered D flip flop using Modified Signal Feed Through Scheme

Asiyah Zainab Naqvi
School of Electronics
CDAC, Noida, India

Abstract: – In this work a low power pulse triggered D flip flop is presented. It can be called a hybrid of two recently proposed pulse triggered D-flip flop, based on signal feed through scheme, utilizing the features of both. Firstly, a comparative analysis on the existing flip flop designs is done. It is followed by the implementation of the above mentioned idea and a brief simulation analysis of the same. The comparative analysis showed that the new pulse triggered design consumed almost 84% less power than the conventional design. Along with that, when various D FF architectures were compared, it was found that the Conditional Discharge Flip Flop consumed maximum amount of power. When the first SFT scheme was introduced it consumed 14.23 % less power, as compared to CD FF and with the modifications of transmission gate introduced in the scheme and new pulse generator even more power saving of about 54% was observed. Then the proposed design was implemented which showed an improvement of 19.6% in power reduction, by utilizing one of the virtues of the original SFT design in the modified design. All the simulations are performed on CMOS 180nm technology, using Pyxis Schematic of Mentor Graphics.

Keywords: Pulse Triggered Flip Flop, Signal Feed Through, Explicit Pulse Triggered Flip Flop

I. Introduction

Power consumption is an important concern in digital systems design. And flip flops, being a basic building block for various functional units, are always a focus for modifications. Several attempts are constantly made to upgrade this elemental unit, so that the improvements can be reflected on the higher modules using them. Flip Flops are classified into various categories. Two of them are Master-Slave Class of Flip Flops and Pulse triggered Flip Flops. In master-slave class there are two stages, a master stage (negative latch), followed by a slave stage (positive latch). On the low segment of the clock the master is transparent, and the data is delivered to the output of master stage, and slave is on hold during this period. Then on the rising edge the data is no more sampled by the master and the slave now accepts the output of the master and provides it to the output. On the contrary a Pulse Triggered Flip Flop is single latch architecture. It is basically a latch which is triggered by a short pulse, so that it seems to operate like an edge triggered flip flop. Due to this major difference of being a single latch design, it seems to have a significant reduction in power consumption. But there needs to be a lot of accuracy in the pulse generator, to obtain the suitable pulse width for proper operation and optimum performance [2]-[4]. Pulse Triggered Flip Flops are also classified into two groups. One is implicit type P-FF in which the pulse generator is a part of the latch itself, and the other one is explicit type P-FF, in which the pulse generator and the latch are separate. In this paper the focus has been put on different designs of explicit type PT-FFs. The paper is organized as follows: A brief description of previously implemented PT-FFs designs is discussed in section II. Section III, introduces an idea which came out of the two schemes, discussed in section II. It can be called a hybrid of those two methodologies, utilizing some design techniques of both of them. The simulation results are compiled and compared in this Section IV. Analysis proceeded in the following way. Firstly, the designs were separately implemented, providing the tools' inbuilt clock feature, to specifically test the working and compare the power consumption of the latch stages only. Secondly, the two pulse generators were separately designed and their power consumptions were compared. Then finally, the complete design of the explicit PT-FFs, including both pulse generators, and data storage latch were combined and their performance was analyzed. Section V concludes the paper.

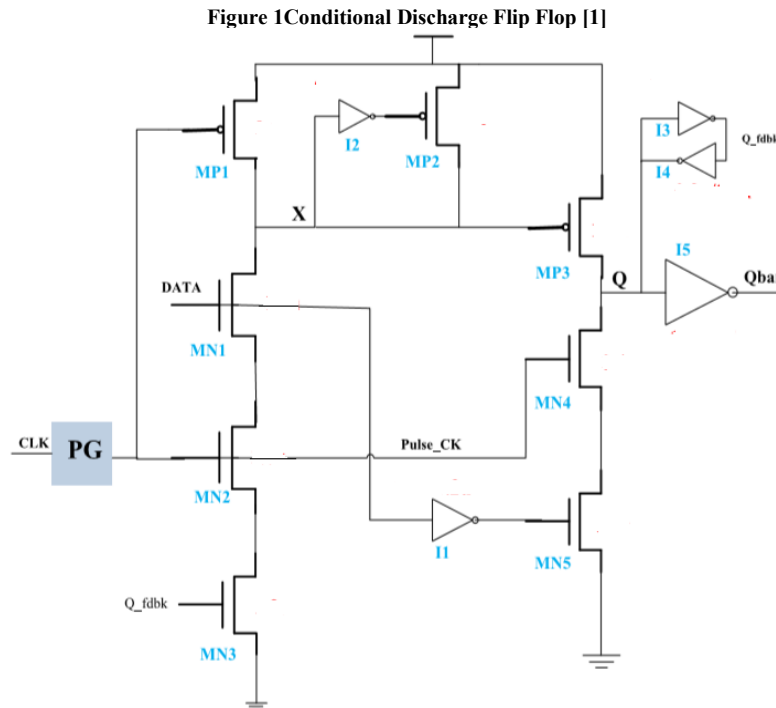
II. Reviewing Existing Pulse Triggered D Flip Flops

There are three designs which are implemented and analyzed. (A) Conditional Discharge PT-FF. (B) Signal Feed-Through Scheme PT-FF (C) Signal Feed-through scheme with a modified input stage and pulse trigger section.

A. Conditional Discharge Flip Flop [1]

This design was introduced as a remedial measure to eliminate the undesirable discharging problem on every positive edge of the clock, even if the input was at static “1”, in a classic explicit PTFF design, known as ep-DCO.

In this architecture a conditional discharge path is provided to ensure that unless the output value changes, the node X is not going to be discharged [1]-[4]. As can be seen in Fig. 1, the design consists of total 17 transistors in the latch stage only, and the pulse generator used here is the conventional NAND based pulse generator circuit.



Describing the operation of the circuit below briefly, the data is sampled when the pulse triggers, the NMOS MN2.

Assuming Qfdbk was initially high, as a bit “1” appears at the data all the transistors MN1 to MN3 turn ON and discharge the node x to ground. This way MP3 raises the Q to a high value. The advantage is that as long as input bit is same, the output is retained without any glitch during pulses, by the addition of MN3. The trade off is that the discharge path of MN4 and MN5 consume more time. This limitation was resolved by another approach discussed below.

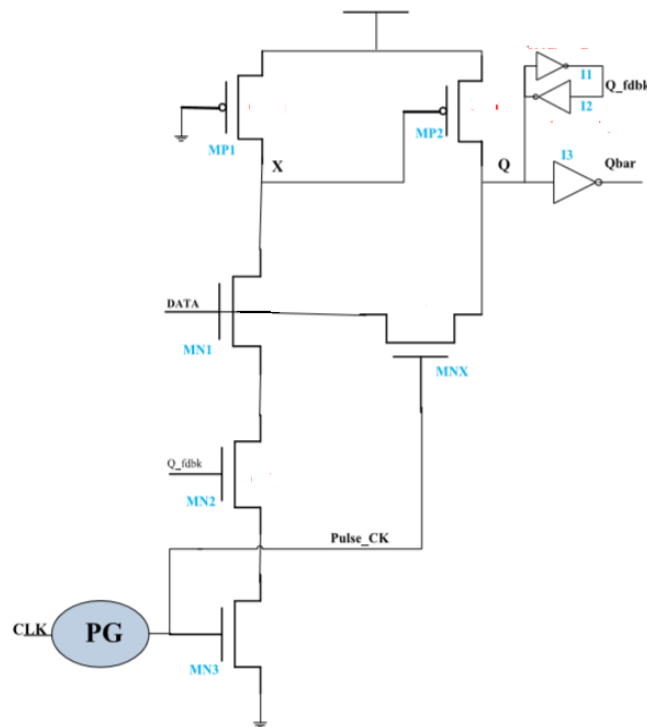
B. Signal Feed Through Scheme [2]

It controls the output directly from the input using a pass transistor MNx as shown in the Fig. 2. In this architecture the transistor MP1, has been put on a weak pull up state, by connecting its gate to the ground permanently, giving rise to a pseudo-nMOS logic design. This way the charge keeper circuit for the node X has been removed [2].

The “0” to “1” charge process remains the same, as mentioned above in Conditional Discharge. But the additional benefit of this signal feed through is the boost which the pass transistor provides in transferring this bit “1” from data to Q. This way a comparatively fast transition is experienced in this design. Here the pass transistor plays a dual role and when “1” to “0” transition occurs, MNx is likewise turned on by the pulse trigger and node Q is discharged by the input stage through this route [2].

Fig 2 shows the circuit for the signal feed through scheme.

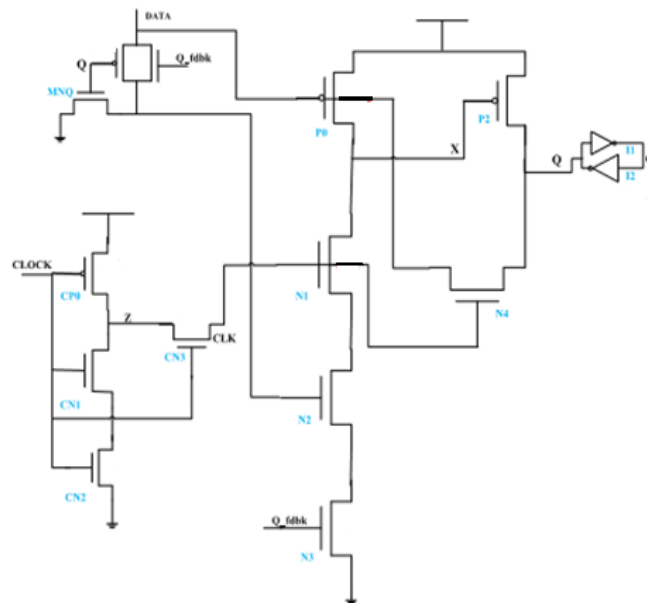
Figure 2 Signal Feed Through Scheme Flip Flop [2]



C. Modified Signal Feed Through Scheme [3]

The few limitations of the above described signal feed through scheme were aimed to be minimized and so this design was recently proposed. The new modifications in the original SFT Flip Flop shown in Fig 3, are as follows:

Figure 3. ModifiedSignal Feed Through Scheme Flip Flop [3]



- (i) A whole new pulse trigger generator was proposed to save power and area both.
- (ii) In original signal feed through scheme the first PMOS MP1 was in a weak pull up mode, as it was connected to the ground. It was permanently ON and consumed some power. This limitation was eliminated by connecting the PMOS P0 to the data input.
- (iii) Transmission gate was introduced in this modified SFT, which helped in reducing the leakage power[5].
- (iv) Another transistor (MNQ) was introduced which provided a secondary discharging path.

The new pulse generator worked on the principle of transistor sizing. The transistor CN3 is designed with larger size, than the discharge path. So, when the clock appears, the CN3 conducts faster and after a short period, the discharge path is activated to ground the signal. Within this delay a short pulse is obtained at node Z and this way the new pulse generator works, with only 4 transistors.

The additional discharge path through MN_q, reduces the delay, when 1 to 0 transition occurs at the output Q. Reduction in the number of transistors is evident with this newly proposed PG.

III. Proposed Design

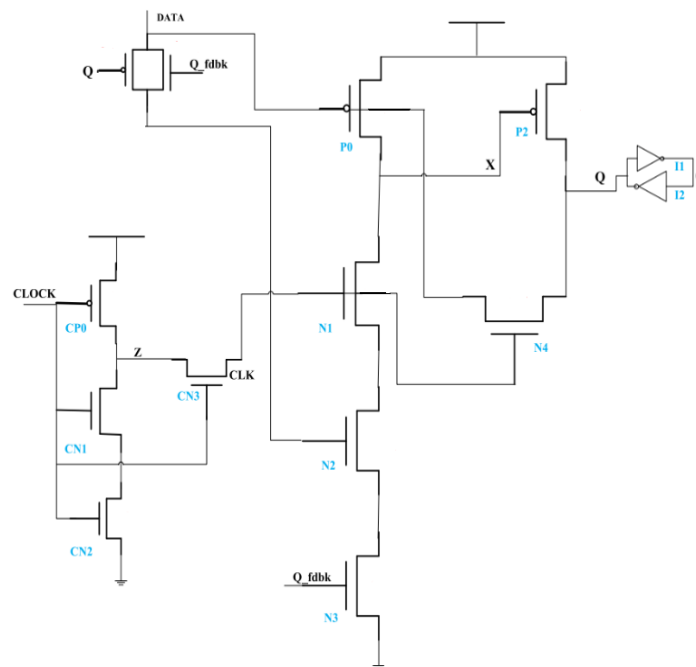
The circuit described in Fig. 3 was tested and proved to be a very power efficient D flip flop. The modification of transmission gate at the input, and the data input, controlling the gate of PMOS P0, proved to be the major reasons behind the significant power saving.

The transmission gate is retained from the previous design. The purpose of TG in the input stage, is to reduce the leakage power in the standby mode. When the nodes X and Q were switching, the transmission gate stage turns off the transistor N2, and so this way leakage power is reduced.

When 0 appears at the data, the transistor P0 is pulled up and X is high. Therefore, transistor P1 is off and Q gives a low output. As long as there is no data change, the discharge path of the input stage is disconnected keeping N2 off [1]. When a “1” appears at the data input the three transistors N1, N2 and N3 are activated. As the pulse triggers the latch region, the node X is immediately discharged to 0, making P1, pull up the output Q. Also, the pass transistor N4, passes the high value to Q.

Fig. 4 below shows the proposed design.

Figure 4. Proposed Design



The idea of modification came from the fact that the original signal feed through had no secondary discharge path, yet managed to make the transition with the trigger controlled pass transistor MN_x only [2][6]. If we remove the transistor MN_Q only from the modified design above and observe the working and operation, the Q would still follow the data input. No doubt the expected limitation, which is the disparity between the “0 to 1” and “1 to 0” propagation delay would be present. But if only the power consumption is a major concern, then this elimination of transistor MN_Q would save even more power.

IV. Simulation Results

The analysis of the above designs proceeded in the following way:

Firstly, the latch stage was designed, excluding its pulse trigger sections, to test the performance of the latch part only. During that analysis, the tools’ inbuilt provision to force the clock was used, to provide a uniform pulse train of 100 Pico second pulse width. Then the pulse triggers were separately designed and tested, and their performance comparison was specifically done. After that both were joined and the complete designs were

implemented and compared. All the designs were implemented and simulated on Mentor Graphics' Pyxis Schematic using 180nm technology.

A. Comparative Analysis of Pulse Generators exclusively

The two pulse generators were separately designed and their power consumption and quality of waveforms were analyzed. The results are shown below.

Figure 5. Implementation of the two Pulse Generators

(a) Conventional NAND based PG

(b) New PG Circuit

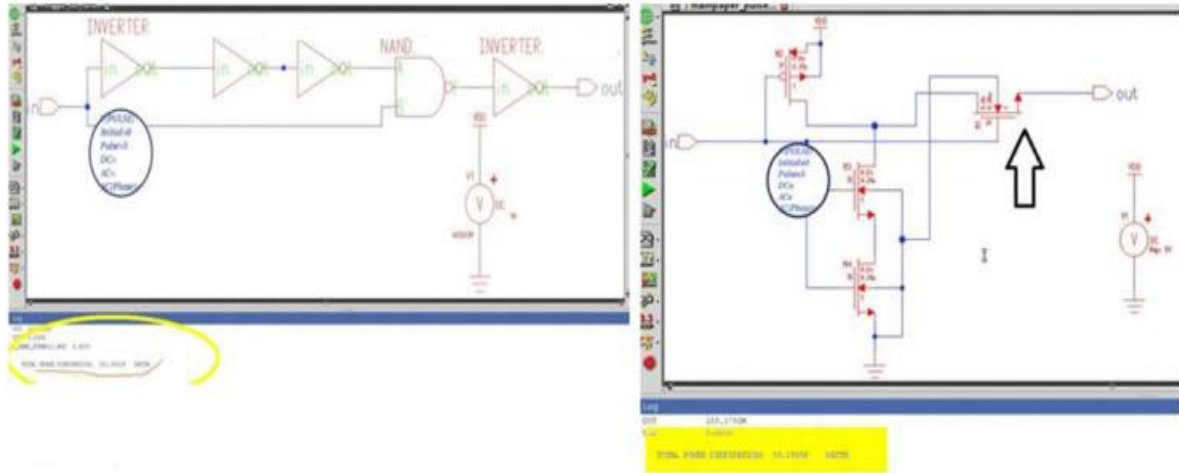


Figure 6. Waveform of Conventional PG

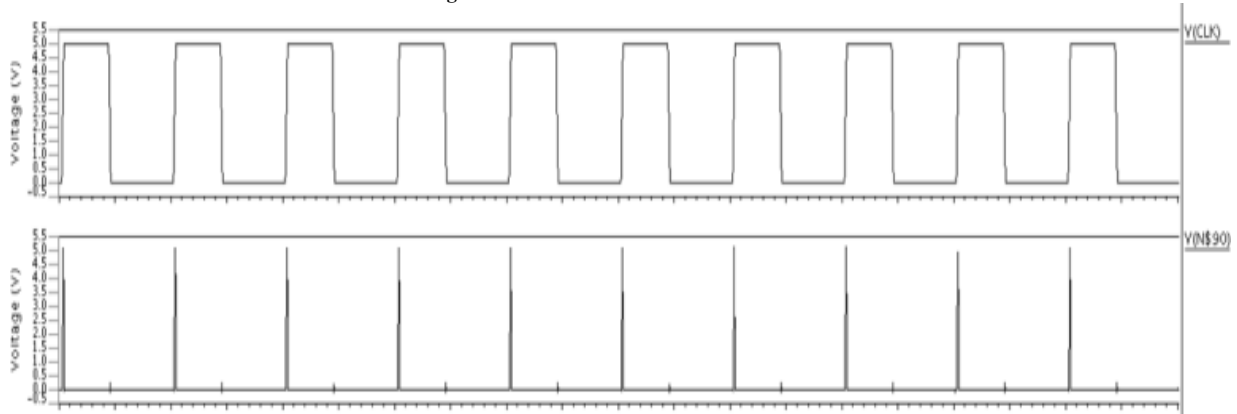


Figure 7. Waveform of New Pulse Generator

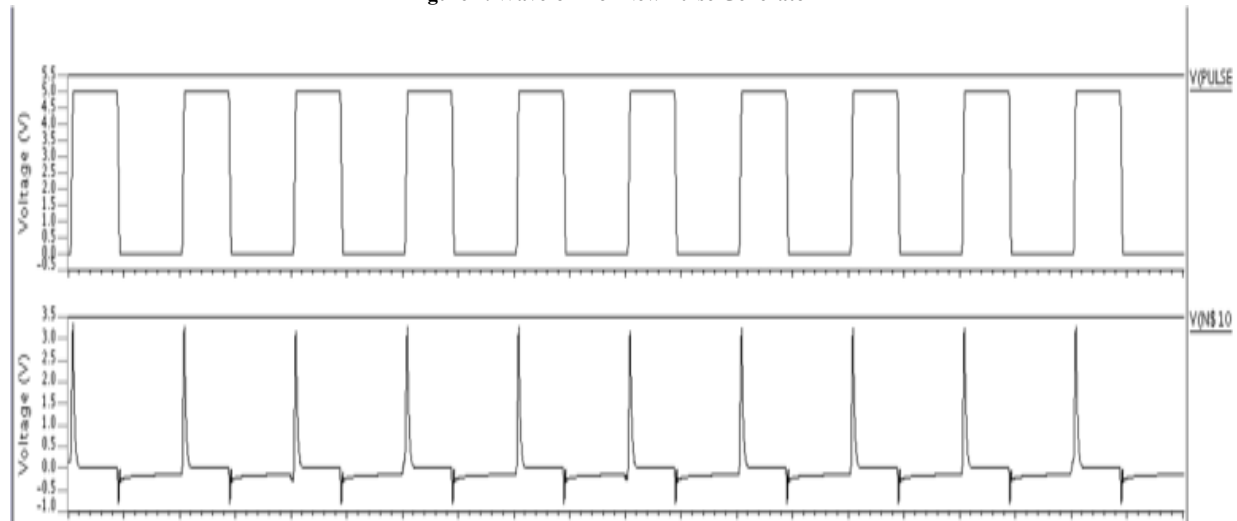


Fig. 5 (a) and (b) show the circuit implementations of the two Pulse Generators. Fig 6 shows the output results for the conventional NAND based pulse generator circuit. It gives a good quality of pulse train, with stability and less glitches. But the power consumption was enormous as compared to the new design.

The tabular representation shows a quick comparison.

Table I. Comparison between two Pulse Generators

Type of Pulse Generator	No. of Transistors	Power Consumption (Pico Watts)
Conventional PG	12	351.955
Proposed PG	4	56.159

B. Comparative Analysis of the Latch Stage exclusively

The Table II below shows a quick comparison of the power consumed by all the flip flops mentioned in the preceding section.

Table II. Comparison between two Pulse Generators

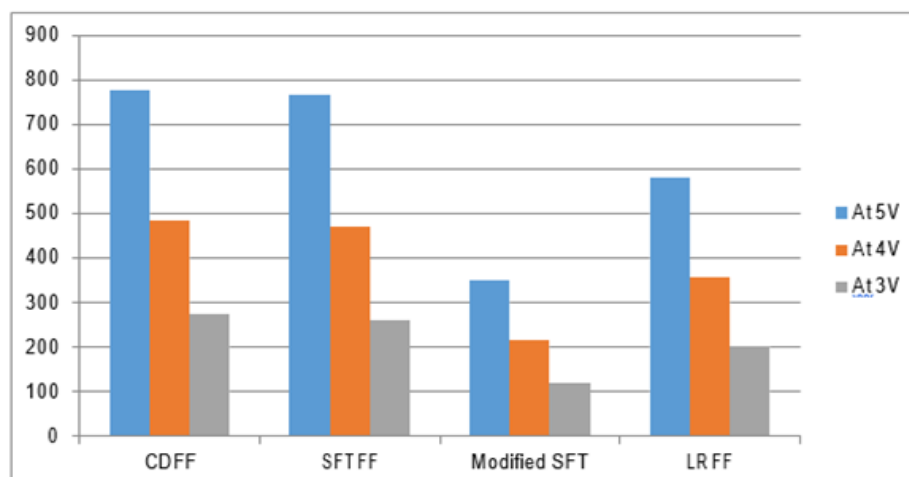
Latch stages without pulse generator	Power Consumption (pW)
Master Slave based Logic Reduction Flip Flop	579.29
CD FF	443.557
SFT	422.665
MODIFIED SFT	292.326
PROPOSED SFT	221.89

The latch stage was separately designed and their performance parameters were compared. Here the feature which is provided in the Pyxis Schematic Software was used, to provide a uniform pulse train of 100 picosecond. This way the working of only the latch stage was observed, without the designed pulse generator. The Table II shows some of the results for the same. These results are also compared with a master slave configuration D Flip Flop circuit which is based on LOGIC STRUCTURE REDUCTION SCHEME [5]. It was recently proposed and had been proved to be a power efficient master slave type D flip flop.

C. Comparative Analysis of the complete PT FF Designs including Pulse Generator and Latch Stage both.

Now, the designed pulse generators, and the latch stages are combined to analyze the working of complete PULSE TRIGGER FLIP FLOP architectures. Their working was tested on different data input sets.

Graph 1: Indicating Power Analysis of all the Flip Flops at different Voltages



The below Table III summarizes the comparisons of some of the performance parameters.

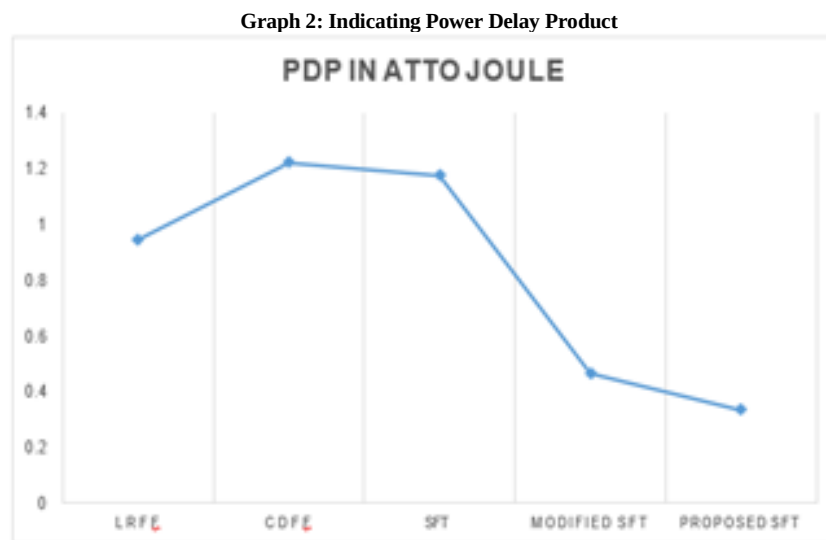
As can be seen in the comparison the Table III, a lot significant reduction in power as well as in number of transistors can be obtained using this new PG.

Table III. Comparisons of various performance parameters of Flip Flop Designs

Type of Flip Flop	No. of Transistors	Average Power Consumption (PicoWatts)	Average Tpd (Nsec)	PDP (atto joule)
LRFF (master slave)	19	579.29	1.6314	0.956
CD FF	30	776.6251	1.572	1.22
SFT	24	764.567	1.535	1.174
Modified SFT	17	351.243	1.321	0.464
Proposed SFT	16	282.358	1.178	0.332

The CD FF and SFT both used the conventional pulse generator. But the modified SFT and the idea which is proposed to further modify it, used the new design of Pulse Generator.

The graph 2 indicates the PDP variation in all the D flip flop designs as shown below:



As can be seen above in the graph 2, the proposed SFT seems to have the best PDP due to remarkable power consumption savings and the conditional discharge flip flop has the maximum value. Generally, master slavebased designs are said to consume more power than PT FF, but since this design is also a recently proposed power efficient one, it seems to outweigh two of the PT FF designs, due to their high power trigger sections.

V. Conclusion

The above results show how different types of pulse triggered flip flops were designed. Among the two pulse generator designs, the conventional one consumed more power, but seemed to have a good quality of pulses, which can also be observed in the Fig. 6. While the recently proposed PG showed significant amount of power saving, yet some glitches were observed in the output, which did not hamper the workingof the module at this stage. But the effect of the same in the higher modules needs to be analyzed.

The addition of transmission gate and the control given to the data input to turn ON the first PMOS at the input stage, in the original SFT design proved to be the major reason for power saving. The expected limitation of the proposed design is the disparity in the 0 to 1 and 1 to 0 transition timings.

Generally master slave based designs are said to consume more power than PT FF, but since this design is also a recently proposed power efficient one, it seems to outweigh two of the PT FF designs, due to their high power trigger sections.

The scope for future work lies in further exploring the proposed design whether the removal of the alternate discharge path would hamper the operation of D Flip Flop in any other way or not.

VI. References

- [1] Peiyi Zhao, Tarek K. Darwish, Magdy A. Bayoumi, "High-Performance and Low-Power Conditional Discharge Flip-Flop", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 12, No. 5, May 2004.
- [2] Jin-Fa Lin, "Low-Power Pulse-Triggered Flip-Flop Design Based on a Signal Feed-Through Scheme", *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*, Vol. 22, No. 1, January 2014.
- [3] Ahmad Karimi, Abdalhossein Rezai, Mohammad Mahdi Hajhashemkhani, "A novel design for ultra-low power pulse-triggered D-Flip-Flop with optimized leakage power", *INTEGRATION, The VLSI Journal*, vol. 60, pp. 160-166, January 2018.
- [4] C. K. Teh, T. Fujita, H. Hara, and M. Hamada, "A 77% energy-saving 22-transistor single-phase-clocking D-flip-flop with adaptive-coupling configuration in 40 nm CMOS", in *ISSCC Dig. Tech. Papers*, pp. 338–339, Feb. 2011.
- [5] Jin-Fa Lin, Ming-Hwa Sheu, "Low-Power 19-Transistor True Single-Phase Clocking Flip-Flop Design Based on Logic Structure Reduction Schemes", *IEEE TRANSACTIONS On Very Large Scale Integration (VLSI) Systems*, Volume: 25, pp. 3033 - 3044, Nov. 2017
- [6] Kirti Gupta, Prayanshu Sharma and Neeta Pandey "Design of Low Power Subthreshold Linear Feedback Shift Registers", *1st IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems*,
- [7] Deepak Berwal, Ashish Kumar and Yogendera Kumar, "Low Power Conditional Pulse Control with Transmission Gate Flip Flop" *International Conference on Computing, Communication and Automation (ICCCA2015)*
- [8] Massimo Alioto, Elio Consoli, Gaetano Palumbo, "Flip-Flop Energy/Performance Versus Clock Slope and Impact on the Clock Network Design", *IEEE Transactions on Circuits And Systems—I: Regular Papers*, Vol. 57, No. 6, June 2010
- [9] Praveen Kumar Y G., Kariyappa B S, M.Z. Kurian, "Implementation of Power Efficient 8-bit Reversible Linear Feedback Shift Register for BIST", *International Conference on Inventive Systems and Control*, (ICISC-2017)
- [10] T. Sharanya, S. Aruna Mastani, "Design of Low-Power Explicit type Pulse Trigger Generator for Flip-Flops" *International Conference on Communication and Signal Processing*, April 6-8, 2016

VII. Acknowledgments

My work would have been incomplete without the support and valuable guidance of Dr. Arti Noor.



Look ahead clock gating with probability driven MBFF integration for dynamic power optimization

Mohd.Adnan¹, Dr. Sunita Prasad²
¹²School of Electronics
CDAC, NOIDA, Uttar Pradesh
INDIA

Abstract: Clock gating and multi-bit flip flop both are dynamic power reduction techniques yet they are treated separately in digital designs. The clock gating techniques are widely used in as a tool for reduction of dynamic power in digital circuits. The technique of gating of clock can be applied at design phase as well as RTL stage. In this brief paper we propose an integrated technique for dynamic power reduction which is Look ahead clock gating (LACG) with multi-bit flip flop (MBFF), driven by probabilistic approach. The power reduction 16%-37% is proposed to achieve. The design is implemented on 32-bit LFSR on Xilinx vertex7. For design synthesis, Xilinx ISE design suit v14.7 is used.

Keywords: MBFF, DDCG, clock gating, LACG

I. INTRODUCTION

Among the major constituent of dynamic power dissipation is due to system clock roughly 30% -70% of the total consumption of dynamic power [1]. Many techniques for dynamic power optimization have been developed over the years out of which clock gating is predominant. In digital circuits the source of power consumption is mainly due to sequential elements when they get switched or toggle they dissipate more power which is termed as dynamic power.

$$P_{\text{total}} = P_{\text{static}} + P_{\text{dynamic}}$$

Static power:

$$P_{\text{static}} = (I_{\text{sub}} + I_{\text{gate}} + I_{\text{junc}} + I_{\text{contention}}) V_{\text{DD}}$$

Dynamic power:

$$P_{\text{dynamic}} = P_{\text{switching}} + P_{\text{shortcircuit}}$$

Where,

$$P_{\text{dynamic}} = \alpha C_L V_{\text{DD}}^2 f$$

In the above equation “ α ” is always referred as activity /toggling factor of the device, the more toggling in the device the more dynamic power will be dissipated. In this paper we are more focused on the techniques which avoid unnecessary switching/toggling of sequential device let say Flip Flops (FF). For that we will discuss different clock gating techniques. The clock gating technique is used at many levels such as system architecture level, block design level, logic design level and the gate level [2][8]. Different types of techniques described in [4]. Mainly we are focusing on major technique of clock gating that is data dependent/ driven clock gating (DDCG) and look ahead clock gating (LACG).

CAD tools are supporting very much in rapid increase in design complexity, system level HDL is being used commonly, although substantially increase in the design productivity, these tools require the employment of automatic synthesis algorithm from RTL down to gate level even at net list level. In this paper we are proposing a comprehensive technique for dynamic power optimization which is the integration of clock gating and multiple bit flip flop (MBFF), here the technique for clock gating is employed in integration with look ahead clock gating (LACG). In this paper we have tried to pull out probability driven design which will keep the toggling rate of FF which is also called activity rate of FF. The difference between the grouping and MBFF emphasize on the idea of physical proximity of individual FF. Whereas the plain grouping is based on

the activity of the FF and switching probability. In the following sections we shall discuss the different clock gating technique with some of the advantages and associated disadvantages.

II. SYNTHESIS BASED CLOCK GATING

There are several clock gating technique discussed by [4]. A simple clock gating technique is shown in the Figure.(1). This technique is referred as simple AND gate based clock gating. This technique is based on the data toggling or activity of the target FF. It compares the past output and current input of the FF using the XOR gate and the output of the XOR gate fed in to the AND gate. The AND gate performs gating of the clock pulse, the gated clock pulse depends on the data toggling in FF. The AND gate output is fed in to target FF as gated clock pulse. This technique suffers from serious glitches in the gated pulse which leads to the timing constraints problem and hazards.

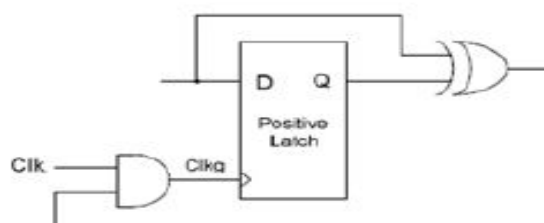


Fig 1. Simple AND gate based clock gating.

III. DATA DRIVEN CLOCK GATING (DDCG)

The problem due to glitches is eradicated by adding an extra latch D₃, in between XOR output, and AND gate input. It provides a stable pulse for a given clock period and hence eliminates the problem of setup and hold time[6].

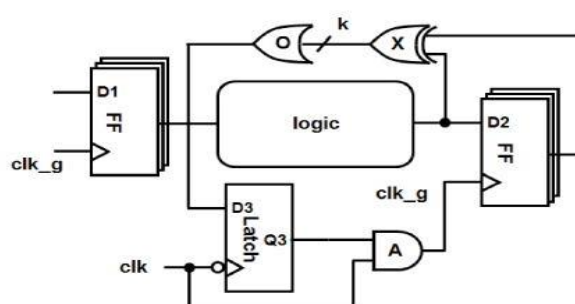


Fig 2. Latch based Data driven clock gate(DDCG)

Moreover, the dynamic power dissipation can be achieved in greater amount keeping in mind the probability of toggling of FFs. Challenges in DDCG are to observe the similar toggling ratio of FFs and their grouping accordingly. General studies shown by [2] which is % activity of FFs and the number of FFs into each activity percentage.

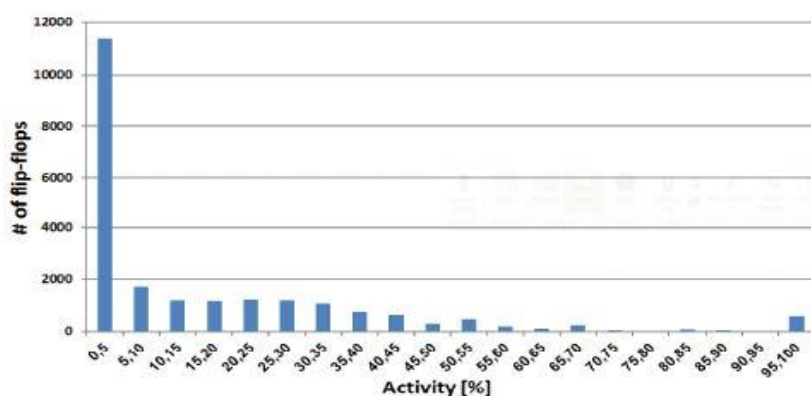


Fig.3 Concentration of FFs in percent activity

IV. AUTO GATED FLIP-FLOP (AGFF)

This is the basic circuit on which LACG is designed. In auto gated FFs a negative clock drives master latch, it holds input during the negative cycles. Till that time slave computes whether the input is different or not, if input is different then XOR output will be high, which will enable the AND gate. Further in the positive clock pulse the slave gets switched to next state.

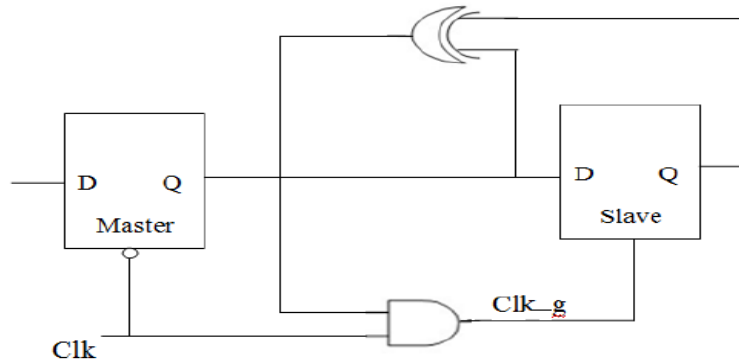


Fig.4 Auto gated flip-flop

The problems with AGFF are gating provided only to the slave latch at half clock in other half clock, circuit is not gated. The FFs which are lying on the critical path get a serious timing constraint which rejects gating.

V. LOOK AHEAD CLOCK GATING (LACG)

It eradicates the problem of auto gated FFs and DDCG; it takes AGFF and design the enhanced gating cell. It addresses three goals. 1) it avoids tight timing constraints by allotting full clock cycle for processing. 2.) Makes it applicable for general bulky design. 3.) stops the clock pulse in master latch too.

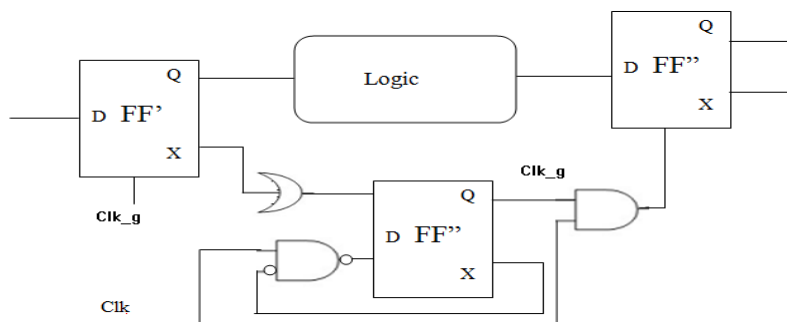


Fig.5. LACG with last and next stage.

The problem with AGFF the output of XOR gate is only valid for $[t_{\text{setup}}, t_{\text{ccq}}]$, which render very narrow window for gating. Where t_{setup} is the setup time of FFs and t_{ccq} is the contamination delay from clock to output, XOR output gets corrupted after delay of t_{ccq} and eventually turns zero. It must be latched for valid entire half period of clock which can solve the problem as 5a. Here the additional latch also consumes power but it can be minimized by gated clock as well. Working of LACG is illustrated in fig 6. Since the name suggest look ahead it takes the input from last module and it compares at $(n \geq 1)$, if It finds changing in data then it switches the next target FFs.

VI. LACG WITH MBFF

In this section the techniques and algorithm for grouping are discussed, how design is driven by clock gating. As suggested by [5][7], performing MBFF at post placement is burden so here we are trying to merge FFs at higher level of design such as RTL, architectural and structural etc. The gating of grouped FFs will be performed at this design level as well, introduction of MBFF at RTL level is also free from timing constraints.

This work focuses on:

- 1.) Paper proposes the fusion of MBFF and LACG.
- 2.) Minimization of power consumption by probability driven algorithm.

A. Integration

Let 1 bit FF consumes energy E_1 and data to clock switching probability of FFs be 'P'

$$E_1(P) = \sigma_1 + \mu_1 p$$

Where σ is internal driver energy and data toggling energy is μ . Taking the case of n-MBFF, let σ_n and μ_n be the n- driver and data toggling of n-FFs in n-MBFF. Hence

$$E_n(p) = \sum_{j=0}^n (\sigma_n + j\mu_n) \binom{n}{j} p^j (1-p)^{n-j} = \sigma_n + n\mu_n p$$

The potential of energy saving is expressed by the ratio $\frac{nE_1(p) - E_n(p)}{nE_1(p)}$ for n-MBFF. In general, for 65nm technology the grouping saving statistics on the basis of switching probability is given by [8].

Ideally P must be 0.1[6]. Lets determine the appropriate no. of flip flop 'n' that must be grouped for optimized version of MBFF, let 'P' ($0 < p < 1$) be the avg. toggling probability of flip flop also called activity factor.

TABLE I: GROUPING VS. SAVINGS

Activity	Grouping	savings
P=0 (0%)	2-MBFF	35%
	4-MBFF	55%
P=1(100%)	2-MBFF	15%
	4-MBFF	23%

The solution for n for optimal grouping is given by [8]. Which is

$$(1-p)^n \ln(1-p) C_{FF} + \frac{C_{latch}}{n^2} = 0$$

Where C_{FF} is the input capacitance of the flip flop and C_W is the wire capacitance in per unit length, latch capacitance is given as C_{latch} . table II shows the optimal n depends on P.

TABLE II: OPTIMAL FF GROUP SIZE DEPENDANCY OF ON TOGGLING PROBABILITY.

p	0.01	0.02	0.05	0.1
K	8	6	4	3

B. FFs group optimization in MBFF

The algorithm which is based on the fact that how many flip flops can be grouped if we know the behavior of the module for illustration a Johnson counter truth table is given in figure 6, it is clear that Q_0 and Q_1 have same output data for almost 6-7 clock pulses which means we can provide same gated clock to both the FF which minimizes the number of drivers used by sharing and power dissipation in those driver as well, same goes for Q_0, Q_1, Q_2 , and Q_3 which can form a 4-MBFF as they are switching correlated.

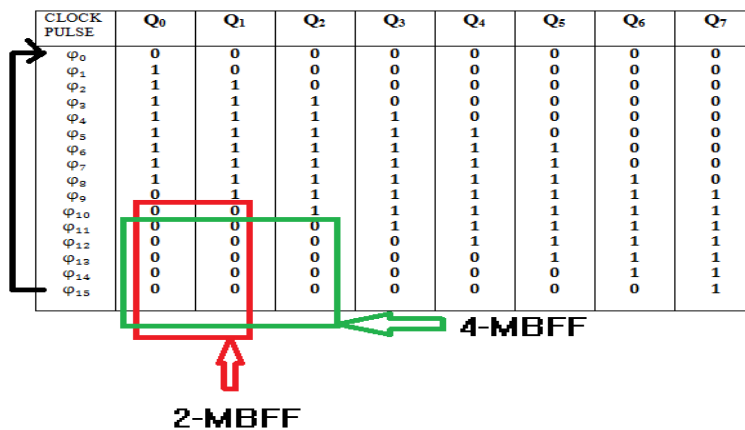


Fig. 6 illustration of MBFF

A very interesting study is given by [1] , which shows the safe and optimized region for the grouping and its impact on power dissipation on the basis of activities of FFs.

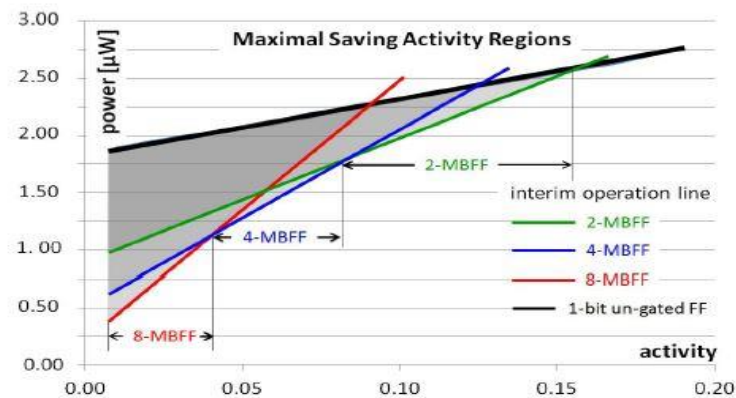


Fig. 7 MBFF vs. power dissipation on the basis of activity.

VII. SIMULATION RESULTS

The above techniques are implemented with 32-bit LFSR. For design synthesis we have used XILINX ISE design suite 14.7 using Verilog HDL and XILINX vertex7 FPGA as target device. We have performed simulation with $f=5\text{MHz}$, 10MHz , 20MHz . with typical toggle rates of 5%,10% and 12.5%(standard data).

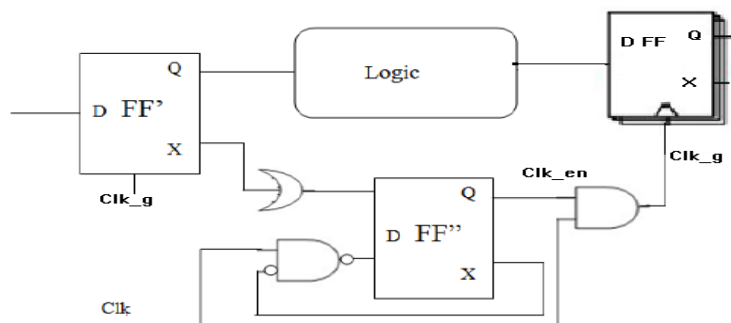


Fig.8 LACG cell with MBFF used in LFSR

After synthesize we have created the RTL Xilinx ISE design suit v14.7 of a Look ahead clock gating (LACG) cell .A wave form is also attached which is LACG based 32-LFSR of polynomial “ $Y=X^{28}+X^{27}+X^2+X^1$ ” where 28,27,2,and 1 are tap values .

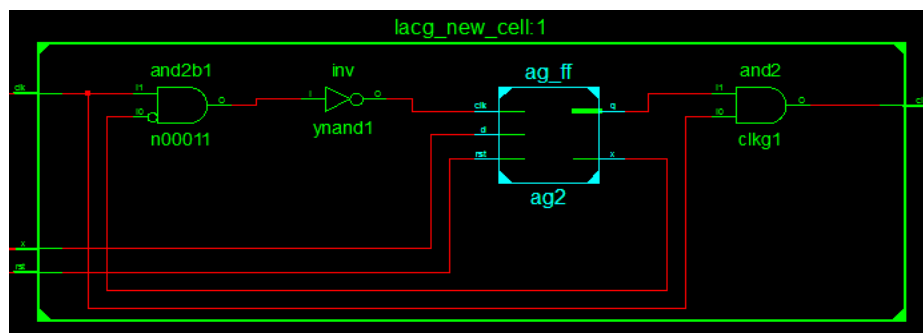


Fig.9: RTL of an LACG cell.

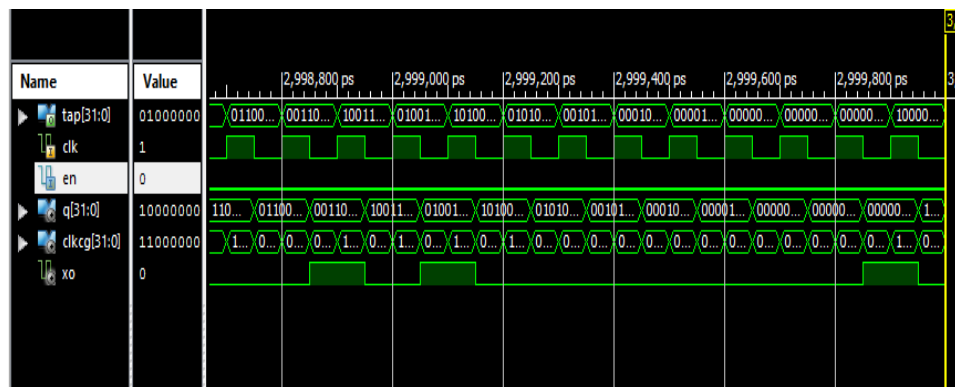


Fig.10 output waveform of LACG based 32-bit LFSR.

VIII. DESIGN COMPARISON

The design power analysis is done on Xilinx power estimator (XPE), the following power dissipation results are obtained at $f=10\text{MHz}$, and the toggle rate of FF is taken 12.5% , there are two results 1) in table 1 only clock gating with different techniques and in result 2) table 4 LACG with MBFF with $n= 2,4$ and 8 grouping are performed .

TABLE III
POWER IN DIFFERENT TECHNIQUES VS. BITS

(Power in μW)	16-bits	32-bits
LFSR	228.4	302.3
With DDCG	215.8	280.8
With LACG	207.8	252.2
savings	9.01%	16.5%

TABLE IV
LACG WITH MBFF ON 32-BIT LFSR

With Look-Ahead Clock gating 32-bit (μW)	Dynamic Power (μW)	Without grouping and clock gating	Saving in dynamic power
For $n=2$	241.7	302.3	20.19%
For $n=4$	228.6	302.3	24.37%
For $n=8$	207.9	302.3	31.22%

IX. CONCLUSION

Discussing the current scenario of power optimization technique in FPGAs and digital circuits then both the technologies i.e. clock gating and multi-bit flip flops are being used separately for optimization purpose, this paper proposes a model of integration of both the low power techniques which are implemented simultaneously. The low power LFSR has wide range of applications such as in BIST, spread spectrum communication and Cryptography etc. It is clearly shown by simulation results that the proposed design is capable of achieving 16.5% dynamic power reduction with LACG and 31.22% reduction by 8-MBFF with LACG. In this paper we have discussed the model of grouping of FFs on the basis of probability of toggling which makes the grouping more logical and efficient, in clock gating we have implemented an advanced gating technique which has nearly zero timing constraint and no lagging in processing.

There is a huge scope for further research in the probability driven section of design and its algorithms, which will make the design more efficient and precise.

References

- [1] Gluzer, D. and Wimer, S., 2017. "Probability-Driven Multibit Flip-Flop Integration With Clock Gating". IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 25(3), pp.1173-1177.
- [2] Wimer, S. and Koren, I., 2014." Design flow for flip-flop grouping in data-driven clock gating". IEEE Transactions on very large scale integration (VLSI) systems, 22(4), pp.771-778.
- [3] Rajan, N., 2017, May. "Dynamic power optimization using look-ahead clock gating technique". In Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017 2nd IEEE International Conference on (pp. 261-264). IEEE.
- [4] Kathuria, J., Ayoubkhan, M. and Noor, A., 2011. "A review of clock gating techniques". MIT International Journal of Electronics and Communication Engineering, 1(2), pp.106-114.
- [5] Jiang, I.H.R., Chang, C.L., Yang, Y.M., Tsai, E.Y.W. and Chen, L.S.F., 2011, March. INTEGRA: "Fast multi-bit flip-flop clustering for clock power saving based on interval graphs". In Proceedings of the 2011 international symposium on Physical design (pp. 115-122). ACM.
- [6] Wimer, S. and Albahari, A., 2014. "A look-ahead clock gating based on auto-gated flip-flops". IEEE Transactions on Circuits and Systems I: Regular Papers, 61(5), pp.1465-1472.
- [7] Yan, J.T. and Chen, Z.W., 2010, June. "Construction of constrained multi-bit flip-flops for clock power reduction". In Green Circuits and Systems (ICGCS), 2010 International Conference on (pp. 675-678). IEEE..
- [8] Yan, J.T. and Chen, Z.W., 2010, June. "Construction of constrained multi-bit flip-flops for clock power reduction". In Green Circuits and Systems (ICGCS), 2010 International Conference on (pp. 675-678). IEEE.



Design and Implementation of Power Efficient Dynamic Circuit for High Fan-In Circuits

Ditipriya Banerjee¹, Dr. Arti Noor²

^{1,2}School of Electronics

CDAC, Noida

India

Abstract: Dynamic circuit designing technique is most commonly used in implementing performance driven high fan-in circuits. In the paper a comparative power and delay analysis of 16-bit OR gate using different dynamic circuit topologies is carried out with respect to parametric variations. A new low power, dynamic circuit technique is proposed for high fan-in gates. By the reduction in voltage swing of the pull-down network power reduction of 45.83 % is achieved with a delay improvement of 13.51% using the proposed topology. Furthermore, using the proposed topology, a high fan-in tag comparator is simulated in 180 nm CMOS technology at 1V supply. The tag comparator is fast with lower leakage power and better noise immunity.

Keywords: Dynamic logic, power dissipation, keeper transistor, evaluation network, tag comparator.

I. Introduction

In critically high performance processors, dynamic logic design is commonly used for high fan-in circuits. Transistor count for high fan-in gates in dynamic logic design is much lesser than static logic design. Though dynamic circuits take lesser chip area and are fast, but poor noise immunity and high power consumption are the major shortcomings for high fan-in circuits [1]. To reduce power consumption, technology scaling is employed by scaling down the supply voltage as well as threshold voltage (V_{th}) needs to be reduced accordingly. Reducing threshold voltage consequently increases subthreshold leakage power which reduces the noise immunity. Higher switch capacitance is accounted for increased power consumption with increasing fan-in in dynamic circuits. Increased number of parallel branches in pull-down network affects the noise immunity. Better noise-immunity and power efficient high fan-in gates are used in high-speed processors.

High fan-in gates are employed in register files, L1 latches, flash memories, PLA, tag comparators, wide MUX and De-MUX in high speed power efficient processors. Tag comparator is used to produce Hit bit for cache memories. Tag comparator is a power consuming component in microprocessors. If the tag comparator does not send hit signal to cache controller, cache cannot operate. In this paper, a significantly low power consuming, high fan-in dynamic circuit design is proposed. By the reduction in voltage swing of the pull-down network, power reduction is achieved with better noise-immunity.

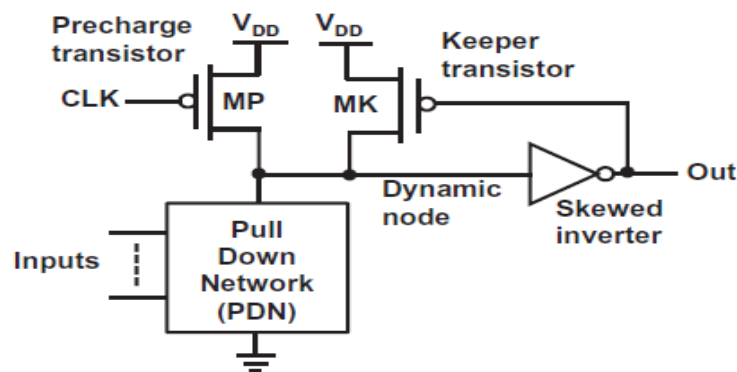
The work done in this thesis combines the different dynamic circuit topologies along with the proposed design for the comparative analysis of power consumption and delay. 16-bit OR gate is implemented using all the mentioned topologies using *Pyxis schematic tool* by *Mentor Graphics* in 180 nm CMOS technology at 1V. The power and delay analysis of the proposed circuit over other dynamic circuit topologies shows a significant power and delay reduction. To demonstrate the effectiveness of the proposed dynamic circuit, a high fan-in i.e. tag comparator is implemented at a frequency of 1GHz at a hot-spot temperature of 110 °C.

II. Standard Dynamic Circuit

A standard dynamic circuit is shown in Fig. 1, keeper transistor is used for maintaining the dynamic node against charge sharing, charge leakage, leakage current and noise [2]. In dynamic circuit the contention between the pull-down network and keeper transistor accounts for the delay and power consumption [3].

With technology scaling, threshold voltage (V_{th}) must be reduced to achieve high performance at the cost of increasing subthreshold leakage current resulting in higher power consumption and reduced noise immunity. Increasing leakage current degrades the noise immunity for high fan-in gates, as number of leaky NMOS transistors increases.

Fig. 1. Conventional dynamic circuit [4].



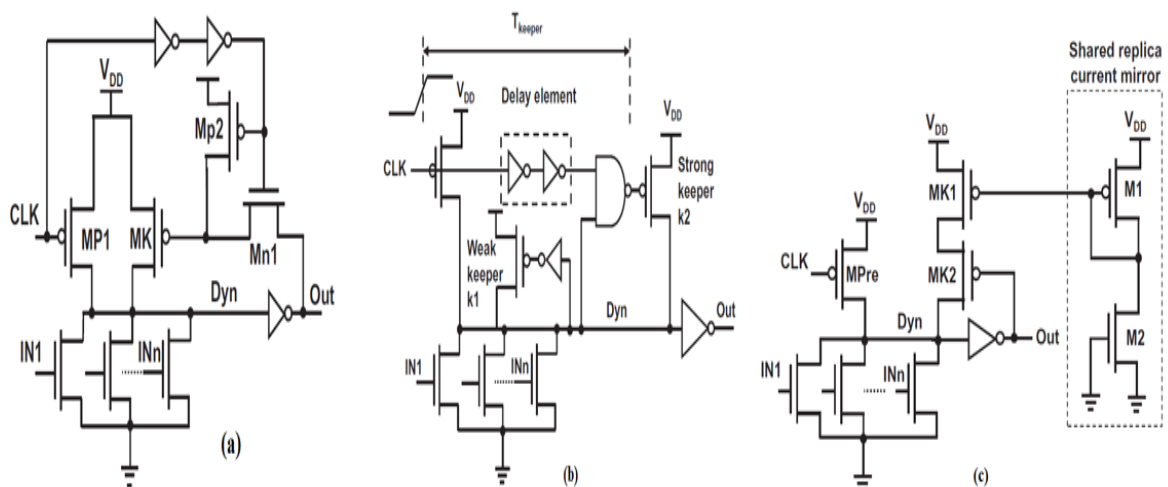
To maintain acceptable noise margin level in deep sub-micron technologies, keeper upsizing is done, though it leads to higher power consumption and delay due to increase in current contention [4]. Keeper ratio K is defined as,

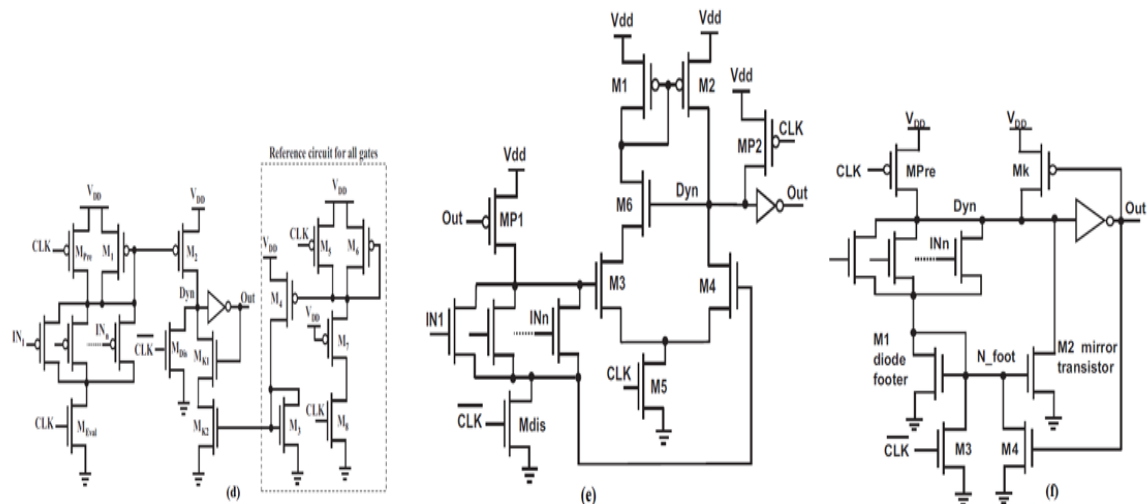
$$K = \frac{\mu_p \left(\frac{W}{L}\right)_{\text{Keeper-transistor}}}{\mu_n \left(\frac{W}{L}\right)_{\text{Pull-Down-network}}}$$

where, μ_n and μ_p are electron and hole mobilities respectively. However, in new generation and technology scaling, traditional keeper approach gets less effective.

Various domino circuit design topologies have been acknowledged in the paper. Basically, the circuit topologies can be divided in two category. First one, changing the keeper scheme such as high speed domino (HSD), conditional keeper domino (CKD) and leakage current replica (LCR). Second group reconstructs the evaluation network such as current comparison based domino (CCD), voltage comparison based domino (VCD) and diode footed domino (DFD). In Fig. 2(a) a clock delay is interposed for reducing the current through the keeper transistor and pull-down network at the onset of evaluation phase. However, HSD suffers the drawback of afloat dynamic node at the beginning of evaluation phase, any noise at inputs could cause the evaluation node to discharge in absence of keeper transistor [5]. In CKD, Fig. 2(b), two keeper transistors are added. The strong keeper increases the noise immunity of domino gate and the weaker keeper holds the state of dynamic node during transition of the NMOS logic [6]. However the NAND gate introduces delay and additional power consumption.

Fig. 2. (a) HSD [5] (b) CKD [6] (c) LCR [7] (d) CCD [4] (e) VCD [8] (f) DFD [9]



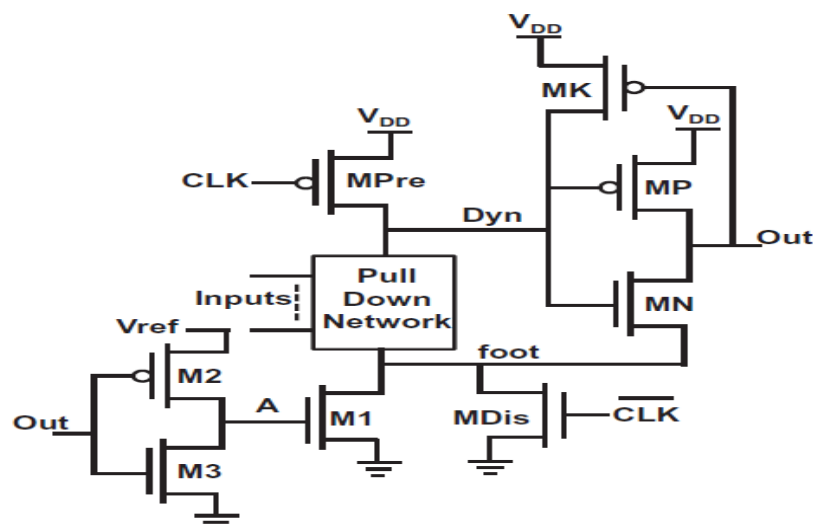


LCR keeper shown in Fig. 2(c), uses a standard analog current mirror for tracking any process, voltage and temperature variation [7]. More than one dynamic gate can share a single current mirror structure. It consumes more power with lower robustness. In CCD as shown in Fig. 2(d) footer transistor reduces the leakage current in the diode configuration, enhancing the robustness of the gate [4]. In Fig. 2(e), VCD topology uses sense amplifier for detecting the voltage difference across pull-down network for generating correct output [8]. The reduced voltage swing results in lower power consumption. As shown in Fig. 2(f), in DFD an additional NMOS transistor in diode configuration is added in series with the evaluation network [9]. It decreases the subthreshold leakage current due to stacking effect. However, it suffers from low robustness.

III. Proposed Design

The proposed design targets in reducing the power consumption by reconstructing the evaluation network [10]. As stated earlier, technology scaling attributes to leakage current which contributes to large power consumption, thus reducing the leakage power is very crucial. The proposed design is shown in Fig. 3 which consists of transistors M_{Pre} for charging dynamic node Dyn and M_{Dis} for discharging node $foot$. Source of transistor M_N is not grounded, but connected to the node $foot$. In evaluation phase, M_{Dis} is OFF so voltage at node Dyn cannot discharge to ground. Therefore, the voltage swing across pull-down network reduces, which results in subsequent reduction in power consumption. It also leads in reducing the contention between M_N and M_P transistors.

Fig. 3. The proposed dynamic circuit design [10]



In conventional circuit, V_{DD} can be reduced to V_{tp} , needed for turning on the PMOS transistor whereas in the proposed design, for turning on the transistor M_p minimum value of V_{DD} is $2V_{tp}$ as voltage of dynamic node is approximately $V_{DD}/2$. Therefore minimum power supply in the proposed design is doubled as in comparison to conventional circuit.

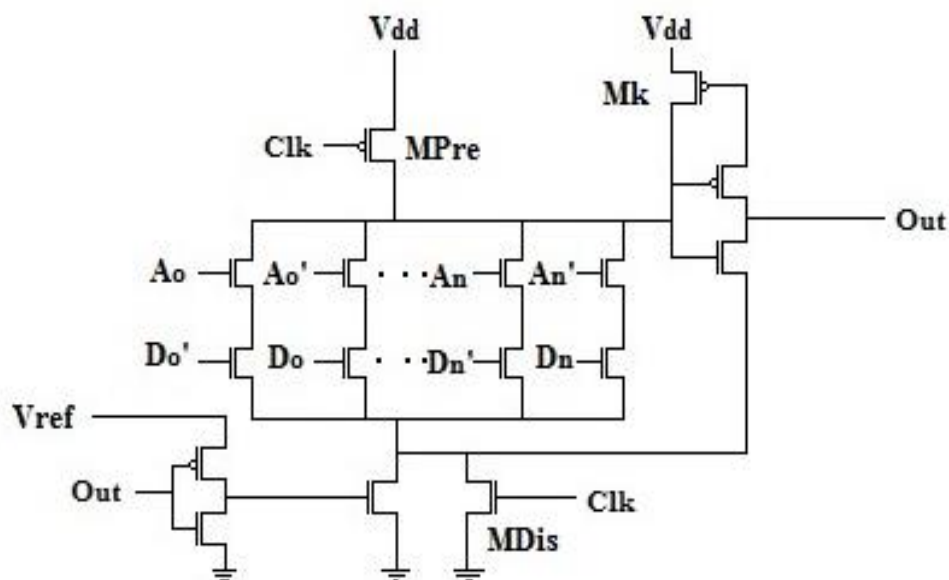
Prechare phase: The input and clock signals are low i.e, CLK=0. Transistors M_{Pre} , M_{Dis} , M_K , M_2 and M_N are ON. M_2 transistor increases the voltage of node *Dyn* to V_{DD} and M_{Pre} increases node *A* to V_{ref} .

Evaluation phase: The clock signal is high, CLK=1 and input signals may also be high, turning M_{Pre} , M_{Dis} and M_2 OFF. Transistors M_K , M_1 and M_3 will be ON/OFF according to the input signals. There can be two scenario depending on the input signals, one, there is no conducting path in the evaluation network and the other that at least there is one path is conductive. In first scenario subthreshold leakage current only conducts through nodes *Dyn* and *foot*, causing slight decrease in voltage at node *Dyn* and increase in voltage at node *foot*. Moreover, transistors M_K and M_1 prevent changing of voltages at *Dyn* and *foot* resulting M_K and M_1 to remain ON and hence there is no change in the output voltage. In the latter case, voltage of node *Dyn* gets reduced and *foot* increases. Both the nodes reach the same non-zero voltage and the difference within them reduces. Hence M_K , M_1 and M_N gets OFF raising the output to V_{DD} .

IV. Tag Comparator

Low speed of off-chip memory against high speed processor cache is compensated using tag comparator [11]. Higher hit rate of larger on-die cache memories for higher speed are needed in modern microprocessors. In every generation with increase in number of tag bits and physical address space, power consumption and delay associated in employing the high fan-in domino circuits also increases. In this paper, a significant low power high fan-in domino circuit is used in designing the tag comparator. For a 64-bit processor, 40-bit tag comparator is used for generating *hit/miss* bit cache controller [12]. Tag comparator implemented using the proposed domino circuit design is shown in the Fig. 4. In order to reduce the power consumption on high fan-in tag comparator due to high switching capacitance the voltage swing on the matching line or dynamic node is reduced in the proposed design. The hit/miss signal is provided by comparing the voltage of dynamic node and the reference voltage V_{ref} .

Fig. 4. Tag comparator design



V. Simulation Results and Comparison

The circuits given in the paper include the HSD, the CKD, the LCR, the CCD, the VCD, the DFD and the proposed dynamic circuit technique are simulated in a 180 nm CMOS technology at $V_{DD} = 1V$ and typical temperature of 110 °C with 5Ff load capacitance. High fan-in OR gate circuits are implemented at different temperature of 27 °C, 55 °C and 110 °C with same delays of 70ps, 90ps and 110ps.

For all the given circuit techniques, transistor sizing in pull down network is set to $L_{min} = 180$ nm and $W_{min} = 270$ nm. For CCD, in output inverter the width ratio of PMOS to NMOS transistor is set to 2. Initially sizing of other transistors is minimum and then changed to get the desired delay. For an equitable comparative analysis only sizes of transistors which determines design specifications are only changed.

In HSD, size of M_K i.e. keeper transistor is varied, in CKD size of both the strong and weak keeper transistors is changed to obtain necessary delay. In LCR, to reduce channel length modulation effect length of transistor M_1 is increased and keeper transistors M_{k2} and M_{k1} is varied. In CCD technique, transistor sizing of M_{eval} and M_2 is varied, to alleviate channel length modulation, length of transistors M_3 and M_{k2} is kept large. For

VCD circuit technique, transistor length of M_2 , M_3 and M_{dis} is increased to get the desired delay. In DFD, size of transistors M_K and M_2 is changed.

For the proposed dynamic circuit design, transistors M_{Dis} , M_N , M_2 , and M_3 are set to minimum. Width of transistors M_{Pre} and M_1 are set twice and thrice the minimum length. Upsizing transistors M_1 and M_P , downsizing M_K reduces delay with power trade-off and vice versa.

Fig. 5. Power analysis for 16 bit OR-gate for different dynamic circuit technique at different temperature

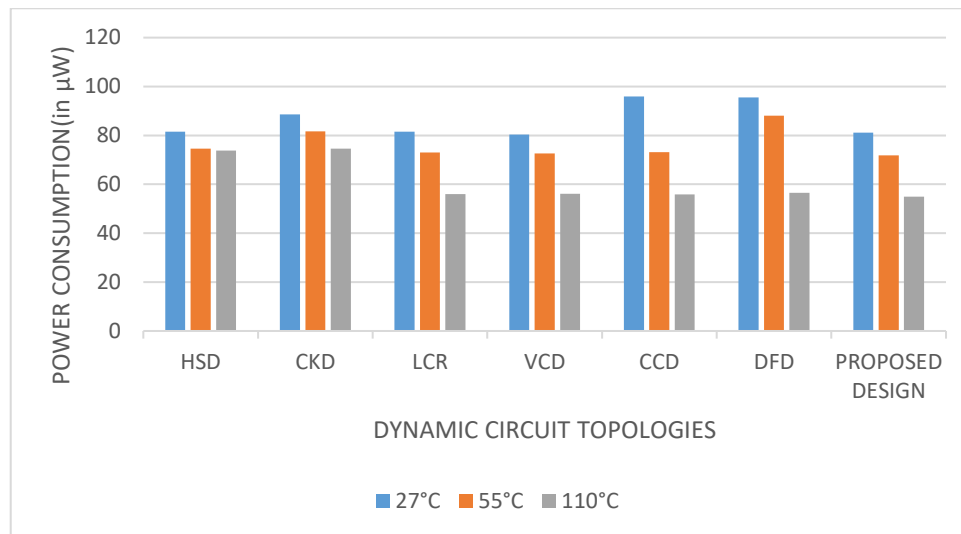


Fig. 4, shows the superiority of the proposed dynamic circuit over the mentioned temperature range, where CCD is the most power hungry circuit and the proposed design shows a significant reduction in the power consumption.

Table 1, summarizes the comparison of power consumption of 16-bit OR gate using different dynamic circuit techniques at a clock frequency of 20 MHz. It can be seen from the table that even on increasing the temperature, power consumption is reducing in every technique, this is due to the phenomenon called inverted temperature dependence [13]. Power reduction is achieved even at higher temperature due to the scaling down of parameters like supply voltage V_{DD} , threshold voltage V_{th} and appropriate sizing of the transistors.

Fig. 6, gives the comparative propagation delay analysis of all the mentioned high fan-in dynamic gate techniques, where CKD and DFD has the maximum delay wherein the proposed design shows reduction in delay timings making it a fast operating circuit.

TABLE 1: Comparative power analysis of 16-bit OR gate using the different dynamic circuit topologies at clock frequency of 20 MHz

Design	Temperature (° C)	Power (μW)
HSD	27	81.5353
	55	74.6164
	110	73.8031
CKD	27	88.6400
	55	81.6405
	110	74.6164
LCR	27	81.4931
	55	72.9608
	110	56.0547
VCD	27	80.3724
	55	72.5851
	110	56.1713
CCD	27	95.9041
	55	73.1855
	110	55.8043
DFD	27	95.4696
	55	88.0695
	110	56.4825
Proposed Design	27	81.0769
	55	71.8350
	110	55.0027

Fig. 6. Propagation delay of output with respect to the clock signal at 20 MHz for different dynamic circuit techniques

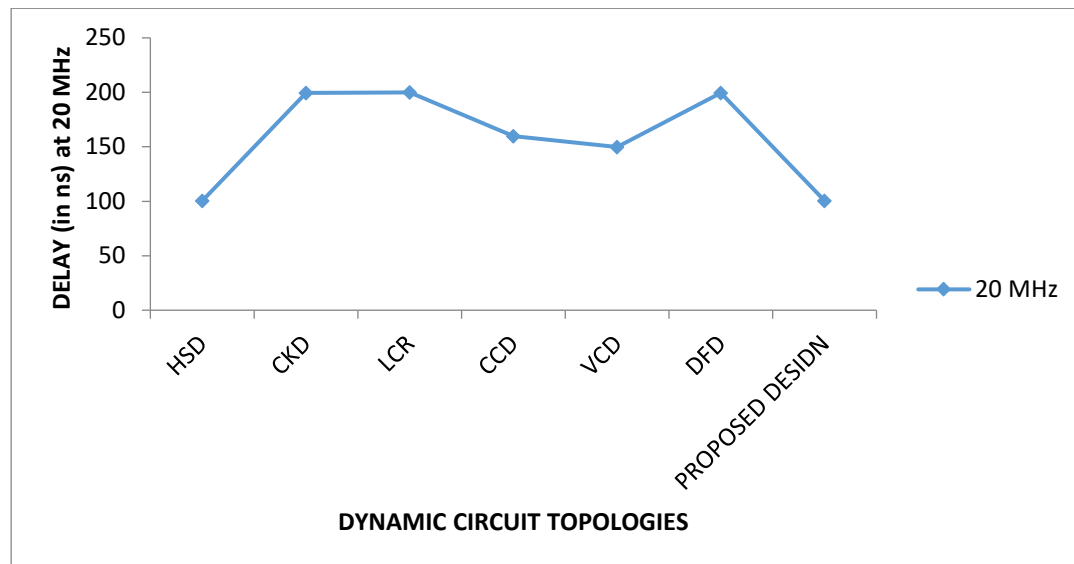
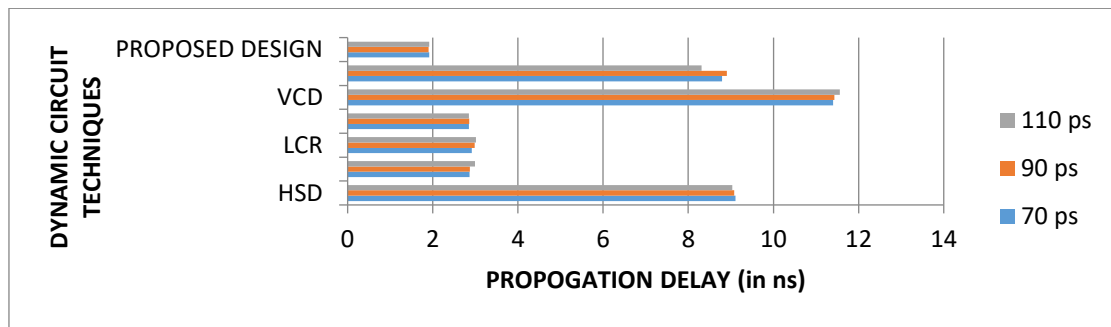


Table 2, summarizes the propagation delay at output node of 16-bit OR gate using the mentioned dynamic circuit techniques due to the clock signal delay of 70 ps, 90 ps and 110 ps. From the graph it can be seen that the proposed design shows significant reduction in propagation, making it appropriate for high performance circuits. Fig. 7, shows the propagation delay of all the mentioned high fan-in dynamic gate techniques implemented on 16-bit OR gate, where VCD and DFD has the worst delay wherein the proposed along CCD, CKD and LCR shows reduction in delay time.

TABLE 2: Comparative analysis of propagation delay at output due to delay in clock signal of different domino techniques at a frequency of 500 MHz

DESIGN	CLOCK DELAY (ps) at 500MHz	PROPAGATION DELAY (ns)
HSD	70	10.044
	90	17.037
	110	17.019
CKD	70	8.837
	90	8.919
	110	8.923
LCR	70	4.846
	90	4.830
	110	4.863
CCD	70	5.842
	90	5.765
	110	5.403
VCD	70	8.848
	90	8.883
	110	8.764
DFD	70	8.719
	90	8.907
	110	9.931
PROPOSED DESIGN	70	0.243
	90	0.372
	110	0.377

Fig. 7. Propagation delay at output with respect to the delayed clock signal at 1 GHz for different dynamic circuit techniques



TAG COMPARATOR

To demonstrate the effectiveness of the proposed dynamic circuit, a high fan-in i.e. 40-bit tag comparator is implemented with the waveform of its operation given in Fig. 5. At a frequency of 1GHz at a hot-spot temperature of 110 °C, the power consumption is 3.5372 nW.

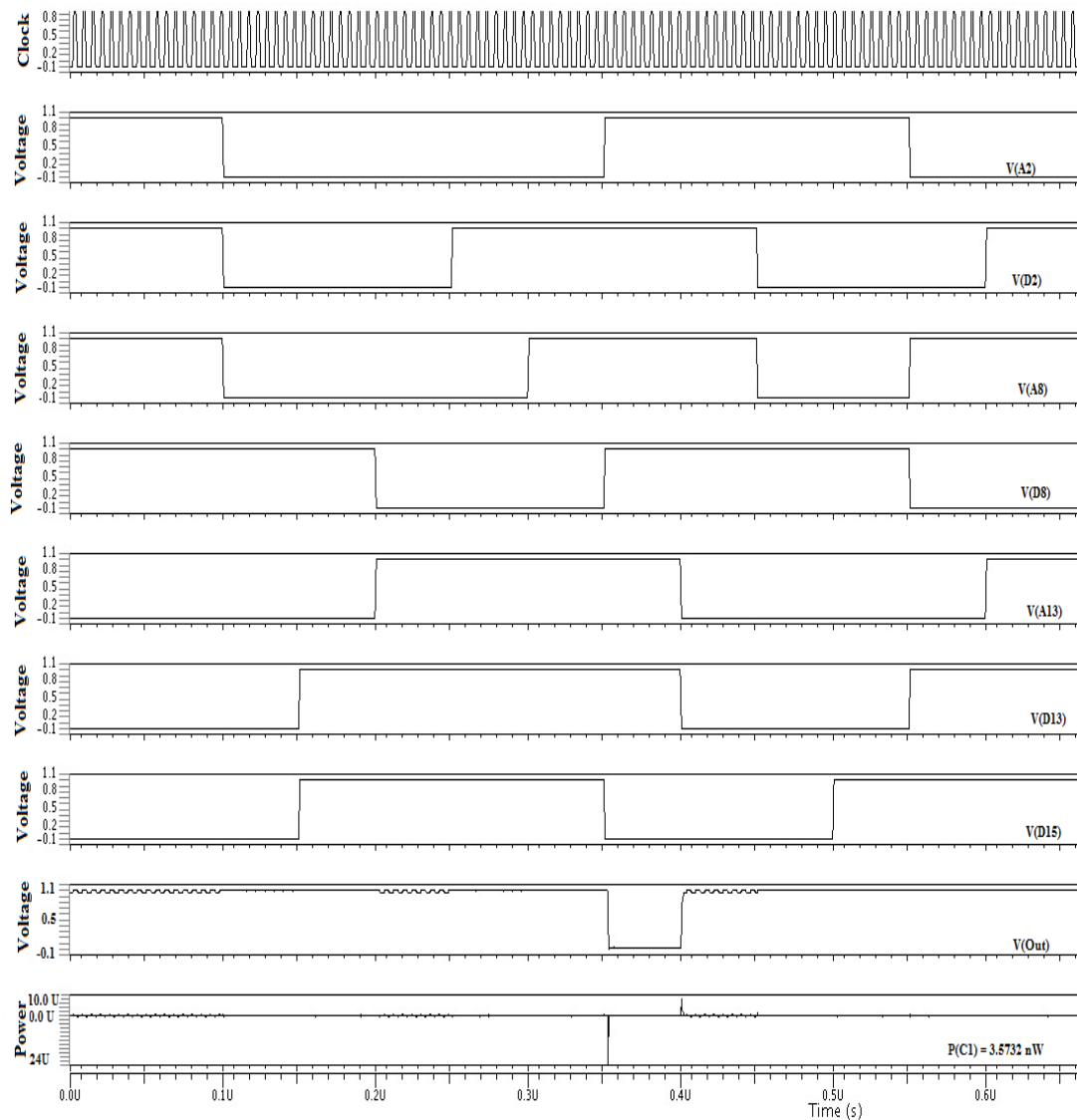


Fig. 8. Waveform of 40-bit Tag Comparator

VI. CONCLUSION

In this paper different dynamic circuit topologies is discussed for high fan-in gates. However dynamic circuit suffer from charge leakage, high power consumption and lesser noise immunity. Different domino circuit approaches is hence used to address the above problems. 16-bit OR gate is implemented using all the approaches and a comparative analysis is carried out for studying the effect of the approaches in respect to power and delay. Furthermore the advantages of the proposed dynamic circuit over others can also be seen.

In the proposed dynamic circuit, there is a significant reduction in power due to low voltage swing on the dynamic node. By using proper reference voltage, the performance of the circuit can be enhanced. In comparison to larger keeper transistor, smaller keeper is sufficient to obtain the desired robustness in the proposed circuit. The leakage current across the PDN is reduced and better noise immunity is attained. The proposed circuit and other circuit techniques are simulated using *Pyxis schematic tool by Mentor Graphics* in 180 nm CMOS technology.

The power and delay analysis of the proposed circuit over other dynamic circuit topologies shows a power reduction of 45.83 % at 20 MHz clock frequency and reduction in delay by 13.51 %.A high fan-in tag comparator is designed using the proposed circuit. The circuit exhibited a fair reduction in power consumption without dramatic degradation in speed even at higher frequencies.

References

- [1] Mohammad Asyaei, "A new low-power dynamic circuit for wide fan-in gates," in INTEGRATION, the VLSI journal, 2018, vol. 60, pp. 263-271.
- [2] J.M Rabaey, A.P Chandrakasan, B. Nikolic, "Digital integrated circuit," 2nd ed., Prentice Hall Englewood Cliffs, 2003.
- [3] FarshadMoradi, Ali Peiravi, "An improved noise-tolerant domino logic circuit for high fan-in gates," in International Conference on Microelectronics, 2005.
- [4] Ali Peiravi, Mohammad Asyaei, "Current-comparison based domino: new low-leakage high speed domino circuit for wide fan-in gates," in IEEE Transaction on very large scale integration (VLSI) systems, vol. 21, no. 5, May 2013.
- [5] Mohab H. Anis, Mohamed W. Allam, Mohamed L Elmasry, "Energy-efficient noise tolerant dynamic styles for scaled down CMOS and MTCMOS technologies," in IEEE Transaction on very large scale integration (VLSI) systems, vol. 10, no. 2, April 2002.
- [6] Atila Alvandpour, Ram K. Krishnamurthy, K. Soumyanath, Shekhar Y. Borkar, "A sub-130-nm conditional keeper technique," in IEEE Journal of solid-state circuits, vol. 37, no. 5, May 2002.
- [7] YolinLih, NestorasTzartzanis, William W Walker, "A leakage current replica keeper for dynamic circuits," in IEEE Journal of solid-state circuits, vol. 42, no. 1, January 2007.
- [8] Mohammad Asyaei, "A new leakage tolerant domino circuit using voltage-comparison for wide fan-in gates in deep sub-micron technology," in INTEGRATION, the VLSI journal, 2015, vol. 51, pp. 61-71.
- [9] Hamid Mahmoodi-Meimand, Kaushik Roy, "Diode-footed domino: a leakage-tolerant high fan-in dynamic circuit design style," in IEEE Transaction on circuit and systems-I, regular papers, vol. 51, no. 3, March 2004.
- [10] Mohammad Asyaei, Emad Ebrahimi, "Low power dynamic circuit for power efficient bit lines," in International journal of Electronics and Communications (AEÜ), 2018, vol. 83, pp. 204-212.
- [11] Mohammad Asyaei, Ali Peiravi, "Low power wide gates for modern power efficient processors," in INTEGRATION, the VLSI journal, 2013.
- [12] MahshidNasserian, Mohammad Kafi-Kangi, Mohammad Maymandi-Nejad, FarshadMoradi, "A low power fast tag comparator by modifying charging scheme of wide fan-in dynamic OR gate," in INTEGRATION, the VLSI journal, 2016, vol. 52, pp. 129-141.
- [13] PreetiVermaa, Ajay K. Sharmaa, Vinay Shankar Pandeya, Arti Noor, AnandTanwar, "Estimation of leakage power and delay inCMOS circuits using parametric variation," in Recent trends in engineering and material science, Elsevier Perspectives in science, 2016, vol. 8, pp. 760-764.
- [14] Massimo Alioto, Gaetano Palumbo, MelitaPennisi, "Understanding the effect of process variation on the delay of static and domino logic," in IEEE transaction on very large scale integration (VLSI) systems, vol. 18, no. 5, May 2010.
- [15] Li Ding,PinakiMazumder, "On circuit techniques to improve noise immunity of CMOS dynamic logic," in IEEE transactions on very large scale integration (VLSI) systems, vol. 12, no. 9, September 2004.



Gesture control using RADAR

¹Ambuj Maurya, ²Isha Gupta (Project Engineer)
School Of Electronics, CDAC Noida

Abstract: This paper depicts the development of RADAR based gesture recognition for multimedia devices using Verilog. This paper proposes how the radar can be used as a control element in place of the traditional button and touch-based interfaces. The proposed design currently have the capability to control the volume of the connected multimedia device. Here in this design Andriod phone is used as a multimedia device.

Keywords: Doppler Effect, HB100, RADAR, IOT, Microwave Motion Sensor

I. INTRODUCTION

In recent years, gesture based recognition increases and these gestures are using either IR sensor, touch gesture or Video gesture for controlling the multimedia device. The IR based recognition are prone to anybody e.g. Mobile front IR is used for picking and rejecting calls. Touch based gesture in the mobile and multimedia devices are having wide range for controlling the parameters like volume and power option. Video gesture are used for controlling robots, mobile application execution and obstacle detection in front of visually impaired people. But these all gesture are having a problem of wear tear after a time. RADAR is the radio detection and ranging these are widely used in getting the exact location, height and speed of the airplane in the wide applications. Radar is also used for lock-target and exact firing. As time passes, technology is becoming modern and process of IC fabrication like RADAR is reduced in its size which can be easily fit into any multimedia device. The advantage of using the RADAR is that it can be operated from inside the multimedia devices without getting exposed the user which reduces the wear tear of the device. Ultimately the RADARs have the capability to get exact location of the object from any device.

II. RADAR

Radio Detection And Ranging is the object detection technique which is used for getting the location and change in the position of the object in the radiation zone/area. The RADAR basically emits the Radio waves and checks the difference in the received and transmitting frequency which gives the information of the object position. If the object is at the same position the difference between receiving and transmitting frequency is same. If the object is moving towards the RADAR then the apparent frequency apparent frequency is called Doppler shift. Doppler shift is given by

$$f_D = \frac{2 \cdot v}{\lambda} \cdot \cos \theta$$

where,

f_D : Doppler Frequency (Hz)

v : Speed of the wave-source

λ : Wavelength

θ : Angle between the direction of the transmitted and reflected signal.

Proposed design use HB100 RADAR module which have operating frequency of 10.25 GHz. The module transmits the 10.25 GHz and the apparent change in frequency is obtained on the intermediate frequency pin of the module.

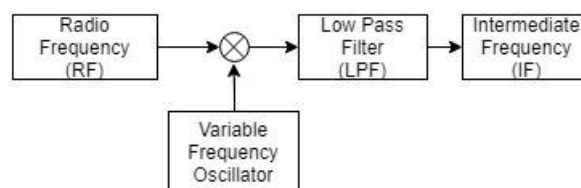


Figure 1: Intermediate Frequency

$$f_{IF} = f_{RF} - f_{VFO}$$

where,

f_{IF} : Intermediate Frequency
 f_{RF} : Radio Frequency
 f_{VFO} : Frequency of Variable frequency Oscillator

The Radar Module have 3 pins V_{CC} , GND and SIG . The V_{CC} and Ground pin is used for the powering up of the module and SIG pin is the signal pin. SIG is the signal pin which is high when there is no motion of the human hand and goes low if there is motion in the human hand in its range.

The output of the RADAR module is observed on the Arduino made Oscilloscope. Analog pin of the Arduino takes input and then using UART it transmits the reading from the Analog pin to the PC using UART communication. Using the Processing Language 3.3 we plot the data received from the Arduino using UART communication.

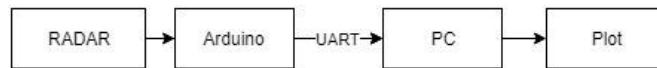


Figure 2: Using Arduino as a Oscilloscope

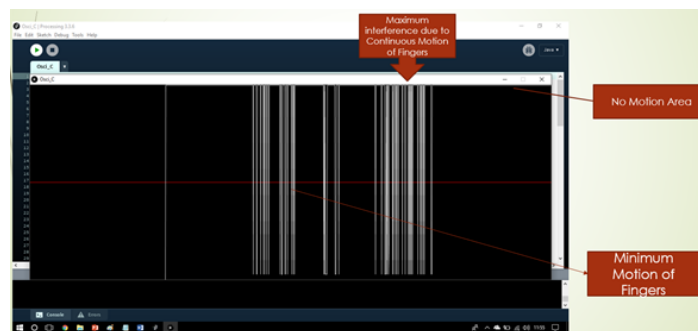


Figure 3: Oscilloscope result

III. PWM

PWM is the **P**ulse-**W**idth **M**odulation is a type of modulated signal in which there is a change in the width of the pulse of carrier pulse train according to instantaneous level of modulating the signal. Here proposed uses the PWM for demonstrating the change in the pitch of the buzzer due to the change in the speed of the motion of the fingers. Using PWM design generates the square wave of different duty cycle which causes the change in the sound of the Buzzer. The duty cycle is decided by the number of low edges from the RADAR module.

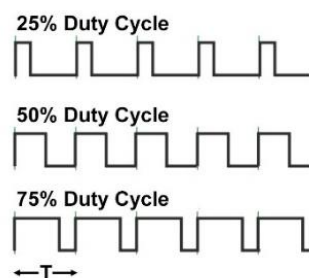


Figure 4: PWM Generation

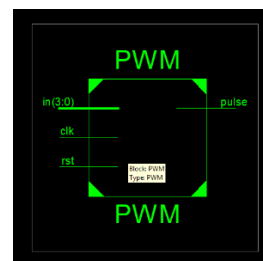


Figure 5: Xilinx PWM RTL Design

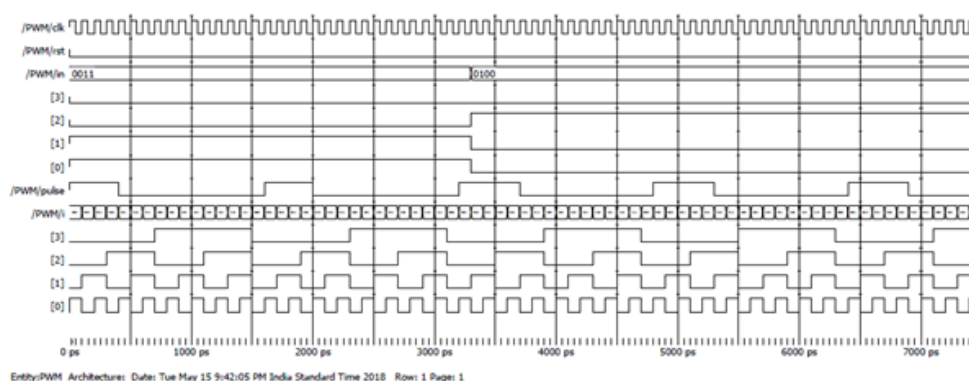


Figure 6: PWM Simulation result

V. BLUETOOTH

Bluetooth HC 050 module is a Bluetooth Version 2.0 Device. HC05 Module is used for transmission of data from the FPGA to Multimedia device connecting it with board wirelessly. The HC05 is in master mode with baud rate of 9600 bits per second (bps), stop bit of one bit, zero parity and 8-data bit. The Bluetooth Terminal application is used for receiving the data from the FPGA board.

VI PROPOSED DESIGN

The proposed design uses the Edge board with Spartan 6. For simulation of design and generation of bit Xilinx ISE 14.7 and ModelSim ALTERA STARTER EDITION 6.4a is used. Motion of fingers causes the Doppler shift in the radio frequency radiated by the HB100 Module. FPGA receives the change in the output of the Radar module signal pin. It counts the number of low edges and according to that it generates the PWM for buzzer and transfers the same data to the Cellphone using the Bluetooth module.

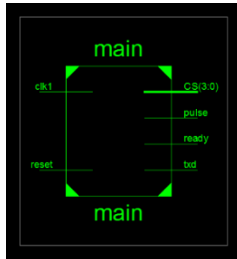


Figure 10: Main Module RTL Design

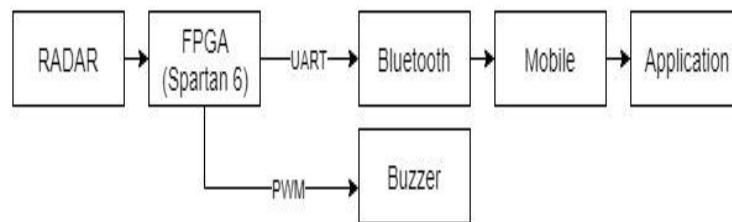


Figure 11: Block Diagram

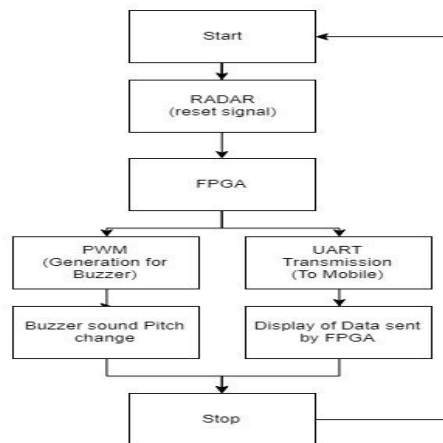


Figure 12: Program Flow Chart

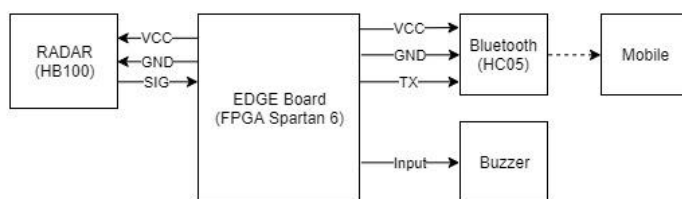


Figure 13: Hardware Connection Block Diagram



Figure 14: Hardware Setup

VII. SIMULATION AND RESULT

The main module simulation result contain the two main output signals transmission (txd) and pulse. The pulse signal is the PWM generated signal and transmission (txd) signal is for the UART transmission of the data from the FPGA board using HC05 Bluetooth module.

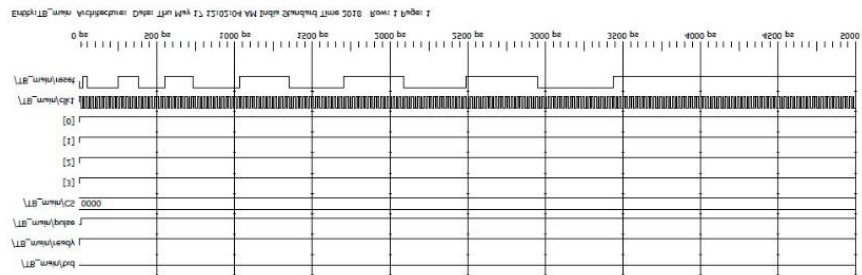


Figure 15: Main Module Simulation result

VIII. CONCLUSION

In the proposed paper which designed for the use of radar in multimedia devices is an efficient way of for the recognition of gesture. The radio waves have unique advantage of responsive to only human bodies as compared to other Video and touch based system without coming in direct contact with human. This advantage reduces the wear and tear in long term and gives it capability to operate from inside the multimedia device without coming into the direct contact. Using the radars we can achieve greater accuracy and more precise movement of the fingers which in turn helps in extracting more feature from the body.

IX. REFERENCES

- [1] Y. Li, Y. Du, X. Ye and Z. Cai, "Doppler radar real-time signal processing based on FPGA," *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, Beijing, 2016, pp. 443-446. doi: 10.1109/SIPROCESS.2016.7888301
- [2] H. M. Waidyasooriya, Y. Takei, S. Tatsumi and M. Hariyama, "OpenCL-Based FPGA-Platform for Stencil Computation and Its Optimization Methodology," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 5, pp. 1390-1402, May 1, 2017. doi: 10.1109/TPDS.2016.2614981
- [3] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network," in *IEEE Access*, vol. 4, pp. 7125-7130, 2016. doi: 10.1109/ACCESS.2016.2617282
- [4] P. S. Chinchansure and C. V. Kulkarni, "Home automation system based on FPGA and GSM," *2014 International Conference on Computer Communication and Informatics*, Coimbatore, 2014, pp. 1-5. doi: 10.1109/ICCCI.2014.6921803
- [5] S. Sharma, J. Boddu, G. S. Charan, S. Sharma, S. Sivanantham and K. Sivasankaran, "Home automation through FPGA controller," *2015 Online International Conference on Green Engineering and Technologies (IC-GET)*, Coimbatore, 2015, pp. 1-4. doi: 10.1109/GET.2015.7453813
- [6] V. C. Nguyen, D. K. Dinh, V. A. Le and V. D. Nguyen, "Length and speed detection using microwave motion sensor," *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*, Hanoi, 2014, pp. 371-376. doi: 10.1109/ATC.2014.7043414



Interfacing of Multiple FPGA using I2C protocol

¹Harshita Varshney, ²Ankur Vats, ³Isha Gupta (Project Engineer)

¹²³School of electronics, CDAC Noida
India

Abstract: – This paper shows the implementation of I2C bus protocol which is used to connect RISC processor, booth multiplier through arbitration logic on Basys 2 FPGA board.

I2C bus protocol is also called Two-wire-interface. Each I2C-transport good gadgets coordinate an on-chip interfacing which enables them to discuss straightforwardly with each other by means of the I2C-transport. The outline was combined by Xilinx ISE Design Suite 14.7. Here I2C transport convention's lord begins information transmission and slave gives an affirmation to every transmission. The protocol's implementation was verified by applying a sample communication to the synthesized hardware.

Keywords: I2C, RISC processor, booth multiplier, arbitration logic.

I. INTRODUCTION

This project consists of implementation of i2c bus protocol for interfacing of RISC processor, booth multiplier through arbitration logic on FPGA. In this project one FPGA is used as a master in which RISC processor and booth multiplier are designed and second FPGA is used as slave memory to store data.

There are different kinds of correspondence conventions like “RS-422, RS-485, RS-232, SPI” (Serial fringe interface), and Microwire for interfacing low-speed and quick peripherals contraptions. The above traditions require more stick relationship on an Integrated Circuit(IC) for serial data correspondence, when contrasted with I2C convention which just needs two pins. USB/SPI/Microwire conventions and the greater part of the UARTS are for the most part simply 'point to point' information exchange transport frameworks. These utilization information way multiplexing and message sending to give benefits between two associated gadgets. To overcome this issue, the I2C tradition was displayed by Phillips, which requires only two lines for correspondence called SDA and SCL lines. [3]

II. I2C PROTOCOL

I2C protocol consists of mainly two bus line known as data line(SDA) and clock line(SCL). SDA is bidirectional and SCL is unidirectional. [1]

The data transfer on the I2C bus can be segregated by four states known as START, Address frame, Data frame and STOP conditions. After the START signal by master, the 7-bit address send with MSB being transferred at first followed by R/W bit (8th bit) as shown in Fig1. These 8-bits together form a BYTE and are referred to as Address Frame. The next bit after Address Frame is the ACK bit which shows that the previous frame was successfully received/or not and another frame may be sent. After transfer of this address frame the 9-th bit on the SCL is used to pull low the SDA line and this phenomenon is also known as “clock stretching” because delay provides more time to slave to store the received byte or to get ready for another frame. After the stretch, data frame transfer is initiated in which 8-bit data is transferred along with an ACK bit that functions same as that of above bit in address frame. This whole data transfer is terminated by the master by generating STOP condition. [2].

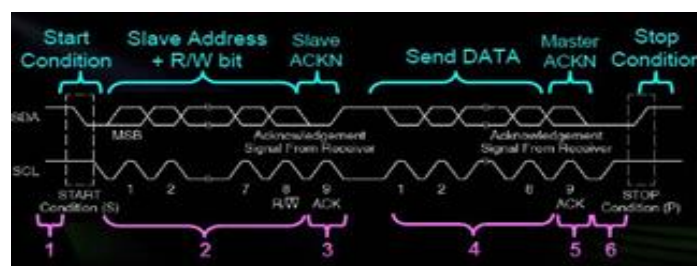


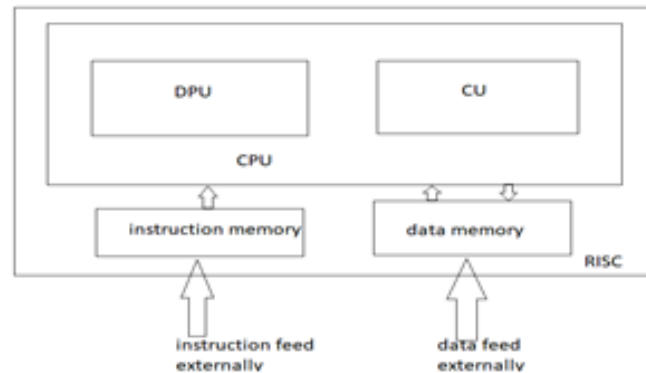
Fig1: I2C start-stop condition

III. RISC PROCESSOR

A processor can be simply defined as a logical or processing circuit that works upon the data fetched through memory to perform a specific task. And can be termed as the brain of the computer.

RISC processor is abbreviation of Reduced Instruction Set Computer. The RISC processor comprises of Central Processing Unit, Instruction memory and Data memory. In the proposed design, I2C protocol has been implemented in multiple master single slave configuration, where RISC processor acts as one of the master to the slave memory.

Central Processing Unit of the RISC comprises of two essential components: Controller Unit and Data Processing Unit as shown in fig 2.



. Fig2: RISC processor block diagram

Controller Unit

It is responsible for co-ordination of various activities like fetching of the data from the data memory, retrieving instructions from the instruction memory, storing of the data after the data processing unit has processed the result either on the address of the data memory or on address of the memory acting as slave.

Data Processing Unit

This unit is mainly responsible for processing the instruction based upon the opcode of the instruction fetched and perform the operation of the opcode with the current value of the accumulator with the value stored on the address of the data memory and store the result on the accumulator for further processing. Accumulator is an internal register on which the computation takes place.

8-bit instruction format of RISC processor

RISC processor instruction format is shown in table, first 3-bit used for opcode and last 5-bit are used as operation.

7	6	5	4	3	2	1	0
Opcode			Address of the values onto which the processor will operate upon				

Fig3: Instruction format of RISC processor

Instruction set architecture

The instructions on the instruction memory and data on the data memory is fed through the switches on the FPGA Board. The result of the accumulator is being written on the slave memory using I2C protocol.

OPCODE	FUNCTION
000	ADDITION
001	SUBTRACTION
010	BITWISE AND
011	BITWISE OR
100	BITWISE NOT
101	BITWISE XOR
110	BITWISE XNOR
111	JUMP AT LOCATION PROVIDED

Table 1: Instruction opcodes of RISC processor

IV. BOOTH MULTIPLIER

Settled width Multipliers are generally utilized as a part of sight and sound, number juggling units of microchips, and computerized flag processors. It is winding up more difficult to additionally enhance execution

and diminish the power utilization of multipliers under the prerequisite of full precision. Nonetheless, the necessities of high accuracy and rightness are not all that strict for some, applications identified with human observation, for example, sight and sound flag preparing and machine learning. High accuracy and rightness in the activities of computerized rationale configuration are identified with the for the most part acknowledged prerequisite of rightness of data.[1] Now a days Booth algorithm has been increasingly employed in the fixed-width multiplier design to speed-up and logic reduction. Booth's algorithm is a multiplication algorithm that is used multiplies two signed binary numbers in 2's complement notation.

Operation flow of booth algorithm is shown in flow chart in fig 3. Procedure to calculate the result of two sign bit is as follows:

1. Assume M is the multiplicand.
2. Assume Q is the multiplier.
3. Consider a 1-bit register Q_{n-1} and initialize it to zero.
4. Consider a register A and initialize it to zero.

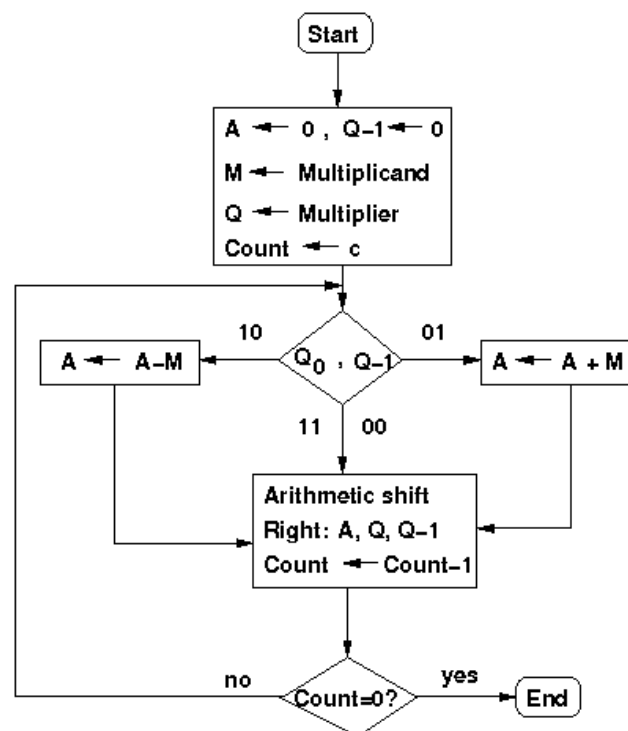


Fig4: flow chart of booth multiplier

V. ARBITRATION LOGIC

I2C protocol supports multi master configuration. In the proposed design, the architecture has mainly two arbitration logic in order to resolve the conflict when multiple master try to access the SDA line at the same time.

Arbitration based on W-AND (Wired AND)

The settlement of the dispute on the SDA line of the I2C protocol is done using the WAND logic. On ANDing the values of the two masters bit by bit, the master responsible for the output of wand to be zero wins the arbitration.

Arbitration based on inclination

In this arbitration logic, whenever both the master are trying to acquire the SDA line at the same time that is, external button for the master on the FPGA are pressed together than start sequence of the RISC processor (m2_data_in) wins the preference due to inclination and acquires the control of the bus and write its processed output on the slave memory and till the master who has won the arbitration completes its operation the master which loses its arbitration goes in the wait state where it checks the availability of the SDA line.

VI. PROPOSED DESIGN

Proposed design for interfacing of FPGA using I2C protocol is shown in figure.

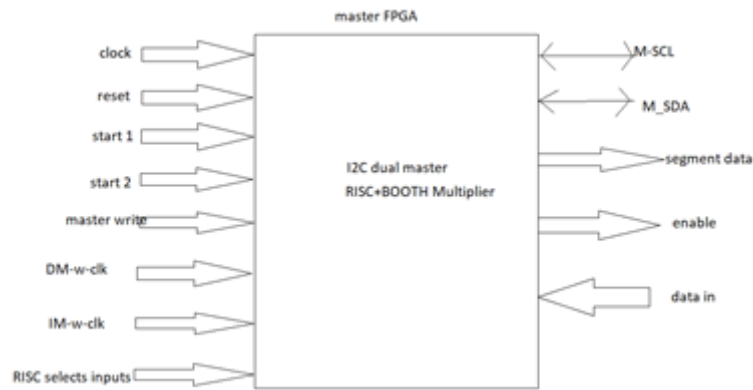


Fig4: Block diagram of FPGA as master

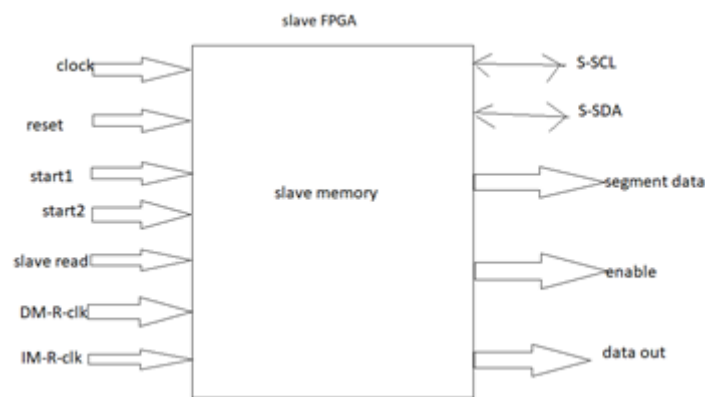


Fig5: Block diagram of FPGA as slave

VII. SIMULATION AND HARDWARE IMPLEMENTATION

In this paper Fig6 shows the hardware implementation of project in which master (FPGA shows data 0001) works as booth multiplier and RISC processor and slave (FPGA shows data 00F0) work as memory.

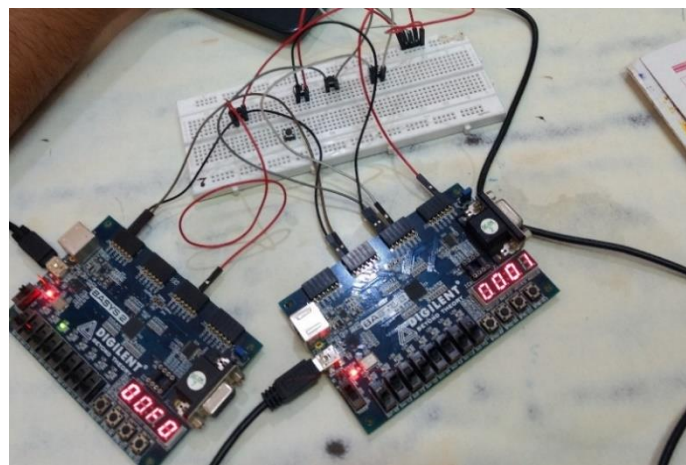


Fig6: Hardware design implementation

Fig7 shows the simulation result of I2C protocol in which master data1 is transferred to slave data out.

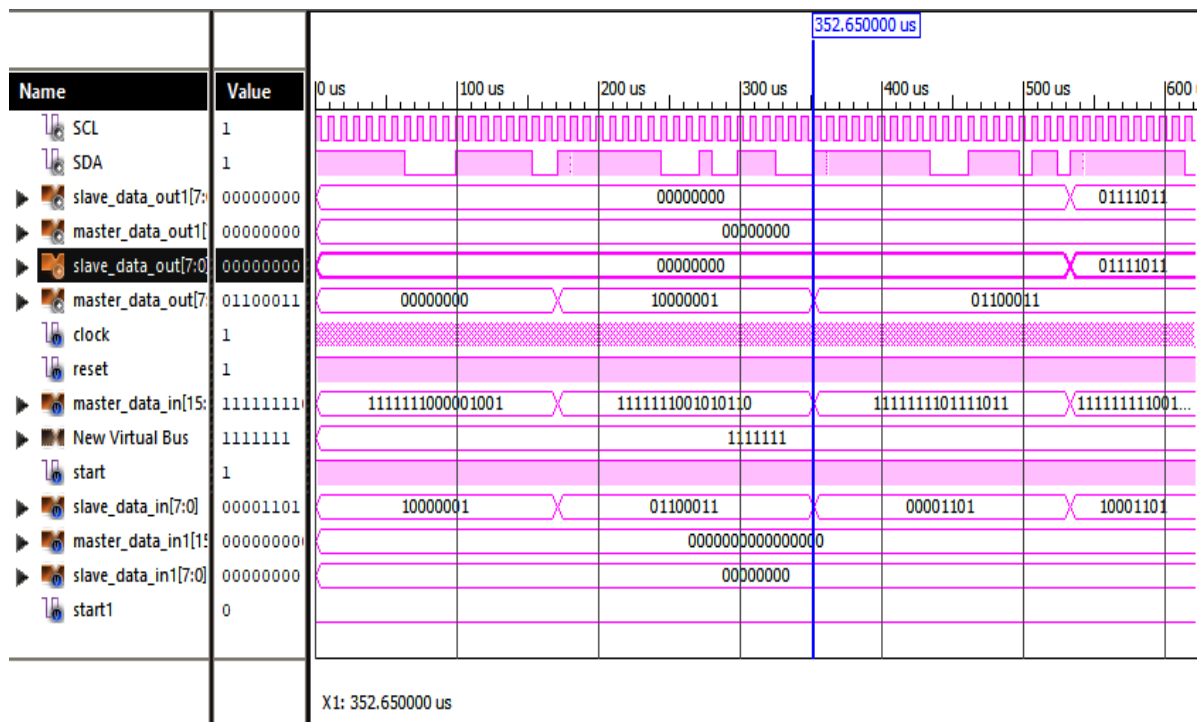


Fig7: I2C protocol simulation

Fig8 shows the simulation result of slave in which when read clock signal is high data is transferred to data out signal.

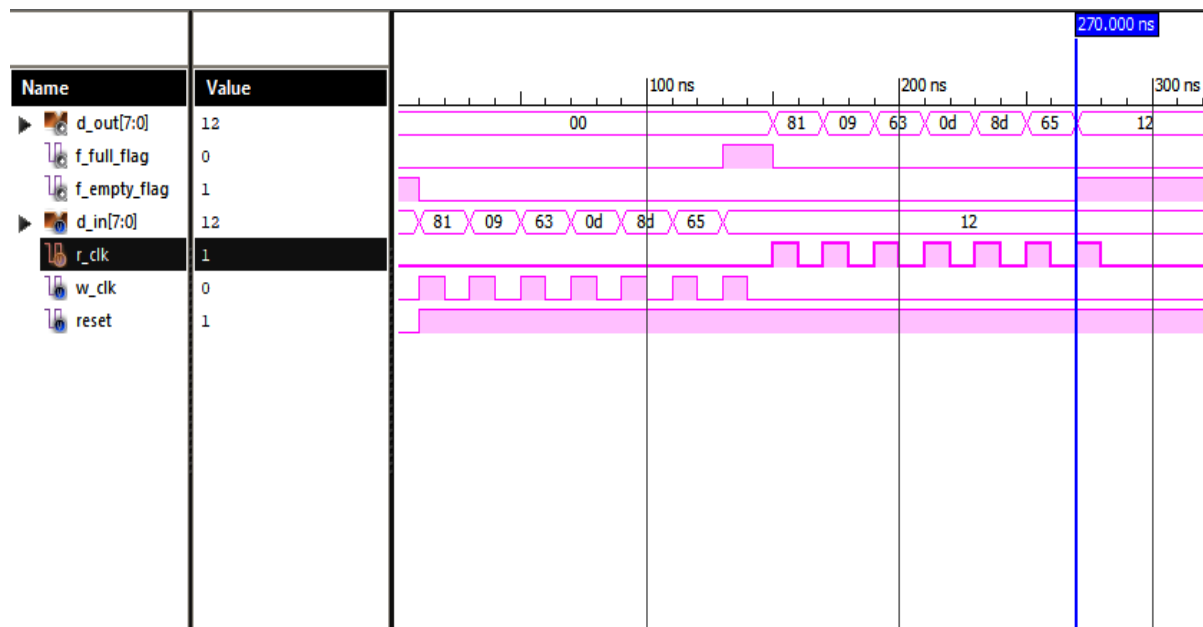


Fig8: simulation of Slave

Fig9 shows the simulation result of RISC in which depending upon the opcode of the instruction, result is generated.

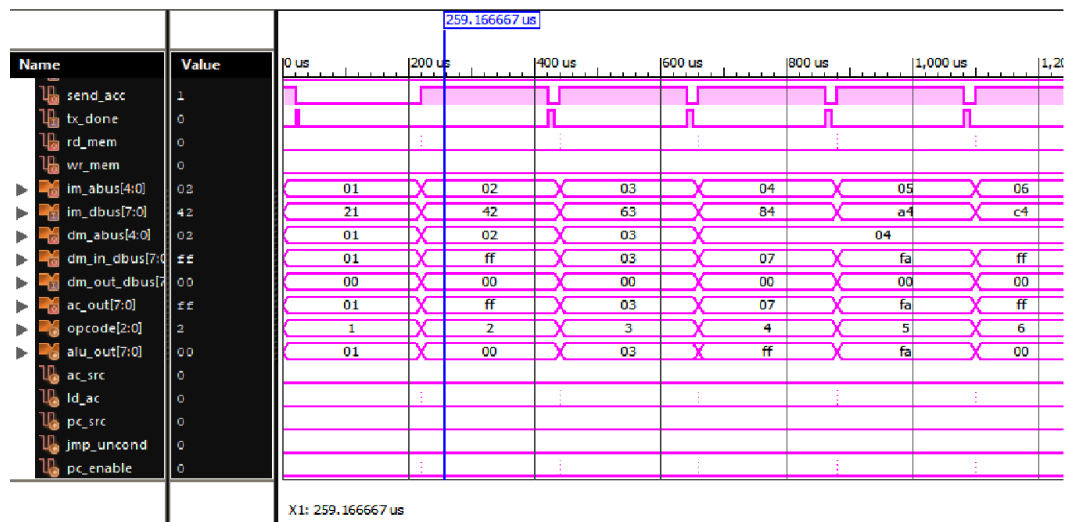


Fig9: simulation of RISC

Fig10 shows the simulation result of booth multiplier in which M is multiplicand, Q is multiplier and mult shows output.



Fig10: simulation of booth

Fig11 and Fig12 shows the arbitration result, when both master call for slave at a time it decides which master get the high signal.

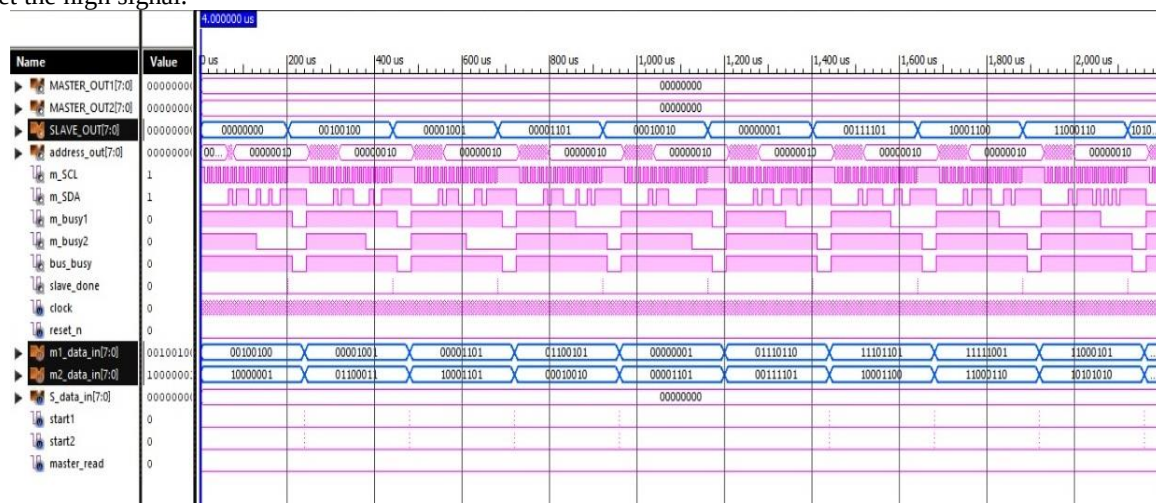


Fig11: simulation of wired AND arbitration logic

Designing and Verification of AMBA AHB LITE Protocol

Manjeet Kumar¹, Dr. Sunita Prasad²

¹ Student, School of Electronics, VLSI Design, C-DAC Noida

² Principal Technical Officer, C-DAC Noida

Guru Gobind Singh Indraprastha University

Dwarka Sector-16C, New Delhi

INDIA

Abstract: Design and Verification play a significant role in the implementation of SoC. Verification Check gives the genuine authorization and usefulness of a DUT and confirms that it meets the given requirements. This paper explain the behavior of Advanced High Performance Lite interface at System level which is a subunit of AHB protocol, an important version of the Advanced Microprocessor Bus Architecture(AMBA). It additionally incorporates the designing and verification of AHB- Lite protocol for sequential and non-sequential (incrementing and wrapping burst of different size) transfers.

Keywords: System On Chip (SOC), Advanced Microcontroller Bus Architecture (AMBA), Advanced High Performance LITE (AHB Lite) , System Verilog, QUESTASIM tool.

I. Introduction

A System on a chip or System-On-Chip (SOC) is an integrated circuit called as IC or Chip. It integrates all computer components or other electronic system mainly consists of a central processing unit (CPU), memory, input/output ports and storage elements on a single substrate. The essential part of System-On-Chip (SOC) is not which segments or components it contains, but additionally how they are connected to each other.

The AMBA (Advanced Microcontroller Bus Architecture) which is given by ARM is an on chip interconnects protocol for connecting and controlling different functional blocks at System-On-Chip design. The transport protocol used by CPU is a vital part of co-check as this is an important correspondence between the Memory, CPU and other custom gears. AMBA introduced many version of protocol for on chip communications with time like APB, AHB, AXI, CHI etc. The requirements of high performance synthesizable design are fulfilled by the AHB Lite protocol of AMBA which is a subunit of AMBA AHB protocol. AHB Lite simplify the design for interconnection which used only single Master, General methodology of verification is identified with SystemVerilog which is latest and all around requested technique utilized as a part of VLSI domain for confirmation of different SOC's and digital system devices.

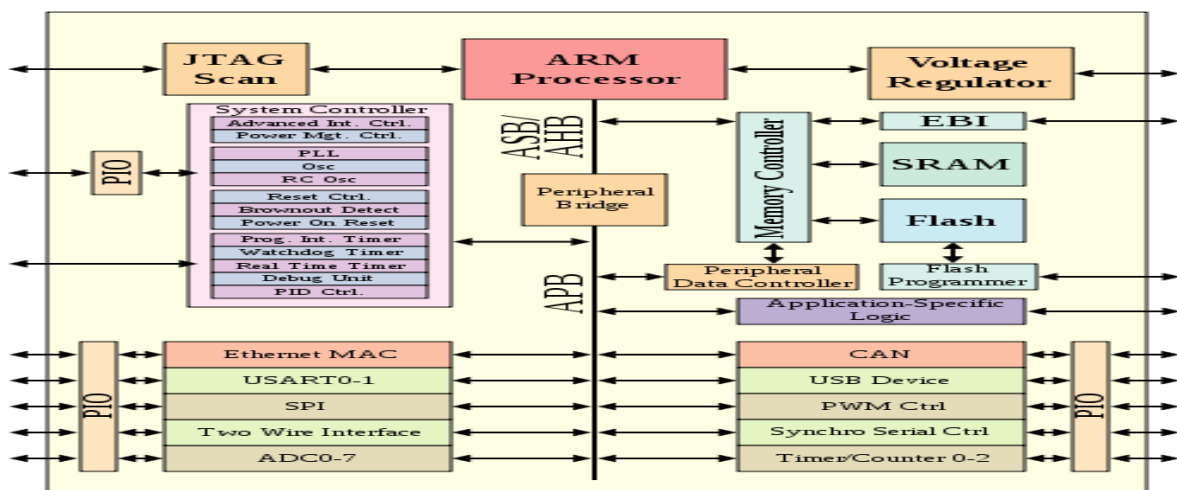


Fig.1 Block diagram of System-On-Chip

II. Protocol System of AHB Lite

The AHB LITE protocol of AMBA is implemented a synthesizable design for high performance. AHB Lite is a vehicle interface that gives a simplex transport mechanism and ensures high speed data transfer limit. AHB Lite highlights all those features which is needed for systems with high performance and high clock frequency including: [7]

- ☐ Transfer of Burst
- ☐ Operation at single clock edge
- ☐ Implementation of non tristate
- ☐ Extended bus configuration for data ex. 64bits,128 bits,256 bits,512 bits, and 1024 bits.

Slaves in AHB LITE protocol are widely recognized as memory gadgets, interfaces and high transfer limit peripherals. While peripherals with Low data transfer capacity can be fused as slaves of AHB Lite, for the reasons of system execution they mostly consist on Advanced Peripheral Bus (APB) of AMBA. The connection between high performance AHB Bus and APB is provided by slaves of AHB LITE. This connection is called APB bridge.[7]

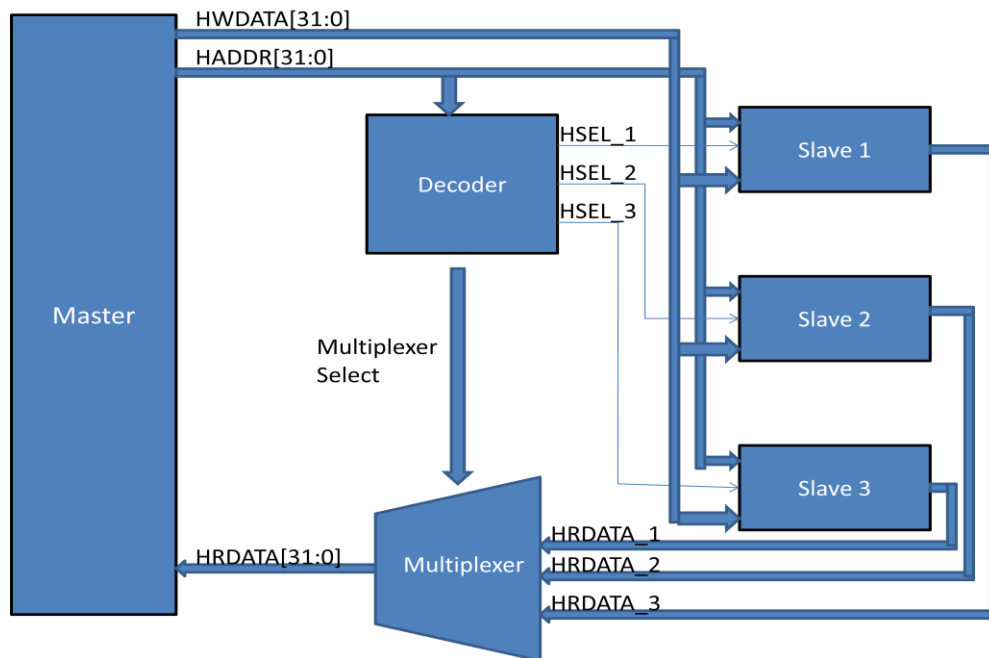


Fig.2 Block diagram of AHB Lite

In Fig.2 Block diagram represent AHB Lite system comprising of single master and multiple Slaves. The interconnect logic consist an address decoder and single multiplexor from slave to master.

The address from master is monitored by the decoder so that desirable Slave is selected and output data routes back to master through multiplexer. The AHB Lite system having following main components:

- ☐ Master
- ☐ Slave
- ☐ Decoder
- ☐ Multiplexer

AHB Lite Master: Address and control information is provided by master component to start read and write operations.

AHB Lite Slave: In system the transfers started by master is responded by slave component. Appropriate slave is selected using **HSEL_x** signal from the decoder to control the transfer. **Decoder:** Address of each transaction is decoded by decoder and a select signal is given for particular slave to start transfer. **Multiplexer:** A multiplexer from slave to master in AHB Lite implementation is used to control data and response signal from different slaves to master.

Working of AHB Lite:

By using address and control information master starts a transfer. Address and control signal gives information regarding address, size, direction of transfer and also mentioning whether it is a part of burst transfer or not. There are three types of transfer:[7]

- ☐ Single Burst (fixed)

- ☐ Incrementing Burst (Not wrap at Address Boundary)
- ☐ Wrapping Burst (Wrap at Address Boundary)

TABLE I. Transfer Types

Transfer Type	Description	HTRANS[1:0]
IDLE	data transfer is not required	00
BUSY	Masters are enable to add idle cycle between the burst transfer	01
NON- SEQUENTIAL	Address and Control Signals are not belong with last transaction .	10
SEQUENTIAL	Address and Control Signals are belong to last transaction.	11

Data is transported from master to slave in write transaction and slave to master in read transaction. Transaction occurs in two phases: [7]

- 1) Address Phase- Single Address and Control cycle
- 2) Data Phase - For data, single and many cycles.

Address phases cannot be increased, so during address phase every slaves must be able to sample addresses. Although, by using HREADY signal slave can make request to increased data phase. When HREADY signal is low wait states are inserted to provide extra time to slave for sampling the data. The success and failure of transfer is represented by HRESP signal.

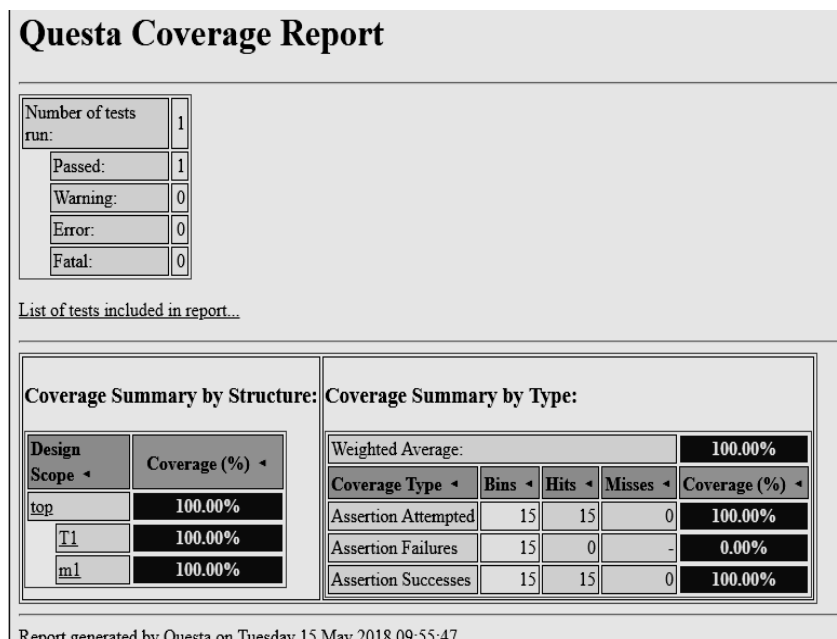
TABLE II. Transfer Responses

HRESP	HREADY	
	0	1
0	Transfer is Pending	Transfer is Successfully completed
1	Respond Error, First Cycle	Respond Error, Second Cycle

III. AHB-Lite Advantages over AHB

Following cases are not required to follow by bus master of AHB Lite protocol: [8]

Fig. 3 Coverage Report without injecting errors into the design



- Arbitration Technique
- Chance to lost bus control. The HREADY signal on bus can be used to drive clock enable signal for master.
- Burst termination in early stages. Burst is not needed to rebuild because of early termination as bus can be used always by the master.

- HBUSREQ Output signal is not exist on master side. If such signal exist left it unconnected.
- HGRANT Input signal is not present on master side. If such signal exist tied to High.
- As Split or Retry responses are absent, Master does not hold the address of last transfer and able to restart last pending data transfer.

IV. Environment Approach

Master Block: Bus Functional Model (BFM) approach is used to write master component (not synthesizable). It consists a no. of write and read tasks- that include the way the DUT is utilized. A read and write tasks would occur on DUT Bus by calling those tasks.

Slave Block: It is just a memory component which is our DUT and is programmed to be synthesizable. We have parameterized the quantity of slaves utilized as a part of our module.

Monitor: This module block monitors the traffic on bus and checks if the protocol is being followed by the ongoing transactions using assertions technique. The results have been mention in an Assertion Scoreboard.

Interface: Reside of all the bus related signals and a slave modport to give an appropriate query to the slave.

V. Simulation Result

Simulation Output is observed on QuestaSim tool of mentor graphics using System Verilog as Hardware Verification Language (HVL). Different test combinations are applied for different Operations. The AHB Lite Slave module is verified by applying multiple test patterns. In this paper configurable address space of slave is between 000H to 3FFH. The address continues for every next slave i.e. the 2nd slave has addresses from 400H to 7FFH and so on. 00H to 03H addresses in every RAM are configured to be read-only addresses. To perform various transactions like write and read operations between master and slave, their values and concatenated input is passed to invoke a function. Bins have been written based on the constraints and 100% functional coverage without injecting errors into the design and 93.33% functional coverage by injecting errors into the design has been obtained. Assertion scoreboard also developed to validate the behavior of DUT. Verilog and SystemVerilog are provided as programming and verification languages. The result is achieved on the latest version of tool i.e., QUESTA SIM 32 10.0 B tool (from Mentor Graphics).

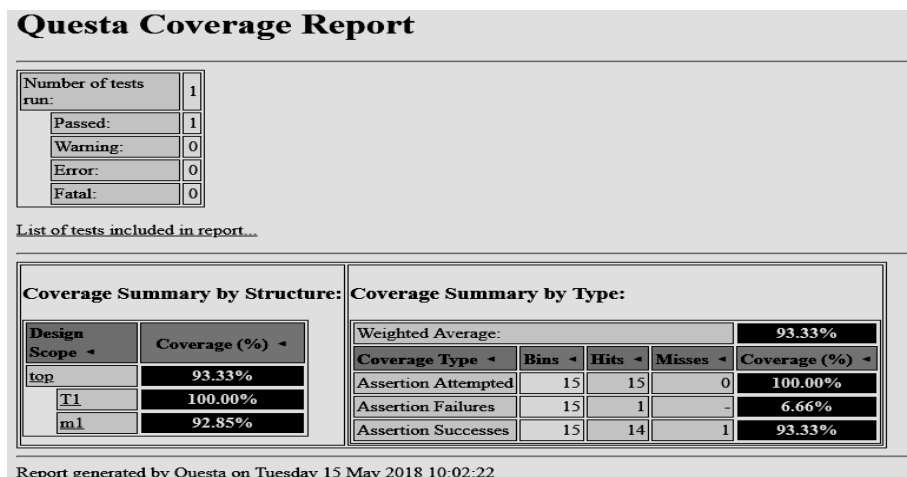


Fig. 4 Coverage Report by injecting errors into the design

-----ASSERTION SCOREBOARD-----			

TYPE OF CHECK	TOTAL COUNT	PASS COUNT	FAIL COUNT
error_check	0	0	0
read_only_error_check	4	4	0
basic_write_error	4	4	0
basic_read_error_check	4	4	0
basic_burst_write_check	2	2	0
basic_burst_read_check	4	4	0
HREADY_check	0	0	0
bursts_count_check4	2	2	0
bursts_count_check8	1	1	0
bursts_count_check16	0	0	0
address_change4	3	3	0
address_change8	3	3	0
address_change16	0	0	0

Fig. 5: Scoreboard

Figure-3 and Figure-4 shows 100% functional coverage without injecting errors into the design and 93.33% functional coverage by injecting errors into the design has been obtained.

VI. Conclusion

In this paper, AHB LITE interconnect environment code is simulated and corresponding output is observed. The functional Verification of the AHB LITE protocol and resultant simulation waveforms are obtained by the mentor graphics QUESTASIM tool and also examine as per the expectation without any change in the features of DUT. The DUT under verification is totally wraps by the environment and observes the optimum functional based coverage. Bins have been written based on the constraints and 100% functional coverage without injecting errors into the design and 93.33% functional coverage by injecting errors into the design has been obtained. Assertion scoreboard also, developed to validate the behavior of DUT. The result is achieved on the latest version of tool i.e., QUESTA SIM 32 10.0 B tool.

References

- [1] Srinivas, M.B.R. and Musala, S., 2016, August. Verification of AHB_LITE protocol for waited transfer responses using re-usable verification methodology. In *Inventive Computation Technologies (ICICT), International Conference on* (Vol. 3, pp. 1-3). IEEE.
- [2] Gurha, P. and Khandelwal, R.R., 2016, September. SystemVerilog Assertion Based Verification of AMBA- AHB. In *Micro-Electronics and Telecommunication Engineering (ICMETE), 2016 International Conference on* (pp. 641-645). IEEE.
- [3] Mahesh, G. and Sakthivel, S.M., 2015, March. Functional verification of the Axi2OCP bridge using system verilog and effective bus utilization calculation for AMBA AXI 3.0 protocol. In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on* (pp. 1-5). IEEE.
- [4] Lam, N.S. and Nguyen, D.M., 2014, July. AHB-master controller formal compliance verification. In *Communications and Electronics (ICCE), 2014 IEEE Fifth International Conference on* (pp. 340-345). IEEE.
- [5] Singh, A.K., Shrivastava, A. and Tomar, G.S., 2011, June. Design and implementation of high performance AHB reconfigurable arbiter for onchip bus architecture. In *Communication Systems and Network Technologies (CSNT), 2011 International Conference on* (pp. 455-459). IEEE.
- [6] Mulani, P.D., 2009, December. SoC level verification using System Verilog. In *Emerging Trends in Engineering and Technology (ICETET), 2009 2nd International Conference on* (pp. 378-380). IEEE.
- [7] Specification of AHB LITE AMBA-3 ,available at <http://www.Arm.com>
- [8] Specification of AMBA (Rev 2.0) that available at <http://www.Arm.com>
- [9] Palnitkar, S., 2003. *Verilog HDL: a guide to digital design and synthesis* (Vol. 1). Prentice Hall Professional.
- [10] Spear, C., 2008. *SystemVerilog for verification: a guide to learning the testbench language features*. Springer Science & Business Media.



Piezo-electric Energy Harvesters: A Battery less Solution for Advance Technology

Garima Bandhawakar Wakhle

Department of Electronics & Communication Engineering

School of ICT(Guest)

Gautam Buddha University, Gr. Noida, UP

INDIA

Abstract: With advancement in technology emphasis is to design low power, low voltage devices to further trim down the device size. Not just the shrinking of device size is a solution but another important aspect need to be focused is the batteries required to run the devices especially for WSN which are considered as back bone of IoT. The batteries still have problem of limited life span and essential to be changed or recharged by external source after an interval of time, which seems to be a difficult task in many applications like medical instrumentation, military applications, Industrial automation and IoT. In the present article the objective is to provide an overview on working principle of Energy Harvesting sensors, materials used for designing, it's properties, applications and future prospects.

Keywords: MEMS, PEH, IoT, Unimorph, Bimorph

I. Introduction

With the progression of technology, MEMS based energy harvesting devices has become the part of our day to day life instead of being a niche technology. With the advancement in research on WSN around 1998, it has become an integral part in computing, communication and micro-electromechanical technology [1,2], they are highly in demand in various fields like healthcare, Industrial automation, defence and now IoT. But the limited life span of batteries is the biggest constraint for WSN, a novel approach to increase the lifetime of batteries preferably unlimited become possible with energy harvesting technique. Vibrations and electrostatic transduction method can be employed to transfer mechanical energy into electrical energy, are one of the promising ways for power generation. The devices designed using this technique will be the self-power generators and replace the batteries perpetually [3].

Various sources of energy are available in the nature as fluid flow, machine rotation, vehicles, bridges etc., which produces a significant vibrational energy that can be utilized to convert in to the electrical energy. Piezoelectric materials have a unique ability to convert mechanical energy into electrical energy using MEMS generators and nanogenerators devices. The devices based on PEH techniques are the novel alternative approach for powering wireless sensor nodes [4,5] and provides incessant source of energy.

Objective of the present review is to provide a detailed summary of the energy harvesting technologies, it's working principle, modeling, designing aspects and future prospects. With recent development and advancement in IoT systems, an emphasis will also be given to discuss the energy harvesting techniques for these systems which are preferably WEH based on RF sensors.

II. Working Principle

The Piezoelectric devices work on the principle of second order mass spring damping system [6-8]. These inertial transducer generators offer the motion based on spring-mass damper, by second order differential equation as follows:

$$m \frac{dz^2}{dt^2} + b \frac{dz}{dt} + kz = F(t) \quad (1)$$

In the above equation m , b and k are seismic mass, total damping coefficient and spring stiffness respectively, while $F(t)$ is an external force applied onto the device. The damping coefficient b comprises of both mechanical losses and coefficient based on energy conversion, which are b_p and b_e , respectively. k_{spring} and k_{piezo} are the spring stiffness and the piezoelectric stiffness that gives the total stiffness of spring. This approximation is good as long as the structure vibrates with little displacements and the mechanical behaviour of motion remains linear.

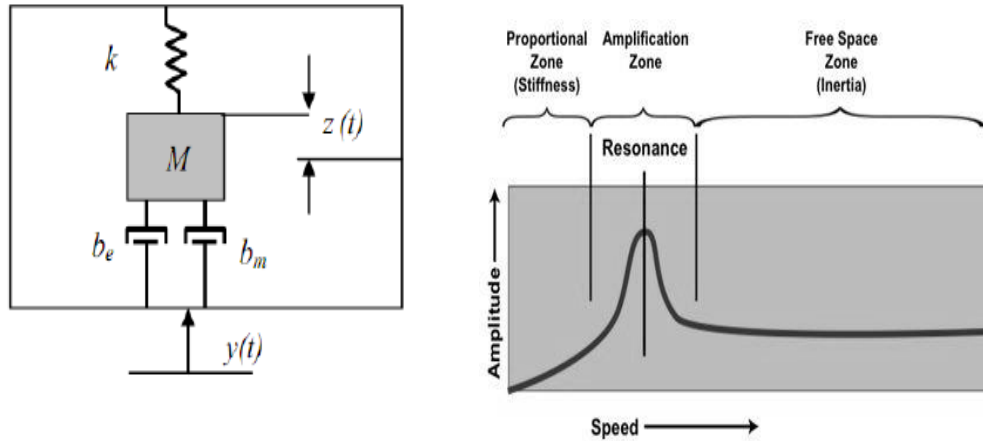


Fig.1: Shows mechanical mass spring damper for PEH and it's resonance response

The working concept of a PEH is based on the basic principal of mechanical damping, resonant vibration in mechanical structures occurs when a natural frequency is at or close to a forcing frequency. The most notable characteristics of resonance are increased vibration. i.e. when a certain operating speed is reached at certain frequency. It has also been observed that as the operating speed is increased beyond the resonant frequency, the vibration amplitude will start to decrease.

The resonant angular frequency for the vibrator is defined as:

$$\omega_n = \sqrt{\frac{k}{m}} \quad (2)$$

ω_n is the resonance frequency at which device provides maximum energy.

The pressure exerted due to vibrations is

$$P_e = \frac{m\zeta_e a^2}{4\omega_n(\zeta_p + \zeta_e)^2} \quad (3)$$

Where $a = \omega_n^2 Z$ is the acceleration for sin wave propagation and displacement is defined by a sine wave equation $z_{fr} = Z \sin(\omega t)$. The mass is m and ζ_p, ζ_e are the mechanical and transducer damping ratio. Output power reaches the maximum value at resonance but the high damping value flattens the value of Power (P_e). [8-11].

The harvesting devices have different modes of operation as d_{31}, d_{33} and d_{15} , performance of the devices significantly depends upon it. The most commonly used mode of operation is d_{31} : this mode of operation has stress along the direction 1 and generates electric field due to polarization in the direction 3. The power density always has less value in d_{31} mode as compare to d_{33} because stress is not applied along the d_{33} direction [12].

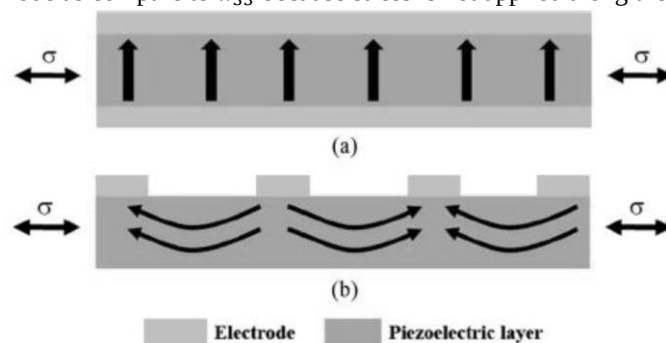


Fig.2: Polarization in piezo-material in various operating modes (a) d_{31} (TBE) and (b) d_{33} (IDE)

The most commonly used design aspects of a cantilever beam have two modes of operation: Unimorph and Bimorph. The bimorph structures are most commonly used for PEH compared to unimorph as the layer of piezo-material on both side of the beam enhances the efficiency and doubles the output power or energy [8].

III. Architectural Design of MEMS based PEH

In this section our objective is to deliberate about the complete electronic circuit of energy harvester using MEMS cantilever beam. The basic block diagram of a PEH shown in figure 3, explains various subdivisions of an energy harvester.

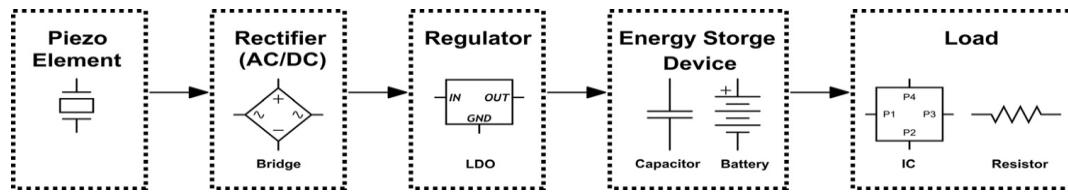


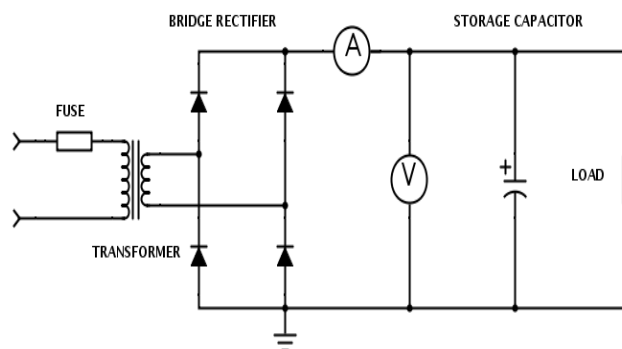
Fig 3: Basic Block Diagram of electronic circuit of PEH

The electronic circuit of PEH consists of several segments like AC-DC rectifier, a voltage regulator and an energy storing device. As the piezo-element or the MEMS based energy harvesters are subjected to vibrations or mechanical stress, a voltage (electric potential) is generated across the layers of the material, due to strain is known as **piezoelectric current**. Each segment of above shown circuit is discussed briefly:

(a). Rectification Circuits

For the rectification of voltage various types of rectifier circuits can be used, most commonly used rectifiers are AC-DC full wave or half wave bridge rectifier. [13,14] A full wave rectifier is a good selection as it converts the entire AC signal into DC with four diodes.

Fig 4: Basic circuit of a diode bridge harvesting rectifier



The SSHI circuits have two type of configurations as parallel and series combination. The Parallel SSHI is one approach for rectification, that allows enhancing the electro-mechanical coupling coefficient and increases gain upto 10 times, in the harvested energy [15-16]. The parallel-SSHI consists of two switches, one for positive half and another for negative half, on the contrary, a series-SSHI rectifier has been designed using an inductor switch branch in series to piezoelectric material in addition with a diode bridge rectifier [17,18].

(b). Voltage Regulator Circuits for Energy Harvesters

After rectification voltage generated from the device still needs to be regulated. Two most commonly used voltage regulators in energy harvesting are step-down and step-up. Step down converters regulate high input voltage to low input voltage. Various rectifier circuits have been designed and used for commercial purposes.

Tyahiet.al. [19] has designed a commercial, step-down converter (LTC1474, Linear Technology) for powering remote sensing network. Once the output voltage of the piezoelectric element had been rectified, a reservoir capacitor is used to charge and store the harvested energy. These capacitor already have a pre-set value and as this value is obtained, it discharged into the step-down converter to provide a stabilized output voltage for the load circuit.

For an in-shoe RF tag systems [20], a linear regulator has been used in a power condition circuit and the output voltage exhibit some ripples during transmission. The excitation characteristics of the application due to the piezoelectric element results in a high impedance source with high voltage and low current. A higher efficiency of approx. 7.6% makes RF tag linear regulator, more attractive and provides 1.3mW power at walking frequency of 0.8 Hz.

Ottmanet.al.[21] designed an adaptively controlled regulator circuit to maximize the power input to electrochemical battery. The power stored in the battery can be maximized by maximizing the current flowing in the battery, the results show 400% improvement compared to battery charged directly. Later on, Ottmanet.al. [22] analyzed a discontinuous conduction mode (DCM), which determines the maximum power flow.

(c). Various type of Storage Device

The regulated output should finally be stored in a compatible storage device, these devices can be either a battery or a super capacitor. The output power obtained from Piezoelectric energy harvesters are low therefore these devices cannot be directly able to power the electronic devices hence first of all the energy is accumulated in the storage device then applied to the load. Both the storage devices have distinct characteristics, many researchers used capacitors as energy storage devices because unlike batteries they don't need a minimum voltage to start charging. M.Umedaet.al.[23] and N.G.Elvinet.al.[24] used capacitors as storage medium basic benefit of these

devices i.e. they don't need minimum voltage to start charging, due to their high power density they accumulate energy instantly. On the other hand, the batteries are the sustainable solution to provide the constant source of energy and power with intermittent vibrations and limited charging cycle [25].

For piezo-electric energy harvesters a new storage device has been designed known as supercapacitors, these capacitors are different from conventional capacitors. The basic design of the capacitor consists of electrodes with very large surface areas and the distance between the two plates is less than 1nm thus achieve much higher capacitance. These electrodes hold very large surface areas therefore a large amount of energy can be stored as compare to the conventional capacitors. Guan and Liao[25] compared the supercapacitors with NiMH and Lithium batteries and concluded that supercapacitors had the highest charging/discharging capacity, around 95% high as compare to these batteries. Due to the larger surface area the basic disadvantage of supercapacitors is, their ownself-discharging rate which is higher than a rechargeable battery. [26,27]. Now a days number of commercial IC's are available such as the LTC 3588-1 by linear technology [28] and Texas instruments BQ25505[29]. Recently, a ZnO based IC LTC3588-1 has been used to power an electric watch [30]. A DC-DC converter includes a high-efficiency boost charger to output voltage that directly used to charge the supercapacitor, is a power management IC as TI's BQ25505. Wensi wang *et.al.* [31] in 2013, a new design of power management circuits is proposed in order to utilize the low ESR characteristics of super-capacitor and the low leakage current characteristics of the TFB in the hybrid energy storage.

IV. Conclusions

With advancement in microelectromechanical systems, it has enabled the micro fabrication in various domains with diverse applications from AI, IoT, cell culture studies, Bio-medical applications, drug therapy, diagnosis and many more. In the year 2016, due to technical innovations, accelerometers are the leading segment with wide applications in bio-medical sector and having revolutionized treatment of cardiac diseases as cardiac MEMS, bionics, ENT implants etc. The period 2017-2025 have been defined as the era of fastest growth in the R&D and application of MEMS market to compete, to augment their foothold in the market. With the fastest growing development in AI and IoT the growth projections for MEMS based sensor has been raised to \$ 40 billion in 2015 to \$ 200 billion in 2025. Various reported work indicates that the new product developments reflect the requirement for smaller and lower-cost sensors with enhanced performance and greater functionality. It is expected that the MEMS business is yet to come from the products that are in early stages of the development or yet to be invented, including integrated optical switching, processing chips, various RF communication and remote sensing products. So, not just the development and advancement in the technology but a gigantic market is looking for a robust, low cost and low power solutions.

References

- [1] G.J. Pottie, "Wireless sensor networks" *Information Theory Workshop*, 1998, 22-26 June 1998, Killarney, Ireland.
- [2] K. Sohrabi; J. Gao; V. Ailawadhi; G.J. Pottie, "Protocols for self-organization of a wireless sensor network" *IEEE Personal Communications* (Volume: 7, Issue: 5, Oct 2000)
- [3] Beeby, S.P.; Tudor, M.J.; White, N., "Energy harvesting vibration sources for Microsystem applications." *Meas. Sci. Technol.* **2006**, *17*, R175–R195.
- [4] P. X. Gao, J. Song, J. Liu, and Z. L. Wang, "Nanowire piezoelectric nanogenerators on plastic substrates as flexible power sources for nanodevices," *Advanced Materials*, vol. 19, no. 1, pp. 67–72, 2007.
- [5] C. Sun, J. Shi, and X. Wang, "Fundamental study of mechanical energy harvesting using piezoelectric nanostructures," *Journal of Applied Physics*, vol. 108, no. 3, Article ID 034309, 2010.
- [6] Priya, S.; Inman, D.J. "Energy Harvesting Technologies" Springer: Berlin, Germany, 2009.
- [7] Ferrari, M.; Ferrari, V.; Guizzetti, M.; Marioli, D.; Taroni, A. "Piezoelectric multifrequency energy converter for power harvesting in autonomous Microsystems" *Sens. Actuators A Phys.* **2008**, *142*, 329–335.
- [8] Williams, C.; Yates, R.B. "Analysis of a micro-electric generator for Microsystems" *Sens. Actuators A Phys.* **1996**, *52*, 8–11.
- [9] Roundy, S.; Wright, P.K.; Rabaey, J. "A study of low level vibrations as a power source for wireless sensor nodes" *Comput. Commun.* **2003**, *26*, 1131–1144.
- [10] Beeby, S.P.; Tudor, M.J.; White, N. "Energy harvesting vibration sources for microsystems applications" *Meas. Sci. Technol.* **2006**, *17*, R175–R195.
- [11] Gafforelli, G. Energy Harvesting: State of Art. Online: http://ftp.stru.polimi.it/incoming/gafforelli/Energy%20Harvesting/Pres_EHState%20of%20art.pdf (accessed on 28 February 2014).
- [12] [Principles of Piezoelectric Energy Harvesting](#), May 12, 2015, By: APC, Posted in: [Piezo Applications](#).
- [56] Ottman, G.K.; Hofmann, H.F.; Bhatt, A.C.; Lesieutre, G.A., "Adaptive piezoelectric energy harvesting circuit for wireless remote power supply" *IEEE Trans. Power Electron.* **2002**, *17*, 669–676.
- [13] Ottman, G.K.; Hofmann, H.F.; Lesieutre, G.A., "Optimized piezoelectric energy harvesting circuit using step-down converter in discontinuous conduction mode", *IEEE Trans. Power Electron.* **2003**, *18*, 696–703.
- [14] Lefevre, E.; Badel, A.; Richard, C.; Guyomar, D., "High-Performance Piezoelectric Vibration Energy Reclamation" In *Smart Structures and Materials*; International Society for Optics and Photonics, San Diego, CA, USA, 2004; pp. 379–387.
- [15] Guyomar, D.; Badel, A.; Lefevre, E.; Richard, C. Toward, "Energy harvesting using active materials and conversion improvement by nonlinear processing" *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2005**, *52*, 584–595.
- [16] Grace, E.A.; Rajan, S.E.; Asis, A.A.C., "Performance evaluation of different rectifiers for piezo-electric energy harvesting applications", In Proceedings of the IEEE 2011 International Conference on Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEECE), Sivakasi, Indian, 15–17 December 2011; pp. 248–252.
- [17] Shu, Y.; Lien, I.; Wu, W., "An improved analysis of the SSHI interface in piezoelectric energy harvesting" *Smart Mater. Struct.* **2007**, *16*, 2253.

- [18] Garbuio, L.; Lallart, M.; Guyomar, D.; Richard, C.; Audigier, D., "Mechanical energy harvester with ultralow threshold rectification based on SSHI nonlinear technique" *IEEE Trans. Ind. Electron.* **2009**, 56, 1048–1056.
- [19] M. B. Tayahi, B. Johnson, M. Holtzman, and G. Cadet, at the Performance, Computing, and Communications Conference, 2005.
- [20] N. S. Shenck and J. A. Paradiso, *IEEE Micro* 21(3), 30–42 (2001).
- [21] G. K. Ottman, H. F. Hofmann, A. C. Bhatt, and G. A. Lesieutre, *IEEE Trans. Power Electronics* 17(5), 669–676 (2002).
- [22] G. K. Ottman, H. F. Hofmann, and G. A. Lesieutre, *IEEE Trans. Power Electron.* 18(2), 696–703 (2003).
- [23] M. Umeda, K. Nakamura, and S. Ueha, *Jpn. J. Appl. Phys., Part 1* 36(5B), 3146–3151 (1997).
- [24] N. G. Elvin, A. A. Elvin, and M. Spector, *Smart Mater. Struct.* 10(2), 293–299 (2001).
- [25] H. Chen, T. Cong, W. Yang, C. Tan, Y. Li, and Y. Ding, *Prog. Nat. Sci.* 19(3), 291–312 (2009).
- [26] R. Kotz and M. Carlen, *Electrochim. Acta* 45(15–16), 2483–2498 (2000).
- [27] M. S. Halper and J. C. Ellenbogen, "Supercapacitors: A brief overview," Report No. MP 05W0000272, 2006.
- [28] Linear Technology, see <http://www.linear.com/product/LTC3588-1> for details of the LTC3588-1 energy harvesting circuit.
- [29] Texas Instruments see <http://www.ti.com/product/bq25505> for details of the BQ25505 energy harvesting power management IC, 05W0000272, 2006.
- [30] Y. Hu, L. Lin, Y. Zhang, and Z. Wang, *Adv. Mater.* 24(1), 110–114 (2012).
- [31] Wensi Wang, Ningning Wang, Alessandro Vinco, Rashid Siddique, Mike Hayes, Brendan O'Flynn, Cian O'Mathuna, "Super-capacitor and Thin Film Battery Hybrid Energy Storage for Energy Harvesting Applications ", *Journal of Physics: Conference Series* 476(2013) 012105, doi:10.1088/1742-6596/476/1/012105.



Face Recognition by Removing Illumination Effect

Shivani Rawat¹, Tushar Patnaik²

C-DAC, Noida

GGSIIP University, Delhi

Abstract—Face recognition is one of the most widely used biometric techniques. Lots of research has been done in face recognition, but it is always a challenge to detect the face. One of the biggest problem is illumination variations. To overcome this problem, facial recognition systems have been proposed by many researchers. The goal of this paper is to improve the recognition accuracy under illumination variations and reduces the false positive rate. Firstly, those images will be taken which contain illumination. Viola Jones algorithm is used to detect the face region. Cascade Object Detector uses, viola jones algorithm function use to detect the mouth, eyes and upper body etc. Next, during pre-processing, Binarization is done. Median filter is used to remove the illumination variation and noise presents from images by still preserving the important facial features. Then, principal component analysis is applied to images to recognize the faces by comparing feature vector of test image with feature vector of images. Finally by different classifiers, we Train the dataset to identify the human faces from the given images irrespective of the illumination conditions.

Keywords – Illumination, pose, occlusions, Face Detection, Face Recognition

I. INTRODUCTION

In the era of artificial intelligence lot of work have been done in frontal face detection system, but it is always a challenge to detect the face in the following conditions:-

- Multiple faces
- Image resolution
- Scale
- Orientation (in-plane rotation)
- Pose (out-of-plane rotation)
- People's expressions and gestures
- Aging,
- Occlusion (partial hiding of features by clothing, shadows, obstructions, etc.

So there may arise some false detected regions of an image, which do not contain any face. Taken care of all these problems there are lots of techniques available. Face are those structure which can change its color, shape and size, the detection of Face become more complicated when the lightning are not good and the person is wearing spectacles, glass and scarfs etc. There are so many applications in face detection area like computer vision communication etc. by use of these applications, it got more attraction towards evolving technology. There are so many problem in face detection like illumination problem which are caused by high brightness, change of contrast, distance from the image, incorrect settings of camera, shadows of the other person on the focused image, etc., the change in pose, multiple faces, aging effect. So by these problem it is very difficult to identify the correct face from the image.

As need of high demands of security raised, the technology grows day by day to fulfill these requirements. New creation, development of any software, enterprise should be uncomplicated and must be acceptable for users. And Biometrics are seeking attention of end user towards itself by making its system more user friendly and save the original identity by not losing the important facial features. Biometrics is the evolving area of bioengineering. In biometrics we can recognize the person by its physical appearance or also by behavioral characteristic. There are so many biometric systems for example signature, finger prints, voice, iris, retina, hand geometry, ear geometry, and face. Among all these recognition system, face recognition system is the most popular and emerging system. This system actually helpful in biometrics, its detection capacity is far better than the other systems. We can also call "Biometric face recognition" as automatic face recognition. This biometrics system focused on to differentiate the one person face to another person. The main goal of face recognition system is to understand the firm structure of human and to get the exact match of the human face from the recognition system in order to get the exact accuracy.

There are two steps in face recognition and these are:

1. To verify the face.
2. To identify the face.

In the first step, first we have to detect the face and locate the facial part of the image. Now in second step we have to extract the facial feature and then train the features and recognized the face from the database by matching the exact person. There are so many methods for face recognition is proposed. There are so many classifier algorithms present for classification of the face. These are some possible high-level classification which is used:

- Holistic Methods: In this method, the input is the whole face of the person is taken as the raw image.
- Local Feature-based Methods: In this method, eyes, nose, mouth etc. facial features are to be extracted. The location and the appearance of face are taken as input in the recognition step. In this paper, an approach is proposed to reduce the false positive rate of the face and to increase the accuracy of the system by preserving the most important features of the face irrespective of illuminations conditions.

II. PROPOSED APPROACH

In this proposed approach certain steps are taken to recognized the face from the images, these are show in below figure 1

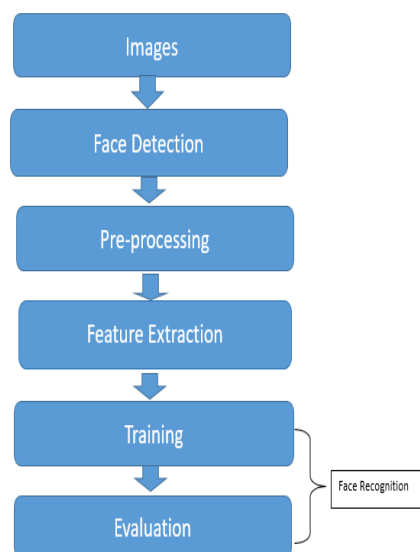


Figure:1 Proposed approach for Face Recognition

To overcome the illumination, the first step acquisition of image must be improved. Images to be taken contain the illumination variations.

1. Face Detection

Take the images containing illumination variations, it uses the viola jones algorithms to extract the eyes, nose, and mouth from the face. The cascade object detector uses viola jones to detect various part of the face. So, it will detect the human face from the images and make a boundary box around the face. Then in the next step, to remove the false detection of human face area, we applied a condition, if eyes is detected in images then it will further take for pre-processing otherwise discard the images.

Step1: Image acquisition is done which contain the illuminated images.

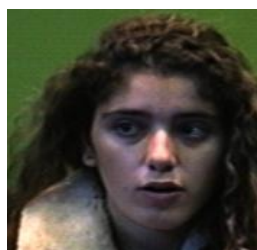


Figure 2: Original image

Step 2: Cascade Object Detector (), it will detect the human face from the images and make a boundary box around the face.

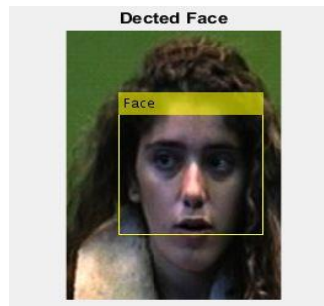


Figure 3: Detected face

Step 3: To remove the false detection of certain area, we applied a condition that if eye pair is detected in the image then only consider the image for further pre-processing otherwise discard the image.

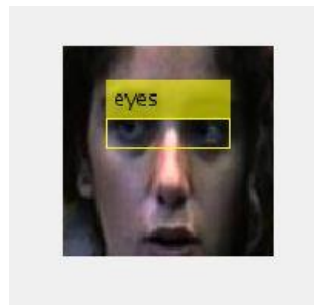


Figure 4: Eyes detection in the image

Step 4: `faceImage = imcrop (A, BBOX(i,:));` Crop the detected face region.

2. Preprocessing

In this step, the image will be converted into grayscale and then it will be binarized. Then every binarized image must be normalized in 112*92 pixels for further step. Then use the unsharp masking for sharpening the image. Now, apply the various filters over the sharper images for removal of noise without losing the important facial features. Noise will be removed so that the result of further steps could be more efficient.

Step1: Conversion RGB to gray image



Figure 5: RGB to gray conversion

Step 2: Conversion of gray to binary image



Figure 6: Binarization

Step 3: Normalized the image into 112*92 pixels

It is done to take all the images in fixed pixels value for further processing

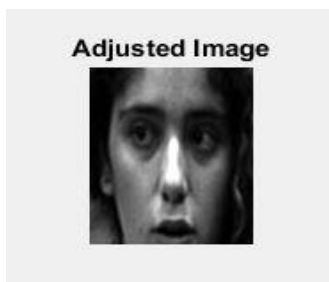


Figure 7: Normalized image

Step 4: Use Unsharp Masking



Figure 8: Sharping of image

Step5: Apply different filters for removal of noise

Then we applied various filter such as wiener filter, Gaussian filter, median filter, Gabor filter etc. , just to remove the noise present in the face by preserving the important facial feature. Gabor filters are bandpass filters which are used in image processing for feature extraction, texture analysis. Median filter is an effective method that, distinguish out-of-range isolated noise from legitimate image features such as edges and lines to some extent. Wiener filter degraded additive noise and blurring. Gaussian filter is a linear filter. It's used to blur the image or to reduce noise.



Figure 9: Median Filter Image

Step6: Median filter gives the better results than the other filter, so further feature extraction, we consider the images came out from the median filter.

3. Feature Extraction

The recognition purpose features of extracted illuminated images are calculated and stored in a feature vector for classification. In feature extraction, PCA is used for feature extraction and stored the feature vectors value. It also uses the Gabor features to extract the features set values from the face. These features are then trained by classifiers in classification

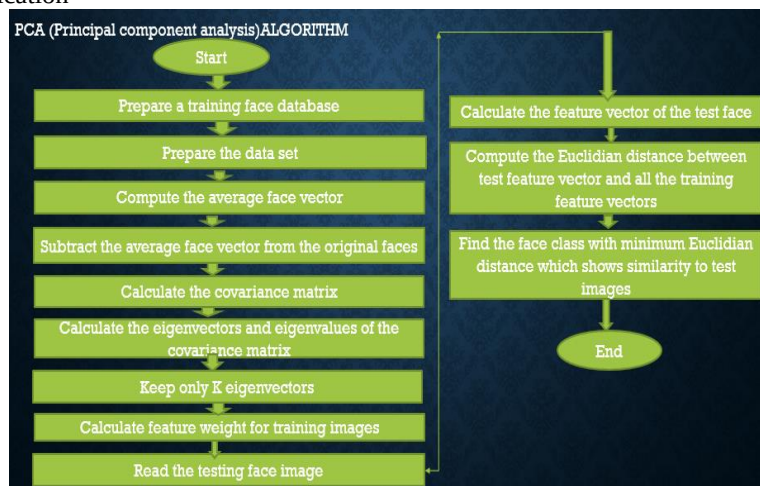


Figure 10: Flow chart of PCA

Step 1: Enter the number of image to be recognized

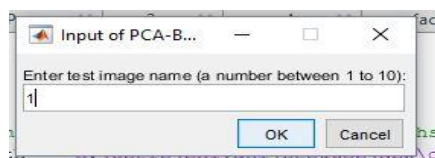


Figure 11: Input of PCA

Step 2: Given Test image

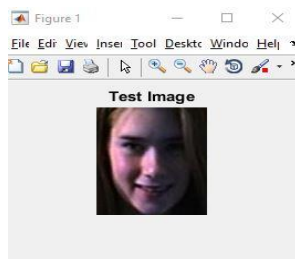


Figure 12: Test image for recognition

Step 3: Recognized image

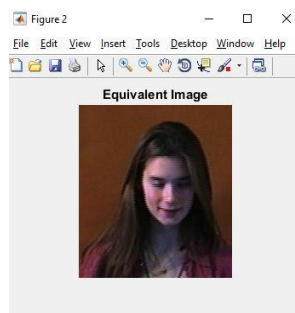


Figure 13: Recognized image

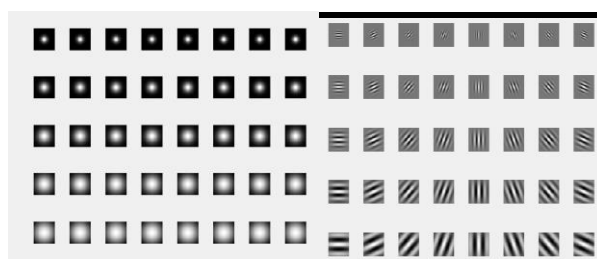


Figure 14: Gabor features of a face

4. Training and Evaluation (face recognition)

In this step artificial neural network is used for machine learning and the images are recognized by passing the feature vector of extracted images in the learning model obtained irrespective of illuminations conditions.

1. In classification train the datasets to detect the exact matching of human face by removing the illumination effect.

2. For training of dataset we use certain classifier like SVM, ANN, Decision tree, Random Forest etc.

- SVM (Support Vector Machine): It is a supervised learning model. SVM is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.

- ANN (Artificial Neural Network): Neural network consists of three steps. All layers are connected to each other's. First layer is of neurons as an input and these input are send to next layer which send the output of second layer to third layer.

- Decision tree: A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

- Random forest: It is a collection of trees, and all trees are slightly different from each other. It uses for classifications, regression and many more tasks.

Step1: In SVM, train the dataset for classification, Multi class SVM is used for different classes of different face.

Step 2: For different classifier, certain images are to be trained and then tested.

III. IMPLEMENTATION RESULTS

We have applied so many classifiers, then observed that the ANN (artificial neural network) gives the better result than the other classifiers. The performance test has been performed using different number of training a test images. 120 illuminated images are taken. There are 5 images per individual from the database is considered so that the effect of different number of training and test image combinations are tested. The output shown here is number of correct images identified, by the total images to be tested.

CLASSIFIER	OUTPUT	ACCURACY
SVM	98/120	81.80%
Random forest	83/120	70.76%
Decision tree	92/120	76.92%
Neural Network	100/120	83.07%

Table 15: Result using proposed approach

IV. CONCLUSION AND FUTURE SCOPE

For this dissertation a no. of approaches have been researched upon and then the approach which best suites the required inputs i.e. illuminated face images, are taken and implemented. The problem in recognition of face under the varying lighting conditions, to improve this lot approaches are to be taken which improves the accuracy rate of the recognition. For reducing the false positive rate and to improve the accuracy, firstly we detect the face regions from the Images. By detecting the face area from the images, it limits the search area for further processing. Using of viola- jones algorithm gives better result of detection. It focused on the initially detection of face from the images cause the improvement in the further pre-processing steps. In pre-processing noise introduced in the image are removed, applying various filter removal of noise is done by preserving the important facial feature. PCA algorithm is helpful in finding the extract face regions from the images. Using of different filters shows best result from the binarized images. For the feature Extraction, PCA is used to train the feature vector of images, which highly increased the recognition rate. Neural network gives the better result by still preserving the important facial feature, in varying the illumination conditions.

➤ Future work

- Different pose variation images can be used under illumination conditions.
- Extreme bright images can be tried, where images are very hard to identify.
- By trying the certain techniques, automatic noise detection from images can be done.
- Use more feature extraction techniques may improve the more accuracy.

REFERENCES

- [1] Richa Mishra and Dr. Ravi Subban, "Face Detection for video summary using Enhancement –Based Fusion Strategy under Varying Illumination Conditions", International Conference on Science, Engineering and Management Research (ICSEMR 2014), IEEE, pp.1-8, Nov. 27-29 2014, Chennai, India
- [2] Mangayarkarasi Nehru and Dr. Padmavathi S, "Illumination Invariant Face Detection Using Viola Jones Algorithm", International Conference on Advanced Computing and Communications Systems (ICACCS-2017), IEEE, pp.1-4, January 6-7 2017
- [3] Rajeshwari, Varsha power and Aditi jahagirdar, "Reliable face detection in varying illumination and complex background", International conference on Communication, Information and Computing Technology (ICCICT), IEEE, PP 1-4, 23-25 Nov 2012., Allahabad, India
- [4] J, K. Karibasappa, Gopal Krishna M.T "Survey on Skin Based Face Detection on Different illumination, poses and occlusion." International Conference on Contemporary Computing and Informatics, (IC3I), IEEE, PP 728-733, 27-29 Nov 2014. Mysore, India
- [5] Priyanka Goel, Suneeta Agarwal, "An Illumination Invariant Robust and Fast Face Detection, Feature Extraction based Face Recognition System", 2012 Third International Conference on Computer and Communication Technology, IEEE, PP 110-115, 23-25 Nov. 2012, Allahabad India
- [6] Shorouq Alotaibi, Nouf Alharbi, Heba Kurdi "Face recognition under varying illumination based on homomorphic filter and local binary patterns" Intelligent system conference, IEEE, 7-8 Sept 2017, London, UK, pp 589-591



Efficient Security in Cloud Storage for Dropbox

Preeti Chhikara¹ and Sanjay Ojha²

School of Information Technology (SoIT)

Centre for Development of Advanced Computing (C-DAC), Noida

B-30, Sector 62 Noida, Uttar Pradesh, INDIA

Abstract: The objective of this research work is to provide security to data which is store in dropbox by users. So to improve security file splitting, encryption and decryption and key share is done for better security. So that user can store data and no other can use the data.

The AES algorithm will help in encryption and decryption, encryption is done while sending the data and decryption is done while uploading the data. While key sharing, key is shared during encrypt the data and while decrypt if key matches then only receiver decrypt the data. So, it can be concluded that with the help of this user can safely store and share the data to other users. So the system will help in better security.

Keywords: cloud storage, dropbox, AES algo , Diffie-hellman algo

I. Introduction

Cloud Computing is a model for allowing convenient, on-demand access from anywhere, to share pool of computing resources. These can include services, storage, networking, applications etc that can be rapidly and easily provision. It includes many essential characteristics: - On-demand self-services, broad networking access, resource pooling, rapid elasticity, metering etc.

Cloud storage is a cloud computing modle, which is used for store data. Cloud storage is made by different – different companies for storing the data, where we can save our data and we have to pay some money for that like google drive,one drive etc.

Companies made data storage systems,data centers where data servers are provided and we save our data. This is the safe method to store data and companies made multiple copies of these data, so that if servers gets cracks then also data remains safe and protected.

Free cloud storage services are:-google drive,dropbox,mege,degoo,box etc

Dropbox is a cloud storage service, where we can save our files, photos, videos, etc and then we can access that data anywhere and we can perform new work on that data and we can download that data at anywhere and anytime. Dropbox provide free storage space for store data. Dropbox can be download on any device like phone, laptop, ios, macs etc. Services in dropbox are simple and designed elegantly.

Dropbox link allows user to download dropbox application on their system and then user need to made their account on dropbox and then only user use the applications. Here user can save the files, photos; videos etc, through dropbox user can share its files to other user who use dropbox. Dropbox provide the features like upload download, share files with other user.

II. Related Work

Dropbox and box is compare on the bases of statistic percentage of upload throughput, available space, download throughput, response time. Each statistic rating is given out of 5 and both dropbox and box have 5/5 for upload throughput but response time is differ 4/5 for dropbox and 3/4 for box. Then by comparing, dropbox is faster than box in both upload throughput and download throughput. [1]

Hash-based message authentication code (HMAC) algo is used for creating one time password (OTP) that send on the user Id or mobile number. OTP is send for the security of the account only registered user get the OTP and account is save from the hacker. [2]

Phishing attacks were told that how attacker attack accounts by sending the fake page link with email id. Here TF-Idf (Term frequent-inverse document frequency) and crawler algo used for finding fake mail and fake URL. [3]

Focus on the file sharing, chunks and synchronization. Rsync algo is used for the synchronization. In file allocation structure, observation is done and there was network traffic problem in dropbox, for solve network traffic waste problem sync mechanism used. [4]

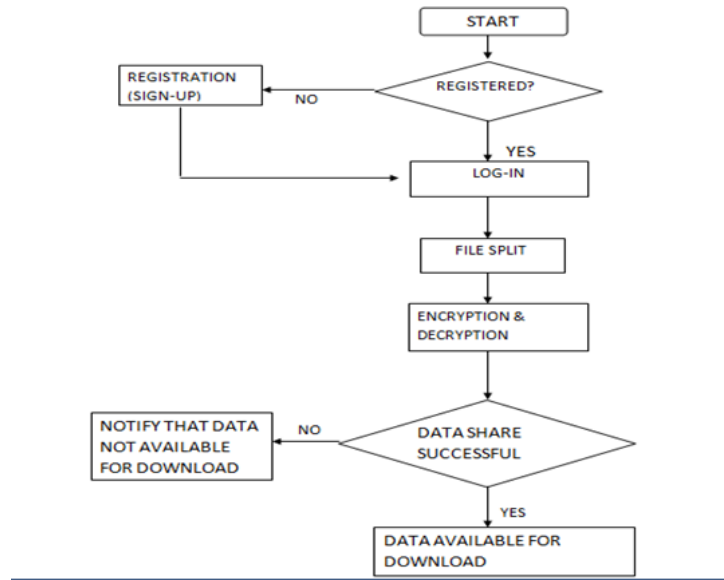
Author tells that dropbox uses the deduplication algo for delete its duplicate data. If copy of data is stored then index is made and algo reduce the storage capacity because unique data stored. At time of internet block attacker did not connect and information did not known by the attacker and he did not able to access the user account. [5]

III. Approach

The objective of our research is to make data more secure during save and transfer in dropbox.

The fig. 1 shows the diagram of the proposed approach of our research work. The approach consists of following steps:

Fig. 1 Proposed Approach



1. Firstly user registers, and while uploads the files split into parts and then save in dropbox.
2. When sender sends file to receiver, file encrypts at user side and decrypt at receiver side.
3. While sharing the file key share takes place and when key matches then data available for download otherwise not.

Step:-1

It contains the Registration steps, log-in steps, split and merge steps.

REGISTRATION:-

Step 1. Enter personal details.

Remarks:-User id created successfully.

Log in :-

Step 1. enter user name and Password

Remarks:-login successful

Here, when user open the application then log-in is required for registered user, if user is not registered the he will register first then he able to log-in and enter the application.

Then user can save its data but the data is saved in the splitting form, it is already mention in the application that in how many parts data is split and save. Split data save in the part extension form so that user knows that in how many parts data is divided. Splitting is done at the user end.

At the receiver end receiver download the receive files and then decrypt the file and save the file, after save the file then only receiver can merge the file and then he is able to access the file data.

Step: - II

Here, when sender want to send the split file then system ask for the encrypt and save the file at sender end after encryption then only sender able to share the file to the receiver.

At receiver end the receiver receive the encrypt file and then decrypt the file and save the file on system then ask for the merge of file then only receiver able to access the data of file.

Here, encryption and decryption is done with the help of AES algo because takes less time for encrypt and decrypt files then other algo.

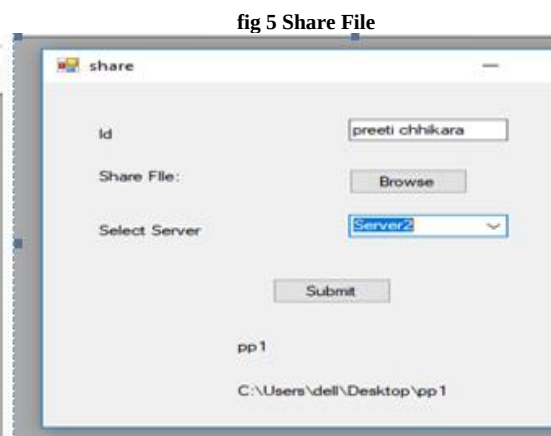
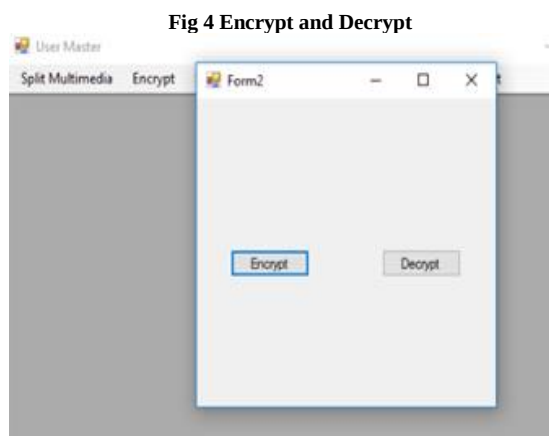
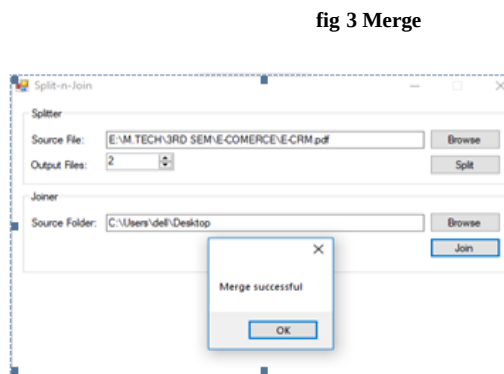
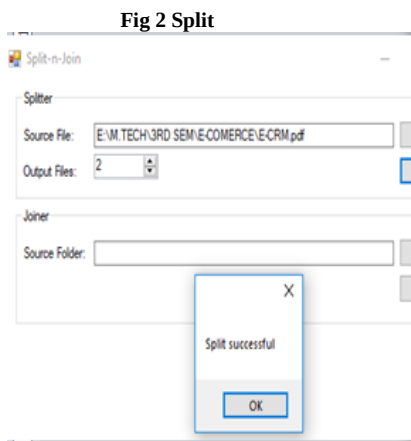
Step: - III

Here, share of data is done. When the sender sends the data to receiver then system ask for the id and path of data which sender want to share. When the sender sends data and receiver receive data they both have a key that ensure that the receiver is right person whom sender wants to give the data.

Key which identifies that receiver is right or not that is generated at the time when sender sends the data to receiver id.

If key matches then only receiver decrypts the file otherwise not.

IV. Results



V. Conclusion

The motive of this research work was to give an improved solution the problem of data theft. Before this work was started many other approaches of some other authors were take as subject of reading to see through the latest trends and practices that were previously followed. After the study it was concluded that there were some loop holes in the previous methodologies which needs to be work on. The main problem that was noticed was that the data is not secured in the user account. In this research work an approach has been proposed that can secure the data. So it can be concluded that if we speak of security nothing is 100% secure only preventive measures can be taken to minimize the risk of data steal by some attacker form the user account.

Reference

- [1] Huu dinh,alexander dworkin,christopher o'neill, scott savage, jimmy leak, mohammad aazam,marc st-hilaire "OmniBox: Efficient Cloud Storage by Evaluating Dropbox and Box ",24th international conference on telecommunications (ICT),2017.
- [2] Aditya Yulianto, S.Kom ,Maria Irmira Prasetyowati, S.Kom., M.T. "One-Time Password Mechanism Implementation for Dropbox Files Security" ,8th International Conference for Internet Technology and Secured Transactions,2013
- [3] N.S. Sudharsan, Dr. K. Latha "Improvising Seeker Satisfaction in Cloud Community Portal: Dropbox", International conference on Communication and Signal Processing , pp: 321-325,201
- [4] Shenglong Li, Quanlu Zhang, Zhi Yang and Yafei Dai" Understanding and Surpassing Dropbox: Efficient Incremental Synchronization in Cloud Storage Services", Global Communications Conference (GLOBECOM), 2015 IEEE
- [5] Bruce Wu, Tung Nguyen ,Mohammad Husain "Implementation Vulnerability Associated with OAuth 2.0 on Dropbox", 12th International Conference on Information Technology,2015

- [6] Zeng, Wenying, Yuelong Zhao, Kairi Ou, and Wei Song. "Research on cloud storage architecture and key technologies."In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 1044-1048. ACM, 2009.
- [7] Harnik, Danny, Benny Pinkas, and Alexandra Shulman-Peleg. "Side channels in cloud services: Deduplication in cloud storage.", IEEE Security & Privacy 8, no. 6 (2010): 40-47.
- [8] Esposito, Christian, Massimo Ficco, Francesco Palmieri, and Aniello Castiglione. "Smart cloud storage service selection based on fuzzy logic, theory of evidence and game theory.",(2015).

Acknowledgements

I take this opportunity to acknowledge all those who have guided me during this project. I express my earnest gratitude towards Mr. Sanjay Ojha, my project guide for his valuable encouragement and guidance.

His motivation and suggestions were invaluable in the project work. I am grateful for his most cooperative attitude and suggestions, without which I would not have been able to do this work.

I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.

Customer Review Rating Using Opinion Mining in Cloud Computing

Madhu¹ and Sanjay Ojha²

School of Information Technology (SoIT)

Centre for Development of Advanced Computing (C-DAC), Noida

B-30, Sector 62 Noida, Uttar Pradesh, INDIA

Abstract: The objective of this research work to provide a platform which can compare various mobile operators because it not possible to read all comments. So to improve this problem we will use boyer moore algorithm which gives total number of positive, negative, neutral review. This algorithm will generate a comparison graph also and in this we can also select feature it means feature wise selection also present here due to this we can select the feature like high internet speed, low cost, better voice quality. So, it can be concluded that with the help of this user can easily compare which service is better in terms of which parameter. So the system will help them to make fair judgement.

Keywords: opinion mining, customer review, boyer moore, voice quality, cost, speed

I. Introduction

Cloud Computing is a model for allowing convenient, on-demand access from anywhere, to share pool of computing resources. These can include services, storage, networking, applications etc that can be rapidly and easily provision. It includes many essential characteristics: - On-demand self-services, broad networking access, resource pooling, rapid elasticity, metering etc. Customer review and rating is play a very important role in companies. The help of these feedback user can decide to buying a product or not. But it is very difficult to read all reviews. This system will take out customer feedback and specific list by the user. These websites helps people to express their comments about different item. Opinion mining is the takeout people's sentiments about certain products.

It will give the number of positive, negative, neutral reviews with the help of boyer moore algorithm. With the help of these reviews we can observe that which review have most occurred in our data and we can easily make opinion which service is better. Here there is a table which generate by the algorithm which tells the number of number of positive, negative, neutral review and here feature selection also provided with the help of this we can select feature like voice quality, call drop, cost, speed. So due to this method we can compare the feature like if we live in Bombay and we want to find that which mobile operator have high internet speed so with the help of this website we can compare like Airtel, Vodafone, Jio in all of these which one is better in Bombay which have high internet speed. Similarly we can compare voice quality, call drop, cost also. So it provide a better way to compare the services and user can make fair judgement.

II. Related Work

Due to increase in number of reviews it is difficult to read all the reviews. Customer review and rating is play a very important role in companies. The help of these feedback user can decide to buying a product or not. Opinion mining is the takeout people's sentiments about certain products. So due to increase use of social media websites a lot of research work is done based on the customer reviews.

It provides a better way to rating a product based on its specification. It First of all it take out the specification and every specification have contain a polarity it may be positive or negative feedback. Then overall product rate is compute. It is helpful for those customers those want particular features. [1]

It is used to study the customer review to arrange comments in positive or negative is called sentiment mining. The opinion means the comment which tell emotion of the customer. Many customers read the comment randomly and it is difficult to read all comments. So this paper proposes the analysis and prediction rating from customer reviews with the help of classifier model. It is started from collection of data. [2].

We know the most important thing is to recognize opinion, aspects after that it do data classification which are depend on the these features and this phenomenon known feature takeout. Here sentiment is done by machine learning algorithms like support vector machine and naïve bayes. In opinion summarization using decision tree firstly data preprocessing is performed to remove redundant data and take the features. In second step tree classification with the help of using machine learning to attain performance for many features are selected based on aspect. It gives better result for the accuracy, precision comparison to SVM and naïve bayes method. [3]

In this author used three algorithms names are Naive bayes, Random tree and Random forest to calculate either the opinion was positive, negative or neutral. And also is not possible to extract opinion by reading all

review so opinion mining play a main role in taking emotions of customers. It use the document level opinion mining around 7000 customer feedback and the random forest hold the highest accuracy.[4]

We see that web search engine increase the search process to a particular corpora (verticals) like sports, images, videos, shopping etc each have its own specialization. But the problem is still available due to different nature of different verticals so how the system combine the vertical results is still an problem. So due to increasing number of vertical attributes require a solution which is effective .It describes a search federation system that make efficient use of user feedback. It works with heterogeneous verticals. It has successfully used in many markets.[5]

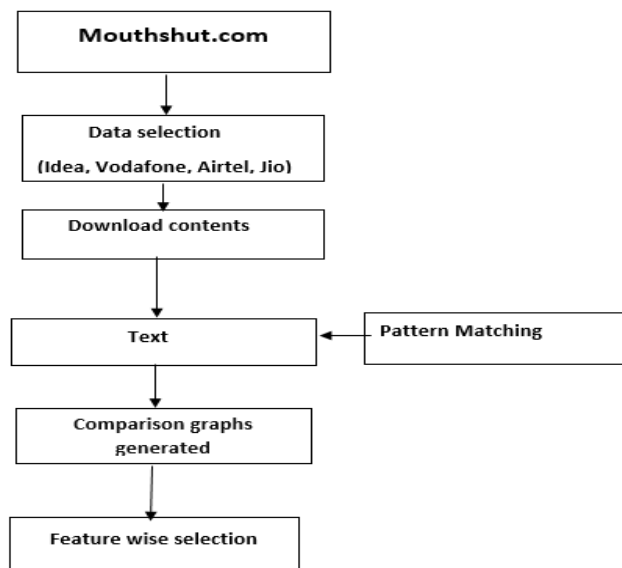
III. Approach

The objective of our research is to provides a single platform for compare the services of various mobile operators. With the help of this we can easily compare which service is better in terms of which parameter .So the system will help them to make fare judgement.

The fig. 1 shows the diagram of the proposed approach of our research work. The approach consists of following steps

1. **Collection of data .**
2. **Applying boyer moore algorithm to define positive, negative, words.**
3. **Our algorithm will generate the comparison graph.**
4. **Feature wise selection is performed by taking parameters like voice quality, call drop, cost, speed.**

Fig. 1 Proposed Approach



STEP 1: First of all we will collect data which is available at mouthshut.com. Here mouthshut.com plays the role of data storage where all types of review are present. Mouthshut.com is a customer review website. We will extract the data of mobile operators (idea ,Vodafone, airtel, jio) . We will get the data in HTML content form. For this we will convert it into plain text form by the help of regex option of class and the property of class (ignore case) which help us to remove tags and extra symbols like smileys, stop words.

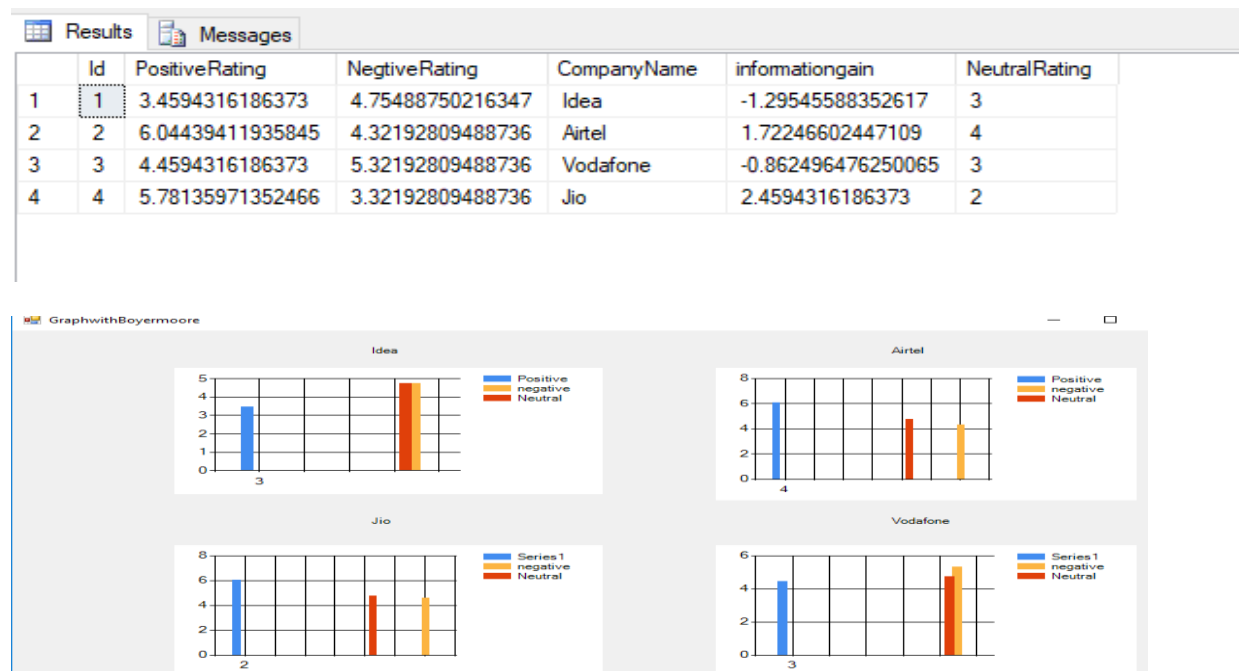
STEP 2: Once we will get relevant information we need to define data into three categories. So for this apply boyer moore technique to classify the data into their relevant class . So, we will train our technique by providing positive, negative, neutral words. With the help of those words our algorithm calculate occurrence of positive, negative, neutral review.

STEP 3: Once we will get the reviews then our algorithm generate the graph .The graph tell us the occurrence of number positive, negative, neutral review.

STEP 4: In last step feature wise selection is performed in this we will select the mobile operator with the help of parameters like voice quality, call drop, cost, internet speed. So, with the help of this we can compare services in a particular location.

IV. Results

We have performed boyer moore algorithm on the customer reviews and it will generate graph of positive, negative, review.



V. Conclusion

The motive of this research work was to provide a platform in which we can compare various mobile operators because it is difficult to read all the comments which are available on the website and the decisions made by them might be biased. So to improve this problem we will use boyer Moore algorithm which gives total number of positive, negative, neutral review. In this research work an approach has been proposed that it generate a comparison graph with the help of this we can identify occurrence of positive, negative, neutral review and in this we can also select feature it means feature wise selection also present here due to this we can select the feature like high internet speed, low cost, better voice quality.

References

- [1] Vamsee Krishna, Archana R, Vimal kumar K, "User specific product recommendation and rating by performing sentiment analysis on product review", IEEE International Conference on advanced computing communication systems, pp: 1-5, 2017.
- [2] Wararat Songpan, "The Analysis and Prediction of Customer Review Rating Using Opinion mining", IEEE International Conference Software Engineering Research, Management and application, pp: 71-77, 2017.
- [3] Rajalaxmi Hegde, Dr. Seema, "Aspect Based Feature Extraction and Sentiment Classification Review Data sets using Incremental Machine learning Algorithm", IEEE International Conference on Advances in Electrical, Electronics, Information Communication and Bio- Informatics, 2017.
- [4] Asma Musabah Alkalbani, Farookh Khadeer Hussain, Omarkhadeer Hussain, "Analysing cloud service reviews using opinion mining", IEEE International Conference on Advanced Information Networking and Applications, pp: 1124- 1128, 2017.
- [5] Chen CM, Shieh CZ. Online Unified Feedback System, IEEE Journal for Research In Applied Science And Engineering Technology, vol.9, no.16, pp: 586-597, 2016.
- [6] Q.Su, X.Xu, H.Guo, Z.Guo, X. Wu, X. Zhang and B.Swen. Hidden Sentiment association in Chinese web opinion mining. Proceeding of the 17th International Conference on World WideWeb, 2008, pp.959-968.
- [7] S.Atia and K. Shaalan, Increasing the accuracy of opinion mining in Arabic. Proceeding of the 1st international conference on Arabic computing linguistics, 2015, pp.106-113.
- [8] H.S. Le, T.V. Le and T.V. Pham, Aspect Analysis for Opinion Mining of Vietnamese Text. Proceeding of International Conference Advance Computing and Application, 2015, pp.118- 123.

Acknowledgement

I take this opportunity to acknowledge all those who have guided me during this project. I express my earnest gratitude towards Mr. Sanjay Ojha, my project guide for his valuable encouragement and guidance. His motivation and suggestions were invaluable in the project work. I am grateful for his most cooperative attitude and suggestions, without which I would not have been able to do this work.

I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Efficient resource utilization using load balancing in Cloud computing

Supriya¹ and Sanjay Ojha²

School of Information Technology (SoIT)

Centre for Development of Advanced Computing (C-DAC), Noida

B-30, Sector 62 Noida, Uttar Pradesh, INDIA

Abstract: Today Millions of people accessing the website daily through cloud computing servers. These requests, known as the load, are processed explain the situations in which servers are unable to process requests. This is a problem for companies that have personal cloud system because it can effect the impression on the organizations. To address this problem, by the servers. Servers might take significant time to process playing YouTube videos, the video buffers in spite of having a good signal strength; or, while buying something through e-commerce, the page does not load due to too many requests. These examples best engineers developed the technique of load balancing to handle the huge load on the servers. Load balancing reduces the overbearing load on the servers by distributing the load to various servers. There are various types load are possible in cloud computing like memory, CPU and network load.

Keywords: Virtualization, load balancing, green computing, static load balancer, dynamic load balancer

I. Introduction

Cloud Computing includes many essential characteristics:-On-demand self-services, broad networking access, resources pooling, rapid elasticity, metering. These can include services, storage, networking, applications etc that can be rapidly and easily provision. Cloud computing offers different services such as IAAS, PAAS, SAAS for which users pay different amount according to their requirements. The aim of resource allocation is to find how many resources are available at particular time interval so that we can allocate resources efficiently when the users are requesting for various services. Load balancing is technique that manage the availability of resources. Load balancing means handling or sharing the workload efficiently across various virtual machine on the server. It is technique that manages the availability of resources.

II. Related Work

There are numerous problems of Load Balancing algorithms. Customer satisfaction is important to distribute the work load among various nodes. Load balanced scheduling demands a number of challenging issues due to heterogeneous and dynamic nature of cloud. A major concern in these issues, LB allows cloud computing to “scale up to increasing demand”.

To achieve the desired goal load balancing provides several algorithms. Some algorithms aim at getting maximum resource utilization, minimum response time and achieve higher throughput etc.

In designing any load balancing algorithm some major issues considered are

- **Geographical Distributions of the Nodes:** It is used in the large scaled applications like Twitter, Facebook etc. To maintain the efficiency of the system and handling fault tolerance well DS of the processors in cloud computing environment is very helpful

- **Static vs. Dynamic Nature of Algorithms:** The issue of design of LB algorithms is based on the nature of the system which may be static algorithm and dynamic algorithm.

The author give an approach to handle the multiple workload in the cloud environment. Adaptive resource allocation of

Service Based System is proposed. It's maximizing the throughput of the services by allocating the resources to the server.

It's provide high QOS service by maximizing throughput of the server. In this machine learning technique is used. Here optimal policies are learned to control the host to be lively or inactive. Yet, it takes considerable time to acquire the best policies. Resource allocation and resource allocation controller is built by running Python 3.4. Resource allocation is calculated through resource allocation controller. [1]

In this central load balancing technique is used to avoid the situation of overload and underload of virtual machines on the basis of priority and state .It's assign the load to various virtual machines .By using cloud analyst tool comparative analysis is done. Through this paper it has been concluded that in the throttle algorithm

we have to check the availability of virtual machines again and again but in synchronized throttle Virtual machine load balancing technique (STVMLB) the checking of the availability of virtual machine starts from next to allocated VM. The author explained the algorithm by using simulation that central load balancing technique algorithm perform better than throttle and round robin algorithm. In this we created one center with five virtual machine running on it. [2]

The author has presented several techniques and methods of load balancing on the basis of various parameter like performance, scalability, resource utilization, response time, fault tolerance in different cloud environment. In this concept of round robin and weighted round robin are discussed.

A Weighted Round Robin Load Balancing Algorithm (WRLBA) follows weight (or preference concept to assign the new request to the server. Here, the authors select the server with a next higher weight as a result it can handles additional traffic than the other servers.

In cloud computing one of major issue is load balancing. In this author has analyse the advantage and disadvantage of static and dynamic technique .The Dynamic technique handles the dissimilar behaviour of cloud whereas static technique fails to handle dissimilar behaviour of the cloud. In this paper various distributed algorithm was studied. [3]

In this paper the author make use of skewness metric to mix VMs workload with completely dissimilar resources characteristics properly so that the servers are properly utilized. Here, the list of virtual machines are sorted based on the resulting temperature of the server. we can sort the virtual machine in descending temperature. Our aim is to migrate away the virtual machine and Eliminate those virtual machine which can reduce the skewness of the server. [4]

Author has chosen virtual machine scheduling policy to increase the application performance when there is limited resources and there is service request variation. The author has discussed various virtual machine scheduling policies and performance is analysed in order to accomplish the better quality of service. There are 2 policies available for analysing the performance: 1) **Space Shared**: In this one processor is using by 1 virtual machine and then it finishes its task then we will give the processor to other virtual machines. They will using the processor sequentially one by one.2) **Time Shared**: In this for some time the processor is using by 1st virtual machine and some other time it will be using by 2nd virtual machine. They will sharing the processor one by one concurrently or parallel. Processor will be switching between these virtual machine. [5]

III. Approach

The main focus of load balancing algorithms in cloud environment is to allocate the incoming requests to all available virtual machines with minimum response time.

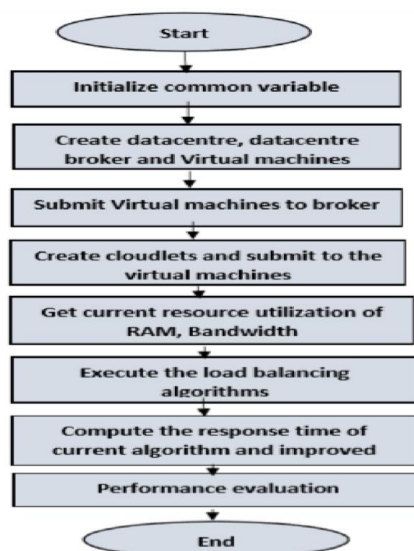


Fig. 1 Proposed Approach

The fig. 1 shows the diagram of the proposed approach of our research work. The approach consists of following steps:

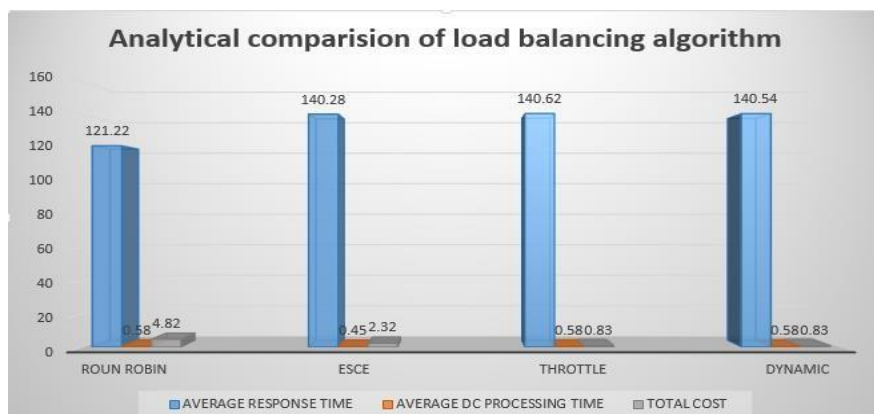
1. The first step is the initialization of common variables. In this step we set the number of users.
2. The next step is the creation of datacenter, this in turn lead to the creation of host along with their characteristics. Datacenter is the resource provider. The characteristics of host can be size of processing element, bandwidth utilization, RAM etc.

3. After datacenter is created next step is the creation of datacenter broker instances. The creation of broker is very important as it is responsible for communication between datacenter and the submission of cloudlets.
4. Next step is the creation of Virtual machines instances. The virtual machines have same characteristics as that of host.
5. Virtual machine then submitted to the broker.
6. After the virtual machine is submitted the next step is the cloudlets which are also created with certain parameter like MIPS requirement, bandwidth requirement etc.
7. The status of each VM is checked.
8. Then schedule the virtual machine on the basis of sorted task.
9. Then algorithm is proposed and then implemented in cloud computing environment using cloud sim toolkit in java language.
10. Execute the existing algorithm and proposed algorithm.
11. Then obtain the current utilization of RAM.
12. Simulation result will compare using different parameter including response time, resource consumption cost and data center processing cost.
13. After that performance is evaluated

IV. Result

We have performed 4 load balancing algorithms with different parameters. The response time is minimum for the User Bases who are nearer to Datacenter.

Output Metrics	Load Balancing algorithm			
	RR	ESCE	THROTILE	DYNAMIC
Avg Response Time	127.22	140.28	140.62	140.54
Avg DC Processing Time	0.58	0.45	0.58	0.58
Total cost	4.82	2.32	0.83	0.84



V. Conclusion

As millions of requests per second are sent to the server, the static load balancing techniques are not suitable. The dynamic load balancing algorithms have been used to increase the performance of the overall network. We have compared and analysed various algorithms in Cloud Computing. We have concluded that according to our requirement or need the algorithm are used. The proposed approach of scheduling the resources using priority based throttle algorithm and equally spread current execution algorithm is a efficient and optimize way to schedule resources to appropriate tasks. Also, the processing time of the algorithm and the average waiting time can be slightly reduced. Here parameter were compared with respect to datacentre processing time, response time.

References

- [1]. S. S. Yau, H. G. An, "Adaptive resource allocation for service-based systems", *International Journal of Software and Informatic*, vol. 3, no. 4, pp. 483-499, 2009 Adaptive Resource Allocation of Multiple Servers for Service-based Systems in Cloud Computing
- [2] Shikha Garg, Rakesh Kumar Dwivedi, Himanshu Chauhan "Efficient Utilization of Virtual Machines in Cloud Computing using Synchronized Throttled Load Balancing", 1st International Conference on Next Generation Computing Technologies (NGCT), pp.77-80, 2015.
- [3] Aarti Vig, Rajendra Singh Kushwah and Shivpratap Singh Kushwah: "An Efficient Distributed Approach for Load Balancing in Cloud Computing India International Conference on Computational Intelligence and Communication Networks "pp.751-755, 2015.
- [4] Zhen Xiao, Weijia Song, Qi Chen "Dynamic Resource Allocation using Virtual Machines for Cloud computing environment", IEEE Transaction on Parallel or Distributed System (TPDS), pp-1107-1117, 2013
- [5] Zhongjiang Yao, Xiaodan Zhang, Yuepeng E, Tong Li, Jingguo Ge, Haina Tang, "Performance Evaluation and Analysis of the Virtual Networks Built by Different Virtual Machine Managers", High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS) 2016 IEEE 18th International Conference on, pp. 1549-1556, 2016.
- [6] S. Patel, R. Patel, H. Patel, and S. Vahora, "Cloudanalyst: A survey of load balancing policies," *International Journal of Computer Applications*, vol. 117, no. 21, pp. 21-24. May 2015.

Acknowledgment

I take this opportunity to acknowledge all those who have guided me during this project. I express my earnest gratitude towards Mr. Sanjay Ojha, my project guide for his valuable encouragement and guidance.

His motivation and suggestions were invaluable in the project work. I am grateful for his most cooperative attitude and suggestions, without which I would not have been able to do this work.

I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Palmpoint Recognition using Local Binary Pattern

Arshi Husain¹, Tushar Patnaik²

C-DAC, Noida,

GGSIIP University, Delhi

Abstract- In present time, Palmpoint recognition has gained significant prominence in biometric identification systems as Palmpoints have enough stable and distinguishing information for separating an individual from a large group of individuals because of its rich features which includes principal lines, wrinkles etc. In this paper, we propose a palmpoint verification approach using Local Binary Pattern. It has four stages, image acquisition, pre-processing, feature extraction and feature matching. In pre-processing stage, enhancement of the images, binarization, segmentation, noise reduction, rotation and translation normalization is done. The ROI is extracted from the whole hand image by generating a mask. After obtaining the ROI, textured feature extraction is performed using Local Binary Pattern. Then finally, SVM has been used for the purpose of classification.

Keywords: Biometrics, authentication, palmpoint recognition, Support vector machine, Local binary pattern

I. INTRODUCTION

In recent time, the development of accurate and reliable security systems is a matter of widespread interest, and in this context biometrics is seen as a highly operative automatic mechanism for personal identification and verification. Biometrics is the technology which examine individual's body characteristics for verification.

Palm print recognition has gained significant importance in biometric identification systems and it is one of that biometric feature which remains the focus of most researchers because of its rich features, uniqueness, high stability of features, public acceptability of such systems and many other reasons. It is very less invasive and easier to use in practice. Compared with fingerprint recognition palmpoint images contain more information, so more features can be extracted, therefore more comparisons and hence more accuracy will be there. Iris input devices are costly and the technique is intrusive as an individual might fear of adverse effects on their eyes. Palmpoint is hence appropriate for everybody and also this method is non-intrusive as it doesn't require any personal data of the user and as compared with face recognition palmpoint is hardly influenced by age and accessories. Thus it has become significant and quickly emerging biometrics innovation in the recent years. Therefore, in this work we build a recognition system for palmpoint biometric of the subjects.

The rest of the paper is arranged as follows: Sec. II describes the proposed approach, Sec. III shows the implementation results, Sec. IV presents conclusions and future scope and finally references are provided.

II. PROPOSED APPROACH

The proposed framework consists of four modules which consists of Preprocessing, Segmentation of ROI, Feature extraction and Classification. Flowchart of the proposed system is shown in Fig.1.

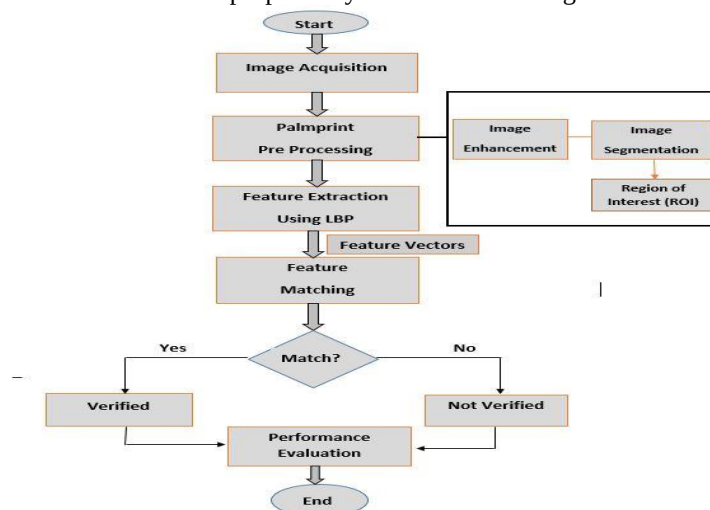


Fig.1. Overview of the proposed system

A. Preprocessing

In this module, the input RGB images are first converted into grayscale images. Then Binarisation is done using Otsu's method. Otsu method is type of global thresholding technique which depend on gray value of the image and is used for the conversion of a gray scale image into a binary image. See Fig.2.

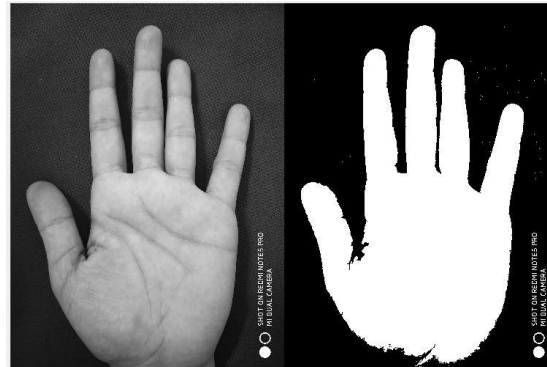


Fig.2. Binarization using Otsu's method

B. Segmentation of ROI

In this step, the region of interest is extracted from the whole image. Input to this step is the Binary image, and output is the segmented palm. See Fig.3.

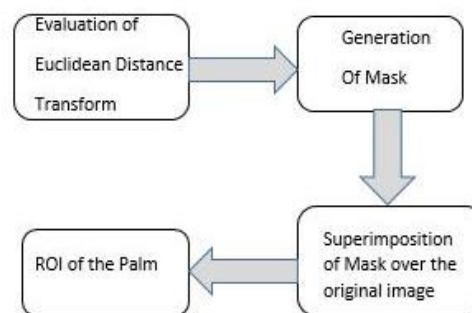


Fig.3.The Proposed System to Extract the ROI of Palm

Thus, Segmentation can be achieved in the following four steps:

Step1: Calculation of Euclidean Distance Transform

In order to extract the ROI from the whole hand image, first of all the Euclidean distance transform of binary image is computed.

The output of this step is shown in Fig.4.

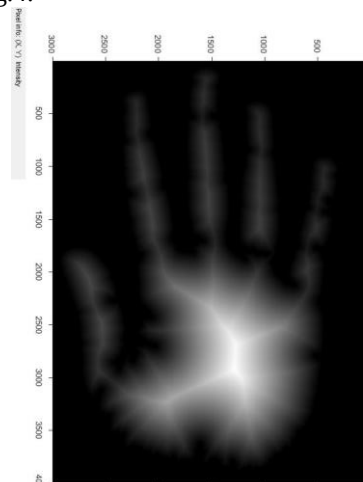


Fig.4. Euclidean Distance Transform

Step2: Generation of Palm Mask

A Palm Mask is generated by extracting the one largest strongly connected components from the whole binary image. See Fig.5.

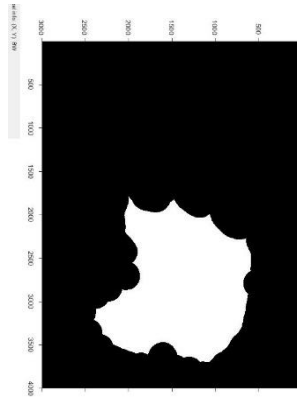


Fig.5.Palm Mask

Step 3: Superimposition of Mask over the binary image

Now the Mask generated in the previous step is superimposed over the Binary image and everything outside the mask is erased, this will give the segmented palm from the whole image of hand which is used for further Feature extraction process. Fig.6. shows the segmented Palm image



Fig.6. The ROI of Palmprint

C. Feature Extraction

In this step, the features are extracted from the preprocessed palmprint images using local binary patterns. The LBP features encode local texture information. The output of this step will be further used to create the training dataset and testing dataset. See Fig.7. The feature vector of each image obtained by the feature points of LBP is of dimension 1X10

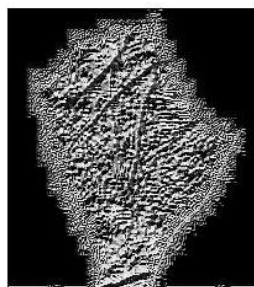


Fig.7.LBP Image

LBP: Local Binary Pattern (LBP) is an influential texture descriptor that is unvarying to any gray-scale or rotational changes. As texture is one of the most clearly perceptible features in low-resolution palm-print images, we consider that local binary pattern based features are very distinctive for palmprint identification.

The LBP operator which was presented by Ojala is an easy technique for the texture analysis. Two values are obtained during LBP Computation, firstly, the LBP value and its contrast measure. LBP feature vector, is returned as a 1-by N vector of length N representing the number of features, here in our case the length of each Feature vector obtained is 1X10. The three steps for obtaining the feature vector using LBP are as follows:

- Create LBP Code Matrix,

- Create LBP Histogram,
- Concatenate LBP Histogram from each segment.

The LBP operator is unvarying to any monotonic transformation on the pixel values. Therefore, it is very appropriate for demonstrating palms which are illuminated from a fixed distance but with fluctuating light intensity.

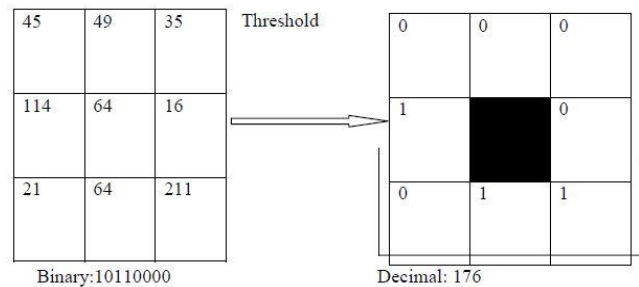


Fig.7. Example of Computing the Local Binary Pattern

D. Feature Matching

In this stage, the features which has been extracted are compared to the features stored in the database to decide if there is a match and decision making is done to accept or reject the query based on match score. i.e. in this stage, it is decided whether those features belong to the same person or not. Classification is done to divide the feature space into several classes based on some decision rule. Classification algorithms have two Modules of processing: training and testing. The classification used in the system is the supervised classification, in which, the system is trained first, using the specific set of features (LBP in this case). And then the testing data is prepared to check the similarity between the training features and testing features. In this system, Linear SVM (Support vector machine) is considered.

III. IMPLEMENTATION RESULTS

Table.1. display the results for images from 4x4 blocks to 8X8 blocks palmprint images. The greatest verification result is 82.9% for left and 83.4% for right hand's palm-print image by using the pattern size of 8x8 blocks. So, it can be observed that accuracy increases with increase in the block size of LBP.

Pattern size	LBP performance for right hand palmprint (%)	LBP performance for left hand palmprint (%)
4X4	74.04	73.19
5X5	77.87	76.17
6X6	80.85	78.51
7X7	82.53	81.27
8X8	83.40	82.97

Table.1. Results using the proposed Approach

Table.1. shows the LBP performance results for both left and right hand palm-print images, 83.4% accuracy being the greatest result using the pattern size of 8X8 blocks.

IV .CONCLUSION AND FUTURE SCOPE

This paper presents a technique for palmprint recognition by convincingly selecting the most discriminative features and extracting the texture using Local Binary Patterns on the ROI.

During the study, a number of approaches have been researched upon and then the approach which best suites the required inputs i.e. the segmented palm images, is taken and implemented.

First, the input RGB images are preprocessed to so that further feature extraction can be done effectively. For preprocessing, the system uses processes like Image Enhancement, Binarization and image segmentation. For binarization purpose Otsu's thresholding method has been used. Then textured feature extraction is performed using Local Binary Pattern technique using different block size on the ROI obtained from the whole image of the hand. ROI is extracted using combination of different segmentation techniques and generating the Palm Mask. The feature vectors which are obtained from the segmented Palm images using the LBP feature points are then

fed into the SVM to perform classification. SVM has advantages such as fast learning speed, ease of implementation with least human involvement. The feasibility of proposed system was verified on the IIT Delhi database. The result shows that using the LBP texture method, the performance accuracy of palm print images is 83.40%, so the LBP is a texture based technique which select the most texture information and distinctive features in the images. It is also observed that accuracy increases with increase in the block size of LBP. Hence, experimental results show that the proposed technique which uses a mix of LBP and SVM gives propitious results, hence this technique can be used for human identification.

In future, we can use different parameters for improving the accuracy of the system. The different feature extraction techniques used in the proposed approach along with other existing approaches, can be used with efficient preprocessing techniques for making a hybrid method in order to obtain an improvised result. The different voting method can be applied for making a hybrid method and feature subset selection techniques with the use of weighted decision combination methods.

REFERENCES

- [1] Saiyed Umer, Bibhas Chandra Dhara, and Bhabatosh Chanda, "A Novel Palmprint Recognition System using Patch Based Filter Response", 2017 IEEE International Conference on Identity, Security and Behavior Analysis(ISBA), INSPEC Accession Number: 16963871, IEEE, 2017.
- [2] M. O. Rotinwa-Akinbile, A. M. Aibinu, and M. J. E. Salami, "Palmprint Recognition Using Principal Lines Characterization", First International Conference on Informatics and Computational Intelligence, INSPEC Accession Number: 12508354, IEEE, 2012.
- [3] Afsal. S, Rafeeq Ahamed. K, Jijo Jothykumar, Shabeer Ahmed and Farrukh Sayeed, "A Novel Approach for Palm print Recognition using Entropy Information Features", 2016 International Conference on Wireless Communications, Signal Processing and Networking(WiSPNET), INSPEC Accession Number:16304367, IEEE, 2016.
- [4] Chun-Wei Lu, Ivy Fan , Chin-Chuan Han , Jyh-Chian Chang, Kuo-Chin Fan and H.Y. Mark Liao, "Palmprint Verification using Gradient Maps and Support Vector Machines", Signal and information Processing Association Annual Summit and Conference(ASPSIPA ASC), INSPEC Accession Number: 13245743, IEEE, 2013.
- [5] Indrayani Awate and Prof. B.A. Dixit, "Palm Print based Person Identification", 2015 International Conference on Computing Communication and Control Automation(ICCUBE), INSPEC Accession Number: 15305474, IEEE, 2015.



Noisy and Low resolution document analysis

Gunjan Sharma¹, Tushar Patnaik²

C-DAC, Noida

GGSIIP University, Delhi

Abstract: Document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic material. Document analysis is required in order to gain understanding about document i.e. quality of document, content of document and so on. In this research, Document analysis is conducted on datasets that are prepared for research proposes that consists of scanned images of text which is inherently inaccessible because the content in the images, not searchable text, highly noisy and low resolution images. And user input such type of images into concerned application i.e. OCR without knowing whether an image is noised which further cannot readable or extractable by the OCR. In this work, noisy and low resolution dataset that aims for identify the readiness of the image for OCR which is used to convert the pdf and scanned document into editable format. This approach provide a method for identifying that given image is suitable for OCR. In this approach, the input image taken is of Hindi document. Analysis on the basis of connected component, histogram analysis. Further low resolution image have enhanced by using contrast enhancement and morphological operations followed by interpolation.

Keywords: document analysis; resolution enhancement; power-law transformation; morphological operation.

I. Introduction

The increasing importance of document imaging devices, Text Extraction from document image, the recognition of words in pictures, is a typical request. Changing images into text document is beneficial for enhanced convenience in editing. This work is usually done by Optical Character Recognition. OCR is an innovation that empowers the clients to change over various sorts of document images into editable and accessible information, for example, paper reports, PDF documents or pictures caught by methods for a scanner or an advanced camera. OCR brings different advantages, Some of information pictures are affected by numerous outside or undesirable environmental impacts, for example, noise, uneven enlightenments which might be readable to human yet it is a real test to existing OCR system since human can't recognize noise by observing the image only however such kind of document does not give reasonable output.

There are various components that influence the exactness of text recognized through OCR. These components include: scanner quality, type of printed document (laser printer or photocopied), paper quality. "Foxing" and "content appear on the other side" found in old paper records, watermarks and non-uniform enlightenment are cases of issues that influence the exactness of OCR compared to a clean text on a white background.

Document analysis is required so that system will identify that given image is suitable for the OCR or not. This analysis on the basis connected component and histogram analysis.

Further, in this work resolution of the low resolution image has improved by using sort of method including contrast enhancement and morphological operations followed by interpolation.

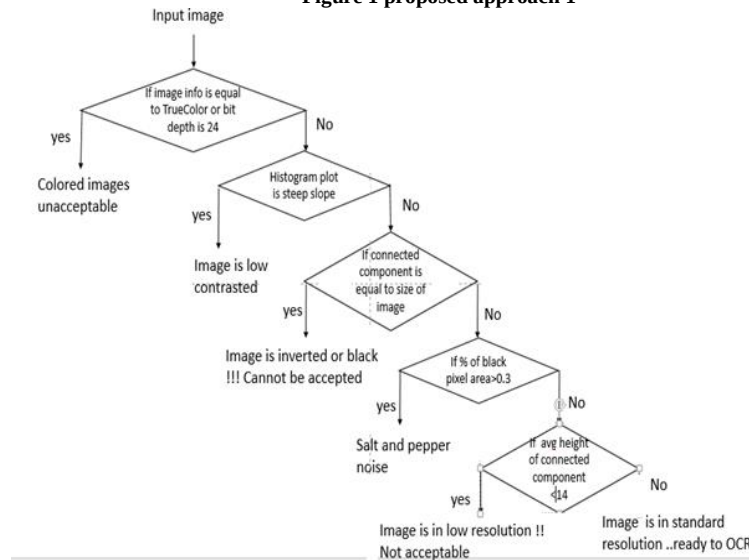
The rest of this paper is organized as follows. Proposed approach in which Analysis of noised image and enhancement of low resolution images are specified in Section II. In Section III, Implementation results is presented. Section IV contains analysis of result. Section V conclusion and future work.

II. Proposed approach

In this research, work has been done on analysis of noised image, images are colored, and low contrasted, and inverted, and salt n pepper noised image and low resolution images. OCR system which plays an important role in this work. Basically OCR is used to change pdf documents into editable format. Here problem is when user unnecessarily input the image without knowing that image is suitable for OCR system or not because human cannot judge the image by seeing only whether it is noised or not.

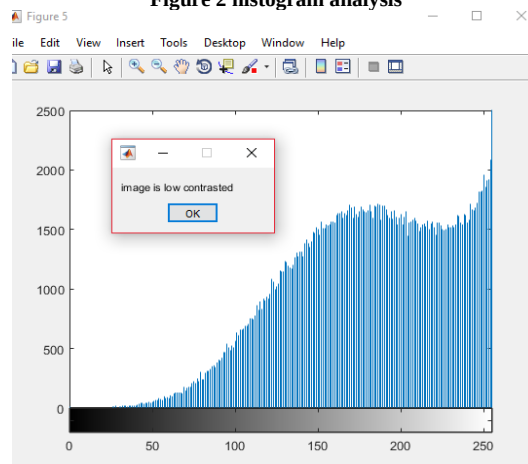
A. Analysis of noised image

Figure 1 proposed approach 1



This analysis has been done on the basis of some condition where it recognize the behavior of noised image. There are some following assumption below:
For colored images, if image property color type is equal to true color i.e. RGB format has not acceptable by OCR. For low contrasted images, if histogram shows steep slope shape for a given image so that image is low contrasted image.

Figure 2 histogram analysis



For inverted and black images, find the connected component and label them using bounding box property
Label the Connected Components

```
[L Ne]=bwlabel(I)
```

Properties of image regions

```
Propied = regionprops(L,'BoundingBox')
```

Find the maximum connected component, and Compare with size of image, if max connected component equal to size of image. Then image cannot be acceptable for particular application.

For salt n pepper images, first applied thresholding to translate the grayscale image into binary image. OTSU thresholding has been used in this approach. Then Find the area, where pixel value is zero. Calculate the percentage of black pixel where area is less than equal to 3. If percentage of area is greater than 0.3 (which is set as threshold value here) image is not acceptable.

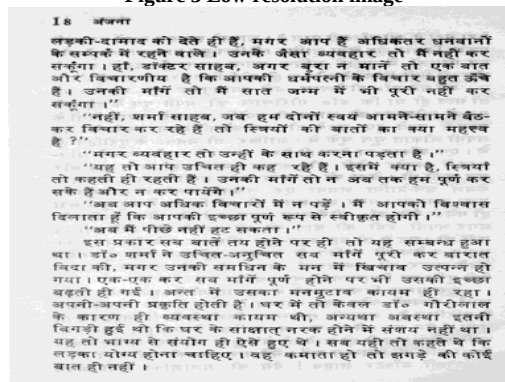
For low resolution images, Find connected component and label them. Calculate height of each connected component. Then, Calculate average height of connected components. If avg height is less than 14 (threshold value) image unacceptable.

B. Enhancement of low resolution images

Low resolution image is defined as less no. of pixel per inches. Enhancement of low resolution images is required when such kind of images pass it to the OCR system. It will process the image and produce output as junk (not a single word matched with original image) shown in fig.5.

In this paper, some sort of algorithm has been used to improve the quality of image.

Figure 3 Low resolution image



1) Grayscale transformation

Grayscale conversion is used to convert the RGB that is colored images to grayscale. Color image are stored in three dimensional array having 65535 different values which is difficult to process them unlike grayscale image they store 0-255 values i.e. shades of gray and can easily work with. This research has used averaging function of MATLAB which takes red, green, blue color intensity value from given image and divide by no of colors which changes the RGB image into grayscale image.

2) Contrast enhancement

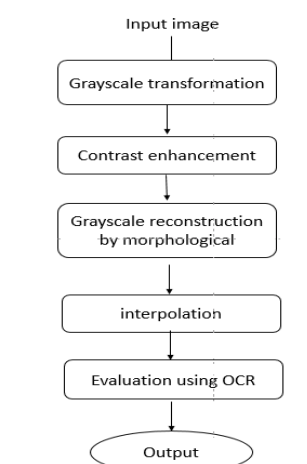
After grayscale transformation component gets connected or merge together which degrade the document and destroy information. So to mitigate this problem power law transformation technique has used.

$$s = cr^{\gamma}$$

This symbol γ is known as gamma, due to which this transformation is also called as gamma correction. Various estimated values of γ fluctuates the enhancement or contrast of the image. Different display devices or screens have their own gamma values. this is the reason they show their picture at various intensity.

In this paper, gamma value 1.02 has used, because gamma value below unity get image darker and above unity will eroded connected component but higher value of gamma will degrade the image badly. So, this chosen value has given the expected result i.e. it enhanced the image.

Figure 4 proposed approach 2



3) Grayscale reconstruction by morphological operations

Morphological operations is used for adding and removing pixels from an image depend on the shape and size of structuring element which is used to process the image. This approach use imerode () function of matlab first, which remove the pixel from an object by using structuring elements. then applied imdilate () function which adds the pixels to an object. Then imopen() function is applied to the eroded image to fill the vertical strokes of the component. Then again imerode () applied to that image and reconstruct the marker and mask image by joining them together. Here, mask image is the original image and marker image is the last eroded image. This process has given reconstructed image i.e. distorted image have regained its original information at some extent.

4) Interpolation

After, morphological operations reconstructed image has restore some sort of information but when OCRd the reconstructed image it doesn't give expected result i.e. it give junk output in which not even a single word match with original image. For that interpolation has been applied to enhance the image more.

This research has used nearest neighboring interpolation, the reconstructed image has used as input image to the Interpolate where pixels are resampled with the help of nearest value of pixel in the image to fill the empty space. Then interpolated image has passed through Hindi OCR which gave observable output.

5) Evaluation using OCR

Figure 5 original image after OCR

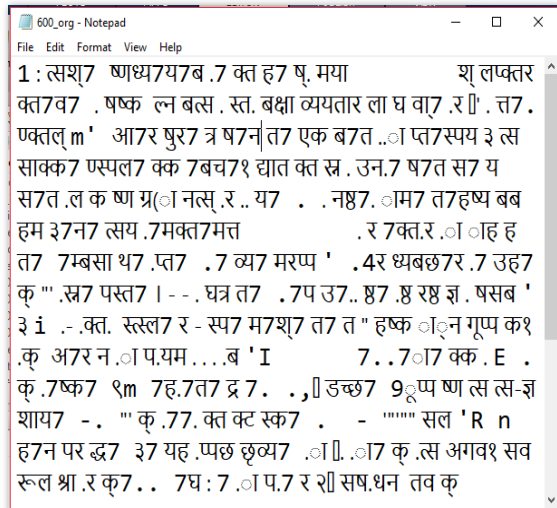
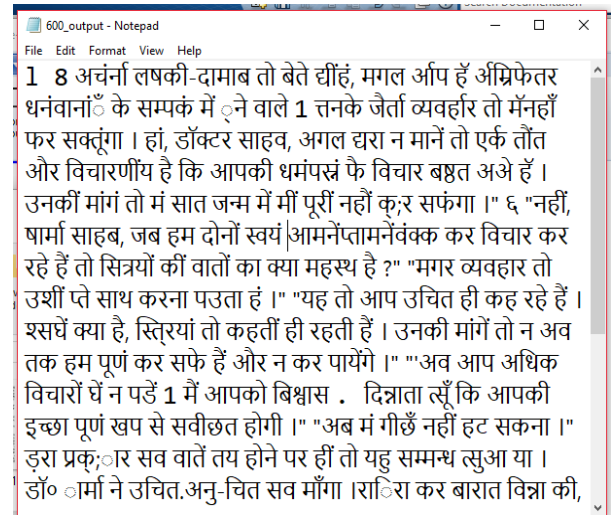


Figure 6 enhanced image after OCR



III. Implementation results

Table 1 document analysis results

TOTAL IMAGE 220	With proposed approach, Total no. of image that correctly analyzed that they are not suitable for concerned application i.e. OCR	Total no. of images that incorrectly analyzed, they are suitable for concerned application i.e. OCR
120 noised image	108	12
100 normal image	100	0

The analysis has been done on 220 images including 120 noised image i.e. colored, low resolution, low contrasted, salt n pepper and inverted images. 100 image are normal images i.e. doesn't corrupted with noise.

Table 2 OCR evaluation

Total images	No of image gives junk output when OCR	No of image gives junk output when OCR	No. of images gives understandable output
220			
120 noised image	67	30	13
100 normal images	NONE	NONE	100

In OCR evaluation 220 image has analyzed with OCR and identified 67 images has given junk output, 30 has given no output and 13 images gave understandable output.

Table 3 proposed approach result

Total images	Using Proposed approach, total images give readable result	Total image images gives still junk output
50	37	13

This approach has applied on 50 low resolution images in which 37 images are perfectly OCR and give understandable output and 13 images gives junk output.

IV. Analysis

Having knowledge about all type of noise models is an added advantage for people working in this domain. In previous papers, gamma value was 8 which lighten the given image extensively so here gamma value 1.02 has used. Enhanced low resolution image has been rectified by observing the OCR output.

V. Conclusion and future scope

For this dissertation analysis has been done for type of images included inverted or black, salt n pepper, low contrasted, low resolution and colored images. In this research all inverted images has correctly identified on the basis of size of connected component, All salt and pepper noised image has analyzed correctly on the basis of black pixel evaluation, low contrasted image has also correctly analyzed by using histogram evaluation, low resolution images has correctly identified on the basis of average height of connected component but some image has incorrectly identified as low resolution image whereas such images has given expected result when they OCRed, and some colored images has acceptably analyzed by identifying the third dimension matrix in the image information but some has wrongly classify as colored images whereas such image has accepted by OCR and gave understandable output. This problem was occurred when value of RGB color have same value then image will look like grayscale image.

A number of approaches have been researched on enhancement of low resolution document and then approach which best suited the required input i.e. low resolution image, is taken and implemented. For grayscale transformation averaging method is best suited apart from image luminosity. Then power law transformation has applied to enhance the contrast instead of considering histogram equalization. After contrasting the image, results are not satisfactory at some places. So to better the performance morphological operation has been used, morphological operation includes dilation and erosion that filled the degraded part and nearest neighbor interpolation has improved the image further rather than applying any other interpolation technique because it gave expected results which rectified by observing the OCR output.

- We can add more noised images for analysis further for example Gaussian noise, blurred image, and marginal noised.
- Frequency domain can be implemented on this analysis further.
- Bilinear and bi cubic interpolation can be applied on the resolution enhancement of low resolution.

References

- [1] Binh Quang Long Mai, Tue Huu Huynh, Anh Dong Doan, "A Study about the Reconstruction of Remote, Low Resolution Mobile Captured Text Images for OCR", Advanced Technologies for Communications (ATC), International Conference on 15-17 Oct. 2014, INSPEC Accession Number: 14947241, IEEE 2014, Ho Chi Minh City, Vietnam.
- [2] Mrs. S. Sangeetha, Mr. Y. Hari Krishna M.Tech Student, ECE Dept. "Image Resolution Enhancement Technique Based on the Interpolation of the High Frequency Sub bands Obtained by DWT" International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue7- July 2013 Hindupur. Anantapur (Dist), Andhra Pradesh, India.
- [3] Akanksha Garg, Shashi Vardhan Naidu, Tasneem Ahmed, Hussein Yahia, Dharmendra Singh "Discrete Wavelet Transform Based Resolution Enhancement for Low Resolution Satellite Images" Industrial and Information Systems (ICIIS), 2014 9th International Conference on 15-17 Dec. 2014, INSPEC Accession Number: 14920332, IEEE 2014.
- [4] Hirotaka Kudo, Yoshihiro Sugaya, and Shinichiro Omachi, "Low-Resolution Character Recognition with Compound Mutual Subspace Method", Signal Processing (ICSP), 11th International Conference on 21-25 Oct. 2012, INSPEC Accession Number: 13430230, IEEE 2012 Sendai, Japan.
- [5] Processing techniques for character recognition <http://cdn.intechopen.com/pdfs/11405>.
- [6] Zhan LI, Sino-France, Jiangao LUO "Resolution Enhancement from Document Images for Text Extraction" The 2011 Fifth FTRA International Conference on Multimedia and Ubiquitous Engineering 28-30 June 2011, Loutraki, Greece INSPEC Accession Number: 12191891, IEEE 2011.
- [7] Document Analysis as a qualitative research method <http://www.academia.edu/8434566>.

Acknowledgment

I take this opportunity to acknowledge all those who have guided me in this paper. I express my earnest gratitude towards P.T.O. Mr. Bhupendra Singh, my co-mentor for his valuable encouragement and his guidance. I am grateful for their most cooperative attitude and suggestions, without which I could not have been able to do this work and My colleagues their motivation and suggestions were invaluable in the project work. I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Age and Gender Prediction for Facebook Profile

Ankita Sharma¹, Kalpana Johari²

C-DAC, Noida

GGSIIP University, Delhi

Abstract: This research contributes to the finding the features which are helpful in predicting the feature age & gender of the profile holder of Facebook from the text data. The data is taken from Facebook profiles. The personal details of profiles holder are already known which helps in verifying the prediction results. This research focuses on the designing of algorithm that is used to train the machine to predict the age and gender of Facebook profile holder. Two machine learning algorithm are executed using WEKA tool for the classification of profile holder in different age group (Young- less than 18 years old, Middle- greater than 18 and less than 30, Adult- Greater than 30 years) and gender (Male, Female) that is Decision Tree and Support Vector Machine.

Keywords: Machine Learning, Age, Gender, Facebook, SVM, Decision Tree

I. Introduction

The age & gender are the identity of the person by which they are known in the society. With the increase in growth of internet the persons around the globe are connected through each other, sitting in one corner of the world person can see, talk and share their views with other person in any place with one touch or click.

The internet has given growth to the social platform sites such as Facebook, blogs, twitter and many others. These social networking sites contains the private data such as name, gender, their educational qualification, date of birth(age), and about work place also they contains the thoughts of the personal which is in the form of video, audio, text and graphics which are shared by the person itself. Since the creation of account on social networking site is free and there is no authentication process that can validate that the personal details provided by the user, hence there is high percentage that misleading details available in their profile. Also the fake profile holder can use others public details for some miscellaneous activities. Thus if there is some system that can judge the profile is fake or genuine.

Our research in one way to reducing the finding the fake profile by predicting the age & gender from the text data.

In our research the data consist of 75 profile data which is manually collected from the Facebook as the Facebook does not provide their data. These profiles are used for training and testing different machine algorithm with different features that are found in the research. Paper is divided into different sections: section 1 describes precious research done in same domain, section 2 describes proposed steps, section 3 describes the data and features used in our research, section 4 is results and 5 is conclusion and section 6 contains references.

II. Related Work

With the increase in the use of social platform a lot of research work is done based on the personal data that are available on their profile. The research [1] was done on Twitter data, [2] on Blog data, [3] on Dutch twitter data. There are different machine learning algorithms that are implemented in many research works such as decision tree, naive bayes, support vector machine, neural network, Deep Convolutional Neural Network (DCNN).

In [4] Cheng et. al. has provided the list of features which can that can be used to predict the gender of the person. These features are broadly categorized in five categories: character-based, word-based, function words, syntactic and structure based. These categories contain multiple features that can be used for prediction.

Also [1] Guimarães et. al. study have provided relationship between the character used in text and the age group of the person belongs to. Also they gave 13 parameter based on twitter data that can be used to predict the age group the person can fall.

Different studies have applied different machine learning algorithms: Decision tree, Regression, SVM, Naïve Bayes and DCCN with different set of data and achieved accuracy range between 75%-85%. Our approach works on Facebook data for which data is manually collected and due to privacy and security issue the data cannot be shared in public.

III. Approach

The objective of our research is to design an algorithm and train a machine to predict the age & gender of the profile holder of social platform web applications (Facebook) using Posts and about information.

The fig. 1 shows the diagram of the proposed approach of our research work. The approach consists of following steps:

1. **Capturing of profiles.**
2. **Subjective analysis of profile.**
3. **Gather the useful features available in literature**
4. **Using the features of previous two steps create list of important features.**
5. **Create machine algorithm for prediction of age & gender.**
6. **Training and Testing of Machine**

Our research approach consists of 6 steps which are listed above. In first step profiles are captures from the Facebook whom we already know to us. Once the profile data (About & Post) is collected then the next step of project starts that is feature extraction which is done by subjective analysis & literature study the subjective analysis come out with the intermediate result that are direct features used for prediction and with literature came with 23 features that are listed later in this chapter. After this hypothesis are made for building the final essential feature list required for prediction. The list of final features contains 23 which also include 4 readability measure score. This hypothesis is base for our research for predicting Age (Young, Middle, Adult) & Gender prediction (Male, Female). Then using WEKA we will be implementing the machine learning algorithms (Decision Tree, SVM).

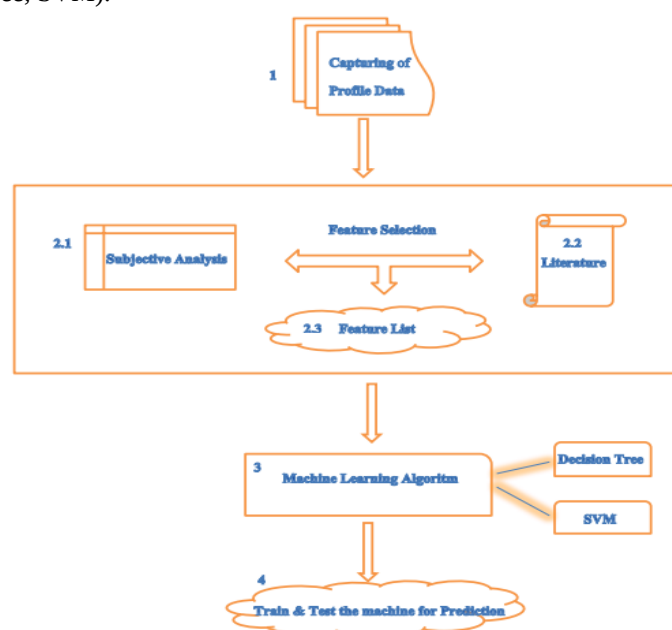


Fig. 1 Proposed Approach

IV. Features and Data

With the subjective analysis and literature study we have come up with 23 features which include 3 direct features from About information available, and 4 readability measure score and others are indirect features.

The data is taken from Facebook profiles. The personal details of profiles holders are already known to us in verifying the prediction results. The data collected is in the form of text only. The text data is from the About section of the Facebook profile that helped in retrieving the direct information from the profile. Another text data was collected from the posts that are mad by the profile holder on which indirect features can be used for the prediction of the age and gender.

Indirect features used in our research are : Work & Education, Interested In, Total number of character, Total number of letters, Total no. of upper characters, Total, Total no. of white space, Number of Special characters, Total no. of words, Average length per word, Words longer than 7, Total no. of short words(1-3 characters), Hapax Legomena, Total number of lines, Total number of sentences, Total number of paragraphs, Average number of words per paragraph, Average number of words per sentence, Number of article words, Number of pronoun words, Number of auxiliary verbs, The number of net abbreviation/N, Flesch–Kincaid readability tests, Gunning fog index, Automated Readability Index, Coleman–Liau index are the final features that are used for age & gender prediction.

V. Results

We have performed 2 machine algorithms with different combination of splits and combinations of features that have been listed before.

We have come up with the different results the result for direct feature for age & gender accuracy is 86.6% and 89.3% respectively and for indirect features 84% and 86% with decision tree. While for indirect features result for age & gender is 88% and 86.66% with SVM.

SNO	ALGORITHM	PREDICTION	CORRECT %	INCORRECT %
1	Decision tree	Age	80%	20%
2	Decision Tree	Age	71%	29%
3	Decision Tree	Gender	81.18	18.81
4	Decision Tree	Gender	80	20
5	SVM	Age	81.18	18.81
6	SVM	Gender	86.66	13.33

Table 1. Result Table

VI. Conclusion

The readability score features individually is not helpful in predicting either age or gender as for same text different score predict different age group for same person hence it require to be joined with other score for much accurate result.

Also features such as Number of lines, Paragraph, and number digits does not plays very significant features for prediction of age & gender that is if these.

If we compare all the results of SVM and Decision tree we can say that SVM performance is better than the decision tree for the data and the features used in our research. The performance of machine may vary with increase in size of data and using more other features which will be handled in our future work.

References

- [1] Guimarães, Rita Georgina, et al. "Age Groups Classification in Social Network Using Deep Learning." IEEE Access 5 (2017): 10805-10816.
- [2] Goswami, Sumit, Sudeshna Sarkar, and Mayur Rustagi. "Stylometric analysis of bloggers' age and gender." Third International AAAI Conference on Weblogs and Social Media. 2009.
- [3] Nguyen, Dong, et al. "" How Old Do You Think I Am?" A Study of Language and Age in Twitter." ICWSM. 2013.
- [4] Cheng, Na, Rajarathnam Chandramouli, and K. P. Subbalakshmi. "Author gender identification from text." Digital Investigation 8.1 (2011): 78-88.
- [5] Barracuda Labs (2011) Barracuda labs social networking analysis. <http://barracudalabs.com/fbinfographic/>. Accessed 2012-5-5.
- [6] Raghvendra, Mr, and Sandeep Sahu. "Gender Recognition & Age Prediction." International Journal Of Engineering And Computer Science 3.09 (2014).
- [7] Count Wordsworth, <http://countwordsworth.com/wordsperparagraph> accessed on 13/05/2018 on 11:28
- [8] Character Count & Word Count , <https://charactercounttool.com/> accessed on 13/05/2018 on 12:20
- [9] Readability test tool, <https://www.web pagefx.com/tools/read-able/> accessed on 13/05/2018 on 12:20
- [10] Crawford M. Talking difference: on gender and language. London: Sage; 1995.
- [11] Yo, Take, and Kazutoshi Sasahara. "Inference of Personal Attributes from Tweets Using Machine Learning." arXiv preprint arXiv:1709.09927 (2017).
- [12] Enron e-mail dataset [Online]. Available, <http://www-2.cs.cmu.edu/wenron/>; 2005, April.
- [13] Reuters corpora [Online]. Available, <http://trec.nist.gov/data/reuters/reuters.html>; 2000.
- [14] Analytics Vidhya, <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> accessed on 16/05/2018 on 17:38
- [15] Jang, Jin Yea, et al. "Generation like: comparative characteristics in Instagram." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015.

Acknowledgement

I take this opportunity to acknowledge all those who have guided me during this project. I express my earnest gratitude towards Dr. Kalpana Johari, my project guide for her valuable encouragement and guidance.

Her motivation and suggestions were invaluable in the project work. I am grateful for his most cooperative attitude and suggestions, without which I would not have been able to do this work.

I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Offline Signature Verification

Mrignainy Kansal¹, Tushar Patnaik²

¹CDAC-Noida

GGSIIP University, Delhi

Abstract - Handwritten Signature Verification is one of the most challenging behavioral biometric measure. Handwritten Signature is widely use in our everyday life as a mean of personal identification. It uses in numerous fields as like in banking, business transaction, document authentication and many more. In this paper I present a novel method for automatic [4] extraction of signature from bank cheque and then perform signature verification. The method use for signature verification consist of image pre processing, feature extraction and verification. Performance evaluated by the K-Nearest Neighbor classifier and SVM [12] classifier.

Keywords - Bank cheque, Handwritten signature verification, support vector machine, K- Nearest Neighbor [12].

I. Introduction

Biometric is the department of technology for identifying or verifying a human identity. Physical and Behavioral traits of a person are tested in the biometrics take a look at. Bodily(Physical) traits include facial expression, features of the eye (i.e. retina and iris), fingerprints impression, hand palm geometry. Then again, behavioral traits consist of handwritten signature, gesture, keystrokes or typing, voice, gait. Handwritten signature is one of the biometric measure.

Automatic extraction of signature from bank cheques has been the high cognizance of researchers worried for document analysis from the past one decade. Bank cheques remains in massive call for regardless of the general fast emergence of e-commerce. One of the maximum important tasks in bank cheque processing is the extraction of handwritten signatures from bank cheques after which feeding them to an off-line signature verification system [11] which tests the signature for identification and authentication.

Signature is a pattern of writing which uniquely identify the individual's identity. As like the other behavioral biometric, signature is one of the widely used as a mean of personal identification. Signature is one among proof to validate a file, that why, if a document that has the authentic signature is consider as a completely precious asset. Signature verification is a process which use to recognize or identify the individual identity. Several strategies are used to authenticate signature. Among the others, handwritten signature is one of the most full-size manner for private authentication. In all advanced nations, people learn to annex the signature inside the early age and practiced over a duration of years. So, handwritten signature is a totally private sample that originates from a complex generation method regarding the instantiation of an action plan stored within the brain of the signer and its execution by means of his/her frame (arm, hand) the use of appropriate writing gadgets (pen, pencil, paper, and many others.). Therefore, it's not surprising that signature evaluation is an exceedingly complex hassle. Signature verification can be divided into important areas depending at the information acquiring technique, off-line signature and on-line signature verification.

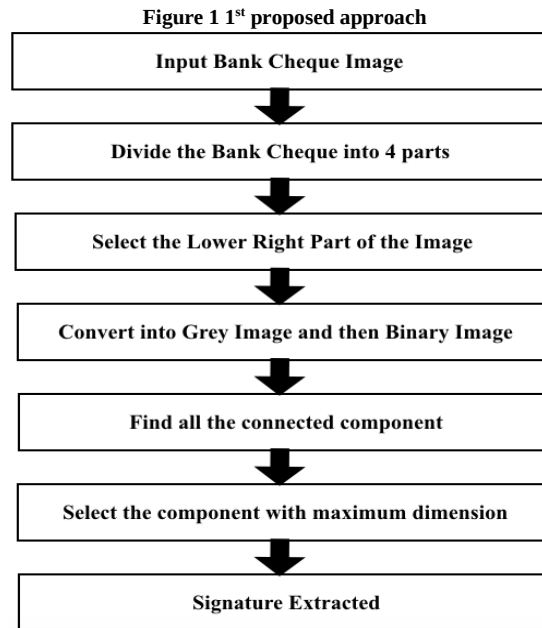
II. Proposed Approach

There are two approach defined in this paper

- Automatic Signature Extraction from bank cheque
- Signature Verification

A. Proposed approach for automatic signature extraction from bank cheque

- 1) Input Bank Cheque Image - The bank cheque database consists of 100 cheque images from 4 different banks (Syndicate bank, ICICI bank, AXIS bank, Canara bank) scanned at 300 dpi resolution.

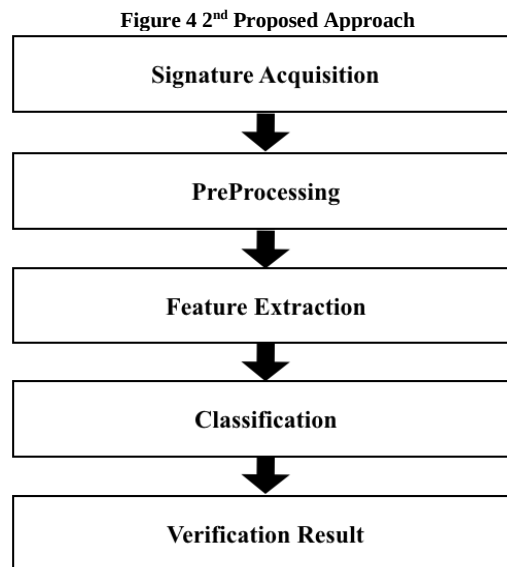


- 2) Divide the bank cheque into 4 part – Bank cheque having number of field in it. Divide the cheque into 4 part and only work over the right lower part of the cheque.
- 3) Convert into binary image – Bank cheque is always colourful. Convert the RGB image into gray image and then convert into binary image. For binary conversion global method of binarization with specific value is used.
- 4) Find Connected component – After getting the binary image to remove the exterior things need to find the connected component in an image. Connected component can be define as we receive the pixel in the 8 direction.
- 5) Select the component with maximum dimension – Now we have all the connected component in an image. Select only that component which having the maximum dimension. And receive the signature part from the bank cheque.

Figure 3 Extracted signature from bank cheque



B. Proposed Approach for signature verification



- 1) Signature Acquisition – Image acquired through scanning or by digital camera. The signature database consists of 210 genuine signature and 220 forged signature. There are 42 authors who give set of 5 genuine signature each on white paper and 220 skilled forger of 42 authors.

Figure 5 Genuine signature images from database

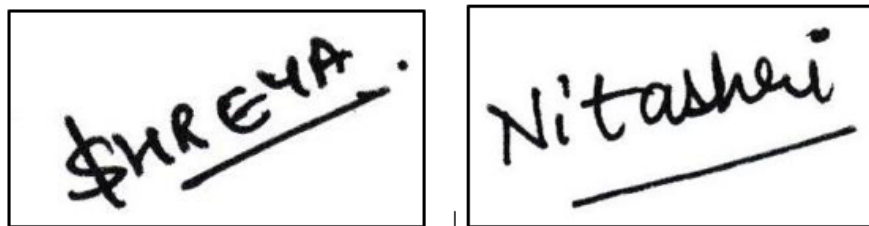
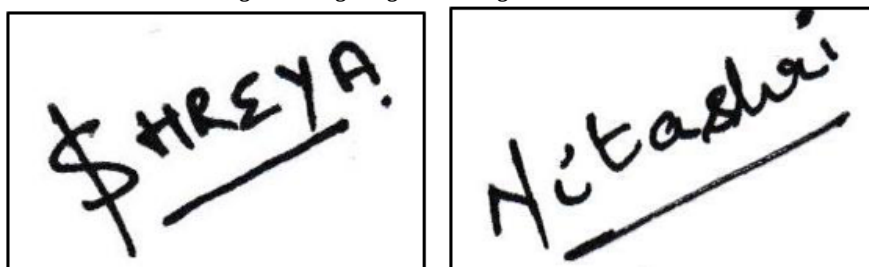


Figure 6 Forged signature images from database



- 2) Preprocessing – To improve the quality of the image pre processing is performed. The image will be converted into grayscale and then it will be binarized. Noise will be removed so that the result of further steps could be more efficient.
 - a) Binarization : Convert the RGB image to gray image and then binarized it. Convert the image into two dimensional matrix of 0's and 1's.
 - b) Noise removal : At the time of scanning the noise get introduced in an image which reduce the accuracy and efficiency of a system performance. So this need to be remove from the image with the help of different filters such as Median filter, Average filter, mean filter, Gaussian filter. Median filter give good result in this study.
 - c) Thinning : Thinning is the morphological operation. In thinning the thickness of the image get reduce from the edge of the image. In signature the thinning is the important part of the pre processing.
- 3) Feature Extraction : In this step the uniqueness of an image get extracted on basis of which we compare the different signature and analysis the result. In this step I applied two different algorithm and then compare their result on the same database.

a) **Global Features** : These features are those which consider the image as a whole and the calculate the uniqueness in the image. In these feature the image as a whole is taken and the feature like height of an image, width of an image, skew angle, aspect ration, centroid of an image, etc. but we need to select the good feature which are capable of clearly differentiate between the forged signature and the genuine signature.

Height : Number of rows in the image is define as the height of the image.

Width : Number of column in an image define the width of an image.

Kurtosis : Pixel distribution in an image and then we fine the average of all the variation column wise that define the kurtosis. Higher value of kurtosis distribution indicates thicker tails, longer and a sharper peak whereas lower value denotes shorter, thinner tails[6].

Area : The coverage pixel of the signature in an image.

Skew Angle : Angle of tiltiness when compare with the straight line as 180° is the skew angle.

b) **Freeman Chain Code Algorithm** : Chain code algorithm is use to find the sequence of chain in the connected pixel. In this algorithm there is a 8 direction arrow which start from the starting point of the signature and continue till it find the connected pixel in an image. This mainly use for the shape recognition. Through this sequence of image the shape of the signature get learned by the classifier. Then the Euclidean distance is use to find the difference between the direction code of the sequence of the two images. In this algorithm the chain code is calculated for the connected component. And then the space code is calculated which is rotation invariant as well as scale invariant also.

4) **Classifier** : In this step I apply two different classifier as K-Nearest Neighbor and Support Vector Machine. And compare the result that which give the good accuracy for the dataset which given to them. Classifier is use to train the classifier and then test them to check the accuracy of the system. Classifier check the accuracy on the basis of the feature set which is given to them. There is not much difference between the K- Nearest Neighbor classifier and support vector machine. In the K- Nearest Neighbor the nearby class are found for the test image. In the SVM a hyper plane is draw which is able to differentiate between the two classes. The SVM is able to classify the linear as well as the non-linear data. SVM give good result when we have the less data for training the classifier. In SVM kernel use to classify, but it is difficult to select the appropriate kernel for the particular data. SVM give good result as compare to the KNN. In my project the SVM give high accuracy as compare to the KNN. As in the dataset there only 5 genuine signature per person which is less for the training. At this case SVM is a good classifier as compare KNN.

III. Result

The first proposed approach there are 100 images of bank cheque from which the approach which is proposed is able to extract the signature from 85 bank cheque. So, the system show 85% of accuracy. The rest 15% are not detected due to the constraint apply on it. The second approach show the best result of 91% with SVM using Global features.

Table 1 Comparison of Result

Classifier	Features	Accuracy percentage
K-NN	Global features	85.71%
K-NN	Chain code algorithm	80.12%
SVM	Global features	91%
SVM	Chain code algorithm	83.33%

IV. Conclusion

In this paper [4], I presented a novel approach for automatic signature [4] extraction and then verify them with the help of offline signature verification. For signature verification the use of two techniques for the feature extraction as global feature extraction and chain code algorithm. Classifier K- Nearest Neighbor and SVM also applied for the performance evaluation. The best performance given by global features using support vector machine as a classifier. Accuracy obtained as 91%.

V. Future Work

- 1) Signature which is not single connected component is also get extracted from the bank cheque.
- 2) Signature verification when combine with other biometric measure as like voice, speech recognition, etc. then may be it give good result.

- 3) Better pre processing method will be designed to improve the accuracy of the global based feature extraction [6].

References

- [1] Digital image processing by Rafael C. Gonzalez, Richard E.Woods, 2nd Edition, ISBN 0-201-18075-8, 2001.
- [2] Aini Najwa Azmi, Dewi Nasien, Fakhrol Syakirin Omar, "Biometric signature verification system based on freeman chain code and k-nearest neighbor ", Multimed Tools Appl, Springer, 2017.
- [3] Hemanta Saikia, Kanak Chandra Sarma, "Approaches and Issues in Offline Signature Verification System", International Journal of Computer Applications (0975 - 8887), Volume 42- No. 16, March 2012.
- [4] Muhammad Imran Malik, Marcus Liwicki , Andreas Dengel, Seiichi Uchida, Volkmar Frinken, "Automatic Signature Stability Analysis and Verification Using Local Features", 14th International Conference on Frontiers in Handwriting Recognition, Heraklion, Greece, IEEE, 2014, INSPEC Accession Number: 14827649.
- [5] Mohitkumar A. Joshi, Mukesh M. Goswami, Hardik H. Adesara, "Offline Handwritten Signature Verification Using Low Level Stroke Features", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, IEEE, 2015, INSPEC Accession Number: 15487119.
- [6] Subhash Chandra, Sushila Maheskar, "Offline signature verification based on geometric feature using artificial neural network", 3rd International Conference on Recent Advances in Information Technology, Dhanbad, India, IEEE, 2016, INSPEC Accession Number: 16140895.
- [7] Harman Preet Kaur, Anmol Sharma, "Offline Handwritten Signature Verification Using Zernike Moments", Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, India, IEEE, 2015, INSPEC Accession Number: 16072007.
- [8] Ahmad Sanmorino, Setiahi Yazid,"A Survey for Handwritten Signature Verification", International conference on Uncertainty Reasoning and Knowledge Engineering, IEEE, 2012, INSPEC Accession Number: 13058680.
- [9] Rania A. Mohammed, Rebaz M.Nabi, Sardasht M-Raouf Mahmood, Rebwar M. Nabi, "State-of-the-Art in Handwritten Signature Verification System", International Conference on Computational Science and Computational Intelligence, IEEE, 2015, INSPEC Accession Number: 15852783.
- [10] L. B. Mahanta, Alpana Deka, "A Study on Handwritten Signature", Interational Journal of Computer Applications (0975-8887), Vol.79, No.3h, October 2013.
- [11] www.aprs.org.au
- [12] K. Tsuda, "Subspace classifier in reproducing kernel Hilbert space", International Joint Conference on Neural Network (IJCNN), Washington, DC, USA, USA, IEEE, 1999, INSPEC Accession Number: 6597524.
- [13] Y. Guerbai, Y. Chibani and B. Hadjadj, "Writer- Independent Handwritten Signature Verification based on One- Class SVM Classifier", International Joint Conference on Neural Network (IJCNN), Beijing, China, IEEE, 2014, INSPEC Accession Number: 14564070.



Analysis of Households Electricity Consumption using Data Mining Techniques

Nitin Singh Bisht¹, Kriti Saroha²
CDAC, Noida
GGSIP University, Delhi

Abstract: - Energy analytics is an emerging area where a lot of research is conducted on electricity consumption analysis like consumer segmentation, characterization, predictions and knowledge extraction from the data collected from smart meter. In this paper, we discuss the analysis of household's electricity consumption is demonstrated using different data mining techniques such as k-mean clustering, decision tree, Naïve Bayes and f-p tree. Household electricity consumption analysis is the evaluation of the consumption requirement of a particular household so that further prediction could be made on behalf of the analysis. The results of the analysis are better demand management and improved energy efficiency. To improve, efficiently plan and develop this new system, a primary challenge is to analyse the electricity consumption data. To analyse the energy consumption and to achieve the objective, best analytic process in data mining techniques including exploratory data analysis and pre-processing, frequent patterns mining and associations, classification characterization, clustering is selected. This paper utilizes the techniques which are applied on the datasets in order to evaluate which techniques are useful for the better understanding of electricity load profile consumption data. As a result, the manner with which they use electricity in the home can be characterised based on individual attributes such as consumption of different appliances for a particular duration.

Keywords- Smart Meter, Electricity Consumption, data mining, classification, clustering

I. INTRODUCTION

The Smart Grid is an electricity supply network that uses digital communications technology to detect and react to local changes in the behaviour of the consumption of energy. It also provides us with a prospect of transferring the energy trade into a refined decade which will contribute to the economic and environmental strength of the society. During the transition period, it will be critical to carry out testing, technology improvements, consumer education, development of standards and regulations, and information sharing between projects to ensure that the benefits that are envisioned from the Smart Grid becomes a reality in near future.

Other benefits related with the Smart Grid are as follows:

- Well-organized transmission of electricity.
- More rapidly restoration of electricity after power failure.
- Reduced operations and management costs for utilities, and ultimately lower power costs for consumers.
- Compact peak demand lower electricity rates.
- Efficient integration of large-scale renewable energy systems.
- Better safety and reliability.

In today's time when an electricity disturbance such as a shutdown occurs which could have an immense effect resulting in sequence of failures affecting different important trades of the society. Increased resiliency of the new improved system which in return would be well prepared to address all the natural as well as manmade disasters. The grid which is smart operates in two-way collaborative volume, which enables the rearranging the direction when grid failure occurs as power shortage could be minimised to the minimum level and these reductions would be helpful in normalising the daily work. During energy outage, the grid which is smart is capable of detecting and isolating the outages and minimizing the damages. This new technology helps in reassuring the power is back on time with minimum delay as much as possible and is one of the important application of the smart grid. Organising these "distributed generation" resources, would provide necessity to public stores or departments which are governed by government or individuals, operating during emergencies is one of the important features on the grid operations. In the age of technology, old and mechanically operated system are widely needed to be manhandled, resulting in many failures which may last for the bulk of hours. It

would increase awareness to consumers how energy should be properly utilised by them and planet could be saved.

This paper presents methodology for clustering and classification of electricity load profile that are reflective of the varied manner with which electricity is used within the home. The optimum clusters were formed so that classes could be found in order to achieve our objective which would match the result by categorization of the clusters formed. Results would be found in the form of frequent devices used and profile class for high consumption and low consumption accordingly different users are classified upon their use.

The task is conducted in 3 phases during which different process are performed on the dataset. During first phase clustering is done and during the second stage, classification technique is applied on the generated clusters. In the final phase, the pattern mining is carried out on the original data available.

II. RELATED WORK

In [3], the authors discuss all the major data mining techniques in details in order to understand the electricity load profile better. Techniques used are beneficial in reducing the peak load and predicting which household devices take more consumption as compared with others. All the techniques were applied on two different datasets out of which one is smart and other one is non-smart. K-mean and the K-medoids clustering algorithms were selected to be applied on the gathered dataset. Both delivered the equivalent result, so only k mean was selected which was later applied on the complete data. The results of clustering discovered the peaks in clustering plot suggesting high consumption of the devices which are being heavily used in the house. Air conditioner, wash dryer and electric heater were the devices consuming the most than the rest. Frequent devices used showed the devices which are being used frequently among the houses. Kitchen devices are mostly used as compared to other devices like Refrigerator, Microwave, Stove and lightning etc. Kitchen appliances were among the devices used more.

In [4], the authors discuss a methodology which is useful in creating series of representations for the profile class load of the domestic sector by using clustering approach. Self-organising maps represented a decrease in Davies–Bouldin index thus suggesting a good technique to use and was therefore selected to segment the data. Nine clusters were formed by using self-organizing maps creating a lattice structure in hexagonal form in the matrix form of three by three. Different profile classes representing dissimilar patterns of the use of the electricity in terms of scale and timing were formed. Based on the manner of which each individual utilizes the electricity it was probable to classify consumers and there means in which they utilize the use of electricity even without the prior knowledge of the consumption of the household electricity. For the classification of customer profile classes, the author used the mathematical techniques in order to find which profile class customers opted to prefer for the majority of the duration for the period. A multi-nominal logistic regression technique was selected to be applied as to determine the pattern with characteristics of certain importance of households using the consumption which is an electrical load in a similar manner to individual profile class.

In [5], Demand response and Energy efficiency are the programs and utility policy used in the paper. By utilizing the customer load shape profile, it is easier to target resident with the potential for benefiting from demand response and energy efficiency. Based on the load shape targeting, the potential of having significant benefits which increases the likelihood of achievement from these utility programs by successfully engaging them. Load profiles, are merged for the utilization of entropy based on the benefit it has to offer. For example, based on lifestyle profile energy reduction and critical pricing when the load is at the peak, it would vary from the devices which is consuming based on the recommendation utilize by most utilities. Result was focused on the shape from morning peak to the afternoon, afternoon peak recommend that activity might be earlier or later on the day. As a mean of energy conservation shapes can also be utilized by analysing the peaks in the different load shape. Entropy within the household data is also capable of yielding the potential for understanding the success between the design and the targeted. Variability in the residence stipulates the high entropy in the particular house whereas utilizing the energy have low potential in demand and response program when energy reduction programs such as discounts for appliances were applied for.

In [1], the authors proposed a processing system for data which analyses the consumption patterns of the electricity, clustering technique such as self-organizing maps and K mean clustering were used. The system which was presented for the complete analysis of the data was being tested and validated using the real data. Clustering technique was applied on this real data resulting in the behaviour of the daily consumption according to the load curves of the different time zones. Identification of patterns from the demand curve was presented by the system, which was not useful in small scale surrounding but very much suited when the region is vast like entire nation. For forming the clusters, K mean algorithm was selected resulting in the formation of five clusters as five is the parameter which was selected as optimal number for cluster formation. Similarities in between the clusters were easily predicted and when difference were compared with those of others are appreciable. After

the clustering was achieved, the system needed to be retrained using decision algorithm. Decision algorithm designed to control when the systems need to be retrained due to changes in the load curve behaviour. Basically, retraining is triggered when the aggregated demand for a day is outside the margins of the aggregate demand of the cluster inside which it has been grouped.

In [2], Incremental summarization and Pattern characterization framework was proposed for the purpose of mining of the data Prediction and analysis of the total electrical energy consumption collected from the different meter reading was proposed in this paper. Interim summarization feature addressed the data stream nature in certain surrounding. The reading was accumulated temporarily from the interim summarization feature, as a result the structure on the dimensions for extraction of subject- oriented digest was quick enough for the information which is being consumed. Algorithm is the second most important feature of the proposed system which actually empowers the knowledge accumulated from the incremental learning and also utilises the historical results to acquire new important knowledge. Interleaving Parallel Compositional Languages produces characterized forms as its outcome. Outcome of Interleaving Parallel Compositional Languages enable to identify cyclic forms, merging forms and splits form from the dataset. In the last, fuzzy patterns were also used as matching in order to consolidate the digest with the outcome of Interleaving Parallel Compositional Languages. Interleaving Parallel Compositional Languages outcomes from the historical results were compared with the Numerical representation of the dimensions of interest of the selected digest. Outcomes with high proximity to the digest are further scrutinized to aid the prediction of the energy consumption. Overall, the Incremental Summarization and Pattern Characterization framework addresses the limitations to data mining in an electricity meter environment with functionality for interim querying, autonomous incremental learning, fuzzy fusion and data-driven prediction. Exploration of granularity and Ad hoc querying and trend analysis were the outcomes of the Interim Summarization. Complex details with multiple scopes were determined by averaging the total consumption of the premise which is opted as apartment during the busy duration of the day. The dimension model is able to organise granularity on the data.

III. PROPOSED APPROACH

Data was collected from publicly available dataset for smart meter data. The data contains home energy consumption per minute level electricity usage data from anonymous homes.

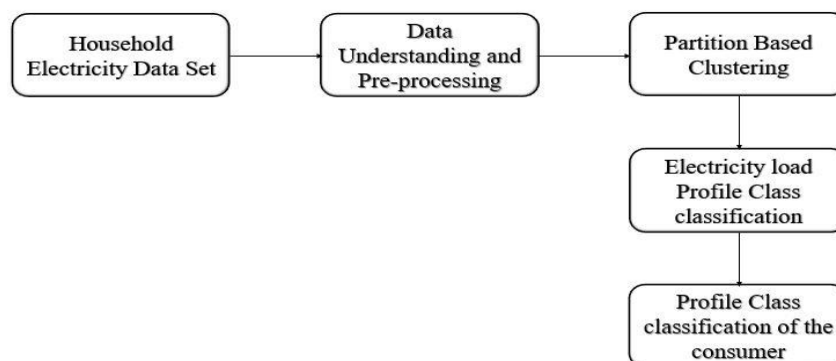


Figure 1: - Proposed approach for Household consumption and its analysis

Data Understanding/Pre-processing: - In data pre-processing, the dataset is pre-processed by considering attributes for analysis and by removing the empty rows. Removal of redundant data is also performed in this step.

Partition Based Clustering: - As shown in figure 1, the clusters are formed based on the data collected from the pre-processing. Optimum number of clusters are selected by using a suitable technique optimal for the dataset. Clusters are formed by selecting the value of k from the selection technique applied. The result would be optimal clusters. Partitioned based was selected because it works better with numerical data and has minimum complexity.

Electricity load Profile Class classification: - After the clusters were created, the profile classes were created based on the clusters which were formed on the above steps. Two profiles were created, one having higher consumption and other having low consumption of electricity. The profile created were used to characterise the data within the home.

Profile Class classification of the consumer: - After the profile classes were created, the classification was performed on the basis of created clusters. For this purpose, two classification technique were selected decision tree and naïve based. The results indicated that decision tree performed better when compared with naïve Bayes.

IV. RESULTS AND DISCUSSION

4.1 Partitioned Based Clustering: - For the partitioned based clustering algorithm, k-mean is selected for the data set and accomplished by applying it successfully on the dataset. In order to implement k, mean, first value of k is to be predicted. The value of K was derived by using elbow method.

4.1.1 Determining number of clusters: - To find the value of k in K- mean algorithm is the challenging task, right value would result in optimum clusters whereas wrong would not do any enhancement in the analysis. Different techniques are available out of which elbow was selected among the group to get the optimum result as clusters.

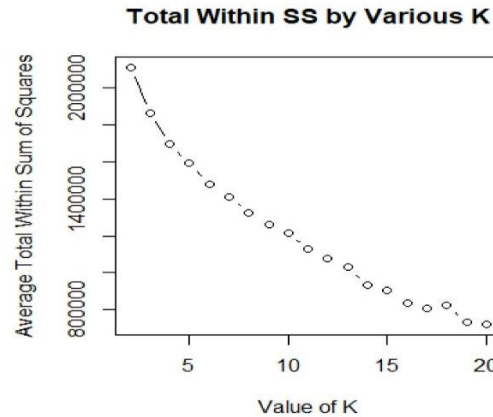


Figure 2: Sum of Squares Plot

As depicted in figure 2, Sum of Squares Plot determined the value of k needed to be considered. The plot represented the decrease in SSE as the value of k reduced and improve upon similarly. The plot showed a strong elbow at K = 15.

4.1.2 Cluster implementation: - After the value of k is predicted, K mean clustering technique is implemented. After implementing the clustering algorithm, the observation was made that the peaks represent the high usages of energy by the appliances or devices by individual of the house.

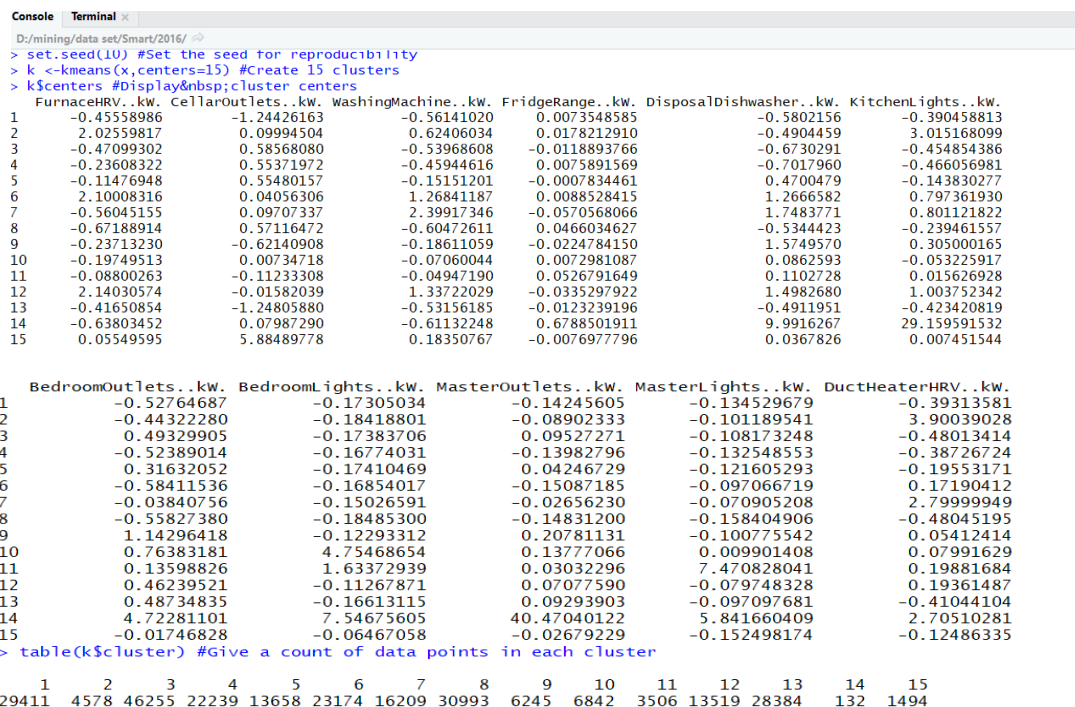


Figure 3: Clustering Result

4.1.3 Cluster Results Interpretation: - As shown in figure 3, Cluster 1 is more dominant in the fridge, which means the appliance fridge is consuming more. Cluster 8 looks to be a heavy at bedroom lights and above the average towards furnace and washing machine but low towards fridge. Cluster 15 is potent in the bedroom lights but soft at kitchen lights.

4.2 Classification implemented and result analysis: - Classification is performed on the clusters which were formed after applying partitioned based clustering. The result of two classifiers j 4.8 and Naïve Bayes were compare in table 1. Classification result showed j 4.8 to perform better when compared with naïve based. J4.8 presented better accuracy on individual meter, individual house and grouped houses.

S.no		J4.8	Naive
1	Individual Meter	96.4481 %	87.1585 %
2	Individual Houses	96.9945 %	91.5301 %
3	Grouped Houses	83.871 %	77.8508 %

Table 1: Comparative analysis of classifiers

V. CONCLUSION AND FUTURE WORK

Analysis of historical data containing the consumption of the devices of different household are being used for better demand management and energy efficiency. Firstly, raw data was cleaned, then clustering was performed on it for which partitioned based clustering was selected and k-mean was applied on the cleans data. The value of k was selected by using the elbow method. After the optimum clusters were created, the result was analysed based on weekend, daily and seasonal which predicted the consumption of electricity. Then based on the clusters, two profile classes were created, high consumption and low consumption. Then classification was performed based on it. Decision tree and naïve based was selected and the result was compared. Decision tree provided better accuracy. Finally, FP growth was applied on the data set to determine the frequent devices used.

REFERENCES

- [1] Hernández, Luis & Baladrón, Carlos & M. Aguiar, Javier & Carro, Belén & Sánchez, Antonio. (2012). Classification and Clustering of Electricity Demand Patterns in Industrial Parks. *Energies*. 2012.10.3390/en5125215.
- [2] D. De Silva, X. Yu, D. Alahakoon and G. Holmes, "A Data Mining Framework for Electricity Consumption Analysis From Meter Data," in *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 399-407, Aug. 2011
- [3] U. Ali, C. Buccella and C. Cecati, "Households electricity consumption analysis with data mining techniques," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 2016, pp. 3966-3971
- [4] McLoughlin F., Duffy, A., Conlon, M. "A clustering approach to domestic electricity load profile characterisation using smart metering data", *Applied Energy*, Vol. 141 March 2015.
- [5] J. Kwac, J. Flora and R. Rajagopal, "Household Energy Consumption Segmentation Using Hourly Data," in *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420-430, Jan. 2014
- [6] *Data Mining: Concepts and Techniques* Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2001.



Flight Crash Investigation Using Data Mining Techniques

Vishali¹, Kriti Saroha²

¹CDAC, Noida

Abstract- Data mining is a broad field of data science that is used in numerous fields like education, medical, investigation, environment etc. In aviation industry, data mining is used to make predictions on future based on patterns found in collected data. Using sophisticated data mining techniques, millions of data can be searched to spot patterns and detect real causes of aviation accidents. This paper gives an insight into various data mining techniques which are efficient in analyzing flight crashes.

Keywords- Aviation, data mining, classification, clustering.

I. INTRODUCTION

Aviation industry is the business sector dedicated to manufacturing and operating all types of aircrafts. Over the last 100 years, there is a drastic change in aviation industry. Aviation industry is stepping up from Wright Flyer planes to Boeing 747 to 787 Dreamliner[5].

As there is a drastic increase in new aviation technologies but still there is a need to improve security measures to avoid accidents. With the increment in aviation accidents there is loss of human life, money in millions of dollars and efforts & time taken by industry. But safety features also change from last 100 years and still the field of aviation is always searching new ways to improve their safety. As an aviation industry collects several case reports including structured and unstructured data and parsing all such data by hand would be impossible. Thus, poses a problem while investigating accidents. But with the relatively new fields of Data Mining it would be able to parse through unmanageable amount of data to find patterns and anomalies that indicate potential incidents before they happened.

II. THEORITICAL BACKGROUND

The probability of crashes caused by pilot error is 50% and causes of such errors can be ranging from checking Flight Management Computer (FMC) correctly to miscalculations of the required fuel. Roughly around 20% of the aircraft losses are due to mechanical failures, despite improvement in engineering and design standards still aircrafts occasionally suffer such failures. The danger by unpleasant weather conditions like storm, snow and fog can account around 5% of aircraft losses. The dangers by sabotage are just 10% and the remaining 15% error occur due to human errors like Air traffic controller(ATC), dispatchers and loaders, all these operations are handled by humans[3].

To overcome the above root causes of aviation industry, patterns are needed to be identified which are major cause of accidents. Such patterns are also known as incidents. Physically finding patterns or causes of accidents in data is not possible because of the large amount of data which is delivered every day. So, with the data mining techniques, can parse through such huge amount of data and find out all those unknown patterns which detect these causes of aircraft accident.

III. RELATED WORK

Several studies that have been conducted using data mining methods in different applications related to aviation industry are as follows:

David Pagels[1]in his research work, applied different Data Mining methods on three different types of aviation data. Three different type of data are_ Flight Data Recorder Data, Synthetic, Incident data and Data Mining techniques used are_ Hidden Markov Model, Multiple Kernel Learning, Hidden Semi- Markov Model and Bootstrapping Algorithm. Flight Data Record contains both continuous and discrete data and to find anomalies in mixed data Multiple Kernel Learning method is used. In detecting anomalies the effectiveness of hidden Markov model versus hidden Semi-Markov model were compared and for analysing the effectiveness of text classification incident reports were used. Previously either only discrete or continuous anomalies were identified in data but with this model heterogeneous anomalies can also be detected.

Christopher and Appavu[2]focused on improving the performance of classification model by applying different feature selection methods. The dataset contained data from year 1970 to 2011 on which five different classification models are applied and results are compared before and after applying feature selection techniques on these classifiers. Result shows that decision tree perform better than other classifiers with accuracy of 97.68% and principal component analysis perform better than other feature selections with 99.76% accuracy.

Shagun and Sabitha[3]aimed to analyse flight crash dataset taken from year 1908 to 2009 by using k-means clustering and cosine similarity. In this model, five clusters are formed containing both numerical and text dataset, in which similarities among different text is examined by using cosine similarity. Result model is able to find operator with maximum fatality and also result shows that fatality of ground is more than aboard.

Feng and Juanjuan[4] focused on pilot related accidents and incidents by using contrast set mining and attribute focussing techniques, applied on dataset taken from National Transportation Safety Board (NTSB) and Aviation Safety Reporting System (ASRS). In this various factors or their patterns were identified which are more likely lead to accidents than incidents.

Nazeri, Donohue and Sherry[5]investigates difference between aircraft accident and incidents. For the analysis dataset is taken from year 1995 to 2004 from National Transportation Safety Board (NTSB). In this work factors are divided into incident factors and accident factors. If only one factor is responsible for problem situation then it is labelled as incident and if two or more factors contributing then it is labelled as pre condition for the accidents.

IV. METHODOLOGY

In this section, we discuss the proposed approach.

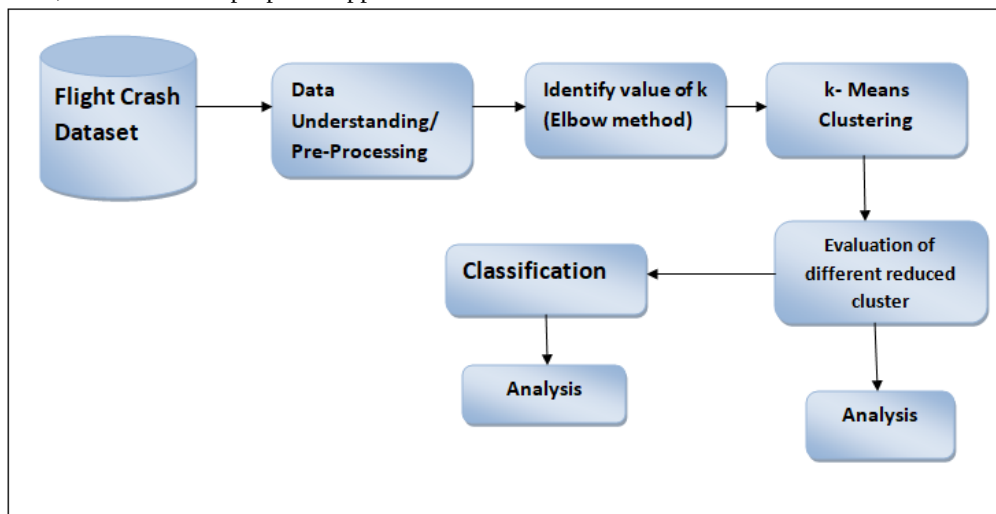


Figure1. Methodology diagram

The methodology for the proposed work is shown in Fig.1. The different modules are discussed below.

Flight crash dataset- It contains flight crash records from the year 1908 to 2008 having attributes like Time, Flight Number, Ground, Date, Registration, Location, Fatalities, Plane Type, Plane, Operator, Aboard, cn/ln, summary.

Data Understanding/ Pre-processing- Flight crash dataset is pre-processed by removing empty rows, Location attribute contain text which is converted into numeric by splitting Location into city and country then assigning individual numeric value to each country. Similarly, Operator attribute is also converted into numeric by assigning individual value to each operator.

Identifying value of k- Elbow method is used to determine the optimal number of clusters for k-means clustering. For flight crash dataset elbow is at $k=4$, indicating optimal number of clusters are 4.

K-mean clustering- It is applied on pre-processed dataset. In k means, the value of k which represents points as initial centroids are determining by using elbow method. Elbow method is used to find the appropriate number of clusters.

Classification- Different classifiers like Decision Tree, Multilayer Perceptron, K- Nearest Neighbour and Naïve Bayes are applied on above formed clusters to explore the use of different classification techniques. In this, items are assigned in a collection to a target classes and the classification is used to accurately predict the target class for each case in the dataset by the model compared with the actual classification in the test data.

V. RESULTS

After Pre-processing, the total number of rows in dataset is 3525 and total number of columns is 6 and attributes for analysis after data reduction are Operator, Aboard, Fatalities, Ground, Route and country. Using R studio k means clustering is applied on pre-processed dataset. In numerical dataset total number of items are 3525 and out of which 915 are in cluster0, 1156 items in cluster1, 1216 in cluster2 and 238 items in cluster3.

A. Clustering Analysis

1. Fatality rate with respect to operators-

The range of ground is 0 to 2750. Operator Pan American world airways/KLM has maximum fatality of 583 at Canary Island while travelling from Tenerife to Las Palmas. Then next operator is Japan Air Lines with next highest fatality of 520 while travelling from Japan from Hong Kong to Kuala Lumpur and both ids resides in cluster1.

The range of aboard is 0 to 644. In this observation also Operator Pan American world airways/KLM has maximum fatality of 583 at Canary Island. The second highest operator is Japan Air Lines with fatality of 520 while travelling from Japan from Hong Kong to Kuala Lumpur and both ids resides in cluster1. Third highest is Saudi Arabian Airlines/Kazastan airline with fatality of 349 while travelling from India to Dehran and this id resides in cluster0. Fourth highest is Turkish Airline while travelling from Paris to London with fatality rate of 346 and this id resides in cluster1.

2. Fatality rate with location-

Table1 describes all clusters with top fatalities rate with respect to the location (country).

CLUSTER	COUNTRY_ID	COUNTRY NAME	FATALITIES
CLUSTER 0	1624	Sri Lanka	476
	2403	Congo	337
	2634	Aichi	264
	3125	Queens	260
	1920	Antarctica	257
	2916	Nova Scotia	236
	2571	Macedonia	210
	2762	Dominican Republic	189
CLUSTER 1	100	Russia	2331
	74	USSR	1332
	80	Colombia	1081
	110	Brazil	1056
	165	India	964
	500	Japan	800
	1146	Canary Islands	761
72	Spain	720	

CLUSTER 2	110	Brazil	1071
	3	France	1006
	18	California	810
	80	Colombia	712
	123	Italy	657
	9	New York	606
	165	India	560
	500	Japan	539
CLUSTER3	2403	Congo	337
	1624	Sri Lanka	293
	2634	Aichi	264
	3125	Queens	260
	2916	Nova Scotia	236
	2571	Macedonia	210
	2762	Dominican Republic	189
	2159	Ukraine	184

Table1. Cluster analysis of all clusters

B. Classification techniques applied on each clusters

The objective of classification is to exactly predict the objective class for each case in the dataset.

Classifier Accuracy - Different classifiers like Decision Tree, Naïve Bayes, Multilayer Perceptron and K-Nearest Neighbour are applied on above formed clusters to explore the use of different classification techniques on flight crash dataset. From Table2, it is observed that Decision tree correctly classified instances are 3123 out of 3527 total instances, which gives 88.59% accuracy. Similarly other classifiers like Naïve Bayes, Multilayer Perceptron and K-Nearest Neighbour at k=3 accuracy is 78.92%, 87.48% and 86.63% respectively.

CLASSIFIERS	CORRECTLY CLASSIFIED INSTANCES	INCORRECTLY CLASSIFIED INSTANCES	ACCURACY(%)
Decision Tree(J48)	3123	402	88.59%
NaïveBayes	2782	743	78.92%
Multilayer Perceptron	3084	441	87.48%
K-Nearest Neighbour (at k=3)	3054	471	86.63%

Table2. Classifiers accuracy

VI. CONCLUSION AND FUTURE SCOPE

Analysis of historical data describing aviation incidents provide important source of knowledge to prevent such events in future and also reduce their occurrences. This prediction model is helpful in find those patterns in

Aviation Flight Crash Dataset that will act as incidents, which can cause the accidents. In conducting such analysis, data mining act as a useful tool to determine the causes of fatality so that preventive actions can be taken.

While evaluating the overall flight crash dataset from year 1908 to 2008, it is observed that aircraft increased from the year 1968-1978. Countries in which maximum number of flight crash were observed is Russia, China, France, USSR, Brazil, Colombia and Douglas DC-3 aircrafts are more prone to crashes as compared with other aircrafts. Among all the operators Aeroflot, military- US air force operator is more prone to accident as compared with other operators. It is observed that flight crash occurred due to pilot error, on the runway, due to bad weather conditions like fog, during takeoff and due to engine failure.

This model can be further improvised by including more characteristics records for incident occurrences and other classification techniques can also be applied to improve accuracy.

REFERENCES

- [1] David A. Pagels, "Aviation Data Mining", (2015) University of Minnesota, Morris Undergraduate Journal.
- [2] A.B.Arockia Christopher & S.Appavu Balamurugan, "Prediction of Warning level in Aircraft accidents using Data Mining Techniques", (2014) The Aeronautical Journal volume 118 No.1206.
- [3] Shagun Sharma, A. Sai Sabitha, "Flight Crash Investigation Using Data Mining Techniques", (2016) IEEE 1st Indian International Conference on Information Processing (IIPCIP).
- [4] Xia Feng & Juanjuan Li, "Analyzing Pilot-related Accidents by data Mining", (2010) International Conference on Computer Application and System Modeling.
- [5] Zohreh Nazeri & G.Donohue, "Analyzing relationships between aircraft accident and incidents", (2008) In Proceedings of the International Conference on Research in Air Transportation.
- [6] Jiawei Han and Micheline Kamber, Morgan Kaufmann, "Data Mining: Concepts and Techniques", Second Edition, 2001.



Analysis of Asthma Disease with Data Mining Techniques

Manmeet Singh¹, Kriti Saroha²
CDAC, NOIDA

Abstract- *Asthma is a disease which occurs due to inflammation of air pathways. Asthma is linked to environmental factors such as pollution, dust etc. This paper focuses on finding the risk factors associated with asthma. Data for this paper was obtained from 2016 survey conducted by BRFSS with 374 variables and 34190 observations.*

Keywords- *data mining, classification, decision tree, regression, Asthma*

I. INTRODUCTION

Asthma is a disease that occurs due to inflammation of air pathways. It is caused by combination of genetic and environmental factors. Factors such as pollution, allergens that produce vigorous immune response and medicines such as aspirin are considered to trigger an asthma attack. Although asthma is an incurable disease, symptoms can be controlled by using corticosteroid and avoiding contact with potential triggers. 358 million people had asthma and 397100 died in 2015. This paper focuses on finding the risk factors associated with asthma such as smoking, obesity, race etc. the data used in this paper is from a telephonic survey conducted in united states by BRFSS (Behavioural Risk Factor Surveillance System) which contains 34,190 observations with over 374 variables. Techniques such as decision tree, k nearest neighbour and multinomial regression was applied to find out which suited the best.

II. RELATED WORK

A lot of data mining techniques have been applied to obtain risk factors of certain diseases [1]. In [1], the author has used various classification techniques such as decision tree, logistic regression, neural network and gradient boosting in BRFSS 2013 dataset using SAS enterprise miner and SAS Enterprise Guide. The two binary target variables used in analysis were if respondents ever had asthma and if respondent still have asthma. They concluded that social and behavioural factors such as race, sex, age etc were associated with asthma. In [2], the association between paediatrics and exposure to mixture of pollutants such as nitrogen oxide, nitrogen dioxide, ozone, particulate matter 2.5. The data related to paediatric emergency department visits was taken from Texas emergency Asthma surveillance which is a network of hospitals of Houston. Pollution levels were gathered from Texas commission on environmental quality monitoring network. The pollution data was averaged to hourly basis. Apriori was applied with minimum support of 0.5 % and confidence of 0%. As a result, the conclusion was that exposure to ozone posed a significant danger to patients under 18 years of age and exposure to combination of chemicals is more harmful. Decision tree was used to find out whether patient has asthma or not on the basis of symptoms and level of control was measured using fuzzy interference system [3]. Accuracy using confusion matrix and Kappa statistics was calculated. Symptoms such as dry cough, sneezing, wheezing, trouble in breathing, suffocation was used to diagnose asthma using decision tree. Attributes such as day time symptoms, night time symptoms, limitation of activities, need for rescue medicine, shortness of breath were to measure level of control. 90 % accuracy rate was achieved for decision tree and 78.32 percent kappa coefficient for level of control using fuzzy interference. The author proposed fusion of naïve bayes and neural network to achieve maximum accuracy in [4]. The data for this paper was obtained from city hospital with 1024 observations and the proposed approach helps patients in diagnosis of asthma. The data included symptoms such as wheezing, sputum production, shortness of breath, blood in cough, history of allergy to common items such as soaps, detergents, pollens, fog etc. 88 percent accuracy was achieved for naïve bayes and 85 percent for neural network. The combination of neural network and naïve bayes achieved the accuracy of 93.5%. Spatial association mining was to find association between allergy and pollution parameters such as ozone, nitrogen dioxide, carbon monoxide, particulate matter 2.5&10 as well as distance from parks and roads [5]. The study was done on data obtained from Tehran's air pollution monitoring station for December 2013 and data of 1000 patients form Tehran's children's medical clinic. Air pollution data was categorized using air quality index and distance from parks and roads was also categorised as association mining does not work on numerical data. The result of the study shows significant relation of living distance from parks and roads to asthmatic allergy prevalence for December 2013.

III. METHODOLOGY

In this section, we discuss the proposed approach.

- **DATA SOURCE**

Data for this work was obtained from BRFSS survey for general health conducted in US in 2016. It had 34190 observations with 374 variables. Factors such as smoking, vaccine, HIV, doctor visits, insurance, race etc were recorded in the data. BRFSS collected this data via landline or cell phone.

- **ANALYSIS**

To identify the risk factors associated with asthma, three classification techniques- decision tree, multinomial regression, K nearest neighbor were applied. This helped in identification of best technique for prediction of risk factor. R studio and WEKA were used to apply the techniques. Two target variables in the data were a) if the person ever had asthma b) if the person still has asthma. Observations that were irrelevant such as file number, date, day, month, state code etc. were removed from the study. Also, the observations for which more than 50% data was missing were removed.

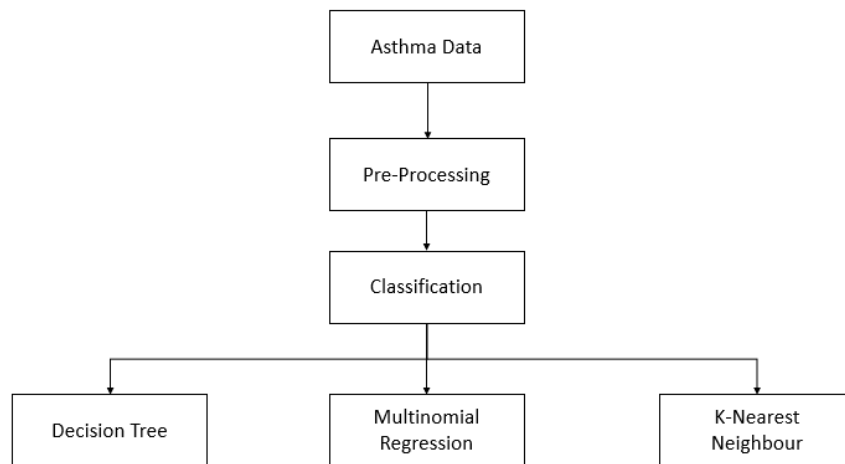


Fig 1

The classification techniques were applied to all the selected variables to find initial accuracy of the models. The methodology for the proposed work is shown in Fig.1.

IV. RESULTS

Three classification algorithms such as decision tree, KNN, regression were applied on dataset containing 34190 rows and 77 attributes. The technique with highest accuracy was decision tree. Regression produced best results for ‘Ever had asthma’ class. But overall accuracy of decision tree was higher. Time for computation was least in case of decision tree and most in the case of regression. Table 1 shows the performance of classifier on the basis of accuracy.

KNN algorithm was used with values of K as 1,2 and 3.

TARGET ATTRIBUTES	DTREE	REGRESSION	KNN		
			K=1	K=2	K=3
Still has asthma	89.22	86.6	86.30	88.8	88.88
Ever had asthma	85.05	85.13	81.10	84.29	84.14

Table 1: accuracy of classifiers with 77 attributes

V. CONCLUSION

Three classification techniques were applied to examine the risk factors associated with asthma on data collected by BRFSS. The results showed that decision tree is the best model for prediction of accuracy. KNN is better than decision tree. The results indicate that the factors such as race, OPD visits, smoking, etc are associated with asthma. These risk factors will help in providing better medical care and preventive actions.

REFERENCES

- [1] “A Data Mining Analysis Of Asthma Risk Factors”, Alma Pochini; Ben M. Williams; Hasanboy M. Isomitdinov; Gongzhu Hu ,2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence, Year: 2015, IEEE Conference Publications.

- [2] “Analysis of Correlation Between Paediatric Asthma Exacerbation And Exposure To Pollutant Mixtures With Association Rule Mining” Giulia Totia, Ricardo Vilaltaa, Peggy Lindnerd, Barry Leferb, Charles Maciase, Daniel Priced Department of Computer Science, University of Houston, Philip Guthrie Hoffman Hall, 3551 Cullen Blvd; Artificial Intelligence in Medicine; Elsevier-NOV 2016
- [3] “Asthma Diagnosis And Level Of Control Using Decision Tree And Fuzzy System” Aman Tyagi and Preetvanti Singh Dayalbagh Educational Institute, Dayalbagh, Agra 282005, India; Int. J. Biomedical Engineering and Technology, Vol. 16, No. 2, 2014
- [4] “Effective Asthma Disease Prediction Using Naïve Bayes - Neural Network Fusion Technique” , Saloni Aneja ,JIIT – Noida, uttar Pradesh, India ; Sangeeta Lal JIIT – Noida Uttar Pradesh, India; 2014 International Conference on Parallel, Distributed and Grid Computing
- [5] “Investigating The Relation Between Prevalence Of Asthmatic Allergy With The Characteristics Of The Environment Using Association Rule Mining”, Y. Kanani Sadata, F. Karimipoura, A. Kanani Sadat; Dept. of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Iran; Dept. of Civil Engineering, Islamic Azad University, Ahar branch, Iran; The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-2/W3, 2014;The 1st ISPRS International Conference on Geospatial Information Research, 15–17 November 2014, Tehran, Iran.
- [6] Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2001.s



Privacy preserving techniques-Outline of Security in Big data

¹Shipra Varshney, ² Ekta Jain

Department of Computer Applications

GGISP University/ Dr. Akhilesh Das Gupta Institute of Technology & Management, New Delhi

(Formerly Northern India Engineering College)

FC-26, Shastri Park

NEW DELHI-110053, India (New Delhi)

Abstract: Big data in last few years become a vital research topic. Big data security concern is increasing because of the use of Big data deployments projects. A single computer malware attack might breach the privacy of individuals which is valuable to think. This might leave the user's big data deployment subject to malware free demand. Now since data is large and collected from various sources processing in Big Data needs outrageous computational power and huge storage along with distributed system. The security's journey and challenges of Big data is very lucid simply by keeping out on unauthorized users and invasions with firewalls, needs robust and well built user authentication, training of end-user, and intervention protection systems (IPS) and interference detection systems (IDS). The prime objective of this paper is to provide an inclusive contraptions and methods of the privacy preservation methods in this demanding area. Also this paper covers the challenges for existing mechanism along with the challenges and upcoming research. The future research and it's inclinations with their orientations related to privacy preservation in big data has also been discussed in this paper. [1]

Key words: Big Data, Data storage, Intervention protection, Interference detection

I. Introduction

Huge amount of data is produced due to high-tech technological development with the various data sources like Internet, healthcare applications, sensor networks, social networking sites and various companies. Bulk of data is triggered from various kinds of sources in variety of structures, appearances along with their designs at fast speed. This bulk of data is known as big data. Various kinds of sources, their structures and appearances is the prime reason why big data security and data collection challenges have been a new and emerging topic from the last few years. Data growing speed and its volume is actually a concerned area and it is getting immensely difficult to store the data using conventional old fashioned methods or systems [1]. There are three categories by which big data can be accumulated and collected and it could be structured, semi-structured, or unstructured. The issues have been increased when we perform operations on data storage and processing tasks. New opportunities through data analysis can be created which can improve the internal decision making which now many big organizations and businesses are aware of. To promote the scientific research and economy many of these concepts are helpful in foreign to business today. Therefore by rebuilding and updating traditional business models and technical values we may reduce this risk [2].

II. Current System

Despite of big data used for various purposes, to better understand the needs, demands and for decision making, it could be effectively utilized for innovation in various fields. This ever growing data has increased the limits of potential invasion of privacy. For example, Google and Flipkart and Amazon can assimilate our buying preferences and scanning habits. Societal and public networking sites like Twitter, Face book store most of our personal information and social relationships of our life. Very commercial shopping websites such as Amazon, Flipkart suggests us the products based on our search. Similarly YouTube forwards us videos based on our explore history. Big Data not only gathers information, it is storing and extracting statistics for the purpose of increasing business and attaining profit, our personal information our privacy is on threat, security is on threat. In spite of removing AOL id and IP address of 20 million search queries in 2006 researchers took only handful of days to classify the users. Below mentioned are few scenarios under which users' privacy may be breached like [4]:

_Users distinctive and idiosyncratic data when combined with outside or extraneous data may classify the conclusion of new features regarding the users. User's actuality and details were not supposed to be revealed to others.

_Users distinctive and peculiar data is sometimes gathered and is collected in a way which adds value to the organization. Like the shopping habits of a user may gives a lot of information of personal data.

_ Data leakage may happen because this delicate data are stored and processed in unsafe locations during processing storage phases.

During recent years several methods have been developed to ensure big data privacy. Grouping can be done based on different methodology i.e., data storage, data generation, and data handling or managing. Falsifying techniques and access restriction techniques are used for privacy protection in data generation phase. One can restrict the access to individual's data. In other falsifying techniques original data is altered before it is accessed by the non trusted party. Privacy protection data in data storage is achieved using Data Encryption techniques. This further can be categorised into Attribute-based encryption (ABE); Identity-based encryption (IBE), and Storage-path encryption.

To save the crucial sensitive information in case, the crucial data can be stored in private cloud using hybrid clouds. In the second phase, Data Processing phase includes knowledge extraction from the data and privacy preserving data publishing (PPDP). To protect privacy of data, anonymization techniques like as generalization and suppression are used. The methods of knowledge producing and rendering have several ways which are used to extract information from complex and vast data using mining based techniques like classification, clustering and association.

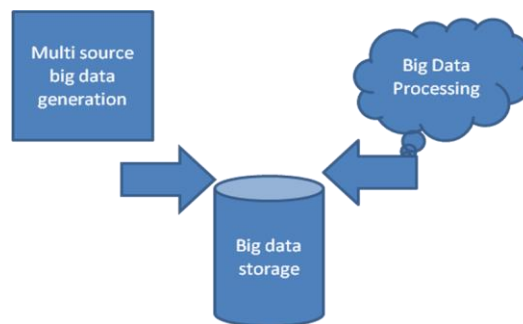


Figure 1. Demonstration of big data life cycle.

Input data has been split into different groups is the basic idea of classification and clustering and this can be done by using the association rules. This helps in finding the useful patterns and relationships in the input data. Now a day's it is a fast growing topic for the researchers "How to protect privacy in big data?" There are so many papers published on the foresaid topic but only few of them are reviewed and surveyed [2], [5].

Very few papers explain the fundamental concept of privacy protection in big data and they failed to enfold several important key points in this area. Namely neither [2] nor [5] provide exhaustive talks and considerations regarding big data privacy with respect to cloud computing. Less research with very fewer papers have been done so far which discusses the future challenges in detail in this area. To secure privacy in big data with technologies, a new but revolutionary concept is the overviewed area of this paper. The phenomenon of big data is discussed in such a way that this thing at each stage has been outlined. Besides in this paper we will discuss about potential research directions. The paper design and its format are defined as follows. The framework of big data and issues related to privacy of big data will be discussed in section III. Issues related to data generation phase and older privacy techniques will be discussed in section IV. Some recent and new ideas and techniques related to privacy in data storage and data processing phase will be discussed in sections V. Eventually upcoming and subsequent research directions are spotted in section VI.

III. Framework

A. Key Concept

The large amount of data in different dimensions by the means of volume, velocity, and variety, of big data needs to outline methodical and systematic systems. Now these systems are used to process at fast speed from a variety of sources. Various phases in big data in its life cycle are shown in Fig. 2.

The storage and processing of large repositories of data nowadays gives rise to new technologies which have been developed. Like cloud computing technologies named as Hadoop MapReduce, are considered for big data storage and processing. These sections briefly outlines the various phases in the cycle of big data contrasting to

this we try to explore how big data are influence from cloud computing technologies. Enough to mark this that certain disadvantages are also related with cloud computing when this is used for storage and processing of big data.

B. The phenomenon of Big Data

- **Data generation:** Various distributed sources generate data and this much amount of data collected by humans and machines has multiplied in the past few years. Some research has shown in their statistics that internet generates 2.5 quintillion bytes of data each day on average and 90 percent of the data in the world is generated in the past few years. YouTube users watch 4.1 million videos. Google delivers results for 3.6 million searches. Wikipedia users publish 600 new edits. Face book social networking site alone is generating 25TB of fresh data day-to-day. Normally, very large, diverse and complex data is generated. And this is the main reason why traditional systems find it difficult and hard to handle it. Some specific domains like business, Internet, research, etc abnormally associated this tremendous data.
- **Data storage:** This phase is the critical phase of the life cycle of big data for storing and organizing extensive data sets. Every Data storage systems consists of two parts i.e., hardware infrastructure and data management [4]. This is the first and foremost important infrastructure -Hardware infrastructure. This refers to minimize the lag time between collecting raw data and performing foreseeing driven actions. Each new generation of hardware continues to deliver remarkable improvements with its predecessors. On the top of hardware infrastructure a set of softwares are deployed for data management to supervise and query huge data sets. To help ensure system availability various softwares layers provides several interfaces to interact with and examines the stored data.
- **Data processing:** Various activities like data collection, data transmission, pre-processing and separating useful information come under this phase. Data which is coming from different and diverse sources needs to be collected and processed. The diverse sources are sites that contain text, images and videos. The data gathering phase needs to ensure that the gathered data are both defined and accurate so that subsequent decisions based on the findings are valid. Data collection technology acquires data from specific data production environment. After collection of data it is transmitted into a proper storage from a specific data production environment in the Data transmission phase. This requires a fast and powerful speed transmission mechanism to transmit data into a proper storage for different types of analytic applications.
- **Finally,** from the data collected, pre-processing phase aims at dealing and using some algorithms to transforms the data to a specific format. This means domain specific analytical methods are used by several applications to extract significant, purposeful information. Extracting, cleaning, and loading of appropriate data to a text file takes place. In data analytics different fields requires different characteristics and few of these fields may leverage similar underlying technology to inspect, transform and model data to extract value from it. Structured data analytics, text analytics, multimedia analytics, web analytics, network analytics, and mobile Analytics are the six broader classifications of data analytics research. This area covers exceptional technologies and algorithms to efficiently process the large quantities of data to obtain information within, a new area to research in near future[9].

IV. Proposed System Architecture

In data handling we can divide the privacy protection part in two phases. In the initial phase, the effectual aim in the collected information is to protect and shelter this information from uninvited disclosure. Why, because the accumulated data may bear crucial, fragile and delicate material instructions about the data owner. The second phase withdraws this meaningful information from the data without violating the safety and privacy because of this it is marked as the premier goal. Fine-grained encrypted data holder is allowed to dispense and share a cipher text of data to others with certain prescribed conditions. Normal encryption and decryption method is used. In existing system the patient any time monitoring the system and not secure. The patient should remember the key. It uses a re-encryption technique to store the data in a secure manner so that no information is lost. It is secure against chosen-cipher text attacks. We use Encrypt-DES and Random Key Generation method for encryption, decryption and key generation. This we can say is more secure compared to existing methods. The Encrypt-DES encryption method used so is acting as a user. There is no need to monitor the system and key any time.

A link between an information systems and the user is the input design. The specifications and procedures for data preparation to put transaction data into a usable form is a mandatory step. The useable data which is thus

collected in this form can be taken by inspecting computer by inputting the data from inputting directly into the system or it can also be collected from printed or written document outside. Some challenges like the how much input is required for controlling the errors, keeping away some delays, avoiding unnecessary or extra steps should be surpassed in the input design and concurrently keeping the process simple and sober. The privacy input is outlined and depicted with a good design for a simple and secure use with ease by keeping the privacy as well. The input design should consider the below mentioned things:

- What details and specifications of the data should be given as input?
- What are the various parameters in the organization or coding in the data? How it can be done?
- What are various operating resources in providing input are chosen?
- How the input authorizations are validated and prepared? What are the different procedures and corrective steps which should be followed when error occurs?

The output that fulfils the requirements of the end user and offers this information after considering the above things can be considered as quality output. Almost all systems process the results which are conveyed to the users and to other system through various outputs. The information which is the most important and direct source in output design is to be secured and moved from the user and also the hard copy output has been generated to the user. The system's user decision-making has been improved by the efficient and intelligent output design. The output formed in these kind of systems also called as **information system** should accomplish one or more of the underneath objectives. [7][5]

- Identification and Tracking current status.
- Analytical Intervention.
- Behavioural manipulation
- Feedback loop to confirm an action.
- Recorded observation.

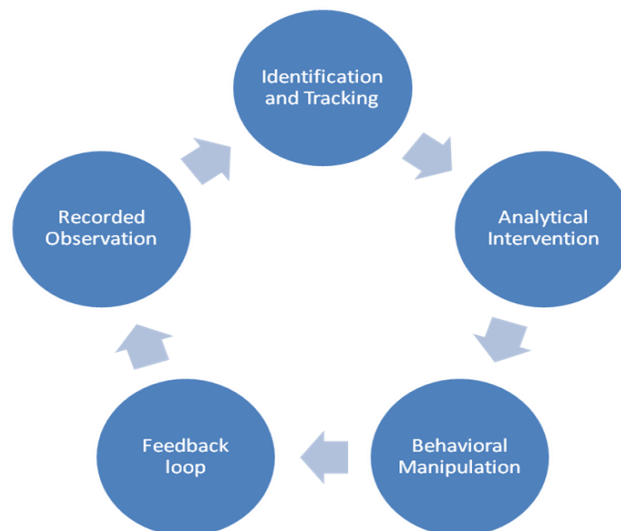


Figure 2. System Architecture

V. System Technique

Algorithm for Encrypt-DES

Step 1: The file of the Data owner will be given as a input file and it is stored in encrypted form.

Step 2: The data owner is required to give one encryption key.

Step 3: Try to encrypt the input file with this key.

Step 4: This input file after encryption has been stored in a big data.

Step 5: A request has been generated from the data consumer.

Step 6: After receiving the request from the data consumer a proxy will be activated immediately.

Step 7: Now the credentiality checking of that particular consumer to identify whether it is valid or not

1) If the consumer is valid

a) The Decryption key will be shared with the consumer

Else

1) If the consumer is not valid

a) The same key defined above i.e. decryption key will not be given to the consumer

Step 8: The consumer obtains the decryption key from the proxy.

Step 9: The input file has been decrypted with the decryption key.

Step 10: Eventually, the consumer can look up the original file of the data owner.

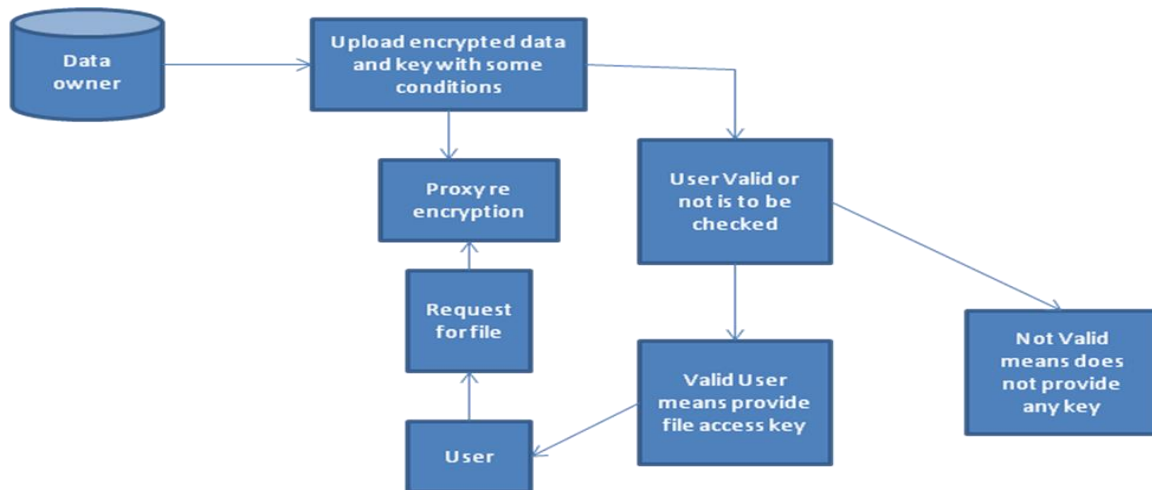


Figure 3. Data flow diagram

VI. Conclusion

The huge amount of data is generating and evolving each day and it is almost impossible to envision the upcoming new era applications without building and implementing data operating algorithms. In this paper a brief encompassing review on the privacy issues has been carried out when we talk of Big data. We have cited out certain privacy problems in each phase of the big data life cycle and there are some pros and cons in the existing privacy preserving technologies. The environment of big data applications is such that it has to safeguard and to preserve the privacy of users from data generation to data processing. Several open issues and challenges in this area still exists where we are going to continuously commit ourselves to work upon for preserving the privacy.

References

- [1] D. Kornack J. Manyika et al., Big data: The Next Frontier for Innovation, Competition, and Productivity. Zürich, Switzerland: McKinsey Global Inst., Jun. 2011,
- [2] B. Maturdi, X. Zhou, S. Li, and F. Lin, "Big data security and privacy: A review," China Commun., vol. 11, no. 14, pp. 135_145, Apr. 2014..
- [3] J. Gantz and D. Reinsel, "Extracting value from chaos," in Proc. IDCView, Jun. 2011, pp. 1_12.
- [4] Big Data Privacy Preservation, Ericsson Labs, <http://labs.ericsson.com/blog/privacy-preservation-in-big-data-analytics>.
- [5] L. Xu, C. Jiang, J.Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," in IEEE Access, vol. 2, pp. 1149_1176, Oct. 2014.
- [6] G. Ateniese, K. Benson, and S. Hohenberger, "Key-private proxy re-encryption," in Topics in Cryptology-CT-RSA (Lecture Notes in Computer Science), vol. 5473. Pp. 279-294, Berlin, Germany: Springer-Verlag, 2009.
- [7] A. Katal, M. Wazid, and R. H. Goudar, "Big data: Issues, challenges,tools and good practices," in Proc. IEEE Int. Conf. Contemp. Comput.,Aug. 2013, pp. 404_409.
- [8] Data Security Based On Big Data Storage S. Ananthi and Anjali Periwal Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamil Nadu, India.
- [9] Protection of Big Data Privacy Abid Mehmood, Iynkaran Natgunanathan, Yong Xiang, Senior Member, IEEE.
- [10] Jul. 2014.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [11] C. Hongbing, R. Chunming, H. Kai,W.Weihong, and L. Yanyan, "Secure big data storage and sharing scheme for cloud tenants," China Commun., vol. 12, no. 6, pp. 106_115, Jun. 2015.Electronic Publication: Digital Object Identifiers (DOIs):

Acknowledgments

This paper is an extended and revised version of three previously published conference papers [5, 8, and 9]. We are thankful to all those who helped us by giving there expert comments and by reviewing the document.



Implementation of Held-Karp Algorithm in Home Delivery System

Rakesh Ranjan¹, Anoop Kumar Yadav²

¹ Assistant Professor, Department of Information Technology, PSIT Kanpur,

² Final Year Student (2017-18), B. Tech(IT), Department of Information Technology, PSIT Kanpur,
Dr. A.P.J. Abdul Kalam Technical University, Lucknow
INDIA

Abstract: We propose an application that uses Held-Karp algorithm to generate "route for delivery man", which makes product home-delivery faster and more efficient. This application is an instance of traveling salesman problem which is classic NP-hard optimization problem. Our proposed application is using GPS Coordinates to deliver the product to the customer. Application has two major functionality, first is allocation of products to the delivery man using "k-means clustering" and second is "generating route for delivery man" using Held-Karp algorithm.

Keywords: e-commerce, delivery system, route, k-mean cluster and google cloud

I. Introduction

Today, home delivery is increasingly becoming a key element in e-commerce. It is observed that home delivery operations already face significant challenges. Difficulty of returning an item and cost of delivery are major factors for the respondents not opting for home delivery services. Unavailability of respondents at the time of delivery is another major factor. There is a high first time delivery failure which increases costs for carriers and lowers consumer satisfaction. That is why current delivery system needs to be modified with the use of current tools and technology. Delivery system problem is similar to Traveling salesman problem defined in the 1800s by the Irish mathematician W. R. Hamilton and by the British mathematician Thomas Kirkman. Traveling salesman problem can be described as: Find a route of N cities such that route contains every city just once and return to the starting point and route should be of least distance^[3]. We are using Held-Karp algorithm for route designing and k-mean clustering for product assignment to delivery man. Our system allows client to give their address as well as GPS coordinate where they want the delivery, then once these data are provided to their respective delivery center then they can use our application to automate product assignment and route generating system.

II. Work done

We have created an android app on which user can buy a product by providing their address and GPS coordinates. These data are sent to respective delivery center. The admin of these delivery center allocate products to their delivery man using k-mean clustering algorithm. Now delivery man login to their android app and see their assignment. These assignments are to be completed by that day only. The delivery man generate the route using Held-Karp algorithm (Which is implemented in the app) to deliver the product. This route is followed by delivery man to deliver the products.

III. Held-Karp algorithm^[3]

Held-Karp algorithm is used to solve traveling salesman problem^[1]. In traveling salesman problem you are given with N cities and you need to find the minimum distance route in which you visit all the cities just once and return to the starting city^[3].

From N cities assume that we are starting from city 1 and subsets $S \subseteq \{2, \dots, N\}$ of cities.

We compute $B(S, v)$ for all sets S and vertices v.

We start with all sets S for size 1, then size 2, etc.

$$B(\{s\}, s) = 0$$

$$B(\{s\}, v) = \infty \text{ if } v \neq s$$

If $|S| > 1$, then

$$B(S, v) = \min_{w \in X - \{v\}} B(X - \{v\}, w) + d(v, w)$$

From $B(V, v)$ for different v we can compute the minimum length TSP tour by adding edge from v to s and 'closing' the tours.

Then we pick the shortest tour.

Total number of sub-problem is $O(n \cdot 2^n)$ and each sub-problem requires linear time, so the time complexity is $O(n^2 \cdot 2^n)^{[1]}$.

IV. k-means Clustering^[4]

Loop through K times

current centroid = Randomly generate values for each attribute

Done = False

All instance cluster = none

while not Done

Total distance = 0

Done = true

For each instance

instance's previous cluster = instance's cluster

measure Euclidean distance to each centroid

find smallest distance and assign instance to that cluster

if new cluster != previous cluster

Done = False

add smallest distance to total distance

Report total distance

For each cluster

loop through attributes

loop through instance assigned to cluster

update totals

calculate average for attribute for cluster – producing new centroid

END While

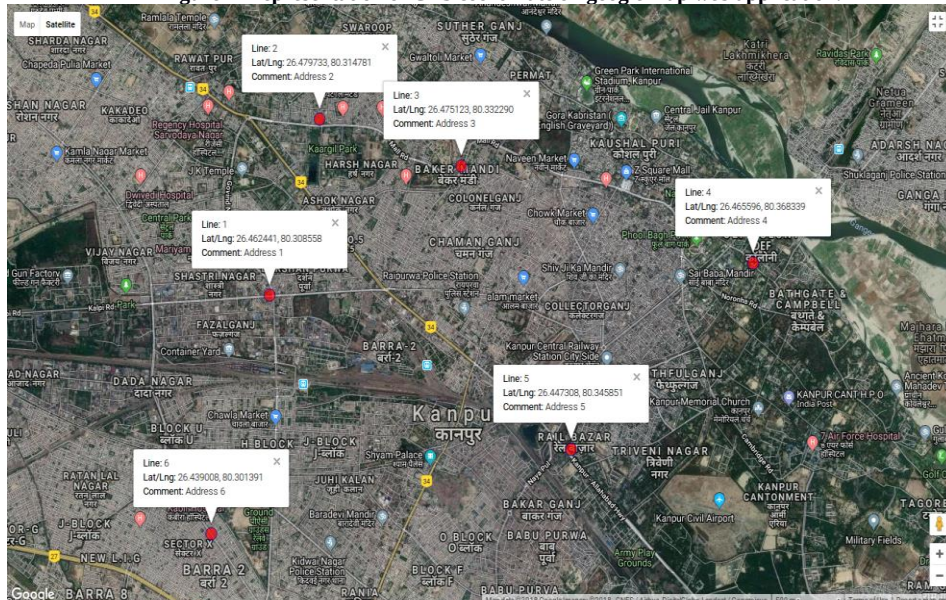
V. Example

When user buys any product all data are saved at the e-commerce web servers and both delivery address and product are transferred to the respective delivery center at each city, we have made a little update in this system when user provides its address it can also give its coordinate if he wants a GPS based delivery. Below is the data in form of table, that shows the product list which arrives at city delivery center for delivery.

Table 1 List of customer with address along with GPS coordinate.

Serial Number	Customer name	Address	Contact	GPS Coordinate
1	Name 1	Address 1	Contact 1	26.462441, 80.308558
2	Name 2	Address 2	Contact 2	26.479733, 80.314781
3	Name 3	Address 3	Contact 3	26.475123, 80.332290
4	Name 4	Address 4	Contact 4	26.465596, 80.368339
5	Name 5	Address 5	Contact 5	26.447308, 80.345851
6	Name 6	Address 6	Contact 6	26.439008, 80.301391

Figure 2 Representation of GPS coordinate on google map web application.



After that, “Delivery Center” apply k-mean clustering, let's say there are two delivery man available, then allocation after the assignment looks like the table below:

Table 2 Assignment of address serial number to delivery man.

Delivery man	Serial Number
Delivery man 1	1, 2, 3, 6
Deliver man 2	4,5

Allocation of each delivery man is added with one extra GPS address of the delivery center form where the route start and end. To get the distance between coordinates using google map axis.[6] Below is the web API which is used to get shortest road distance between two coordinate.

<https://maps.googleapis.com/maps/api/directions/json?key=<APIKEY>&origin=<source>&destination=<destination>&sensor=false&units=metric&mode=driving>

<APIKEY> can be obtained from google cloud^[6] and <source> and <destination> are the two coordinate between which we want to find the shortest road distance, above API return the result in JSON format then we need to parse the JSON and extract distance data. Once all the distances between all the coordinates are obtained it is stored in a 2D matrix and shortest route is generated by using Held-Karp algorithm.

VI. Methods to Improve

Current Implementation can also be improved to offer client exact time of delivery with high accuracy and this system can also easily adapt itself when client changes the delivery address. Rating based system can also be introduced to rate deliver man performance, so that delivery quality can be enhanced. Current delivery system can also be use for return of the product and other address based services.

VII. Advantage

Administrator: This system allows delivery center to easily assign the new products to the delivery man and keep a proper track of all the products and it also reduces the number of product return which increases the efficiency of the delivery centers.

Client: This system allows client to buy products by allowing both normal delivery as well as GPS based delivery on a click of a check box and this system also increases customer satisfaction as product reaches to customer without disturbing client for any landmark or product pickup from a random site. Clients can also track the location product in real-time.

Deliveryman: This system allows delivery man to just follow the path generated by the application and it allows delivery man to reach at destination more quickly as compared to conventional method, this system also saves fuels and time as the system generates the shorts path for the delivery man to take.

VIII. Limitation

Current limitation of this system availability of internet is required at all ends client, delivery center and deliver man and it is heavily dependent on google APIs to do its tasks.

IX. Conclusion

Delivery system is the backbone of a e-commerce industry but they are still using the conventional technology to deliver the products. Our current system allows delivery center to upgrade and use modern technology to deliver products with high-efficiency and high-customer satisfaction, our system will reduce the complexity of current system by removing manual “product assign system” and make easier for delivery man to find the address where the product needs to be delivered. With implementation of above system, it will help in saving a lot of time and resources, which are wasted in conventional delivery system and helps in improving customer satisfaction and trust on e-commerce industry.

References

- [1] ‘Dynamic programming treatment of the travelling salesman problem’, Richard Bellman, Journal of Assoc. Computing Mach. 9. 1962.
- [2] ‘A dynamic programming approach to sequencing problems’, Michael Held and Richard M. Karp, *Journal for the Society for Industrial and Applied Mathematics* 1:10. 1962
- [3] <http://www.cs.man.ac.uk/~david/algorithms/graphs.pdf>
- [4] <https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm>
- [5] <http://www.mafy.lut.fi/study/DiscreteOpt/tspdp.pdf>

[6] <https://cloud.google.com/maps-platform/>

Acknowledgment

We are really grateful to all the people who directly or indirectly helped us in this research work. We are also thankful to our family member who has supported, as their restless and encouraging support has really pulled out the ship of our project/thesis from the storming thunder to beautiful sea shore.



Comparison and Analysis of routing protocols in Simulated and Real Time Environment based on the performance of DTN (Delay/ Disruption Tolerant Networks)

Jatin Nagpal

CDAC, Noida Affiliated by
Guru Gobind Singh Indraprastha University Delhi
Sector - 16C Dwarka Delhi - 110078
India

Abstract: In today's emerging scenario challenged networks such as TCP/IP and related protocols do not provide satisfactory performance as when deployed in mobile and extreme terrestrial environments because of the scarcity of physical and operational limitations.

These limitations are described as long delays, communication errors, high error rates, asymmetric connection speeds, and end-to-end connection loss. That is why the concept of the Delay Tolerant Network (DTN) was used to integrate the development and implementation of such applications in different environments. The delayed tolerant network is an infrastructure-free wireless network that has a function of temporarily storing packets in intermediate nodes until a defined end-to-end route is established. This paper basically intimates the comparison of routing protocols in Simulated and Real Time Environment based on their performance.

Keywords: Delay Tolerant Networks, Routing Protocols, SenseNuts, ONE Simulator

I. Introduction

Mobiles AdHoc networks (MANETs) provide an excellent infrastructure less connectivity solutions. The routing protocols in MANET work on the underlying assumption of end to end contemporaneous path between message source and destination. The routing protocol has to first establish a complete path from message source to its destination. Once the path has been established, a message is forwarded either directly to the message destination, if it is in the direct transmission range of the sender or a message is forwarded through several relay nodes until it reaches its destination. Due to mobility, it may so happen that one of the node may move away from the path, in such a scenario, a very small delay is acceptable but eventually to handle longer delays a new path has to be traced. In case no path exists between source and destination the message cannot be forwarded.

In most of the practical scenarios, this assumption of existence of end to end path may not always be true. There are many applications where the end to end path is not necessary and a substantial delay is acceptable to the applications. Consider a scenario a zebra net (sensor network), where a sensor node is attached to zebras and a base station at a distance is meant to collect the information from zebras. Each zebra with a sensor node act as a mobile node. Periodically, these nodes sense the information and it is stored in the buffer available with zebra. Due to mobility, zebras may come in contact with base station and the sensor nodes may transfer the information to the base station acting as message sink. In this example, the delay with which the sensed information will be transferred to the base station is unknown. The approach of store carry and forward is the only viable approach in such applications. DTNs has their own routing protocols, and several of them have been developed in the recent years. Our approach, called Epidemic Routing is to transmit the packets to hosts, called carriers, within connected parts of ad hoc networks. In this way, packets are quickly flooded in the network. Epidemic Routing then rely on carriers coming in contact with another connected portion of the network through node mobility. At this point, the message spreads to an additional island of nodes. Because of this transitive transmission of data, messages have a high probability of reaching their destination.

The proposed Protocol has been mostly tested in simulated environment and therefore, there is a need to verify the results (performance) of the protocol in real time environment.

A. Architecture of DTN

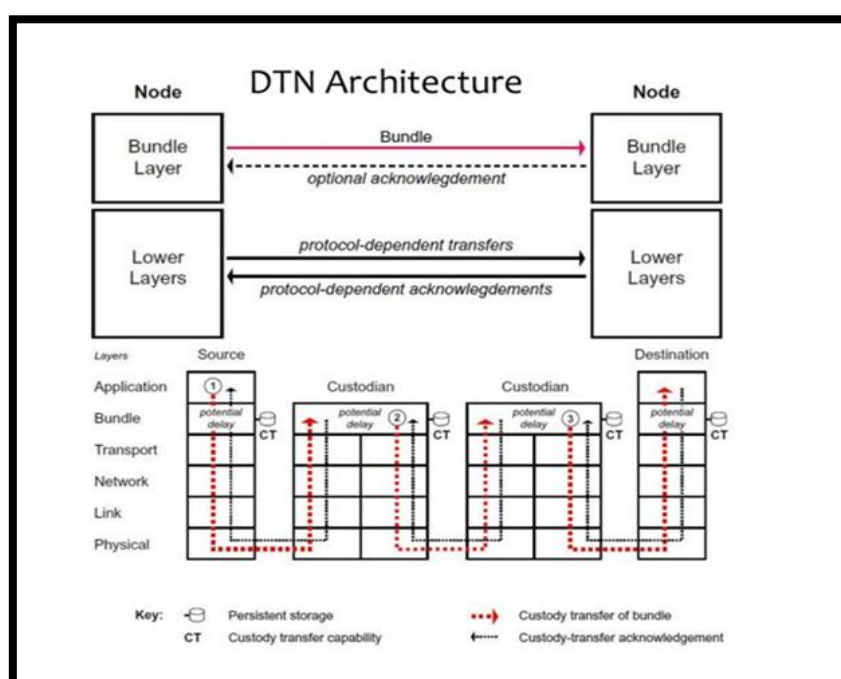
The concept of the DTN architecture is based on the abstraction of the message exchange by connecting different networks asynchronously. Each network can have all underlying technologies, different administrative controls;

The infrastructure may not be accessible. To make communication possible and efficient, the bundle protocol is introduced in the DTN architecture. This layer is located between the applied layer and the bottom layer, providing a general route rather than a number of specific solutions that are limited in their field of application.

This architecture is useful in a way as

- During the session the end-to-end path between source and destination is determined and maintained.
- Loss is relatively small.
- The retransmission property achieves reliable communication which ensures timely feedback from packet receivers
- In the area of security endpoint-based security mechanisms are used that protect the infrastructure by removing traffic.
- The transmission of Messages in Delay Tolerant Networks is done by either replicating the message or forwarding the message depending upon the type of algorithm implemented.

Figure 1: Delay Tolerant Network Architecture



B. Routing in DTN:

Routing is a mechanism of transferring of data from source to destination with an established end to end connectivity. However, as in case of delay tolerant networks there are chances of lack of connectivity, thus the routing protocols such as AODV & DSR which is used in traditional networks fails in this scenario. Therefore, Store and Forward approach is implemented to establish end to end path. This mechanism of store and forward process makes the data packets to be directed step by step and storing them in the intermediate nodes in the network till it reaches the destination. Replication of messages in intermediate nodes is the approach used to make the packet reach destination. The main objective of sending many copies in network is that at least one copy of message reaches to the intended destination node. Thus, it can be concluded that the routing depends upon these features:

Basic Network Assumptions which includes knowledge scheduling, mobility etc.

- There is a possibility that the connection will be lost at any time. Therefore, they are essentially time-dependent because they require long-term storage, which can lead to buffer competition.
- Finds optimal Routing solutions.

C. Brief Introduction to SenseNuts

As we know, Internet of Things (IoT) has made smart communication possible between the sensor devices via internet. It creates a network of things that can sense, communicate, and interact with each other in precise manner. In this paper we are using SENSEnuts platform as a real time environment for implementing one of our proposed routing algorithm i.e. epidemic routing and sensing various parameters like light, temperature as a message. SENSEnuts Gives us a environment that enables to perform WSN and Internet of Things Sensations and many

other functionality to use some Pre defined API's , a user friendly Graphical User Interface , real time environment including sensors, and C based codes for flashing them in nodes

SENSE nuts provides a package with variety of sensors like light, temperature, pressure, humidity along with communication modules and software tool chains that helps in providing an integrated environment for developing the application. Sensed information from this device are transferred to the internet via SENSE nuts Wi-Fi module. Once the information is transmitted to the internet, it gets stored in cloud where the monitoring of sensor data can be done. Thing Speak is an Internet of Things (IOT) analytic platform providing an open source cloud for updating the sensed information on real time basis. Data transmission from device is initiated using HTTP request which is processed, aggregated by Thing Speak cloud and data gets viewed, stored and analyzed.

Different Protocol Stacks

In order to communicate between the nodes application requires a wireless protocol like Wi-Fi, Bluetooth, Zigbee, 3G etc. Thus depending upon the following characteristics the protocol stack is chosen and implemented:

1. Bandwidth Requirement
2. Range
3. Behaviour of Nodes
4. Battery Consumption

Thus IEEE 802.15.4 Protocol is selected as it has Following Characteristics: Low Bandwidth, Low Range, On Off Kind of Behaviour and Low Battery Consumption.

Sensor Nodes used for creating Real time Scenario

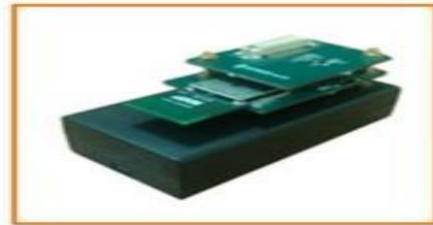


Figure 2: Wi-Fi Vs 802.15.4

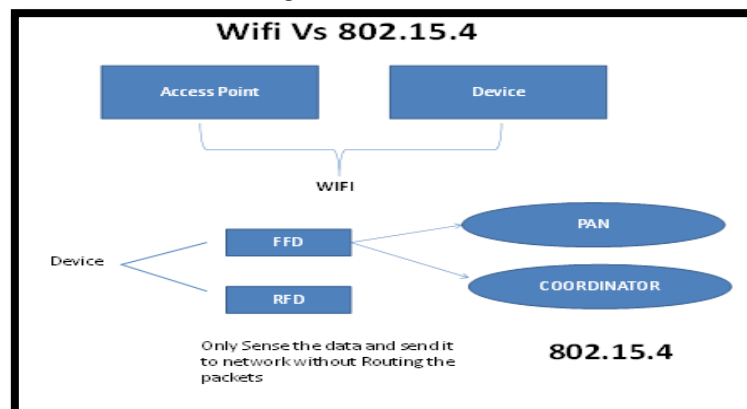
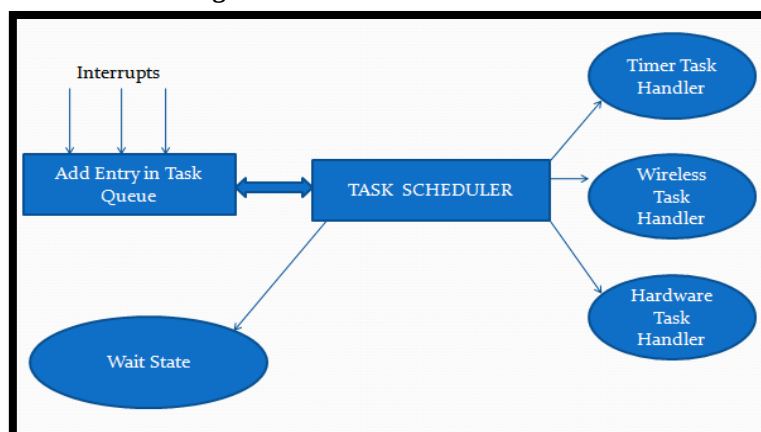


Figure 3: Task Scheduler in SenseNuts



II. PROPOSED ALGORITHM AND APPROACH

1. Creating a physical network using Sensor Nodes (Sense Nuts) and implementing the Routing Algorithm Majorly Epidemic by flashing the Code.
2. Create Messages in nodes and use above Routing Schemes to spread it so as to reach its destination. Generate a file showing established Connections and Message Creation Reports for every Timestamp.
3. Configure the same Environment in ONE and simulating the results obtained at previous step through a CSV file as External Input.
4. Generating the Reports I.e. Message Statistics in ONE and comparing it with Our Physically Created Network.

III. IMPLEMENTATION

Real time Scenario:

A network of 5 nodes was deployed and each node acting as a source of message generator and a relay for transmitting the message to others connected in a same Network using the Epidemic Routing Algorithm.

One node was kept Stationary and connected to PC acting as a destination node and other mobile which senses the data and Sends to the destination Node.

The Log file is captured giving the Node ID and message information i.e. Light and Sensor details.

Figure 4: Communication Between the Nodes

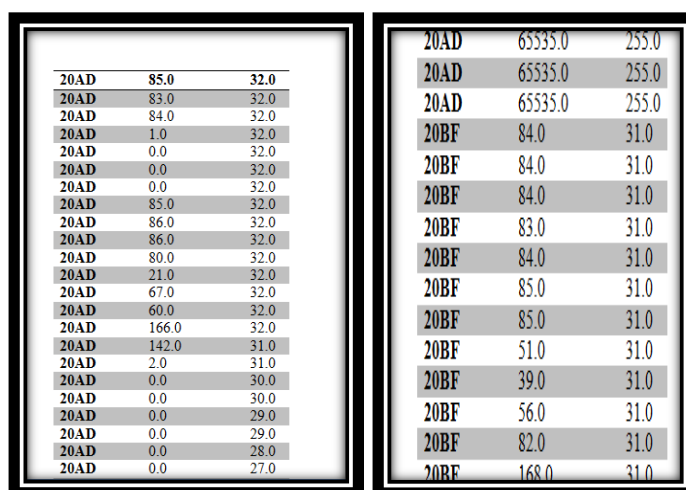
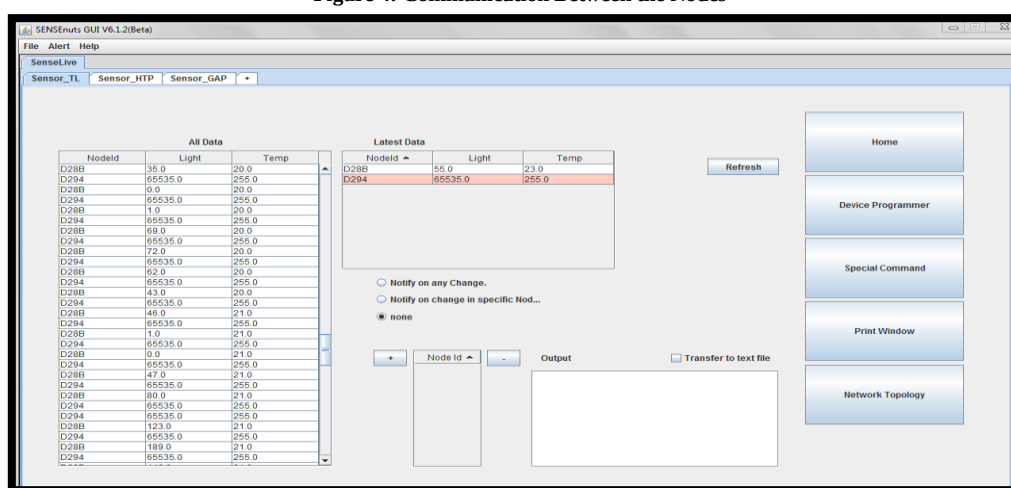


Figure 6: Connection and Log File

Simulation and Analysis

There are many user friendly simulation tools available for simulating the environment for implementing the algorithms of MANETS (mobile ad-hoc networks). But DTN Algorithms cannot be implemented as tool cannot

work properly because of the frequent disconnection environment of nodes in Delay Tolerant Networks. So we are using ONE Simulator for implementing our proposed DTN Algorithm.

A. Introduction to ONE Simulator:

ONE stands for Opportunistic Network Environment, a Java-based simulator for simulating delay and disruption networks. The simulator includes certain sample source codes of some famous DTN algorithms such as Prophet, Spray and Wait, etc. It offers a way to change different types of moves by selecting different models and their maps. Some TXT files are available in ONE, where two files are important. One is README.txt and others are defaults.txt. When the Run button is activated, the settings.txt file is invoked by default to initialize the values of the variables specified in the Java files of the classes used by routing and motion packages. The user can direct to README.txt file to get the information related to Working of ONE Simulator. The External Queue also allows the users to provide the logs for connection and message creation externally and with simulations “off” generate the results.

Thus we are using this utility for implementing our proposed routing on our connection list and getting the results. Connection and message Creation file generated from previous step is used as a file for External Event for simulations. **The configuration for simulation is kept same as the network created in real time.**

```
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : ffffffff9f : 17 : 3 : 4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : ffffffff9f : 17 : 3 : 4 : 17
received :
: 4 : 17 : 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : ffffffff9f : 17 : 3 : 4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : ffffffff9f : 17 : 3 : 4 : 17
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received :
: ffffffff : 3 : 4 : 17
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : ffffffff9f : 17 : 3 : 4 : 17
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : 1 : 17 : 3 : 4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received :
: ffffffff : 3 : 4 : 17
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : 0 : 17 : 3 : 4 : 17
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff94 : ffffffff : ffffffff : ffffffff : 3 :
4 : 17
received size : 5
received :
: 1 : 2 : 10 : 45 : ffffffff2 : ffffffff8b : 0 : 0 : 17 : 3 : 4 : 17
received size : 5
```

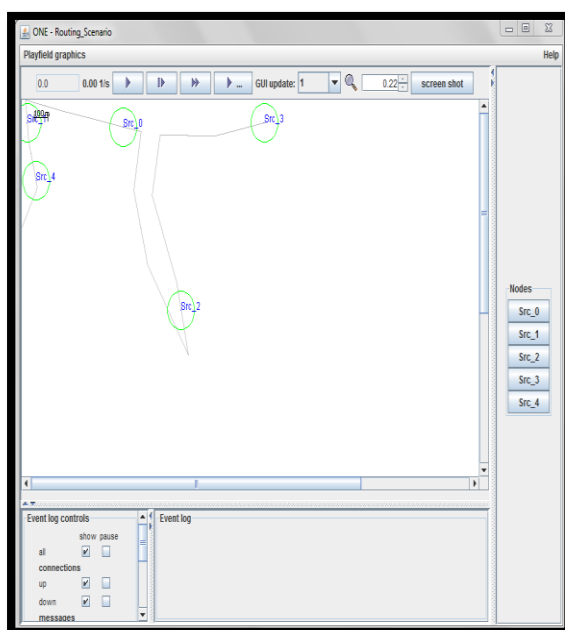


Figure 7: Simulating Environment

#	time	ID	size	hopcount	deliveryTime	fromHost	toHost	remainingTtl	isResponse	path
1719.8000	M20	24	1	717.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1719.9000	M44	24	1	715.6000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.0000	M22	24	1	717.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.1000	M19	24	1	718.3000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.2000	M68	24	1	713.5000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.3000	M18	24	1	718.6000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.4000	M25	24	1	718.0000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.5000	M47	24	1	715.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.6000	M4	24	1	720.3000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.7000	M27	24	1	718.1000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.8000	M64	24	1	714.5000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1720.9000	M1	24	1	720.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.0000	M46	24	1	716.5000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.1000	M6	24	1	720.6000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.2000	M43	24	1	717.0000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.3000	M48	24	1	716.6000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.4000	M66	24	1	714.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.5000	M5	24	1	721.1000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.6000	M23	24	1	719.4000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.7000	M49	24	1	716.9000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.8000	M69	24	1	715.0000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1721.9000	M2	24	1	721.8000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.0000	M28	24	1	719.3000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.1000	M21	24	1	720.1000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.2000	M7	24	1	721.6000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.3000	M29	24	1	719.5000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.4000	M70	24	1	715.5000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.5000	M65	24	1	716.1000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.6000	M45	24	1	718.2000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.7000	M26	24	1	720.2000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.8000	M67	24	1	716.2000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1722.9000	M63	24	1	716.7000	Src_0	Src_4	n/a	N	Src_0->Src_4	
1723.0000	M8	24	1	722.3000	Src_0	Src_4	n/a	N	Src_0->Src_4	

Figure 8: Message transmission

Now Comparing the Results of Real time Environment and Simulator on majorly two parameters Message Delivery Ratio and Latency Rate.

IV. RESULT AND ANALYSIS

Simulation Environment Results

```

Message stats for scenario Test
sim_time: 125.1000
created: 124
started: 113
relayed: 102
aborted: 11
dropped: 0
removed: 0
delivered: 102
delivery_prob: 0.8226
response_prob: 0.0000
latency_avg: 3.8559
latency_med: 0.1000
hopcount_avg: 1.0000
hopcount_med: 1
buffertime_avg: NaN
buffertime_med: NaN
rtt_avg: NaN
rtt_med: NaN
    
```

Real Time Results

Message Delivery Ratio: -

Total No. of Messages Delivered/Total No. of Messages Created

No. of Messages Created: 124

No. of Messages Delivered: 83

Delivery Ratio: $83/124=0.6693$

Latency average

$\sum \text{Delivery Time} - \sum \text{Creation Time} / \text{Total No of messaged Delivered}$

$1708 - 1176 / 83 = 6.4$

VI. CONCLUSION AND FUTURE SCOPE

The Statistics shows that the delivery ratio decreases due to interferences in the environment due to sacristy of physical and operational constraints. These constraints are described in terms of battery consumption, Buffer space, communication disruptions and many other factors. The latency rate also increases as the delay between the time of creation of message to its delivery in destination is stretched. The Expected delay is due to same environmental constraints as discussed.

Thus it gave an insight of comparison of Resource consumption of one of the Routing technique of Delay Tolerant Networks in Real Time Environment and simulated one. This will help us in scaling the Simulator Results to some Threshold values in future if needed.

REFERENCES

- [1] Network Working Group, V. Cerf Google/Jet Propulsion Laboratory S. Bur Leigh Intel Corporation H. Weiss Intel Corporation H. Weiss SPARTA, Inc. April 2007
- [2] CFDP Protocol Specification, CCSDS 727.0-B-1, Jan 2002, <http://www.ccsds.org>
- [3] Network Working Group, V. Cerf Google/Jet Propulsion Laboratory S. Bur Leigh Intel Corporation H. Weiss Intel Corporation H. Weiss SPARTA, Inc. April 2007
- [4] Ying Zhu, Bin Xu, Member, IEEE, Xinghua Shi, and Yu Wang, "A Survey of Social-based Routing in Delay Tolerant Networks: Positive and Negative Social Effects", IEEE Communications Surveys & Tutorials ,Volume: 15, Issue: 1,2013
- [5] Vahdat, Amin, and David Becker. "Epidemic routing for partially connected ad hoc networks." Technical Report (2000).

ACKNOWLEDGEMENT

I take this opportunity to acknowledge all those who have guided me in this paper. I express my earnest gratitude towards Dr. Kalpana Johari , my project guide for her valuable encouragement and guidance. Her motivation and suggestions were invaluable in the project work. I am grateful for her most cooperative attitude and suggestions, without which I could not have been able to do this work. I would also like to express my gratitude to the faculty members for their kind assistance and cooperation during the development of our project.



Cloud Computing Security and Integrity

¹Sri Nath Dwivedi, ²Shivendra Verma, ³Rajat Awasthi, ⁴Tushar Sharma

¹Department of Computer Engg.,

^{2,3,4}Department of Information Tech

Dr. A.I.T.H. Awadhpuri Kanpur U.P.

Awadhpuri Kanpur U.P.

India

Abstract: In today's technological era the Cloud Computing is becoming the first choice of user because of the facilities in a very pocket friendly manner. But, a major problem that arises with cloud computing is security and integrity of the data. In this paper, we have tried to explain what is cloud computing, what are the types of cloud computing and what are the major security issues with the cloud computing and also about data integrity, confidentiality and the privacy issues. We also discuss important solutions developed to tackle with the security issues in cloud computing such as data encryption, secure information management etc. Finally, we have suggested a way that will be useful for the creation of a secured cloud computing server so that the data of user can easily be transferred to the cloud server. We have tried to represent the aspects of the cloud computing security and integrity according to user's issues. In this paper we have also discussed about the research going on in the areas of security for cloud computing. The paper also contains the future work that are important to resolve issues of cloud computing.

Keywords: Cloud Computing; Security; vulnerable; integrity, confidentiality; data;

I. Introduction

Cloud computing generally refers all the computer software and hardware that a person can use but provided by any other company over the Internet in a seamless way. Here, the required software or hardware is really placed or exists is not a matter for the user, the user only knows that it is provided by this company over the high-end Internet connectivity. Cloud computing is simply interpreted as – provider of different things to different people or we can say that it is just like renting the required facility from any source [1]. Cloud computing is a growing idea because in the cloud computing the user can share the hardware, software, application etc that is not available or stored on their computer. A normal person who does not know much about the technology can also use the cloud services as a normal person is able to use other services like electricity etc. [2]. As the use of cloud computing is increasing, the security issues are growing rapidly and the services provided by service providers are like the traditional IT services. This includes protecting user data from leakage, theft, hijacking, etc.

Security in cloud is very easy and less risky. The cloud security consists of software-based management tools to secure the servers rather to maintain the high facilities and hardware. Cloud security delivers a new way of delivering security that prevent, detect and takes place corrective actions in very smooth manner [3]. Cloud computing is becoming a popular option for renting of computing and storage infrastructure services is known as Infrastructure as a Service or IaaS, for remote platform building and customization for business processes is called Platform as a Service or PaaS and for renting of business applications as a whole is called Software as a Service or SaaS. The cloud infrastructure has been divided into three parts:

1. Public cloud- “where the infrastructure resides totally outside of the tenant /enterprises”
2. Hybrid cloud- “where the infrastructure and business processes reside partly within the enterprise and partly consumed from the third party; and Private cloud”
3. Private cloud – “where IT services are mounted on top of large-scale conglomerated and virtualized infrastructure within enterprise firewall and consumed in “per transaction” basis”. Technology consulting firm Gartner has estimated market size for these clouds [4]

The basic things for cloud security are infrastructure, software platform and user's data, access control and identity management. These include broader issues of data integrity and security. Researchers are working for developing different ways and methods to deal with the cloud computing security. The most prominent security issue is vulnerability in the platform layer of security [5]. In this paper we have discussed all the major security threats in the cloud computing, some basic issues of security and some of the techniques or methods developed to overcome these security threats [5].

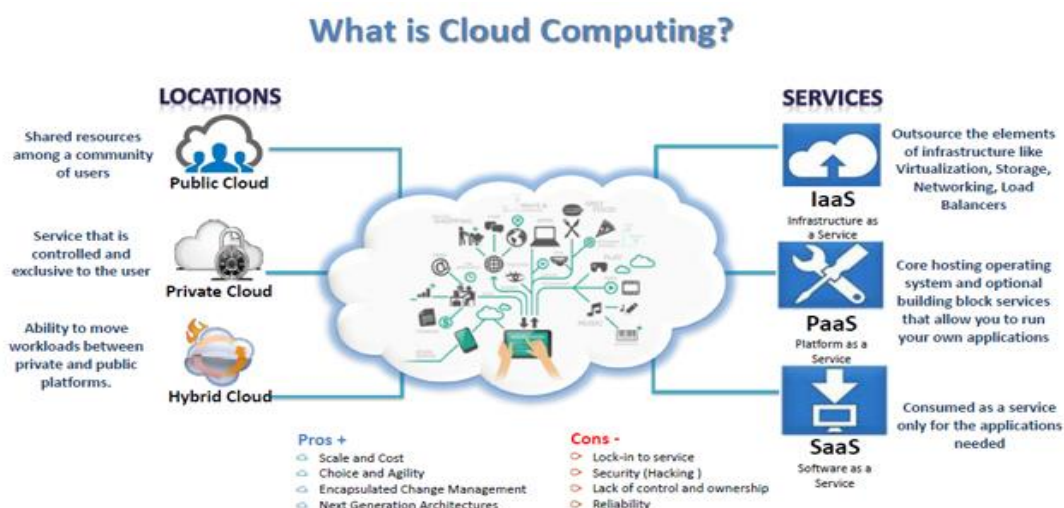


Image 1. What is cloud computing and types of cloud computing

	2016	2017	2018	2019	2020
Cloud Business Process Services (BPaaS)	39.6	42.2	45.8	49.5	53.6
Cloud Application Infrastructure Services (PaaS)	9.0	11.4	14.2	17.3	20.8
Cloud Application Services (SaaS)	48.2	58.6	71.2	84.8	99.7
Cloud Management and Security Services	7.1	8.7	10.3	12.0	13.9
Cloud System Infrastructure Services (IaaS)	25.4	34.7	45.8	58.4	72.4
Cloud Advertising	90.3	104.5	118.5	133.6	151.1
Total Market	219.6	260.2	305.8	355.6	411.4

Source: Gartner (October 2017)

Table 1. Public Cloud Revenue Worldwide (Billions of U.S. Dollars)

COMMON CONCERNS ABOUT CLOUD SECURITY AND IMPLICATIONS:-

1. Data Breaches/ Data Leakage:- Cloud computing is new for users and so the many factors are still undiscovered such as data breaches existed in all formats for several years. So the question remains same “that my important/sensitive data is stored in on line server rather on my hardware or premises is safe?” A study conducted by the Ponemon Institute entitled “MAN IN CLOUD ATTACK” reports that over 50 percent of the survey believe that the organization security measure is low. This study was mainly done on total nine real life situations where security breach occurred and on checking it is found that the survey is showing the fact. We can say that the cloud has vulnerability as a unique characteristic.

2. Hijacking of Accounts:- The increase in server based data storage opens a easy portal for the attackers to access the data stored on the cloud. The attackers simply use the login detail of the user and hijack the account to stole, delete or manipulate the data. The hijacking and unauthorized access of the data leads to a panic stage for the users. Attackers can also falsify the data and can sell the sensitive data of one firm to other. The other methods of hijacking are scripting bugs, reused password which allows the attackers to get easy access without the detection of the data breach. A real life example is in year 2010 when amazon cloud computing faced a scripting bug that targeted customer data and all other details as well as phishing, key logging and buffer overflow all are the current threats. The most new threat that was discovered is man in cloud that uses the theft of user token.

3. Insider Threat:- The attack that is done by a person belonging to inside the organization, the intention of these insider threat is not very dangerous (in general their intention is not so much harmful). The insider threat attacker can use their authorized access to an organization cloud based services to misuse the customer detail, financial report and other sensitive data. Sometimes these insiders don't have any dangerous intention. A survey done by imperva named, “INSIDE TRACK ON INSIDER THREAT” leads to a conclusion that the insider threats are the wrong use of data through malware and malicious intent. On the basis of this survey, they provided four best guidelines for companies to implement a secure strategy. The four guidelines are (i) business partnerships, (ii) prioritizing initiatives, (iii) controlling access, (iv) implementing technology.

4. Malware Injection:- The malware injection is a technique in which a simple script or code is injected into cloud services that runs as SaaS and act as valid instance for cloud servers. No one can predict that, this malicious code is formed and injected in to the cloud and that code also looks like that it is integrated in the system from the time of manufacturing and is running with the servers. Once an injection is executed and the cloud begins operating in tandem with it, attackers can eavesdrop, compromise the integrity of sensitive information, and steal data. A report presented by the East Carolina University titled “SECURITY THREATS ON CLOUD COMPUTING VULNERABILITIES”, reviews all the data breaching occur due to malware injection and states that “ The malware injection technique to breach the security in cloud is becoming a very big concern for cloud computing”.

5. Abuse of Cloud Services:- The hosting of a vast amount of data by both the enterprises and small level organizations has been made possible by the growth of cloud based services. Hosting and spreading of malware, illegal software, data loss, and other digital properties can easily be handled or restricted by the authorized user. For an example, the risk to the security can be increased by the users and as a result in fringe upon the terms and condition of the service provider. These risks include the data sharing such as documents, movie, pirated software, and it may result in the forms of penalty and settlements with the U.S. Copyright Law reaching up to a penalty of \$250,000. Depending on the damage, these fines can be even more cost prohibitive as a part of legal consequence. It can reduce the exposure to risk by monitoring the usage and setting guidelines for what your employees host in the cloud. It must be defined as that follows under the guideline “WHAT IS ABUSIVE OR INAPPROPRIATE BRHAVOUR” along with methods of detecting such behavior has been defined by service providers and legal entities.

6. Insecure APIs:- Insecure “Application Programming Interfaces (API)” give users the ability to customize their cloud facility. However, the very nature of APIs can be a threat to the cloud security. Not only do they give companies the ability to authenticate, provide access, and effect encryption, but they also define some features of the cloud services to fit business needs. As the infrastructure of APIs advances to provide better service, so do its security dangers. APIs give programmers job-critical software with other tools to build their programs, to integrate their applications. YouTube is a simple example of an API, where users have the facility to integrate their uploaded video on YouTube with their own websites. The success of an API lies in the better exchange of information between applications.

7. Denial of Service Attacks:-In comparison to other kinds of cyber-attacks, the typically used method to hijack sensitive information or data is termed as “DENIAL OF SERVICES”. The invasions never tried to violate the security parameter. Rather, the invaders try to stop the communication between the server and website so that the website is unavailable to appropriate users. The denial of service is also used as a smokescreen to stop the security appliances such as web application firewalls, and other vicious activities.

8. Insufficient Due Diligence:-In this category the most of issues that happen are generally technical and these type of security mistakes occur because the organization doesn't have a clear propaganda about the target, available/required resources and policies for the cloud. In simple words, IT can create a major security risk when any firm decides to transfer his data on the cloud without checking and testing that the service that was provided to the customer is really according to their expectations or not and these types of security issues are totally because of the insincerity. This becomes much important for the companies that lie under the regulatory laws or those handle the financial data of the customers.

II. Advanced issues in cloud computing security

• **Lack of execution controls:** Remote execution environment does not has fine-gained control by the external cloud user. Many issues like the management of the peripheral devices, storage devices, input/output demands are not under the preview of the user. The user always wants to invigilate the all process execution to get assurance that the illegal operations are not performed on the servers with the data of the user.

• **“Third-party control of data”:** In cloud computing, the important data of the user is in the hand of another company that is called as third party control. The company says that the data is secure but the user is not totally satisfied until he verifies all the facts by himself.

• **“Multi-user processing”:** In multi-cloud scenario, one provider provides a part of data and another provider uses that data. And the most important part of this is that the data security is the responsibility of both of the providers who jointly tried to secure the cloud.

III. Areas of security research

Trusted computing: The trusted computing involves the process of developing a new technology to check the concern of the distrusted execution environment for the modules that was developed, execute a strong approval key that does a work of verification between the user and host. It validates the user for the host and the host for the user. This process is termed as remote server attestation. All the upcoming process on user-host pair will be checked and verified by the trusted path. A strong virtual machine layer is allowed by the trusted virtual

machine. The checksums are the useful medium to access the data and check the authenticity. These checksums are also used for checking the data integrity and for doing classification of data stored in the cloud. The checksums are costly to reckon so they can not be applied for all the process it only used after installation of full data to the client [7].

Information-centric security (ICS): As the data of the user is kept far away from them somewhere in the cloud, we need content specific access meta data in the data itself. The Strong encryption of all the data is not possible because the data is always processed after sometime in the form of deciphered which can cause it to be exposed. The another way to get the information centric security is to use policy based or role based access controls which can be explained using “Extensible Access Control Markup Language (XACML)” which supervise context-based access rules in policy reinforcement point of the data. Now if the user made any request for the data then it must be authenticated by a pronouncement or by invigilating with the help of central server. Another way is to link the access control module in meta data of the “cryptographic Message Syntax (CMS)” which is message syntax much compact than the XML.

Privacy preserving models: In cloud computing, the sources that have its own set of data require the data processing system. In multi-user processing system, the data sharing is performed and the users have mutual trust on each other. But both the users try to get extra data from other user. The research around multi user computation is the way to develop a random partition (bitwise) that will help in the multi user computation. The random data does not associate any useful content at the other user site. In other cases such as the content received from any user and encrypted with the key of the customer. If this data was passed to the user B then the user B must have some information about the key of encryption of A in order to search the activity so that he can fulfill the request in better way. For example if we have to search a keyword in the given message and that keyword is “computer” then for searching this keyword involves a distinct process logic.

“The research in the search able encryption model is useful here”. Suppose if a user is downloading his own data from the server then another user available on the server may found a interest in the activity of the previous user and he wants to know what information he has received and what information he has got about the data set so it is possible for him to get access on that data. If we want to explain this problem in other language then we can also say that no database is private in theoretical way until the user doesn't receive the original full copy of the database and makes an update. We all know that this thing is impossible [9].

Techniques to secure data in cloud:-

Data Encryption:-If we have to secure very important or sensitive data to store on server and to prevent it from data breach, then we can apply the data encryption techniques to secure the data. Having the security setups like firewall and advanced security system is good but the attackers can penetrate these security setups and can damage the data. When the data encryption is using the encryption techniques then that data will only be accessed by the authorized person only that have the appropriate encryption key. The encrypted data is totally useless to the intruder. This is a technique to translate the data into a secure code if anyone wants to read the data then the person must have the original encryption key [10].

Information integrity and Privacy:- In Cloud computing, the information or data stored on the server, can only be accessed by the authorized user. The server provider allots a secure login gateway to the user. The data can be protected with the help of multi step login system. The multi step login system portal protects the users data by using different techniques like biometrics, retina scan etc [11].

Secure Information Management:-It is a method to secure the data from the attackers. In this, data is stored into a central unit. In the same way as a “security console”. This security console can be handled by the administrator who is a human being. The company tries to provide some great security features like PCL, DSS etc. to secure the data of the user[12].

Malware-injection attack solution: In this kind of solution main storage is used to store a number of client virtual machines. It consists of virtual operating systems which utilizes FAT (File Allocation Table).The FAT table also contains the application to run by a client. The Hyper visor handles and schedules all the instances. Integrity checking is done by IDT (Interrupt Descriptor Table).[13]

IV. Conclusion and future work:-

Cloud Computing is rapidly growing and becoming the choice of the users because the cloud is convenient to user in many ways. As renting a cloud service is less costlier and we can also scale up or down the services very easily. The security issues are becoming the hurdles of the adoption of the cloud computing.

In this finding, firstly we have tried to label the major types of the cloud services provided and then discussed the benefits of those categorized services. We have also covered the issues of technical implementation of the cloud. Work also explains some issues regarding the security related compliance in cloud including few of

advanced concerns related with security. Further we have taken a huge view of the areas of research that are currently going on to ensure the security in cloud. At last we have tried to give some techniques to secure data in cloud. So in this paper we have tried to cover the major portion of the cloud computing starting from what is cloud computing and what are the types of cloud computing; the major security concerns with the cloud; advanced security concern with the cloud; areas of research; techniques to secure data in cloud. The focus of current research work is to make cloud more safe. So the threats to the user data can be minimized. The cloud computing research is also trying to find a way to connect the whole world's system and use the unoccupied space of any user.

We tried to represent a broad level view of the currently emerging security issues in cloud and the main research challenges.

References

- [1] <https://www.explainthatstuff.com/cloud-computing-introduction.html> INTRODUCTION TO CLOUD COMPUTING
- [2] <https://appenda.com/library/cloud/> CLOUD COMPUTING OVERVIEW
- [3] <https://aws.amazon.com/security/introduction-to-cloud-security/> What Is Cloud Security?
- [4] <https://www.gartner.com/newsroom/id/3815165> Gartner Forecasts Worldwide Public Cloud Services
- [5] Chow,R.,Gotlle,P.,Jakobsson, E.S.,Staddon,J., Masuoka,R.,and Molina,J.;2009, Data controlling techniques in the cloud
- [6] <http://www.trustedcomputinggroup.org> Top security concern
- [7] Proceedings of ACM Symposium on OperatingSystems Principles. 2003 Research on the trusted computing
- [8] Proceedings of the 14th ACM conference on Computer and Communication Security,2007 Information-centric security
- [9] Privacy preserving model <http://www.ietf.org/rfc/rfc2630.txt>
- [10] Lindell, Y., and Pinkas, B., Privacy Preserving Data Mining,Proceedings of 20th Annual International CryptologyConference. 2000
- [11] Yashpal sinh jadeja & kirti modi (2012) cloud computing- concepts, architecture and challenges
- [12] Akhil Behl & Kanika Behl (2012), An Analysis of Cloud Computing Security Issues.
- [13] R. Balasubramanian, Dr.M.Aramuthan (2012) Security Problems and Possible Security Approaches In Cloud Computing]

Acknowledgement

This work is supported by Faculty members of Department of Computer Science and Engineering and Department of Information Technology of Dr. Ambedkar Institute of Technology for Handicapped, Kanpur. We would like to thank all of them.



Prediction of Flight Delays using Data Mining Techniques

Neema Chaudhary¹, Kriti Saroha²
CDAC, NOIDA

Abstract: *It is proven fact that Airports are the focus of attraction of any country which forms an everlasting impression. Flight delays possess an adverse impact on airlines and passengers. Airlines have pay to huge amount to passengers as compensation for various inconvenience caused. So, prediction of flight delay is important for airlines and passengers and, to deal with this problem various prediction models have been built.*

Keywords: *Data mining, Artificial neural network, Random forest, Support Vector Machine, delays.*

I. Introduction

Airport is an aperture to a country. According to a previous studies these flight delays causes considerable amount of loss to passengers, airlines and NAS (National airspace system). Airlines have to pay huge amount to passengers as compensation for various inconvenience caused. So, to lessen the unnecessary expenses, analysis of flight delays and prediction of flight delays is important. In this paper, various data mining techniques were used to predict flight delays and it would help airport and airline authorities as they would be able to work on better allocation of resources on the day of maximum delayed flight. More preparations should be done in advance to tackle weather delay in the months of bad weather conditions. It will reduce the inconvenience caused to passengers due to flight delay as they can avoid flying on days of maximum delayed flight. This can also be useful while booking multi-stop trip, involving multiple companies. Knowing that the first flight is likely to be delayed means passenger should book the connecting flight a bit later, so that they are likely to have a shorter time to transition.

II. Related Work

Recently many researches were done on air traffic analysis such as flight delay prediction. In Paper [1] Recurrent Neural Network (RNN) technique is used. In this a two stage model is used. In stage 1 daily delay status of airport is predicted using deep RNN. Flight dataset is collected from US transportation department which includes origin and destination airport including flight schedule. Two different threshold and sequence values were used in the experiment i.e. one is 15 min threshold and 7 days sequence other is 9 days sequence and 30 min threshold and it is observed that larger threshold achieves higher accuracy. So it is concluded that larger threshold is classifying day delay better.

In stage 2 individual flight delay is predicted using output of first stage i.e. daily delay status. In this stage accuracy is checked for different hidden nodes in each layer and it is observed that as the number of layers increasing accuracy improved. So it is concluded that accuracy of airport delay is improved with “Deeper Architectures”.

In Paper [2] performance of different techniques is compared with combination of different feature vectors and datasets. Different techniques used are Markov Jump Linear System (MJLS), Classification and Regression Trees i.e. CART and three candidate Artificial Neural Network. 3 different datasets A, B, C were used. Dataset A was used to evaluate the performance of the OD-pair delay classifier and delay level predictor and for each OD-pair, only those data points in which a delay was non-zero and not during overnight hours were included. Dataset B was balanced dataset means there are an equal number of high and low delay data points. This dataset was used to evaluate the performance of the OD-pair delay classifier and delay level predictor under more challenging conditions. Dataset C was used for evaluating the aggregate airport delay predictions. Among all techniques MJLS model was not appropriate to use for balanced dataset B and neural network became the best performer for balanced dataset. It was observed that ANN performed well for OD-pair delay classification problem and MJLS for OD-pair delay regression problems.

In Paper [3] regression model was used to predict the fuel consumption and binary classifier for predicting flight delay and binary classifier uses flight information and reservation information to achieves an area under the ROC curve i.e. AUC of over 0.75 for predicting flight delay one-day before. Author used various features according to time of prediction like one-day before the flight departure (D), one-week before the flight departure

(W), five-month before the flight departure (M). In one-day before both passenger dataset and flight dataset is available, in one-week before flight dataset and reservation dataset is available and in five-month before only flight dataset is available.

3 prediction models are applied on above dataset- Random forest, XG-Boost and deep neural networks and it was observed that AUC score was greater than 0.6 for all methods in case of one-day before the flight (D) and less than 0.6 in case of one-week and five-month before flight prediction. So using information available one-day before flight prediction is more effective.

In Paper [4] Delays in airlines caused by severe weather situations were predicted using data mining techniques and supervised machine learning techniques such as decision trees, random forest, AdaBoost, k-nearest neighbours classifier. Then prediction accuracy and receiver operating characteristics of each algorithm were compared. To overcome the effect of imbalanced training dataset sampling technique were applied. Impacts of weather data and impact of sampling techniques were investigated and it was observed that weather data improve the predictive ability of each model.

In Paper [5] selected data mining techniques were reviewed and combined into intelligent airport system. Airport retail sectors, airlines and airports itself are reflected for this reason and data mining techniques were useful to improve existing airport schemes such as flight delay prediction, segmentation, passenger profiling to discover better approaches for intelligent airport system. Various machine learning classifiers such as Regression, Support Vector Machine, Bayesian Approaches, KNN and random forest were used to train data.

III. Methodology

In this section, we discuss the proposed approach that incorporates the following.

Supervised Learning Classifiers

- 1) Artificial Neural Network (ANN)
- 2) Random Forest and
- 3) Support Vector Machine (SVM)

Feature Selection

Feature selection is the procedure of selecting only relevant attributes from all the available features. Feature selection is useful because of various reasons such as shorter training times, reduce overfitting, avoid curse of dimensionality, simplification of model.

Model Evaluation

Confusion matrix is used to check the performance of classifiers on test data and it is also known Error matrix. It allows us to identify the confusion between classes as e.g. classes are mislabelled as other. Various performance measures are calculated from confusion matrix such as accuracy, sensitivity, recall, specificity.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Receiver Operating Characteristic (ROC) Curve: ROC graphs are a useful technique for visualizing binary classifier's performance and it is plotted between TPR (true positive rate) and FPR (false positive rate) with TPR on y-axis and FPR on x-axis. TPR is also known as Sensitivity and FPR is 1-Specificity.

TPR=TP/Actual YES and FPR=FP/Actual NO and it means when it's actually no how often does it predict yes.

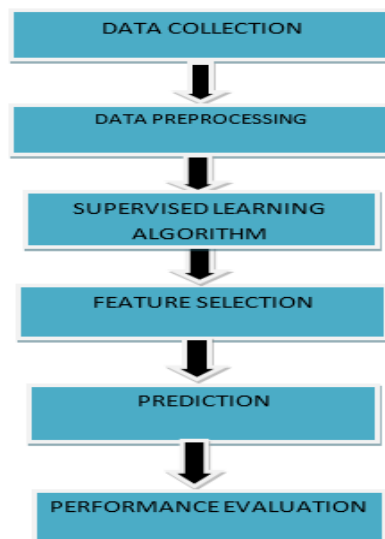


Fig.1. Proposed model

Figure 1 shows the summary of prediction model. Training procedure starts with historical flight dataset collection. In the previous section methodologies used are discussed and remaining explanation about proposed model is discussed below.

Data Collection

Dataset has been collected from US transportation department for year 2017 and it comprises data for non-stop domestic flights by major air carriers, and it provides information such as origin airport, destination airport, departure time, CRS departure time, flight date, flight number, distance, day of week, day of month, weather, tail number.

Data Pre-processing

In data pre-processing, missing values are handled and categorical variables are converted to numerical variables since SVM and ANN work with numerical variables.

Additionally, normalization is done to scale range of features to avoids model from overruled by few features.

Classification

Classification model is trained with flight data using supervised learning algorithms such as Random forest, ANN and SVM. Number of trees in random forest were 500 and the best value of split at each node is 9.

Results

The initial results of classifier on monthly data and half yearly data are shown in table 1 and table 2 respectively.

CLASSIFIER	ACCURACY(%)
Random Forest	89.45
Support Vector Machine	87.64
Artificial Neural Network	87.45

Table 1: Accuracy for monthly data

CLASSIFIER	ACCURACY(%)
Random Forest	97.14
ANN	98.57

Table 2: Accuracy for half yearly data

Accuracy after applying feature selection on attributes is shown in table 3 and table 4. Furthermore, it is observed that initially dropping least important variable i.e. TAIL_NUM increases accuracy but again dropping next least important variable decreases accuracy, so TAIL_NUM is the only variable that is dropped in the feature selection step.

Random forest	Accuracy
Without feature selection	87.18%
Dropping TAIL_NUM	88.27%
Dropping DAY_OF_MONTH	87.91%

Table 3: Accuracy after feature selection (Random Forest)

ANN CLASSIFIER	ACCURACY
Without feature selection	87.45%
Dropping TAIL_NUM	89.09%
Dropping DAY_OF_MONTH	87.82%

Table 4: Accuracy after feature selection(ANN)

IV. Conclusion & Future Work

In this study, the proposed model is used to predict flight delays i.e. flight is on time or delayed. The proposed model found 89% accuracy in case of monthly data and 98% in case of half yearly data. Furthermore, it is found that the feature selection improves the accuracy and reduces the cost of computation using all attributes. The model's prediction performance on attributes after feature selection is better than performance on all attributes. Data analysis shows some useful trends such as which Origin, Destination, Carrier, Day of week, Day of month shows maximum flight delays and must be avoided. Some probable directions for future work are delays

because of hardware failure can be considered and this information can be used in addition to flight information to yield important patterns.

References

- [1]. Y. J. Kim, S. Choi, S. Briceno and D. Mavris, "A deep learning approach to flight delay prediction," 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, 2016, pp. 1-6.
- [2]. K Gopalakrishnan, H Balakrishnan, "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks", Twelfth USA/Europe Air Traffic Management Research and Development Seminar, 2017.
- [3]. Horiguchi, Y., Baba, Y., Kashima, H., Suzuki, M., Kayahara, H., Maeno, J., "Predicting Fuel Consumption and Flight Delays for Low-Cost Airlines", Innovative Applications of Artificial Intelligence, 2017.
- [4]. S. Choi, Y. J. Kim, S. Briceno and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, 2016, pp. 1-6.
- [5]. Juan Jose Rebollo, Hamsa Balakrishnan, "Characterization and prediction of air traffic delays", 2014 Elsevier/Transportation Research Part C, Cambridge, USA, 2014.
- [6]. C. M. Ariyawansa and A. C. Aponso, "Review on state of art data mining and machine learning techniques for intelligent Airport systems," 2016 2nd International Conference on Information Management (ICIM), London, 2016, pp. 134-138.



Punjabi Unicode Inscript Keyboard for MS Word

Ashish Batra¹, Dr. Suneet Madan²

Human Computer Interface Division

Centre for Development of Advanced Computing (C-DAC)

Mohali (Punjab)

INDIA

Abstract: Many Inscript based standalone keyboard applications are available online for typing Punjabi but they are restrictive in nature and most of these do not offer formatting of the keyed-in content in the text area provided in the application. The main problem regarding these keyboard applications is that in the absence of an audio feedback for the keys pressed on the Punjabi Inscript keyboard, unless trained to use that keyboard, the visually impaired cannot make out whether the content is being typed correctly. In order to overcome this problem, we developed the Punjabi Unicode Inscript Keyboard with sound embedded on every keystroke. This keyboard can be used for typing directly in the Microsoft Word. The keyboard is useful for normal users as well because clicking on the keys of the keyboard gives out aural acknowledgment, which clarifies and confirms what is being typed by the user. The Punjabi Unicode Inscript Keyboard for MS Word, developed by C-DAC Mohali, is implemented using Java technology, and has been categorized into four layers.

Keywords: Inscript keyboard, Punjabi, Punjabi Keyboard, Unicode support, Keystroke sound, Microsoft Word, Visually impaired, Punjabi Unicode typing

I. Introduction

With each step towards preparing India to become a developed nation, there has been a growing requirement to make information accessible to the common man in the Rajbhasha Hindi and regional languages as well. This requires development of keyboards that incorporate the characters of Indian regional language alphabets that include consonants, vowels, nasals, and conjuncts. The INSCRIPT (INDian SCRIPT) Keyboard, as the name suggests, is the keyboard for Indian languages (written left to right), standardized by the Government of India.

A number of Inscript keyboard options available online were explored for different languages to find out the features or deviations they provided. More or less, the keyboards offered limited variation. They provided a restrictive text area in the application, in which the content could be typed. At the most, some websites gave the option to keep the keyboard floating or fixed. Once done, this content could be copied and pasted in other applications, as desired. This arrangement is rather inconvenient and time consuming as the typed in content has to be formatted later; the text areas in the keyboard applications do not offer formatting option for the content.

Working on the regional language of Punjab, we checked for all the available keyboard options for Punjabi language. For the benefit of the global audience, Punjabi is one of the twenty-two officially recognized languages of India, as recognized by the Eighth Schedule to the Constitution of India [1]. As per Ethnologue, an authoritative resource on world languages, Punjabi ranks at number 10 worldwide, with at least 50 million first-language speakers [2].

Amongst the available Punjabi Inscript keyboard options, the prominent ones being used are the “Ramington” and “Normal Keyboard based on Amrit Lipi Fonts”. Layouts of both the keyboards are in two layers - Upper character with Shift key press and lower with normal keyboard - making it difficult to arrange all the consonants, vowels and modifiers. Besides, these do not contain some of the special characters used in Punjabi. There are keyboard applications available in three layers as well, and yet they do not cover all the special characters. One keyboard option had four layers, but the content has to be typed in the text area of the application. Another keyboard application provides a basic editor to change the font size, style and color, but there is no direct option to save the text; the typed content in the text area has to be copied to another application in order to save it or to further format it.

II. Application Design

The expectation from the keyboard to be developed was to be able to type the Punjabi content in Unicode and not use the font based typing as done in some keyboard applications, to make typing easier for everyone including the visually impaired, to be able to format the typed in content like indents, spacing, bullets, etc., and to incorporate the alphabets and special characters used in Punjabi.

One of the main objectives to develop this keyboard was to overcome the restrictive typing in the text area that generally comes with other online keyboard applications. The Punjabi Unicode Inscript keyboard provides keyboard for typing in Microsoft Word. The keyed-in content will be in Unicode, making the content accessible and readable across the platforms and applications. The developed keyboard covers all the Punjabi Language characters including special symbols covered on its four layer architecture. The Unicode values behind every keyboard character are as specified under The Unicode® Standard Version 11.0 [3].

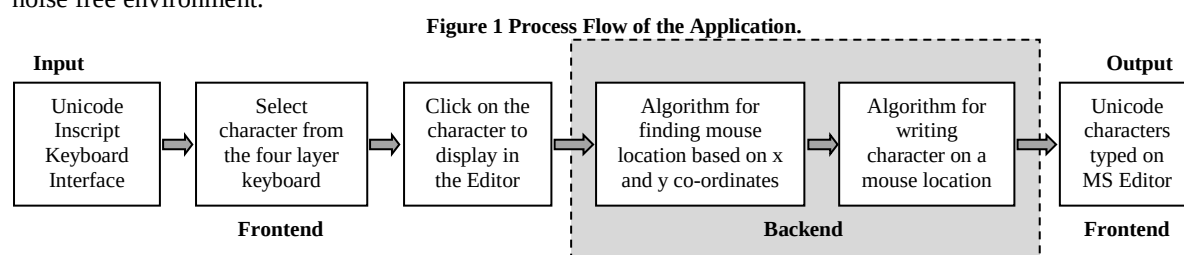
A. Keyboard layout:

The keyboard layout design has four layers structure, with Normal, Shift and Alt-Shift modes. It is designed to accommodate all the Punjabi alphabets and signs. There are a total of 35 letters in Gurmukhi alphabet, out of which the first three in the *varmala* are distinct because they form the basis for vowels. The other online keyboard layouts which we explored are mostly based on two layers, making it hard to make available and manage all the characters on these two layers.

The main advantage of the Punjabi Unicode Inscript keyboard layout over other keyboards is that we have arranged the Punjabi Characters on different layers. On the four layers, Layer1 contains the Matras and Consonants, Layer 2 contains the Vowels along with Consonants, Layer 3 has Nukta characters and special symbols and Layer 4 shows Nukta characters, special symbols along with the Vowels.

B. Methodology:

The Java technology has been used to implement the keyboard. Designing part of the keyboard interface is implemented in Java Swing. Java technology was chosen for its multi-platform compatibility and for its platform independent features. The sound files for every Punjabi character are recorded in .WAV format, in a noise free environment.



For flexible layout, we chose GridBagLayout for aligning keyboard buttons in the GUI. GridBagLayout class is used for aligning components vertically, horizontally or along their baseline without requiring that the components be of the same size [4]. Point class is used to represent location in (x,y) coordinate space. In our application, Point is used for representing three points in our application. First point indicates the window which contains the MS Word. Second point indicates keyboard location corresponding to the MS Word in x and y co-ordinates, and the third point shows the last mouse location. In addition, Line and Clip interfaces are used while playing the .wav file for every keyboard character.

Robot class is used to write characters on MS Word by using mouseMove, mousePress, mouseRelease, keyPress and keyRelease events. mousePress is used for getting focus and mouseMove puts mouse back on the button. Similarly, Dimension class is used to calculate the last mouse location on both x and y coordinates so character on that particular location can be displayed. In addition, Math class is used to calculate where to write characters on MS editor.

For finding mouse location on the basis of x and y co-ordinates, we used the Component and Dimension classes from the open source code. The mouse and keyboard events were used for writing character on a mouse location on the basis of x and y co-ordinates. With the help of these events, the exact position of mouse can be found by calculating the x and y coordinates (see figure 2).

Figure 2 Point and Dimension Classes used from the open source code.

```

Point p = resizingComponent.getLocationOnScreen();
int width = resizingComponent.getWidth();
int height = resizingComponent.getHeight();

p.x += width/2;
p.y += height/2;

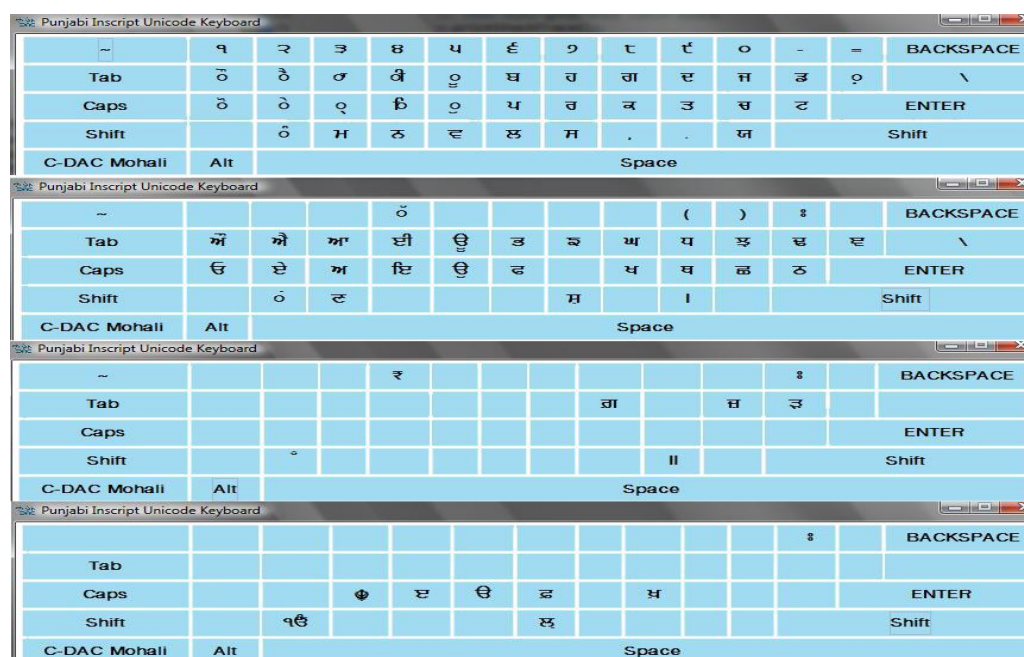
lbl.setText( width + " x " + height ); //NOI18N

SwingUtilities.convertPointFromScreen( p, this );
Dimension size = panel.getPreferredSize();
p.x -= size.width/2;
p.y -= size.height/2;
panel.setLocation( p );
panel.setSize( size );
    
```

III. Results

The Punjabi Unicode Inscript keyboard includes Unicode character input on MS Word, which can help increase the typing speed for visually impaired as well as for the normal users, accuracy of which is improved by providing sound on every character. The keyboard structure is given as under:

Figure 3 Structure of the Punjabi Unicode Inscript Keyboard.



In the figure 3, the four-layered structure of the keyboard is given, which is accessible through the Normal, Shift and Alt-Shift modes. The first layer shows all the Matras and Consonants. Consonants are divided on two layers, on the first layer of the keyboard there are 18 consonants and remaining are on the second and the third layers. There are three lagakhar signs used in Gurmukhi Script – Bindi, Tippee, and Adhak. These are used with the Matras. One lagakhar sign is on the first layer and the remaining two are on the second layer. The second layer shows Vowel Signs, Diacritics, along with consonants. All Nukta characters are placed on the third and the fourth layers. In the third layer, Nukta characters and special symbols are placed. The fourth layer shows Nukta characters, special symbols and Vowel signs.

IV. Conclusion

The Punjabi Unicode Inscript keyboard developed by C-DAC Mohali provides Unicode typing with sound embedded on every character which gives the pronunciation of that character. The sound on every keystroke of the Inscript keyboard helps ensure that the correct character is written in the editor. This gives it more accuracy and typing speed. This application holds wide scope for its user base. It could be used by everyone, young and old. The application would also be very useful for the differently-abled populace, including the visually impaired. We intend expanding on the present work by integrating speech synthesis with the Inscript keyboard. Using the speech technologies, the user's speech will become the input method for typing in Microsoft Word, enabling the user to operate the keyboard hands-free for Punjabi language. The Speech based Punjabi Unicode Inscript Keyboard will provide keyboard for Microsoft Word in which both the input options will be given, speech as well as normal keyboard input. The speech recognition part in the keyboard is being done using Sphinx with DNN techniques.

In the age of voice applications like Siri and Alexa, it is imperative that we develop applications based on speech technologies. This application holds a wide scope for its user base. It could be used by everyone, including the visually impaired, or those who do not wish to type out the words.

References

- [1] Constitution of India. EIGHTH SCHEDULE, Articles 344 (1) and 351. Languages. Page 330.
- [2] G.F. Simons and C.D. Fennig (Eds.). 2018. *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- [3] The Unicode® Standard Version 11.0 – Core Specifications. Unicode, Inc. Online version: <https://www.unicode.org/charts/PDF/U0A00.pdf>
- [4] Java® Platform, Standard Edition & Java Development Kit Version 9 API Specification. Online version: <https://docs.oracle.com/javase/9/docs/api/java/awt/GridBagLayout.html>

Acknowledgement

We are grateful to C-DAC Mohali for supporting this research work and providing us all the required resources and assistance.



Prediction of Survival in Thyroid Cancer using data mining techniques

Jasvinder Kaur and Kriti Saroha

Information technology and Computer Science& Engineering
Centre for Development of Advanced Computing (C-DAC),
B-30, Sec-62, Noida-201309,
INDIA

Abstract- Cancer is the most prominent cause of death throughout the world these days. Oncologists are exploring the appropriate treatment and measures to ensure the quality of care to these cancer patients. The fundamental goals of cancer prediction and outcomes are different from the goals of cancer detection and diagnosis. In cancer prediction is majorly concerned with three predictive perspective: 1) the prediction of cancer in terms of risk assessment 2) the prediction of cancer in terms of recurrence and 3) the prediction of cancer in terms of survivability. Success is not always obvious in machine learning . With any of the method, a proper understanding of the problem domain and recognition of the limitations of the data is important. If a machine learning experiment is correctly implemented and the results stoutly validated, then might has a good chance of success.

Keywords – Machine learning, Cancer prediction, Survival, Oncologist.

I. INTRODUCTION

Data mining refers to analyzing huge amount of data for finding unknown patterns, facts that are not easily supposed to extract from summarized data. Now a days, as we know the amount of data and data complexity is increasing rapidly and from the raw data, then mining of data and extraction of knowledge is becoming more crucial and beneficial. There are mainly four tasks of data mining : Classification - classify the data which belongs to class which are already defined in the training data, Clustering – works same as the classification do but the class labels are not defined, so the clustering algorithm groups the similar data altogether, Regression – is used to compute a function with the minimum error and model the data and Association rule learning – is used for finding relationship between attributes associated with the data. Now-a-days, the medical data has not been available easily and if available then it has complex structures and high dimensions.

A. Overview

Thyroid cancer

- The thyroid gland is known to be located low down in front of the neck, right below the Adam's apple.
- Thyroid cancer occurs approximately in every age group, even though its malignancy might increase with an increase in age, usually after 30 years of age. The malignant types of thyroid cancer are more often found in adult patients.
- Occurrence of thyroid cancer is three times more often in women.

Thyroid Cancer Symptoms

The symptoms of these patients are as follows:

- Majorly oncologists find a lump or nodule in the neck as the symptom
- Patients may suffer from much pain in neck and ears.
- Difficulty in deglutition.
- Difficulty in respiration.
- Croakiness in throat.
- If patient having cough that lasts for longer period of time and is not related to a cold.
- If patient is not diagnosed with any of these symptoms, then they are diagnosed during physical examination by the oncologist.

Major types of thyroid cancer are as follows:

- **Papillary thyroid cancer-** It is the most common type of the thyroid cancer. 80% of the patients are diagnosed with this type cancer. This cancer grows slowly with the development of follicular cells either in the one or both lobes of the thyroid gland.

- **Follicular thyroid cancer-** It is also called differentiated thyroid cancer. Only 10% patients are affected with this cancer. Consumption of inadequate amount of iodine in the diet is the major cause of this cancer.

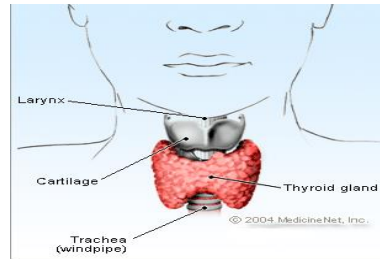


Fig 1:- location of thyroid cancer

- **Medullary thyroid cancer-** This cancer is the most malignant form among the above ones. Only 4% patients are affected with this cancer. It develops from the C cells in the thyroid gland.
- **Anaplastic thyroid cancer-** This type is also called the undifferentiated thyroid cancer. Only 2% patients are affected with this cancer. It is most malignant form of cancer which spreads to the neck and to the other parts of the body quickly.

II. Previous work

This section provides information about various approaches used for survival prediction. These papers include various approaches and techniques for prediction of survival. The approaches are discussed below:

- i. **Prediction of survival in thyroid cancer using data mining techniques**
M.Jajroudi et al. (2014) [1] proposes a method to predict the survival in thyroid cancer using data mining techniques. They used the multilayer perceptron algorithm to predict the survival. This paper has considered 7706 records and 16 features from Surveillance, Epidemiology and End Result from National Cancer Institute. They selected 16 features by reviewing literature and under the guidance of two expert oncologists. The Multilayer Perceptron algorithm applied on the data set. They selected three measurement criteria to analyze the result which were Accuracy, Sensitivity and Specificity. The accuracy, specificity and sensitivity for 1-, 3- and 5-year using multilayer perceptron was (92.9, 92.8%, 93%), (85.1%, 87.8%, 82.8%) and (86.8%, 96%, 74.3%) . According to the research study multilayer perceptron algorithm gives better results for the survival prediction of the thyroid cancer patients.
- ii. **Tree based model for thyroid cancer prognostication**
Mousami Banerjee et al. (2014) [2] proposes a tree based model for predicting thyroid cancer survival. The authors presented the 4 distinct prognostic groups for the thyroid cancer survival prediction. They used the random forest tree algorithm for predicting the survival for above 4 groups. This paper has considered dataset from the SEER database repository, 43392 cases of Differentiated thyroid cancer were recorded.[2] There were 4 distinct prognostic groups . Group I (red) consist of female patients 49years of age. Group II (blue) includes male patients 49years of age, patients 50 to 61 years of age with localized/regional tumor stage, or female patients 62–70 years of age with localized/regional tumor stage. Group III (green) includes female patients 71 to 75 years of age with localized/regional tumor stage and male patients 62–75 years of age with localized/regional tumor stage. Group IV (gold) includes patients 50–61 years of age with distant tumor stage, patients,62–75 years of age with distant tumor stage, patients_76 years of age with localized/regional tumor stage, or patients 76 years of age with distant tumor stage. According to the study the 5 year survival percentage for 4 groups were 100%, 98%, 91%, 64% and similarly for the 10 year survival percentage for 4 groups were 100%, 96%,85%, 50%.
- iii. **A Population based analysis of survival factors in differentiated and Medullary thyroid cancer**
Neil Bhattacharyya et al. (2003) [3] purpose was to determine the survival for differentiated and medullary thyroid cancer. They used the Kaplan meier analysis to predict the survival of differentiated and medullary thyroid cancer. This paper has considered dataset from the SEER database repository out of which 18,118 cases of differentiated thyroid cancer were recorded, in which 15,820 were papillary thyroid cancer, 1799 were follicular thyroid cancer, and 499 were medullary thyroid cancer.[3] The survival percentage for papillary ,follicular and medullary thyroid cancer were 87.7%,80.2%,73.7%.
- iv. **Comparison of Survival Rates Between Papillary and Follicular Thyroid Carcinomas Among 36,725 Patients**
Sam L. Oyer et al. (2014) [4] purpose's of the study was to compare the survival rates of the 36,725 patients between Papillary and Follicular Thyroid Carcinomas. They have used the Kaplan Meier survival analysis with Mann Whitney statistical analysis. The patients who having papillary thyroid

carcinoma or follicular thyroid carcinoma were identified from the Surveillance, Epidemiology, and End Results Database repository. The Kaplan Meier survival analysis and Mann Whitney statistical analysis test applied into the dataset and the results being recorded for the considered dataset. The analysis shows the better result in papillary thyroid cancer.[4] The comparison shows follicular thyroid carcinomas have worse survival than the papillary thyroid carcinomas.

- v. **A Population based analysis of survival factors in differentiated and Medullary thyroid cancer**
 Claire Schwartz et al. (2012) [5] purpose's of study was to estimate the survival benefit of Radioactive iodine for the thyroid cancer patients. This paper used the Kaplan Meier analysis with Mann Whitney test and fisher exact test. They collected the data from the French cancer registry. The Kaplan Meier analysis along with Mann Whitney test and fisher exact test applied onto the dataset and the results being recorded for the considered dataset. They estimate the survival benefit using these techniques. They computed the 10 year survival benefited percentage of patients in differentiated and Medullary thyroid cancer. [5] The 10 year survival benefited patients percentage is 88.7%.

III. The Proposed Approach

Presented below is the flow diagram of our proposed approach:

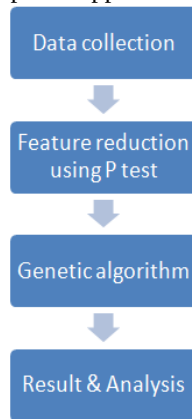


Fig 2:- flowchart of proposed approach

Data collection :- From the figure 2 in the first step of the proposed approach data set is collected from National Cancer Institute . In this dataset 9 attributes are there with 60 instances. Age, Year, Treatment method, tumor size, Gender, Stage, Histology, Race, Survived.

Feature reduction:- From the above figure 2 in the second step of the proposed approach , attribute reduction is be done using p test. In p test first step is to State the hypothesis statement. Then define the significant level (alpha value) and for each attribute compute the test statistics. Then compare the p value from the original p value table. Make the decision based on the stated hypothesis, accept and reject the attribute accordingly. In our study after p test Age, Year, Treatment method, tumor size, Gender, Stage, Survived were left.

Genetic Algorithm:- From the figure 2 in the third step of the proposed approach, genetic algorithm is to be applied on reduced attribute dataset. After performing the genetic algorithm on reduced attribute dataset then it is being performed on all attributes. Then result is recorded in both the cases. In genetic algorithm load the dataset first i.e initialize the population. Then calculate the fitness function for the entire population. Perform the crossover and mutation techniques to generate the new test data for the given population. From the given pool of population find the best fittest individual.

Result & Analysis:- from the figure it is the fourth step of the proposed approach , in this evaluation of the results for the algorithm is done and then we calculate the accuracy .

IV. RESULT

In our study multilayer perceptron and genetic algorithm were performed with all attributes and also with reduced attributes to achieve the better results. As a result, the patients with papillary thyroid cancer has maximum chance to survive with stage 1 with cohort method. The patients who were affected with Medullary and follicular thyroid cancer have minimum chance to survive.

	Existing approach	Proposed approach
Algorithm	Multilayer perceptron	Genetic algorithm
Attributes	9	9
Accuracy	86.8%	90%

Table 1:- Results with all attributes

	Existing approach	Proposed approach
Algorithm	Multilayer perceptron	Genetic algorithm , p test
Attributes	7	7
Accuracy	78.6%	80.3%

Table 2:- Results with reduced attributes

V. CONCLUSION

The basic aim of this paper was to develop a method to predict the survival in thyroid cancer. It reviews various methods and Data mining techniques used by the researchers and gain knowledge how they had used those methods and techniques. This paper also focuses on the challenges faced by researchers while performing those methods and collecting data. In our study Genetic Algorithm and multi layer perceptron being used and then achieved results were compared .Genetic Algorithm yield more efficient and beneficial survival results for the thyroid cancer patients. In our study the patients with papillary thyroid cancer has maximum chance to survive with stage 1 with cohort method. The patients who were affected with Medullary and follicular thyroid cancer have minimum chance to survive. As a future work, the work can be extended by collecting the large number of instances, attributes and also with other Data mining technique in order to achieve the better results for the survival prediction in thyroid cancer patient.

REFERENCES

- [1]. M.Jajroudi,T.Baniasadi,L.Kamakar,F.Aababi,M.Sanei,M.Ahmadzade "Prediction of survival in thyroid cancer using data mining techniques ", Journal of surgical oncology 2014.
- [2]. Mousami Banerjee, Daniel G Muenz, Joanne t. Chang, Maria Papaleontiou and Megan R. Haymart "Tree based model for thyroid cancer prognostication", Journals of Clinicals Endocrinol metabolism 2014.
- [3]. Neil Bhattacharyya , Boston, Massachusetts, "A Population based analysis of survival factors in differentiated and Medullary thyroid cancer", Journal of Otolaryngol Head Neck surgery 2003.
- [4]. Sam L. Oyer, MD; Valerie A. Fritsch, MD; Eric J. Lentsch, MD, "Comparison of Survival Rates Between Papillary and Follicular Thyroid Carcinomas Among 36,725 Patients",Journals of Otolaryngology & Laryngology ,2014.
- [5]. Claire Schvartz, Franck Bonnetain, Sandrine Dabakuyo, Me' lanie Gauthier, Ade' le Cueff, Sandrine Fieffe' , Jean-Marie Pochart, Inna Cochet, Elodie Crevisy, Audrey Dalac, Dimitri Papathanassiou, and Michel Toubeau , "A Population based analysis of survival factors in differentiated and Medullary thyroid cancer", Journals of Clinical Endocrinol metabolism,2012.
- [6]. Shanu Shroff , Siddhi Pise ,Pratiksha Chalekar,Suja S. Panicker"Thyroid disease diagnosis: A survey" ,International Conference on Intelligent Systems and Control ,2015 IEEE.
- [7]. Jiawei Han and Micheline Kamber, Morgan Kaufmann, "Data Mining: Concepts and Techniques", Second Edition, 2001.
- [8]. "Machine Learning Repository"<http://www.ics.uci.edu/~mllearn/MLRepository.html>
- [9]. "Cancer dataset repository"<http://www.cancer.gov/>

ACKNOWLEDGMENTS

I take this opportunity to acknowledge all those who have guided me in this project work. I express my earnest gratitude towards Ms. Kriti Saroha, my project guide for her valuable encouragement and guidance.

I would like to thank the faculty members of C-DAC, NOIDA for their constant support and guidance during my research work. Their motivation and suggestions were very valuable for my project work. I am grateful to them for their most cooperative attitude and suggestions, without which I could not have been able to do this work.

Lastly, I wish to thank my mother and father who were always there for me by giving everything they have, my brother and friends for their love and support.



Data Security and Data Privacy Issues Based On Internet of Things

Dr. Prabhat Mathur¹, Manisha Mathur²

Department of Computer Science & Training and Placement officer¹,

Lachoo Memorial college of Science and Technology,

Jodhpur, Rajasthan, INDIA

Scholar, Mewar University, Chittorgarh, Rajasthan, INDIA²

Abstract: Advancement of Information Communication Technological systems makes life simpler as they would be able to monitor and manage their all appliances remotely in present time. It has twisting towards critical juncture in its evolution. The evolutions of technologies are like fly in the sky, cloud computing, data analysis, and mobile smart phone technologies boost other technologies also. Internet of things is the result of augmentation of technologies to reaching their greatest imaginable. Consequently, Digital world environment has been rapidly change, prominently dramatic mushrooming enormous online or digital or cybercrimes. This paper will illuminate technologies of Internet of Things, security and privacy It presents an outline of widespread applications of IoTs and effects of IoT on security and privacy. Internet of Things technologies, challenges has to face and to propose feasible competence that would grant to keep away from cyber-attacks and upshot in a strengthening cyber security in the internet of things.

Keywords: Cyber Security, IOT Internet of Things, cybercrimes, XSS (cross-site scripting), RFID, SOAP, REST, MQTT, OTrP, security and privacy

I. Introduction

Internet of things is a wide range set of technologies and use instances has not a perfect definition. IOTs as the assist interconnected network devices, embedded in the physical environment, to nurture prevailing progression or to facilitate new set-up. Internet of things is made up of enormous internet connected devices such as printer, webcams, DVRs, Furthermore connected to network to provide information gathering and data exchanging from the environment through sensor or by the actuators provide for other systems to extend and tackle world. It has boost marketing and global economy. Every device is able to convert valuable information from real world into digital data that increased visibility. This visibility has two coins one is boon for humanity and other is curse for humanity especially for females and children. Cyberspace seems like becomes a big earning and amusement park for cyber criminals. Attacker used malware to exploit computing power of Internet of things. Always online devices are at more risk, they increase exposure to would be extortions. As well as the device itself, the web or app based site that allows the user to interact with their IoT device can be affected by persistent cross-site scripting (XSS) vulnerabilities, weak default credentials, and poor session management.[1] www.paratuspeople.co.uk 21 Jul 15:28, 2017, Criminals can exploit personal data or information quickly and easily acquires, sell and imitating by the use of wi-fi or blue-tooth device.

II. Componentes

2.1 Components: Internet of things has key components of Internet of things are:

2.1.1 Devices: Electronic equipment or a set of equipment attached with computer, smartphone, etc. including hardware and software by which user can directly interact the world and directly or indirectly connected to the internet that supportive to communicate with each other as well integrated applications called devices. Data collected by the device is called *telemetry* [2] [wikipedia](https://en.wikipedia.org/wiki/Telemetry). Telemetry is read-only data about the environment, usually collected through sensors. Each source of telemetry results in a *channel*. Telemetry data might be preserved as a stately variable on the device or in the cloud. Although each device might send only a single data point every minute, when you multiply that data by a large number of devices you quickly need to apply big data strategies and patterns[3] *Google Cloud, Solutions*

2.1.2 Gateway: Gateway is flexible hardware device that act as a bridge between two dissimilar networks and used to connect devices with the cloud. It may be routers, firewalls, server, etc, they accomplish the task transformation of protocols for all server layers of OSI model, though the term gateway has a specific function in networking, it is also used to describe a class of device that processes data on behalf of a group or cluster of devices.[4] *Farhad Mohdipoure, Bahman Javadi, Aniket Mahanti,*

2.1.3 Cloud: The data from each device is sent to Cloud Platform (Google Cloud Platform, offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search and YouTube [5]. *Java, Python, Go, Ruby*, where it is processed and combined with data from other devices, and potentially with other business-transactional data. Most important component of Internet of things cloud is distributed database management system.

2.1.4 User interface: Advanced Information communication technology provides simple touch panels controls by the easy interactive design to make complex tasks easier.

2.1.5 Real time analytics: it is an analytical approach. As soon as data flows through the system that capacity to use visualize, accessed of data and related resources. Real time analytics provide authority to availability, monitoring and handling devices, realize enactment, usage configuration and dedicated identify privacy issues and insecure conditions. Hence, can say it is a safeguard for security along with performance resourcefully likely.[6] *September 3, 2015, Greg Knowles, Real-time analytics is the discipline that applies logic and mathematics to data to provide insights for making better decisions quickly.*[7] *Gartner IT Glossary, Real Time Analytics*

III . THE INTERNET OF THINGS TECHNOLOGY

Online world connected with off line world by the expansion of the capacity this can say so Innovative technologies. Connected devices are usually designed to connect in a network, develop activities, and interact with their environment.

Table I Capacity and their description that supported by Advancement technology
[8] *Mattern and Floerkemeier (2010)*

Capacities	Description
Capacity of response	The devices can be configured remotely to develop the activities for which they have been designed.
Communication and cooperation	Several devices connect in a single network that can be connected to the Internet, enabling data sharing and update performance. The technological development of wireless networks, such as GSM, UMTS, Wi-Fi, Bluetooth, ZigBee or WPAN, is essential for this functionality.
Identification	Once the devices are connected to the network, each of the connected objects has an identifier that provides the specific information and the current status of each of those connected objects. The technology that supports these objects is RFID, NFC, or barcodes. This technology also requires a reading device that identifies the object.
Detection	Each of the connected objects obtains information from its environment with the help of sensors. Based on that information, they can respond or only register or share their status.
Actuation	The connected objects contain actuators that allow the object to transform the signals it receives and thus respond to the stimuli from the environment.
Embedded Information Processing	The information received through the sensors can be interpreted through processors or microprocessors with storage capacity allowing the object to remember precedent use of this information.
Location	The connected devices have their location identified. The technologies supporting this capacity are GPS, mobile telephony network, ultrasound, UWB, radio beacons (neighboring WLAN or RFID readers), and optical technologies.
User interfaces	Given the connected objects communicate with people in some way, there are remarkable advances in flexible screens, as well as voice, image, and gesture recognition.

3.1 RFID (Radio Frequency Identification): RFID is a wireless effective identification and sensor technology that reads or captures information stored on a tag [9]*Gubbi, Buyya et al. 2013*. RFID readers read information on the tag passed it to host computer for further analysis. RFID Technology enables mobile phone handsets to make payment once credit card information is embedded into cell phone. RFID and Wireless Sensor Networks,[10] *Dr. C.P. Jain and K.P. Vijayagopalan, National RFID program, C-DAC, Noida*. It has a variety of applications. It used to detect presence and location of objects. A RFID system includes three components: an antenna, a transceiver and transponder [11] *Domdouzis, Kumar et al. 2007*

3.2 WSN (Wireless Sensor Network) technologies: WSN is a network made up of sensor-equipped devices that use sensors to monitor the physical and environmental conditions WSN can communicate with RFID systems for better tracking on location, movement and temperature [12] *Lee and Lee 2015*. The system operates for periods varying from weeks to years in an autonomous way. The sensor network is composed of large number of sensor nodes that can be deployed on the ground, in the air, in vehicle, inside building [13] *I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on wireless, sensor network", IEEE Communication Magazine* [14] *D. Cutler, D. Estrin, M. Srivastva, "Overview of sensor networks", IEEE Computer,2004*. The Sensor Node comprises of sensing (measuring), computing, and communication elements. It works on node operations. To support node operation, open source operating system (OS) is designed

especially for WSN. It includes network protocols, distributed services, sensor drivers, and data acquisition tools. It is event driven execution model, which enables fine grained power management, yet allowing scheduling flexibility needed for unpredictable nature of wireless communication and physical world interfaces. [15] *Dr. C.P. Jain and K.P. Vijayagopalan, National RFID program, C-DAC, Noida.*

3.3 Communication protocols: smart devices are connected through the internet protocols. Protocols are the set of rules, to regulate the communication between two devices occurs. Example: HTTP, Telnet, SOAP, REST, MQTT, Web protocol CoAP (Constraint Application Protocol) etc.

3.4 Domains: Internet of things is a perception applied to any domain. Domain is a group of computers and devices that directed by a set of common rules and procedure. In the context of Internet it comprises of computers, users, workstation, devices, resources, database and servers sharing different type of data via internet resources. It accomplishes all domain functions and network security task.

3.5 IEE 802.15.4: This protocol is a standard compatible with many Internet applications. However, it has limitations in making communications between different devices separated from each other by large distances.

3.6 Bluetooth: Bluetooth has always been so present that it seems strange that it could have evolved during the last several years. Overall, Bluetooth is a wireless connection standard whose purpose is to connect devices while always sharing data at short distances. *Source: based on [16] Mindtribe 2015*

IV. INTERNET OF THINGS APPLICATIONS

IOT Analytics market survey report measured what the really popular Internet of Things applications are right now on the basis of three things: What people search for on Google, what people talk about on Twitter, and what people write about on LinkedIn. The highest score received a rating of 100%, the other Internet of Things applications were ranked with a percentage that represents the relation to the highest score (relative ranking). Smart Home, Wearables, Smart city, Smart Grids, Industrial Internet, Connected car, Connected Health (Digital health/Telehealth/Telemedicine), Smart Retail, Smart Supply Chain, Smart Farming, [17] *Mrs.M.Gandhimathi, 02-03 Feb 2017.* Applications are not only 10 applications that are mentioned above but enormous in all field of life.

V . DATA PRIVACY AND THE INTERNET OF THINGS:

Fascinating technological revolution gives the birth to Internet of things. IoTs comprises with all web enables devices to collect, send or modified data which they acquire from surroundings environment. Operating systems, protocols, endpoints and so many devices are available in the market for monitoring and controlling devices connected through the internet for getting of things. In device-to device architecture replica alternative ways are open for someone to easily break the security or can bypass the security. In this replica IoT devices are directly connected with Bluetooth or IP Address via peer-to-peer connection for maintained relationship between two devices. Hence, device-to device replica provide less security. However, architecture of internet cloud server is an Infrastructure as a Service (IaaS), it deliberating logical as distributed by server visualization. So it provides two different types of credentials one is network credential and another one is cloud access credential, after exchanging and controlling data traffic by the application service provider. Hence, Internet dedicated cloud server architecture becomes more complicated. For collecting data regarding its processing, analysis and usage from unauthorized third parties to control provide to an individual by spreading earlier detached data. Several attacks of five phrases of Internet of things, data leakage, data breaches, authentication, Sovereignty. Data leakage can be internal, external, intentional or unintentional, authorized or malicious [15] *Fei Hu, Apr 2016.* Deprived of security in mind internet connected devices designed could be easily used by the cyber attackers as an object. IoT devices are poised to become more pervasive in our lives than mobile phones and will have access to the most sensitive personal data such as social security umbers and banking information. IoT systems are not usually designed for cyber-security, said Marc Blackmer, product marketing manager for industry solutions at Cisco. The sophistication of cyber-criminals is increasing, and the data breaches that are becoming increasingly familiar will only continue [16] *Marc Blackmer, September 08, 2016 00:01 ET*

5.1 Impact of IOT on Data Privacy and security

Several attacks of five phrases of Internet of things, data leakage, data breaches, authentication, Sovereignty. Data leakage can be internal, external, intentional or unintentional, authorized or malicious [18] *Fei Hu, Apr 2016.* Deprived of security in mind internet connected devices designed could be easily used by the cyber attackers as an object. IoT devices are poised to become more pervasive in our lives than mobile phones and will have access to the most sensitive personal data such as social security numbers and banking information.

IoT systems are not usually designed for cyber-security, said Marc Blackmer, product marketing manager for industry solutions at Cisco. The sophistication of cyber-criminals is increasing, and the data breaches that are becoming increasingly familiar will only continue [19] *Marc Blackmer, September 08, 2016 00:01 ET*

VI. CYBERSECURITY AND THE INTERNET OF THINGS

Very few security apps are available can alert users about possible security threats or compromise arising in their IOT enables home or working place. Malfunctioning or re-programming is the cause for inadequate

secured IOT devices are to prone cyber-attacks. Regularly update security is the sign of IOT products/services are designed in unsuitable for security and need to make secure equipped.

Solution: data encryption needs to be encouraged.

Numerous studies have investigated the impact of the Internet of Things on cybersecurity. Since the IoT is a disruptive trend, a comprehensive understanding of its impact on security requires the awareness that the security of IoT transactions will depend on both the existing business models and those that will emerge in the future [20] *Ehret and Wirtz, 2017; Keskin and Kennedy, 2015; Novales et al., 2016; Porter and Heppelman, 2014; Rong et al., 2015; Verdouw et al., 2013*. As such, IoT is altering previous business models. Therefore, a thorough investigation of business models is necessary in terms of both revealing their weaknesses and finding solutions before the occurrence of cyber attacks and cybercrimes.

This study emphasizes that there are three strategic categories of opportunities derived from the Internet of Things. The first is the Enablers, who will develop the technology behind the Internet of Things. The second category is the Participants, who will design, develop, and integrate the services related to the IoT. Finally, the third category embraces the Breeders who will design their own Internet Services of Things.

6.1 Description of security issues in IoT [21]Tomic and McCann, 2017

Table 2 Shows Description of security issues in IoT

Issues	Description
Authentication	The information received by a reader must come from an electronic tag.
Access control	These are the access rights of different connected devices. Some questions related to this access are related to the management of the data or the identification of the devices with access rights.
Policy enforcement	This includes performing rules that come from order, security and consistency of data to the system.
Trust	It is a concept related to the access points and the criteria measured to identify whether or not to trust a device.
Mobile security	Mobile nodes in IoT are usually transferred from one cluster to another, so the security protocols should be specifically designed contemplating this fact.
Secure Middleware	Given the large amount of data collected in a system of connected devices, security in the middleware is a pivotal concern. The use of <i>ad hoc</i> gateways to ensure such security is common practice.
Confidentiality	Sensitive information must be protected with an RFID electronic tag to prevent it from being read during the transmission process.
Integrity	It ensures that the information received by a device has not been modified or altered during the transmission process.
Privacy	This refers to the protection of user identity

6.2 IOT Security at different layers

The IoT architecture security requirements are different at different layers. In an IoT set up several ubiquitous devices are connected simultaneously and exchanging data of different kinds so that security is always a matter of importance. Due to this reason, the development of an IoT security set up requires an approach founded on a layer-by-layer basis.

6.2.1 Security of the Perception Layer: The security of the perception layer is a challenge, since devices in this layer do not usually have sufficient computational capacity or memory to ensure complete security technology. Therefore, attacks usually come from external sources at the nodes where the collected data are transferred to the transport layer. It is usual that the attacks come from the gateways, external networks, or the Internet connected to wireless sensor networks. The aim of the attacks in this layer is usually to disrupt object identification through the dis of Mac addresses of devices and jamming of networks that connect sensor nodes. Hence, the data cannot be efficiently collected.

6.2.2 Security in Network Layer: The attacks at this layer can directly target the sensor and actuator nodes and alter their ability to collect and share data. Some of the attacks in this layer include a change in the destination of information, actual information sources or records and the information itself. Other alterations as the block of connections between the perception layer and the application layer may result in a complete breakdown of the service.

Since the devices in this layer do not have much memory or computational capabilities, it does not take much bandwidth to achieve flooding in a network of simple nodes. Flooding originates from the network layer, but targets the perception layer, its storage and processing capabilities of the devices. Selective forwarding is where an intermediary node only transmits selective data packets while blocking or dropping other packets.

6.2.3 Security in the Application Layer: Attacks at the network layer can modify source of information and data destination. The exposure of data during the exchange between the different entities is a challenge for security

as databases and applications are the departure point and data is exposed to all kinds of attacks before it gets to the user system.

As data are being exchanged between different entities like databases and applications, they are exposed to all kinds of attacks before they get to the users of the information. Security threats can be from within the layer through mainly unauthorized access, data theft, supply of fake data, worms, and viruses.

In the upcoming years, the development of cybersecurity should prioritize the context, as not all situations are equal and the circumstances surrounding each of the attacks must be deeply studied with all their particularities. There is also an urgent need for an automatic evaluation of the security and privacy needs of the data contained in the devices belonging to Internet of Things systems.

Further reinforcement of security in the cyberspace should entail an adequate development of policies pertinent to various fields, such as education, workforce management, investment, entrepreneurship, as well as research and development. The activities needed for adequate response to cyberattacks should include software development, law enforcement, intelligence, incident response, and national defense.

6.2.4 Description of cybersecurity protocols for the Internet of Things [22] Tripathi, Veena 2017

Table 3 Shows Description of cybersecurity protocols for the Internet of Things

Cybersecurity Protocol	Description
Open Trust Protocol (OTrP)	It is a protocol that manages security in a Trusted Execution Environment (TEE) through the installation, update, and removal of applications. [20] M. Pei A.Atyeo, M.Yoo, Solacia, May2018
X.509	It enables management of digital certificates and public-key encryption. It is known as a standard for public key infrastructure (PKI) that forms the protocol of the Transport Layer Security used to ensure communications by email and on the web. [21]Wilfred Nilsenz, 2016

In future literature for unlike threat confront at each layer of interconnected devices solution recommended by the author aimed at the security of the perception layer, when transferred to the transport layer the protection of data is being target. To improve the security in this layer intrusion detection and wireless encryption mechanisms is benefitted.

6.2.5 The Security Network Layer It is complex layer among all layers of OSI model. It described simple and effective another four level model to reduce complexity. Collecting and sharing data ability is protected the sensor and actuator nodes. Requirement of implementing of Intrusions detection systems and encryption of data view of the fact that it ensures that by the authorized and authenticated entities, data and routers are accessed, by this way security can be improved.

6.2.6 The Security in the Application Layer Application Layer security is turn the corner by the using access control management and privacy protection. When data acquire to the user of information till that it protects data. Noticeably, subsequently point out that cyber-attacks will occur can be understand by the security parameters and specified protocols implementation.

Above description of security frameworks and implementation of protocols specified need of advancement of security from cyber invaders, always.

VII. Conclusion

There are various issues concerning data and privacy with respect to IoT devices. Although risks can be reduced By the proper security and privacy measures, and there will be fewer Susceptibilities. It is important to be cognizant that cloud storage and computing is very important to ensure security risks, as data will not be stored on presumptions. Moreover, to ensure data privacy, companies must be hand on and diminish the data that is collected. Internet of things devices is not secure and safe. Security of these devices is a big demand of today's world. Same testing procedure followed by all companies because securities of devices are depending on testing. Security programs must be opens to connect them. Software must be update regularly.

True security of a device depends on securing software programs, network connections and tools that connect these IoT devices. Profile data safety and security with adequate privacy must be ensured by the IOT service provider as well as user. Read and access data monitoring is a very difficult task. Very quickly one can delete information. So Storage of extensively data becomes difficult for the companies. Hence, cloud technology must be adopted for storing big data.

References

- [1] Paratus People, 21 Jul 15:28, 2017, *Why the Internet of Things is making us vulnerable to cyber attack*, New post, <https://www.paratuspeople.co.uk/blog/posts/5091-why-the-internet-of-things-is-making-us-vulnerable-to-cyber-attacks>.
- [2] Wikipedia, the free encyclopedia, 26 May 2018, at 21:59 (UTC), <https://en.wikipedia.org/wiki/Telemetry>
- [3] Posted by: Google Cloud, Solutions, Last updated March 19, 2018, overview of Internet of Things, <https://cloud.google.com/solutions/iot-overview>
- [4] Mohdipoure Farhad, et al., 2016, FOG-engine: Towards Big Data Analytics in the Fog, Article in Conference Proceeding, presented at IEEE 14th Intl Conf, Conference 8-12 Aug. 2016, IEEE 14th Intl Conf on Dependable, Autonomic and Secure

- Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), publishing IEEE explore, Digital Library 2016, DOI: 10.1109/DASC-PiCom-DataCom-CyberSciTec.2016.116 <https://www.semanticscholar.org/paper/FOG-Engine%3A-Towards-Big-Data-Analytics-in-the-Fog-Mehdipour-Javadi/278d8c8c5888e31c99bde252829cc7077e71de0>
- [5] Knowles Greg, September 3, 2015, IBM IoT Real-Time Insights – Analytics designed for the Internet of Things, What’s New <https://www.ibm.com/blogs/bluemix/2015/09/iot-real-time-insights/>
- [6] Gartner IT Glossary, Real Time Analytics, Continuous real-time analytics is more proactive and alerts users or triggers responses as events happen, <https://www.gartner.com/it-glossary/real-time-analytics>.
- [7] Source: Authors, based on Mattern F. & Floerkemeier C. 2010, “From the Internet of Computers to the Internet of Things”, In Sachs K, Petrov I., Guerrero P. (eds) from Active Data Management to Event Based Systems and more, Lecture Notes in Computer Science, Vol. 6462, Springer, Berlin, Heidelberg, pp 242-259, https://link.springer.com/chapter/10.1007/978-3-642-17226-7_15
- [8] Jayavardhana Gubbi et al., 2013, Internet of Things (IoT): A Vision, Architectural Elements, and Future Direction, *Journal Future Generation Computer Systems archive*, Volume 29 Issue 7, September 2013, Pages 1645-1660, Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, <https://arxiv.org/ftp/arxiv/papers/1207/1207.0203.pdf>
- [9] Domdouzis, K., et al., 2007, “Radio-Frequency Identification (RFID), applications: A brief introduction”. *Journal of Advanced Engineering Informatics*, 21, (2007), 350-355, <https://dl.acm.org/citation.cfm?id=1296610>
- [10] I.F. Akyildiz, et al., 2004, A survey on wireless, sensor network, *IEEE Communication Magazine*, Vol.40, Aug., 2002, pp 102-114
- [11] D. Cutler, et al., 2004, Overview of sensor networks, *IEEE Computer*, Vol.37, no.8, pp41-49, 2004
- [12] Dr. Jain C.P. & Vijayagopalan K.P., National RFID program. applications in the country. This is a multi-institutional project involving C-DAC, Noida, IIT, Kanpur and Sameer, Mumbai. http://www.pareadox.com/foswiki/pub/Luichart/RFIDandWirelessSensorNetworks/RFID_andWireless_Sensor_Networks.pdf
- [13] Mindtribe 2015, Mindtribe Product Engineering , Wi-Fi, Bluetooth and Internet of things. Accessed 1 February 2018 , Whitepaper, mindtribe 25 Kearny St, Suite 200 415.813.1200, <https://mindtribe.com/wp-content/uploads/2015/04/Wi-Fi-Bluetooth-and-the-Internet-of-Things-0.5-edited.pdf>
- [14] M. Gandhimathi 2017, Internet Of Things Application Ranking, Using Architecture And Usage, 02-Days International Conference on Resend Trends in Engineering Science, Humanities and Management, Sri S. Ramaswamy Naidu, Memorial College, Satur, TM India, RTESHM=17, 02-03 Feb 2017, ISBN-978-93-86171-18-4
- [15] Fei Hu 2016, “Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and implementations”, Page no. 7, CRC Press, 05-Apr-016, https://books.google.co.in/books?id=kr_1CwAAQBAJ&dq=data+privacy+and+IOT&source=gbs_navlinks_s
- [16] Marc Blackmer 2016, Security, Privacy and Safety, SOURCE: SecuritySolutionsWatch.com Security, Privacy and Safety, Security Experts at Cisco, Siemens, ImageWare and CerbAir Discuss Threats & Solutions in Exclusive Interviews on SecuritySolutionsWatch.com NEW YORK, NY--(Marketwired - Sep 8, 2016) - G4S (LSE: GFS), ImageWare Systems (OTCQB: IWSY); Siemens (FRANKFURT: SIE, <http://www.marketwired.com/press-release/security-experts-cisco-siemens-imageware-cerbair-discuss-threats-solutions-exclusive-2156463.htm>,
- [17] Keskin & Kennedy, et al., 2017, 2016, 2015, , Strategies in Smart Service Systems Enabled Multi-sided Markets: Business Models for the Internet of Things, Published in: 2015 48th Hawaii International Conference on System Sciences, 5-8 Jan. 2015, Date Added to IEEE Xplore: 30 March 2015, Electronic ISBN: 978-1-4799-7367-5, INSPEC Accession Number: 15024960, IEEE, Conference Location: Kauai, HI, USA, Print ISSN: 1530-1605, Ehret & Wirtz, 2017, *Journal of Marketing Management open access Journal*, Volume 33, 2017 - Issue 1-2: The Internet of Things (IoT) and Marketing: The State of Play, Future Trends and the Implications for Marketing, Pages 111-130, Published online: 11 Nov 2016, <https://doi.org/10.1080/0267257X.2016.1248041>, Novales et al., 2016, Credit Risk Decomposition for Asset Allocation, The Capco Institute *Journal of Financial Transformation*. N° 43, 16 Pages Posted: 28 May 2016, Date Written: May 27, 2016, Chamizo, Alvaro and Novales Cinca, Alfonso, Credit Risk Decomposition for Asset Allocation (May 27, 2016). The Capco Institute *Journal of Financial Transformation*. N° 43. Available at SSRN: <https://ssrn.com/abstract=2785535>, Rong et al., 2015, Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors, *International Journal of Production Economics*, Volume 159, January 2015, Pages 41-55, <https://doi.org/10.1016/j.ijpe.2014.09.003>
- [18] S.Sicari et al., 2015, Security, privacy and trust in internet of things: the road ahead, power point presentation, <http://www.parkjonghyuk.net/lecture/2015-1stlecture/networksecurity/paper1.pdf>
- [19] Tripathi Veena (2017) Internet of Things: Opportunity and Security Issue. *International Journal of Emerging Technologies in , Engineering Research*, 5 (10), pp. 82-86
- [20] M. Pei A.Atyeo et al., May 2, 2018, The Open Trust Protocol (OTrP), *draft-ietf-teep-opentrustprotocol-00.txt* Internet Engineering Task Force, Internet sementec draft, Intended status: Informaional intercede, Arm Ltd, H. Techofenig, ARM Ltd., Expires November3, 2018, <https://tools.ietf.org/id/draft-ietf-teep-opentrustprotocol-00.html>
- [21] Ifred Nilsen, October 3, 2016, IoT security: when X.509 certificate authentication may not work, Electronic Blog post, embedded cracking code to system development, <https://www.embedded.com/electronics-blogs/say-what-/4442799/IoT-security--when-X-509-certificate-authentication-may-not-work>



Data Analysis on Cyber Crime Motives

Anmol Mittal¹, Neha Bajpai²

Maharaja Agrasen Institute of Technology, GGSIPU¹,
Centre for Development of Advanced Computing²
Rohini, Delhi¹, Sector 62, Noida²
INDIA

Abstract: Data Analysis has been performed on dataset of “State/UT-wise Cyber Crime Motives during 2016”, downloaded from the government website “data.gov.in”, several graphs and tables have been made to visualize and study the data. Heat maps and network have also been made in order to show the correlation between several crimes in states/UTs. A user input system was made in order to show the motives least related to each other. The content based recommender system has also been proposed in order to help a user to select a state/UT with less cybercrime. Many helpful conclusions and predictions have been drawn out from it that may help to devise and move forward with new customized and modified strategies.

Keywords: Cyber Crime Motives, Data Analysis, Correlation between different motives, Correlation between different States/UTs, Scikit Learn.

I. Problem Statement and Significance

Data in its raw form is meaningless but if we are able to normalize and process the data then we are able to draw meaningful conclusions from it. Specific machine learning algorithms and data analysis tools can be used to draw out predictions which would increase the benefits provided by the software/service. In today's age data science is becoming one of the most promising fields in terms of career growth as well as that of research. Even basic analysis on raw data helps us to draw out meaningful conclusions which help us to devise future predictions/solutions. This is also called predictive analysis.

As the amount of data is increasing everyday so is the number of cybercrime throughout the world. In order to keep a person's data private and safe, new cybersecurity laws, precautions have to be formed in order to control cybercrime. In the coming age, data is going to be the most powerful way of earning money, immobilizing big companies and countries. If cybercrime is not kept in check it is going to cause a lot of turmoil in the near future. The data taken into consideration is strictly related to India, cybercrime is increasing throughout the Indian subcontinent as well and it is being performed because of various different motives like stalking, robbery and even mass violence. The data has to be studied in order to understand the major motivator/reason for cybercrime in India.

II. Introduction

Researching large amount of data provides with insight and deep knowledge about the scenarios and interrelation among different features in the data and thus helps to devise solutions/actions which couldn't have been proposed without studying the dataset. The data to be used can be either collected using data gathering techniques or by using already collected existing data.

Data analysis is a process used to inspect, clean, transform and remodel data with a goal to reach to a certain conclusion. Analysis can be performed in 2 ways that is either qualitative or quantitative^[1]. In qualitative, any non-numerical data like text, symbols are analysed. In quantitative analysis, main focus is on measurement of the data. The results are numerical. In some cases, both forms of analysis are used hand in hand.

Fig 2.1 shows a basic step by step process of how data analysis is carried out.

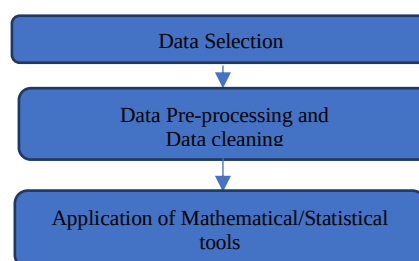


Figure 2.1: Steps to perform Data Analysis

The data taken consists of mainly numbers hence quantitative data analysis would be applied on it. Data Analysis is useful because analytics assist humans in making decisions. Therefore, conducting the analysis to produce the best results for the decisions to be made is an important part of the process. Large organisations and MNCs use this for developing future strategies and providing better content or services to their customers in order to increase profits etc.

III. Data Selection

The dataset was downloaded from 'data.gov.in' [2] (uploaded by Department of State, National Crime Record Bureau), the dataset consists of 36 rows and 20 columns (36×20). The dataset contains 17 attributes which are motives for cybercrime. An attribute "Category" which divides the place from State and UT. The table had limited data and was used to carry out some analysis but then it was transformed in order to make a new customized table on which other tasks were carried out.

IV. Data Preprocessing

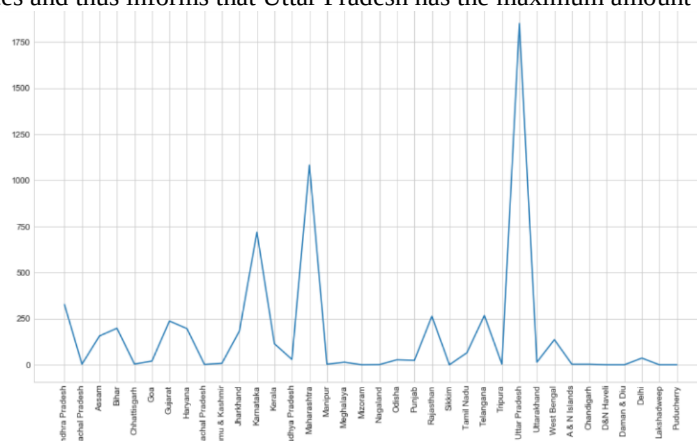
Data pre-processing is a technique that is used to convert the raw data into a clean data set [3]. The data was already in CSV format hence it was directly read by using Python script and since the data didn't have large number of entries, therefore there was no need to normalize the data as there were no redundancies in the dataset.

V. Application of Mathematical/Statistical tools

The process of evaluating data using analytical and logical reasoning to examine each component of the data provided [4] is termed as data analysis. The analysis was carried out using the packages present in Python. These packages helped in conversion and extraction of data from the CSV file and thus in making of different tables etc. to carry out the analysis.

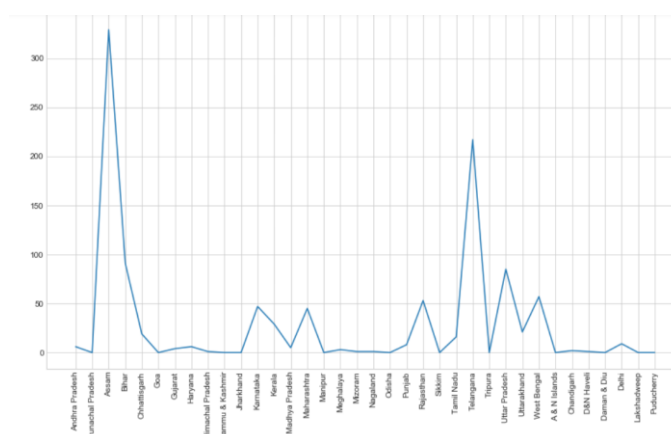
Several graphs were made in order to show which State/UT has the maximum-minimum number of a cybercrime motive out of all the States/UTs. These graphs help the user in order to understand the crimes spread out in the country by visualization.

Graph 5.1 showcases a graph between Illegal Gain and States/UTs, this graph helps to visualize the Illegal Gain in different places and thus informs that Uttar Pradesh has the maximum amount of Illegal Gain.



Graph 5.1: Illegal gain vs States/UTs

Graph 5.2 showcases a graph between Revenge and States/UTs, this graph helps to visualize the Revenge in different places and thus informs that Assam has the maximum amount of Revenge.



Graph 5.2: Revenge vs States/UTs

The mean, standard deviation, minimum and maximum were found for the dataset using python commands. By finding out these statistical values many meaningful inferences can be derived by observing the maximum and minimum value of mean.

	Sno.	Illegal Gain	Revenge	Insult to Modesty of Women	Extortion/ Blackmailing	Sexual Exploitation	Causing Disrepute	Inciting Hate Crimes against Community	Developing own Business/ Interest	Prank / Satisfaction of Gaining Control	Political Motives	Disrupt Public Services	Steal Information for Espionage	Sale/ Purchase of Illegal Drugs/ Other Items	Serious Psychiatric Illness viz. Perversion, etc.	Inciting Hate Crimes against Country	Others
count	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000
mean	17.500000	166.305556	29.333333	19.055556	15.861111	15.805556	12.444444	4.138889	3.916667	3.361111	1.111111	0.916667	0.500000	0.388889	0.250000	0.083333	68.027778
std	10.535654	362.221938	65.970989	63.040551	29.676496	32.958692	32.966168	11.069012	7.587113	16.105284	2.201010	2.309917	2.00713	0.934353	1.052209	0.368394	137.903172
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	8.750000	3.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	17.500000	22.000000	3.500000	2.000000	2.500000	2.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	16.000000
75%	26.250000	186.250000	23.000000	10.000000	13.000000	15.750000	8.000000	1.250000	4.500000	1.000000	1.000000	0.250000	0.000000	0.000000	0.000000	0.000000	48.000000
max	35.000000	1850.000000	329.000000	372.000000	144.000000	144.000000	177.000000	59.000000	39.000000	97.000000	10.000000	10.000000	11.000000	4.000000	6.000000	2.000000	593.000000

Figure 5.1: Description of dataset taken into consideration

This table (fig 5.1) helps to observe the basic statistical values of the data which in turn helps out to draw some basic conclusions.

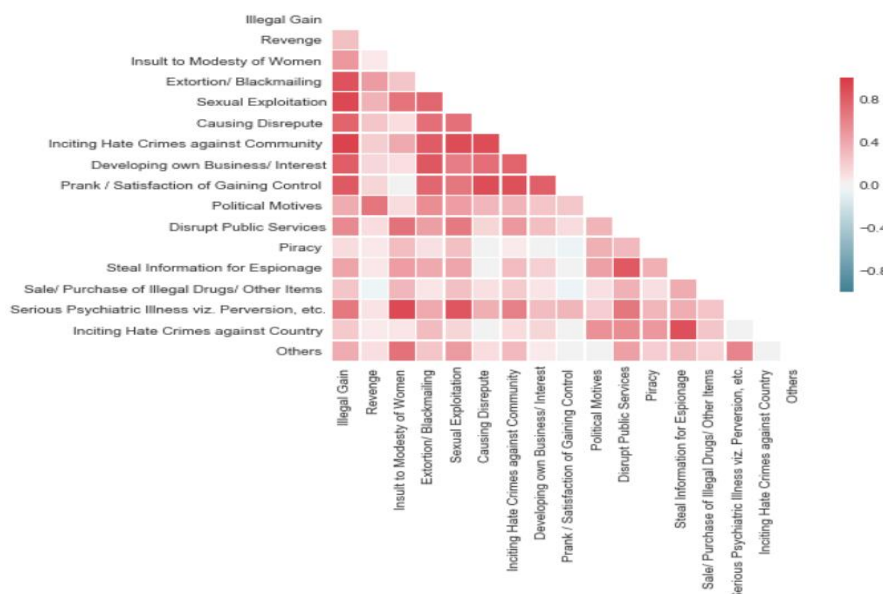
The term "correlation" refers to a mutual relationship or association between values. Correlation table was made for the cybercrime dataset and thus it helps to find the magnitude of interrelation among the crimes.

- Correlation can help in predicting one value of a variable from the given value of another.
- Correlation can indicate the presence of a causal relationship.

A very helpful way of visualizing complex data such as correlation can be done by the help of a heatmap. A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colours [5].

Heatmap was also made using correlation table using several commands to give out a visual representation of the correlation among the different crimes as the correlation table alone was quite tedious to understand. (Graph is shown below as Graph 5.3)

The correlation table is drawn out using the mathematical formula of correlation for each pairs of data possible in the dataset. The formula (fig 5.2) consists of 2 variables and their respective means as shown below:



Graph 5.3: Heatmap for Cybercrime Motives

$$\rho_{X,Y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

X = Variable 1

Y = Variable 2

$\bar{X}$ = Mean of Variable 1

$\bar{Y}$ = Mean of Variable 2

Figure 5.2: Formula of correlation between 2 Variables

A network is a set of objects (called nodes or vertices) that are connected together. The connections between the nodes are called edges or links ^[6].

Using the correlation table, a network figure was made which helps to understand the amount of correlation by displaying it in form of nodes and edges to form a map/network. Fig 5.3 showcases the derived network. The network only connects those motives which have a motive greater than 0.8. Also the case where a motive is perfectly correlated to itself has been eliminated.

Here nodes represent the cybercrime motives and edges describe the amount/measure of correlation among each other.

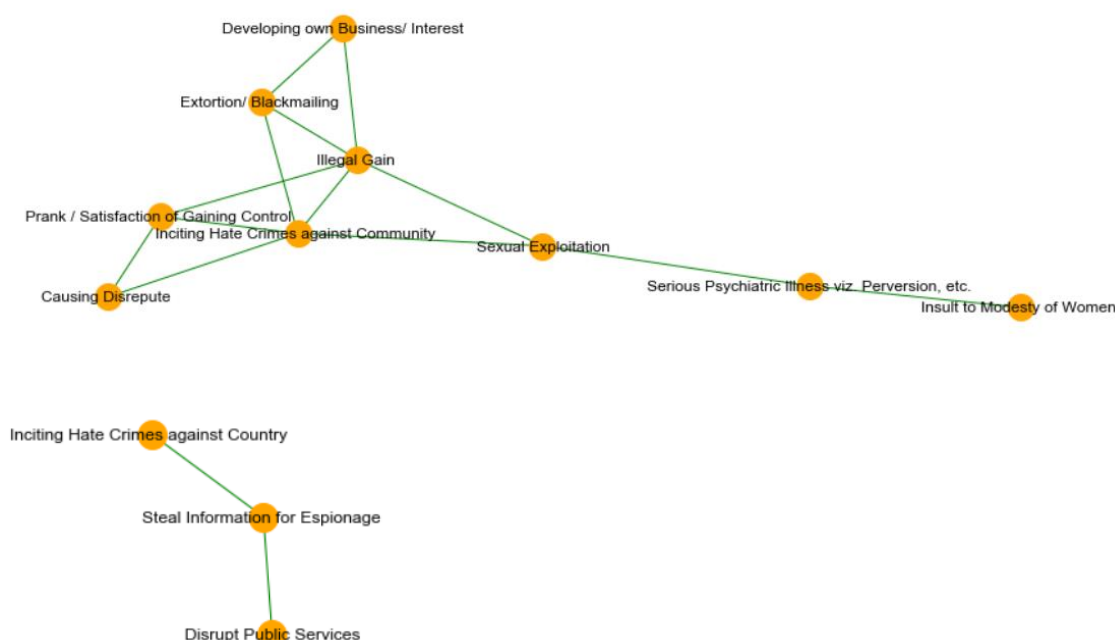


Figure 5.3: Network between different Cybercrime Motives

By developing an algorithm, a user input was made to suggest the user a list of some States/UTs which are better to live than the entered place, this list is strictly made in terms of cybercrime motives.

This list has many real life applications as if a person wants to shift from one place to another in the country and also wants his/her new residence be in a State/UT where cybercrime is less as compared to the present place of residence. Hence by using this recommender system the machine by studying the given dataset would give out a list of some States/UTs where cybercrime is either negatively correlated or has zero correlation in comparison with the present State/UT.

The table (Figure 5.4) shown below showcases a list of States/UTs as output when the current residence was input as Rajasthan. The recommender system used is content based type as it recommends an output using the present information in the dataset. Some content based filtering systems are Google, Wikipedia etc.

State/UT	Correlation
Sikkim	-0.101594
D&N Haveli	-0.004319
Mizoram	-0.004319
Nagaland	0.275171
Assam	0.285421

Figure 5.4: Table showcasing a list of States/UTs given in output

VI. Conclusion

Cybercrime is a very big motivator to a lot different crimes. It can now be said that cybercrime is no longer a problem to repudiate. The country covets a solution to fight cybercrime. After studying the dataset taken into consideration many inferences can be derived from it.

The highest mean, that indicates the most amount of cybercrime motive throughout different States/UTs were found to be for:

Illegal Gain, Others, Revenge, Insult to Modesty of Women respectively.

The lowest mean, that indicates the least amount of cybercrime motive throughout different States/UTs were found to be for:

Inciting Hate Crimes against Country, Serious Psychiatric Illness Viz. Perversion, Sale/ Purchase of Illegal Drugs/ Other Items respectively.

The maximum number of crimes due to motive for cybercrime was found to be in:Maharashtra, Uttar Pradesh, Karnataka respectively.

The States/UTs where cybercrime motives were found to be the least were:

Himachal Pradesh, Madhya Pradesh, Sikkim, Daman & Diu, Lakshadweep, Puducherry.

A user input system has been successfully made which showcases a list of crimes which are least related to each other and suggests the user a state/UT which has less cybercrime in comparison with the State/UT entered by the user.

References

- [1] <http://www.dissertationindia.com/blog/importance-of-data-analysis-in-research>
- [2] <https://data.gov.in/resources/stateut-wise-cyber-crime-motives-during-2016>
- [3] <https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/>
- [4] <http://www.businessdictionary.com/definition/data-analysis.html>
- [5] <https://www.surveyanalytics.com/system/heatmap.html>
- [6] <https://mathinsight.org/definition/network>
- [7] <http://www.labri.fr/perso/nrougier/teaching/matplotlib/>

Acknowledgement

Mrs. Neha Bajpai, Senior Technical Officer & Cyber Security Expert, C-DAC Noida.



Brief Outline & Comparison of Various Energy Efficient Routing Protocols in MANET

Mona Malik¹ and Rekha Saraswat²
C-DAC, Noida
GGSIIP University, Delhi

Abstract: MANET is a wireless mobile ad-hoc network that does not contain base station. In MANET nodes are connected with each other in their transmission/receiving range where nodes can move freely. As the nodes are battery operated therefore energy conservation is a serious concern in the MANET. Energy deficiency of a node can affect the whole network thus to increase the network period, routing within the network must be energy efficient. This work gives an overview of different protocols available for energy efficient routing and gives an insight to these protocols.

Keywords: - MANET - Mobile Ad-hoc Network, AOMDV- Ad hoc on- Demand Multipath Distance Vector Routing protocol, DoS- Denial of service

I. Introduction

MANETs now-a-days are becoming popular as ad-hoc networks are infrastructureless networks therefore economy mode of communication but there is security concern. Rather nodes rely on the power of batteries which have limited range. Power letdown of a mobile node disturb the node and its aptitude to advancing of packets towards destination and therefore the overall network lifetime. To increase the network lifetime maximum no. of nodes in the network need to be alive. Nodes that forward the data are selected dynamically as per the network connectivity.

In MANETs any node can act as host as well as router therefore needs to be active all the time for better connectivity and hence power consumption is the major alarm that needs to be considered.



Figure: Mobile Ad hoc Network

MANET Characteristics

- Dynamic topologies: It means nodes are free to move i.e. can change their links frequently having unidirectional and bidirectional links.
- Self-Configuring: It means that nodes change their path according to their assortment i.e. radio frequency range.
- Battery-Constrained: It means wireless networks have less capacity as compared to hardwired matching part due to fading, interference, noise and multiple access etc.
- Energy-Constrained: It means all nodes have battery that may exhaust due to continuous usage therefore requires energy optimized algorithms.
- Physical Security Threats: It means more prone to physical security threats as fixed matching part. Eavesdropping, spoofing and DoS should be taken care off.

The main motive of this work is to resolve the problem of route retention as done by selecting nodes with maximum residual energy. As nodes in MANET are highly dynamic and therefore have high energy requirements. Due to continuous usage of a node, energy gets drained as nodes have battery and nodes have no other option of battery replacement or charging.

Therefore, for better network connectivity energy of nodes need to be preserved for long lasting network. Various energy aware routing protocols have been proposed but they have issues like hop count not considered with battery residual energy, nodes having maximum energy have been considered, or average residual energy is considered. These approaches do not consider the minimum hop count paths with remaining energy in multipath routing protocols.

II. Literature survey

To perform an analysis on existing routing protocols that focuses on energy efficiency in MANET, brief description of the protocols is given below.

P. Periyasamy [1] et.al 2014 As MANETs have confined battery life so energy considered as the important design method to have been concerned. Hence, the author defines an improved energy efficient routing scheme by making certain modifications to MMRE-AOMDV update rules of routes by making energy adjusted paths over MMRE-AOMDV to form OMMRE-AOMDV protocol. protocol focuses on power control during data transmission. The author focuses on minimal residual energy of nodes according to their increasing order path gets selected , not considered hop count or any threshold value.

AQEEL TAHA [2] et.al 2017 projected a protocol known as per FF-AOMDV. The function defined is able to invent the most improved path between source and destination with minimum energy requirements. The proposed protocol consumes more energy as the packet size increases i.e. energy consumption increases with increased packet size. protocol consumes more energy as the packet size increases i.e. energy consumption increases with increased packet size.

Mukesh Kumar [3] et.al 2016. In projected paper link stability is concerned to extreme energy with multiple paths to enhance energy proficiency. In this the node having highest energy among all neighbor gets selected and intermediate nodes gets continuously exhausts their energy due to regular forwarding and accepting of packets. Source have to decide the node which node gets selected to forward the packet according to the maximum energy for data sending. Route retention problem is not handled efficiently as path with maximum hop count can also be selected and focus is only on maximum energy.

Yang YU [4] et.al 2015 projected a method named DNDR (Dual Node-Disjoint Paths Routing) to modify reliability in nodes. It combines R-AODV and AOMDV having node disjoint paths by selecting only two routes for controlled routing overhead, it sums the backup routing policy to make the process of data recovery effectual for link disconnection. Suitable for only small or medium mobility networks.

M.M. Goswami [5] et.al 2017 given protocol consists of both reactive and proactive components. Reactive route setup is done to find multiple paths to destination while paths are monitored, maintained and improved proactively using different control packets called FM packets. Existing approaches lack the maintenance of alternate paths therefore employs precomputed routes. Paper anticipated an approach i.e. AODV based distributed hybrid multipath routing(ADHM). Paths are build multiply and data is distributed stochastically over the multiple paths found according to their quality. Paths are monitored continuously in a proactive way.

As routes are maintained proactive therefore increases the overhead in the network because of control packets

III. Comparative Analysis

In this part, Comparision of the existing schemes/ protocols has been done, that need to be discussed.

S. No.	Title	Authors	Work details	Pros	Cons
1.	A Node-Disjoint Multipath Routing Protocol in MANETs.	Yang YU, Mangui LIANG	Overhearing of node is considered i.e. packets directed to particular node are also recorded by its neighbor node.	(a)Shows better results than novel approach i.e. AOMDV. (b)Has more reliability than AOMDV.	(a) Suitable for only small or medium mobility networks. (b) As backup strategy for data recovery can increase the channel dispute and transmission collision inside the network.
2.	AODV Based ADHMR for MANET	M.M. Goswami	Given protocol consists of both reactive and proactive components. Reactive route setup is done to find multiple paths to destination while paths are monitored, maintained and improved proactively using different control packets called FM packets	(a)Control packets are considered. (b)Packets transferred readily to destination.	(a) As routes are maintained proactive therefore increases the overhead in the network because of control packets. (b)small networks works efficiently here.

3.	EOAMRP for MANET	P. Periyasamy Dr. E. Karthikeyan	The proposed protocol reduces energy requirement by conserving power by employing power control mechanism for data transmission.	(a) Residual energy consideration. (b) Improves the life expectancy of the network.	(a) As the projected protocol focuses on power control during data transmission. The path selected among multiple paths if has maximum residual energy then threshold value or hop count factors not considered
4.	EEMR with Selection of Maximum Energy and Minimum Mobility in MANET	Mukesh Kumar, Ghansyam Prasad Dubey	Selects the path in which the neighboring nodes have maximum energy.	(a) Picks the route with higher measurements of energy of path. (b) No. of departed nodes will be less. (c) Boosts the network period.	(a) Route retention problem is not handled efficiently as path with maximum hop count can also be selected and focus is only on maximum energy.
5.	EEMRP for MANET Using the Fitness Function	Aqeel Taha, Raed Alsaqour, Mueen Uddin, Maha Abdelhaq, Tanzila Saba	The fitness function is used to search for ideal path from source node to destination node to minimize the energy requirements in multiple routes routing	(a) Increase the sending rate of packet delivered by avoiding nodes from vanishing out because of energy loss. (b) Improves network lifetime	(a) The proposed protocol consumes more energy as the packet size increases i.e. energy consumption increases with increased packet size.

IV. Conclusion

The research work explored above concluded that MANET basically plays a very important part in efficient and improved communication, however the main concern is only the energy of the nodes i.e. battery life of the node that limits its range of communication. Therefore, the main concern is for energy proficiency and routes that are efficient enough that reduces the requirements of resources, in all above mentioned the main focus is on discovering the routes that consumes the minimum energy so that energy constraints can be maintained and routes can be maintained for longer duration. In [4] and [5] as both suitable for small networks but [5] performs better in terms of PDR and throughput with variable node speed.

References:

- [1] AQEEL TAHA, RAED ALSAQOUR, MUEEN UDDIN, MAHA ABDELHAQ, AND TANZILA SABA." Energy Efficient Multipath Routing Protocol for Mobile Ad-Hoc Network Using the Fitness Function." Effat University, Jeddah, Saudi Arabia, under the Internal Research Grant Scheme 2017.
- [2] M.M. Goswami," AODV Based Adaptive Distributed Hybrid Multipath Routing for Mobile Ad-Hoc Network". International Conference on Invention Communication and Computational Technologies 2017.
- [3] S. M. Benakappa, M. Kiran. "An energy-aware node disjoint multipath routing protocol for MANETs with dynamic transmission range adjustment", 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECOT), 2017.
- [4] Mukesh Kumar, Ghansyam Prasad Dubey "Energy Efficient Multipath Routing with Selection of Maximum Energy and Minimum Mobility in MANET". International Conference on Invention Communication and Computational Technologies 2016.
- [5] Yu, Yang, Mangui Liang, and Zhiyu Liu. "An optimized node-disjoint multipath routing scheme in mobile ad hoc, International Journal of Modern Physics C, 2016.
- [6] "A Node-Disjoint Multipath Routing Protocol in MANETs". International Conference on Computer Science and Mechanical Automation 2015.
- [7] P. Periyasamy Dr. E. Karthikeyan "Energy Optimized Ad hoc on-Demand Multipath Routing Protocol for Mobile Ad hoc Networks". I.J of Intelligent Systems and Applications, 2014.
- [8] Hrishabha Raj, Jain Sanjay Kumar Sharma "Improved Energy Efficient Secure Multipath AODV Routing Protocol for MANET" IEEE International Conference on Advances in Engineering & Technology Research (ICAETR) 2014.
- [9] Periyasamy, P., and E. Karthikeyan. "Survey of Current Multipath Routing Protocols for Mobile AD Hoc Networks", International Journal of Computer Network and Information Security, 2013.



IMPORTANCE OF SEO

Dupen
MBA(IT), CDAC NOIDA
Guru Gobind Singh Indraprastha University
Sector 16 C, Dwarka, Delhi, 110078
INDIA

Abstract: Search Engine Optimization (SEO) are the platforms that proves to be beneficial for users as well as the companies. In past few years, SEO has come into light for various companies. SEO helps users to get the relevant information they are looking for. These days, competition is at peak in every field and at that point SEO acts like an Ace of Spades for the companies giving them an edge in the competition over other companies. This research concentrates on the importance of SEO in the current world of marketing. Even though SEO has become an important part of the companies yet, there is the lack of awareness about SEO among users. This paper focuses on how exactly SEO works and proves to be beneficial for the companies and users.

Keywords: SEO, Keyword Research, On-page, Off-page, White hat, Black hat.

I. Introduction

SEO provides the best relevant results to the users. Earlier Yahoo was at the top in the list of search engines but since the Google has arrived it has taken over the maximum number of market and has been constantly maintaining it over the years. It's because it has made it quite simple. Yahoo front page involved number of advertisements and given the search box limited space whereas Google has given the maximum space to its search box. These minute details create big differences. For example, a user say Mr. A makes a search about "Best Digital Marketing Company in India", Now Google will search through all the link and will provide him the most authenticated website at the top.

Google has Crawlers. These crawlers keep on running at the background and creates a database. Suppose, there is a website that shows the reviews about the cars. Now, that website claims that Mr. A is the best reviewer. Crawler will look up to Mr. A and in case it finds a link to Mr. A social handle then it will visit to those links and stores all the information about him in its database. When more than 10-12 websites say you provides the best service then it becomes authenticated and will be at the top in the search results.

There are various sub-topics in SEO which are important. These sub-topics are listed below:

- (i) Keyword Research
- (ii) On Page and Off page optimization

Research Objectives:

The paper mainly identifies the importance of SEO with the main research objective of understanding how it can prove to be beneficial for the businesses. It can be further sub divided into :

- Importance of SEO
- How SEO is beneficial for a business?

II. Literature Review

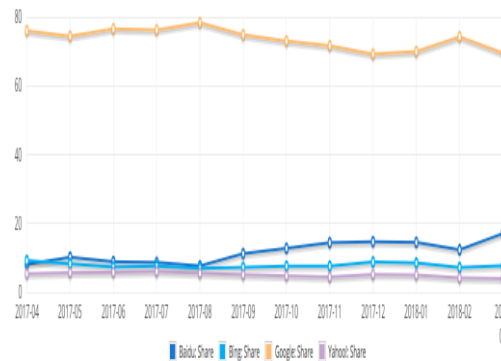
In 1990's, along with keyword stuffing people were using "Spammy backlinks" in an excessive manner to improve authentication of their websites. Since at that time there was no existence of rankings and by the time algorithms were upgraded the practices of black hat SEO was up in practice. Keyword stuffing refers to writing same keywords again and again in the texts to improve visibility of website. But, in 2000's when Google takeover the search engines were made less advertising centric. And Google began to circulate the guidelines of white hat SEO. Since the guidelines initially didn't have any impact on the ranking, so people chose to ignore those guidelines. At that time the ranking was based on inbound link and there was no method available to check the authenticity of those links. So, there was a scope of back linking method to improve the ranking. But by November 2003, with the update in the algorithms of Google many page lost their rankings. 2005 was the biggest year for SEO as in that year Google collaborated with Yahoo and MSN to minimize the link spamming. Since 2010 the evolution of SEO started and since then the conditions have changed immensely. Michael David

stated in his research paper that businesses must understand the importance of SEO as it provides them the edge in the competitive market. As per the paper published in IJERA, Dr. Raman Maini focuses on the keyword research. He shows the importance of keyword research in today's competitive world.

As keyword research do plays a vital role in placing a website to the top. Before creating the content for any any website choosing keywords comes first. Keywords defines your website. For example: You want to target a particular audience for your website. If you are a shoe retailer selling shoes on the internet you would want those people to visit your website who want to buys shoes not the ones who want to buy any cosmetic.

A.S. Hussain stated in his paper that google has become an important link between the clients and companies. It's true as internet is the most convenient way these days to connect whether it's for personal use or professional. And as per the NetmarketShare, 2018 Google is at top in the list of search engines.

Figure 1. It shows the percentage of market share captured by companies over the period of different years.



Hypothesis:

H1: SEO is important for business to generate more traffic.

H2: Right use of keyword research has direct impact on the traffic generated to your website.

III. Methodology

The awareness of this topic is quite less. It's changing year by year. Therefore, with the help of previous case studies this study is an add on in learning about the importance of SEO in today's world for the businesses to gain a competitive edge. SEO is the best possible way for businesses to reach larger audience and take their business to towering heights. Methodology used in this paper is more of exploratory research. More than posing a conclusion I'm exploring the importance of SEO. With the help of previous research papers I'm trying to get into the skin of SEO. As, In how really important SEO is? And what impact it posts on the business. My study might not provides the conclusive evidence of the solutions of my problems but, more than that it will act as hypothesis for further investigation giving deep insights of SEO.

IV. Findings And Discussions

Importance of SEO For Business:

Importance of SEO is unquestionable. As, SEO provides the business visibility and branding. When a user searches for the products in which you are dealing you would want your website to appear on the top and for the the best way is getting good ranking and authentication. That's possible only with the help of SEO. SEO also provides your business much needed credibility. As any searcher makes a mental notes of the rankings. Even though it's out of their conscious but they do. Maximum number of users don't go past the first page. So it's important that your page is listed on the first page of search results.

SEO brings the traffic to your business. The main aim of every business is to earn money and undoubtedly traffic won't bring money to you but it will make your business more recognizable in the market among the users. And the popularity eventually leads to money spinning.

Since SEO brings you valuable traffic therefore, the google analytics tracks that traffic and stores their data in the database. That data will provide you useful insights of your customers.

As per econsultancy.com, In 2011 in total 3 tests were conducted to see how much in store sales of Vodafone has been impacted due to SEO. The tests showed that there was 2.90% increase in the sales. This happened because Vodafone invested in generic keywords. Test showed that 51% increase in spending of generic keywords produced 0.95% rise in sales where as 59% of investment could rise only 0.88%.

As per business2community.com, IGN (Imagine Games Network) joined hands with a SEO company in the year 2003. With the help of this tie up search engine traffic of IGN increased by fifteen hundred percent and their revenue also increased.

Right use of keyword research has direct impact on the traffic generated to your website:

Keywords are the most important part of SEO. It's the basic step before creating the content of a website. You want specific customers to reach your website. Suppose you are a shoe retailer so you would want users who want shoes to visit your website and not the ones who want cosmetics. When a user search for shoes there are plenty of sites that pop up like Amazon.com, Flipkart.com, Shopclues.com Now, it's logically very difficult for you to beat these sites as these sites have already build up their market. There are various tools for keywords research but let's talk about the one which is free and most common that is; "GOOGLE ADWORDS." These tools also help you to know how much number of monthly searches has happened. The most important thing is educating your customers about your product rather than just imposing your terms. The keywords that user uses during the search will depicts his/her mindset. Whether he is there for education, buying or research. Before choosing the keywords it's important to understand the consumer buying process. The following Fig 1 explains this.



Fig 1: Consumer buying process

1. INTEREST

Mr. A is bored from his daily routine life. So, to bring a change he can either go for a spa or for a vacation.

2. GATHER

Out of his number of interests he ends up making a decision to go for a vacation.

3. RESEARCH

Now, it's time to research and explore number of places he can visit to.

4. BUDGET

Also act as a problem definition. He wishes to visit Dubai But his budget is 30 thousand. So he can't visit to Dubai in that budget.

5. COMMIT

He has finally decided that it's best to visit Goa in this budget.

Now, the role of Marketer begins, suppose there was always one website that popped up in each step and helped him cross that stage. So, definitely at the end he will choose to that website and will further refer that website to others also. Therefore, Keywords must be defined in a manner that your website pops up in every kind of search a user makes.

Let's again look at the example of Vodafone. The two keywords focused from them to generate more traffic were 'mobile phone' and 'phone contract'. These two generic keywords were important from the perspective of Vodafone as it represents their direct product mix. With the help of these keywords they were able to generate useful traffic to their sites resulting in the increase of their sales.

V. Conclusion

Search engines is an important tool in today's competitive world. There are number of search engines available out of which Google is on the lead with maximum percentage of market share. Keeping that in mind it would be beneficial for the Businesses to tie up google to generate more traffic to their business. SEO has an important impact on the success of business as happened with Vodafone and IGN. Vodafone and IGN are just two examples showing how much effect SEO has on the business. To get the topmost and desired results for your business SEO is the key. Even though white hat SEO techniques will take time but it will definitely produce results for you. The percentage between on-page and off-page has widened but still the importance of both the optimization is equal.

References

- [1] Kaur, G.(2017). Role and Importance of Search Engine Optimization. International Journal Of Research – Granthaalayah; Vol.5 (Issue.6).
- [2] Jain, A.(2013). Role of off-page Search Engine Optimization in search engine ranking. International Journal Of Advanced Research in computer science and software engineering; Vol.3 (Issue.6).
- [3] Londhe, S.(2017). Review Paper on Search Engine Optimization. International Journal Of Engineering Science and computing; Vol.7 (Issue.4).

- [4] Choudhry,N.(2014). Search Engine Optimization(SEO)- A study on scope of SEO in India. International Journal of Advanced Research in Computer Science and Technology; Vol.2 (Issue.2).
- [5] Khraim, H.(2015). Impact of Search Engine Optimization Dimensions on companies using online advertisement in Jordan. American Journal Of Business and management; Vol.4 (Issue.2).
- [6] David, M.(2013). Search Engine Optimization in 2013. Internetrix Research.
- [7] Maini, R.(2013). Search Engine Optimization for Improving Page Rank and Image Search Accuracy. International Journal of Engineering Research and Applications; Vol.3 (Issue.4).
- [8] Thesis on An Analysis of the Application of Selected Search Engine Optimization Techniques and their Effectiveness on Google's Search Ranking Algorithm by Edgar Damain Ochoa.
- [9] Thesis on Search Engine Optimization and it's implications in Internet Marketing by Kiran Neupane.
- [10] Searchenginejournal.com
- [11] econsultancy.com
- [12] business2community.com
- [13] Scholarworks.csun.edu



Training Need Assessment: Creating Skill Inventory At C-DAC, Noida: An Empirical Study

Ms. Pooja Atul Kokanthankar¹, Mrs. Priti Razdan², Dr. M. Mary Jacintha³,
Mrs. Shipra Sharma⁴

¹ Trainee, ² Joint Director, ³ Principal Technical Officer, ⁴ Project Engineer
Centre For Development of Advanced Computing, Noida

Abstract: Organizations spend great deal of money, time and efforts by joining hands with several training institutions in order to provide employee trainings to enhance their skills and knowledge. These skills, knowledge and abilities tend to obsolete with time which may affect various business processes of the organization. Employees at all levels require trainings, keeping this in mind this paper seeks to understand the Training Needs Analysis (TNA) for C-DAC, Noida so as to identify the existing skill inventory and assess the training requirements of the employees.

The results of this study can be used in future by researchers to identify training priorities and to guide the development of training activities in organizations.

Keywords: Training Needs Assessment, Technical Skills, Interpersonal Skills

I. Introduction

Nowadays the major aim of all Information Technology (IT) organizations is to enhance employee learning and to train them in accordance with the expectations of the current industry and workplace, in order to fulfil their mission, organizations spend great deal of money, time and efforts by joining hands with several training institutions in order to provide employees training in skills that they can further reflect in both personal and professional life.

1.1 Training Needs Assessment (TNA)

Training needs assessment or training needs analysis (TNA) is carried out to determine whether there is gap between present and future levels of performance and if training is required. It even prevents and ensures the appropriate resources, time and effort required towards targeted solutions.

Brown (2002) stated TNA is used to resolve workplace problem. It is a process of collecting the data to understand the training requirements in employees at workplace. Training Need Assessment helps in identifying the needs and propose trainings for meeting organizational competency (Barbazette 2006).

TNA helps in determining whether training is needed, causes of poor performance, content and scope of training, desired training outcomes, provide a basis of measurement, gain management support. According to Rossett (1987) a company conducts a training needs assessment to seek information about optimal performance or knowledge, actual or current performance or knowledge, feelings of trainees and other significant people, causes and solutions to the problem.

II. Literature Review

There are few terms that can be used to refer to the process of identification of training needs. The most common term is Training Needs Analysis (TNA). Reed and Vokala (2006) defined "TNA as a process of gathering, assessing and analyzing data to determine the training needs for an organization". A more detailed meaning of TNA is explained by Chiu, Thompson, Mak, and Lo (1999) "as an ongoing process of gathering data to determine what training needs exist so that trainees can be developed to help the organization accomplish its objectives". TNA refers to an effort where training data is systematically gathered for analysis and solving issues concerned in the performance (Bucalo, 1984). Chiu, et al., (1999) insist that performance problems, lack of knowledge and skills can only be resolved by training (Goldstoin 2014). Training needs assessment is recognized as the first step in any Human Resource Development intervention (Leigh, et al., 2000). However, Desimone et al., (2002) contested that in analysing HRD needs, four levels of needs has to be analysed- needs of the organisation, individual employees skills, knowledge and attitudes and their functional responsibilities as well as departments needs. Peter Musau Syanda et al (2014) observed that there was need to carry out TNA at Kenya Polytechnic University College (KPUC), Nairobi in order to transform it into a University College. This was because most of the staff had not attended any TNA before, thus they needed training to enable them to

cope with the new ranks of the college and its quality assurance demands. According to Shulgana Sarkar (2013) competency mapping model is able to identify training needs of individual employees with significant difference in the median of gap level of competencies after a need based training. According to Rossilah Jamil et al (2007) thousand Malaysian companies had positive opinions regarding their training and TNA practices, though they used several data collection methods and techniques to analyze their training needs but were highly relied on performance appraisals to collect overall TNA data. Syed Kamran Sherazi et al. (2011) stated that due to lack of comprehensive managerial TNA approach used by corporate sectors in Pakistan, their training is mostly subjective based and objectivity of the training is less emphasized during their training session. Hussein Alkinani (2013) opined that due to lack of expertise and support from top management Iraqi construction companies failed to approach TNA systematically, they conducted it based on only one level, neglecting the other levels of needs assessment; this is termed as standalone process of conducting needs assessment. Obeng, G.Y et al. (2013) indicated that Metalwork engineering enterprises possessed diverse skills; they were limited in knowledge-based skills and training, which adversely impact on productivity and competitiveness. Hands-on and a combination of hands-on and use of basic engineering tools are the priority training tools that are required to effectively build the capacity of entrepreneurs. Ali Hamad Badi Ghufli (2011) revealed that TNA in the Abu Dhabi Police (ADP) is generally conducted formally on a regular basis, but they do face certain challenges that disrupt the training selection process in the ADP; lack of expertise of the trainers, nepotism, kinship and personal relations between supervisors and the employees. P. Akilandeswari et al. (2014) revealed that Indian banks both private and public embark on training and developmental programmes for enhancing employee efficiency, knowledge and skills such that they could perform and meet customer requirements. Aarti chahal (2013) stated that private sector banks have better training programmes i.e. staff of HDFC bank are more satisfied with the training program than compared to PNB staff. V.M. Nandakumar et al. (2013) indicated that almost all the hotels are following more or less the same training need analysis methodology, but there is a marginal difference between these hotels which makes these hotels different from one another. According to Blaž Rodič et al (2012) pinned that most of the employees understand the importance of training need assessments and its objectives. V.M.Nandakumar, S.Nitesh (2013) indicated the three main factors that organizations have to remember while adopting or choosing any TNA methodology are how; to understand the current situation, determine the outcome and analyze the gap in employee's performance. Tahmina Ferdous, B.M. Razzak (2012) stated in their study that most of the banks in Bangladesh have their own training institutes for imparting regular trainings to employees, implementation of training requires regular support from management in terms of budget and other physical requirements. The management also have opted for motivational incentives for trainees for taking up training sessions. However, it is found that there is a lack of needs assessment.

On the basis of these studies, it can be concluded that training is most widely used sub-system of HRD and it is an effective tool of employee development as it helps to acquire those skills which are needed to fulfil the requirements of the job. However, many organizations do not make training need assessment and evaluation of training programs in a scientific way. Many organizations use external training and behavioural training for their employees (Dr. V. K. Jain, 2013).

III. Research Methodology

The main aim of this study is to identify the existing skills of the employees, help them to acquire new skills and eliminate the skill gaps by conducting the Training Needs Assessment (TNA) for employees in CDAC, Noida. Therefore the research objectives for this study are:

- Identify the existing skill inventory.
- Assessment of training requirements of the employees.

The study comprises of both primary and secondary research. The research has been conducted using questionnaires with sample size of 160 employees working in the organization. To represent the results of the study use of pie chart, bar diagram and tables have been done.

Participants

Group Coordinators (GC's) and employees working in Academic block and Research & Development Block of C-DAC, Noida were the main participants for the study.

IV. Results and Discussion

The sample size included CDAC employees, out of which 160 employees participated in the research. The questionnaires of Academic block were analyzed and the ratio of employees working in the block were divided into following four major sections- Project, Teaching, Project and Teaching and Administration. Academic Block stresses on synergy between Education, Research and Product Development, thus the ratio composition in figure 4(a) also signifies that majority of employees working in Academic block were indulged in Teaching (22 employees) followed by Project and Teaching (14 employees), Project (12 employees) Administration (8 employees) and Project and Administration (3 employees).

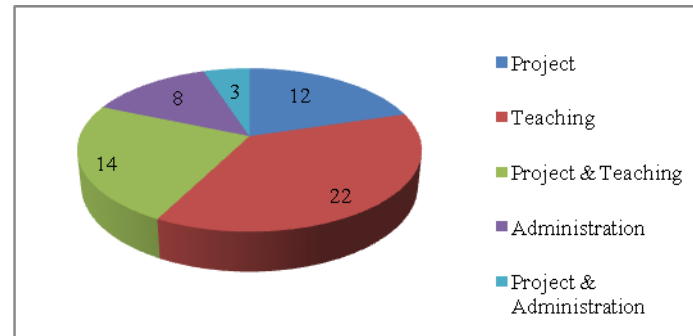


Figure 4(a): Composition of Academic Block

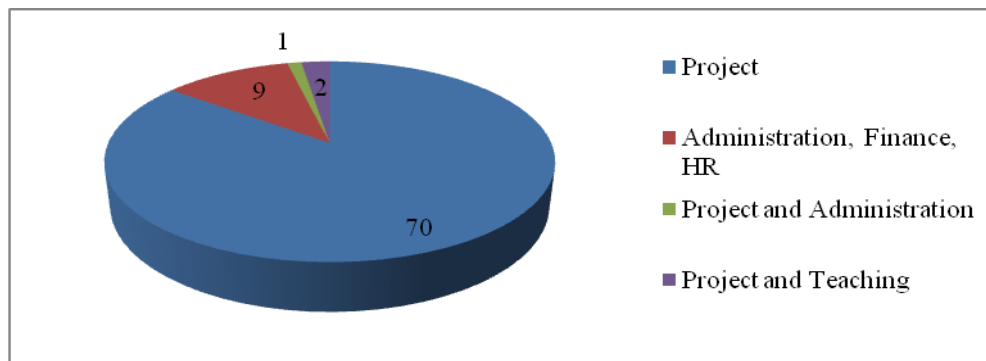


Figure 4(b): Compositions of R&D Employees

The questionnaires of Research and Development (R&D) block were analyzed and the ratio of employees working in the block were divided into following major sections- Project, Administration/ Finance/ HR, Project and Administration, Project and Teaching.

Research & Development Block is mainly associated with creation of technology aspects, thus the ratio composition in figure 4(b) also signifies that majority of employees working in this block are indulged in Project (70 employees) followed by Administration/Finance/HR (9 employees), Project and Administration (1employee) and Project and Teaching (2 employees).

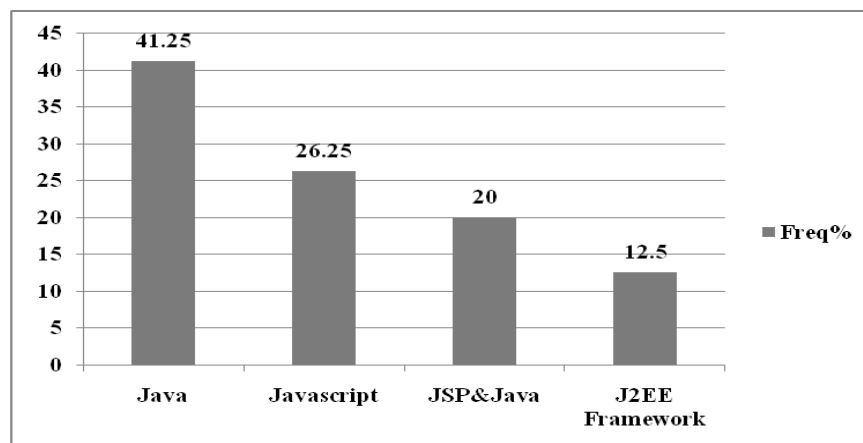


Figure 4(c): Technical Skills possessed by employees working in C-DAC, Noida

Figure 4(c), showcases Java, Java script, JSP & Java and J2EE Framework are the top major technical skills possessed by the employees of C-DAC, Noida. According to the graphical representation it can be concluded that Java is the prime major technical skill that is possessed by 41.25% of employees followed by Java script (26.25%), JSP & Java (20%) and J2EE Framework (12.5%).

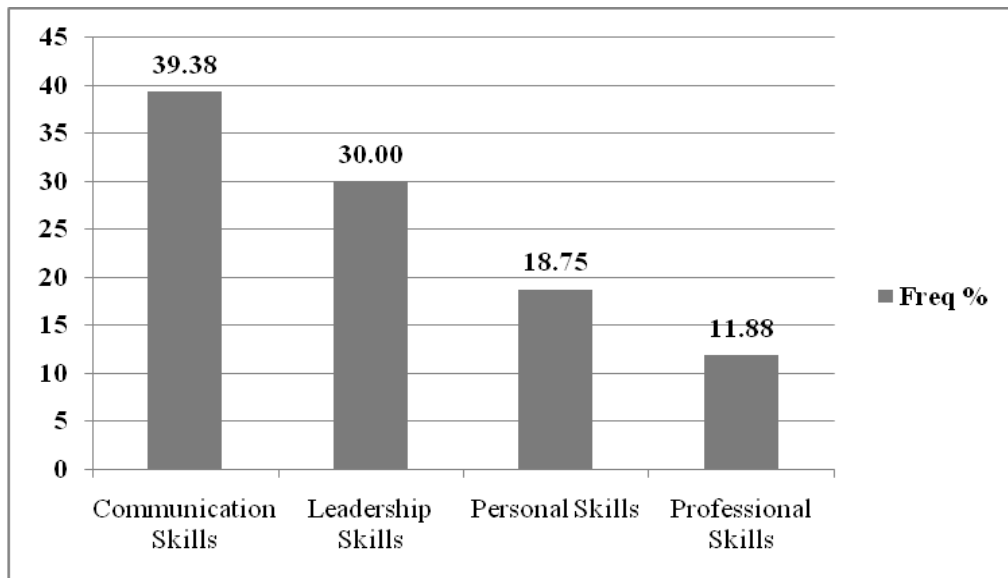
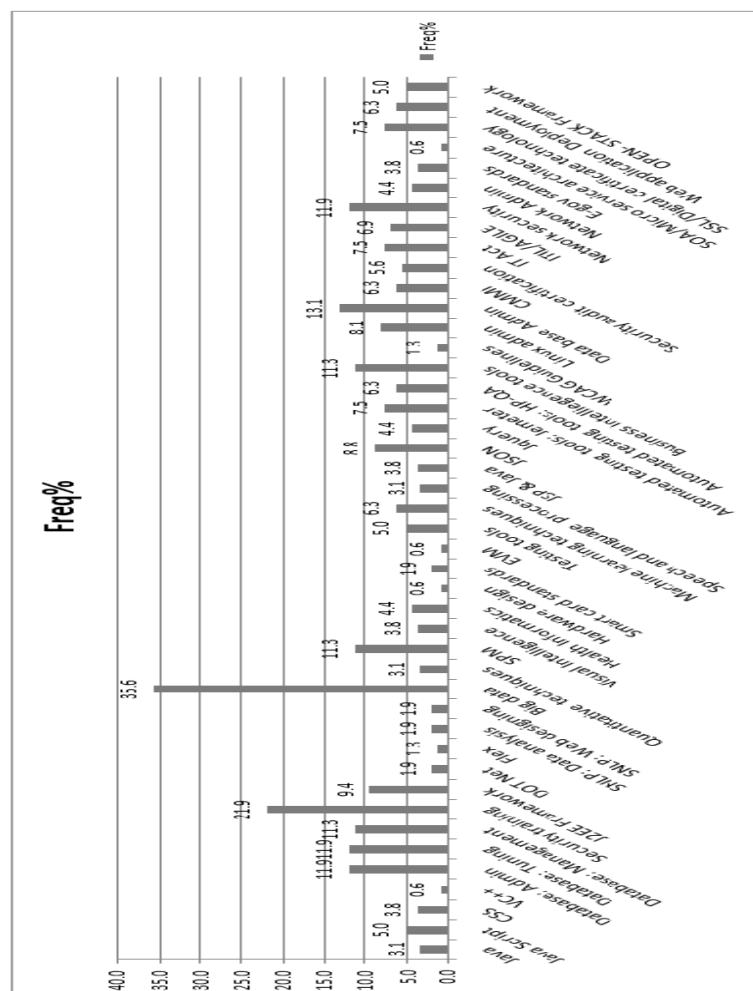


Figure 4(d): Interpersonal Skills possessed by employees working in C-DAC Noida

Figure 4(d), showcases that Communication skills, Leadership skills, Personal skills and Professional skills are the top major interpersonal skills possessed by the employees of C-DAC, Noida. According to the graphical representation it can be concluded that Communication skills is the prime major interpersonal skill that is possessed by 39.38% of employees followed by Leadership skills (30%), Personal Skills (18.75%) and Professional skills (11.88%).

Figure 4(e):TNA For Technical Skills



After receiving the employee responses, the below graphical representation figure 4(e), represents the percentage of employees who require training following Technical Skills.

Java, Java Script, CSS, VC++ Database: Administration, Database: Tuning, Database: Management, Security Training, J2EE Framework, DOT Net, Flex, SNLP: Web Designing, SNLP: Data Analysis, Big data, Quantitative Techniques, Software Project Management, Visual Intelligence, Health Informatics, Hardware Design, Smart Card Standards, EVM, Testing Tools, Machine Learning Techniques, Speech and Language Processing, JSP & Java, JSON, JQuery, Automated Testing Tools: Jemter, Automated Testing Tools: HP-QA, Business Intelligence Tools, WCAG Guidelines, Linux Administration, Database Administration, CMMI, Security Audit Certification, IT Act ITIL/AGILE, Network Security, Network Administration, E-Governance Standards, SOA/Micro Service Architecture, SSL/Digital Certificate Technology, Web Application Deployment and OPEN- STACK.

According to figure 4(e), it can be observed that under the Technical Skill category 35.6% of employees working under C-DAC, Noida needs training in Big Data, followed by Security Training (21.9%) and Database Administration (13.1%). Around 11% of employees require training in technical skills like Database: Administration, Database: Tuning, Database: Management, Software Project Management, Business Intelligence Tools and Network Security.

Finally only a small percentage of employees i.e. between the range 0.6%-1% require training in technical skills like VC++, DOT Not, Flex, SNLP (Web Designing, Data Analysis), Hardware Design, EVM, Smart Card Standards, Testing Tools, WCAG Guidelines, SOA/Micro Services, Architecture.

After a regress consultation from the Group Coordinators (GC's) of various departments of the organization following are the Interpersonal Skills which have been finalized i.e. Communication Skills, Leadership Skills, Influencing, Personal Skills, Creativity and Professional Skill to be included in the TNA questionnaires distributed to the employees.

After analyzing employee response, figure 4(f) represents TNA for Interpersonal Skills and the percentage of employees who require training in various categories.

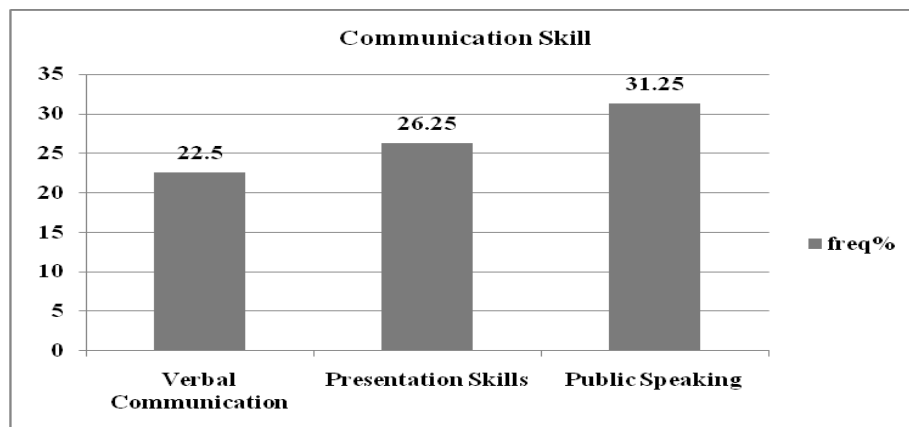


Figure 4(f): Communication Skill TNA

According to figure 4(f), it can be observed that under the Communication Skill category 31.25% of employees working under C-DAC, Noida needs training in Public Speaking, followed by Presentation Skills (26.25%) and Verbal Communication (22.5%).

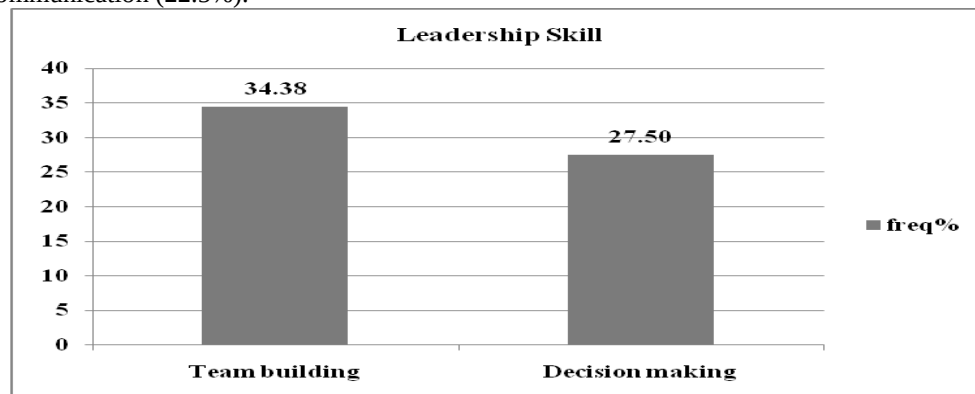


Figure 4(g): Leadership Skill TNA

According to figure 4(g), it can be observed that under the Leadership Skill category 34.38% of employees working under C-DAC, Noida needs training in Team Building, followed by Decision Making (27.50%).

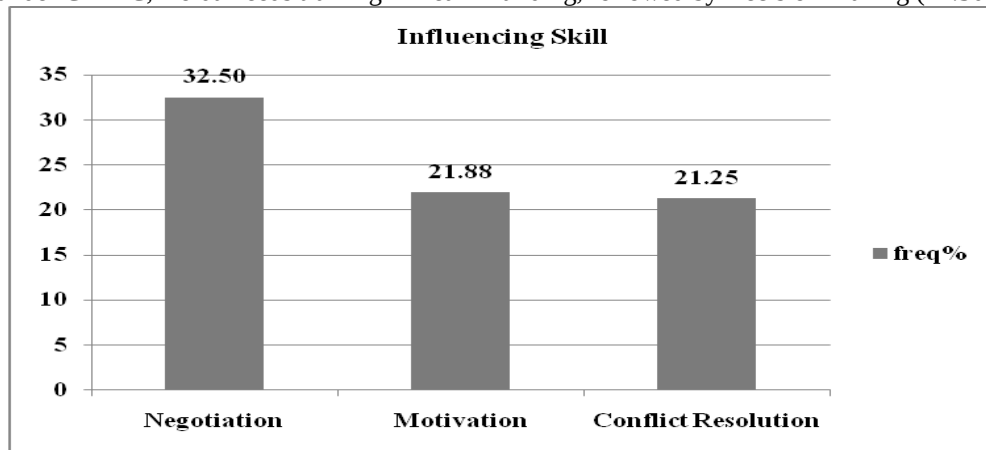


Figure 4(h): Influencing Skill TNA

According to figure 4(h), it can be observed that under the Influencing Skill category 32.5% of employees working under C-DAC, Noida needs training in Negotiation, followed by Motivation (21.88%) and Conflict Resolution (21.25%).

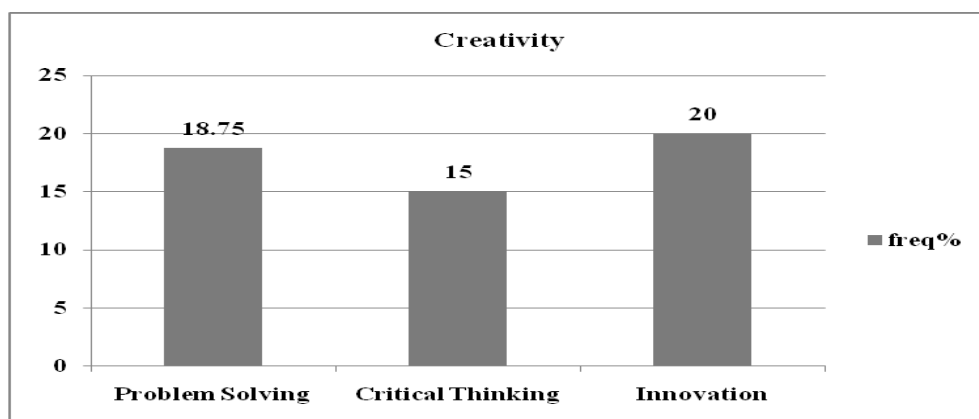


Figure 4(i): Creativity Skill TNA

According to figure 4(i), it can be observed that under the Creativity skill category 20% of employees working under C-DAC, Noida needs training in Innovation, followed by Problem Solving (18.75%) and Critical Thinking (15%).

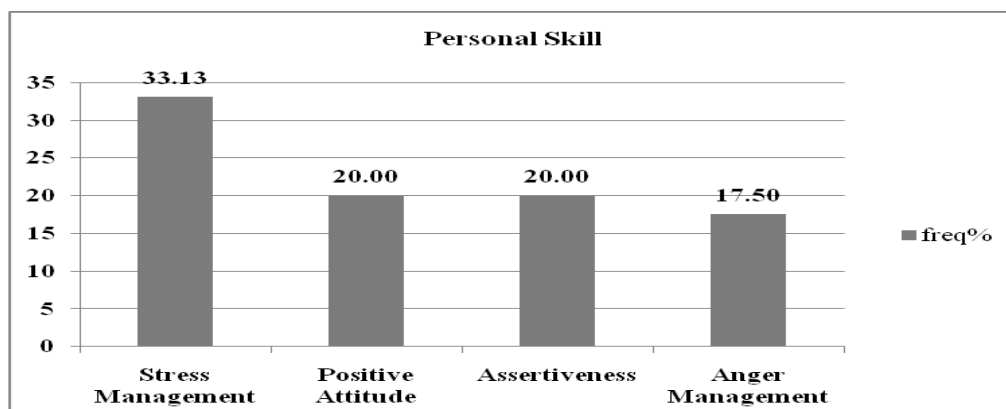


Figure 4(j): Personal Skill TNA

According to figure 4(j), it can be observed that under the Personal Skill category 33.13% of employees working under C-DAC, Noida needs training in Stress Management, followed by Positive Attitude and Assertiveness (each 20%) and Anger Management (17.50%).

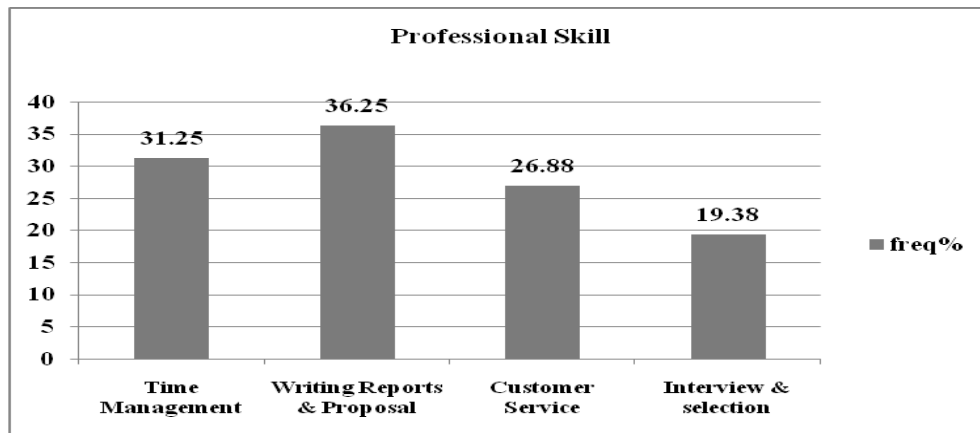


Figure 4(k): Professional Skill TNA

According to figure 4(k), it can be observed that under the Professional Skill category 36.25% of employees working under C-DAC, Noida needs training in Writing reports and Proposals, followed by Time Management (31.25%), Customer Service (26.88%) and Interview and Selection (19.38%).

The responses received by the 160 employees of C-DAC, Noida have been presented in a table format which showcases a skill set for each department. The responses provided by the GC's of all departments have been given the main preference over the responses of the employees. Once the responses of the employees of a specific department were received, I classified the employees on the basis of their positions and concerned departments and later presented the data in a priority matrix chart.

The table includes two major components- Employees request as per designation and GC requirements common with employee's requirements.

V. Conclusion

An organization is just a small element of the big picture; the remaining part is covered by the employees of the organization who work as a team with dedication and hard work in order to reach organizational and individual goals. This paper aimed to study Training Needs Assessment practises of C-DAC, Noida. The Training Needs Assessment of 160 employees of the organization revealed that Java and Communication skills were the prime main technical and interpersonal skills possessed by majority of employees, on the other side Big Data, Security Training are the two main technical skills while Writing Reports and Proposals and Team Building are the two main interpersonal skills employees of C-DAC, Noida require training in. A skill inventory has also been presented in a Priority Matrix table featuring both technical and interpersonal skills for all departments of the organization keeping the Group Coordinators (GC's) request as the major priority. The results of this study can be used in future by C-DAC, Noida to identify training priorities and to guide the development of training activities.



Impact of E-Commerce Marketplaces on Business Performance in reference to Retail Industry

Aditya Kumar

School of Information Technology

CDAC Noida

B-30, Sector-62, Institution Area, Noida – 201307 (U.P.)

India

Abstract: The paper focuses on the brick-and mortar sellers and small businesses that have adopted the E-commerce approach and have expanded their business operations to E-Commerce marketplaces like Amazon, Flipkart, Snapdeal etc. to sell their products online and services that are already available. The study was done to analyse that what were the effects of selling online at marketplaces on the current business performance. E-Commerce is a booming industry and the opportunity to utilize this medium is the need of the hour. The respondents are interviewed asking about their businesses and how their adoption of E-Commerce selling affected their business performance and what were the implications and changes faced by them setting up online.

Keywords: E-Commerce, E-Commerce Marketplace, Brick-Mortar Sellers, Business Expansion, Business Performance

I. Introduction

E-Commerce is the activity of buying/selling of products and services online and over the internet. It terms for any type of business, or commercial transaction, that involves the transfer of information across the Internet. E-Commerce allows consumers to electronically exchange goods and services with no barriers of time or distance. The advantages offered by E-Commerce is online shopping of anything at any time and at anyplace, customers find the products on e-commerce websites which may or may not be available in physical markets. It reduces cost and time.

There are broadly 3 types of E-Commerce Websites- 1). *Transactional*: An electronic storefront that sells directly to public. People search, order, pay for products online. (Walmart), 2). *Static Content*: It provides information about the product or service & contact information on how to proceed with a purchase. (Bajaj, Honda, Yamaha, Ford), 3). *Online Marketplace*: These sites are hosted platforms that bring buyers and sellers together to facilitate transactions, these types of sites include auctions (eBay), sites for local artisans (etsy), sites for retailers & wholesalers (Amazon, FlipKart, Snapdeal, IndiaMart, Alibaba).

The E-commerce marketplace is a place or a website where one can find different brands of products coming from multiple vendors, shops or person showcased on the same platform. The marketplace owner is responsible for attracting customers and the processed transactions, while the third-party vendors deal with the manufacturing and shipping. In terms of revenue, these companies take a percentage of the sales on any product sold across the platforms.

There are many benefits associated with E-Commerce marketplace such as a marketplace allows the business or platform operators to charge a cut on the products that the vendors sell therefore exponentially increasing the profit opportunities and provides a transparent purchasing process like the prices of the product and availability of the stock are all accessible on a single platform in an open environment. It is beneficial to buyers as the buyers gets a wide variety of options on products to compare & choose on a single website helping them find the most reasonable price for the best quality product. Sellers get benefitted as the platform acts as an additional sales channel to market and sell their products and enabling opportunity of overseas sales.

Objective: The objective of this paper is to understand how E-Commerce marketplace implementation affected the brick-mortar sellers in their ways of doing business and what were the challenges faced by them in adopting to this platform.

II. Literature Review:

Lots of research has been done in the field of E-Commerce by various researchers and authors. To support my paper, I had been through some of the early published papers and collected data that was relevant to my research. In 2017, Dahiya researched on E-Commerce and it's impacts on market and retailers in India which focused on the market scenario of India and what are the different types of markets and different types of retailers how e-commerce can be an attractive platform to expand the business.

In relation to how E-commerce can have effects on the economic growth of country, Liu S. in 2013 published paper relating to an Empirical Study on E-commerce's effects on Economic Growth which discussed the development of computers, technology, ICT and how these can be used measure the impact of mechanism of e-commerce development to the national economic growth. Similar study was done by Devaraj P. in 2016 in a hypothesis on assorted e-commerce performance in online market. He focused on the digital commerce and the working of e-commerce technically and classified the various e-commerce scenarios and types like B2B, B2C, some to be named and told that e-commerce can have lots of pros and cons and can be challenging in countries. Similar research to this was done by Shahjee R.in 2016 which focused more on the organisations having e-commerce operations. It was more about ICT and MIS, accounting on business, and various sectors like health, education etc. The paper discussed more of the technical aspects related to e-commerce and its benefits in core functioning of the organisations and how this can also be an issue for people belonging to rural areas to use this platform. Kahai R. et al in 2000 published a journal showing the impact of e-commerce on the entrepreneurs and small businesses and how they can make strategies to sustain and grow in this booming potential world and face the challenges and how changes need to be kept coming to retain the customers.

Research Framework: A conceptual model of the system:

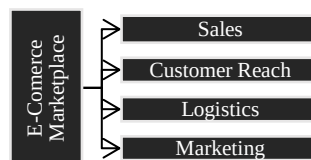


Fig 1: Conceptual Model- E-Commerce Market Place

Hypotheses: The E-Commerce marketplace is significant to business performance.

- E-Commerce marketplace is significant to sales.
- E-Commerce marketplace is significant to customer reach
- E-Commerce marketplace is significant to logistics
- E-Commerce marketplace is significant to marketing
-

III. Methodology

The data for the relevant information is based upon primary data collection by interviewing business respondents that have an already established business. To analyse and understand the respondents a descriptive cross-sectional and exploratory research was conducted with retrospective-prospective way approach to carry in-depth interviews with each one.

The sampling method used was Non-Probabilistic Purposive sampling. Respondents were selected across PAN India who all are using E-Commerce platform to conduct their business online and out of which those were sampled that are using E-Commerce Marketplace to expand their established business & sell their products there. The respondents interviewed were mostly the owners of the business or of managerial level.

The research was primarily qualitative as focused interviews were conducted and responses were taken based on some questions that are:

- a). What was your business & how did you got to know about Ecommerce marketplace?
- b). How is this adoption impacting the current business?
- c). What were the issues faced in implementing this strategy?
- d). How marketplace is helping in running the business?

The collected data was processed and analysed in a statistical tool called NVivo. NVivo is a software program used produced by QSR International for qualitative and mixed-methods research. Specifically used for the analysis of unstructured text, audio, video, and image data, including (but not limited to) interviews, focus groups, surveys, social media, and journal articles.

IV. Findings

Based on the questions, some related factors among all sellers were observed and the following findings were obtained from respondents:

Sales: More than 95% sellers told that their sales of the products were increased using E-Commerce marketplace. There was a tremendous growth in sales figures as compared to offline selling. The financial returns were good and increasing.

Logistics: More than 90% sellers were happy with the delivery service and the effective logistics provided by the marketplace for timely delivery of products. They didn't have to manage the shipping as it was already being done which saved them time and money. Also, the customer returns can be delivered back to seller from customer's end and if there is any damage to the product that would be beared by the marketplace for which compensation is given to seller. There is also provision for self-shipping if sometimes the marketplace is unable to deliver it.

Customer Reach: More than 85% sellers said that there was a global reach to customers and they can also send their products outside their country that was a big challenge in offline selling. Anybody working that do not have time can order 24/7 at any day and get the product at their destination. People that need something specific or unique can find their products online which may not be available near them. There was a dip in percentage as many customers prefer buying from local stores and there are also people that are not computer literate that cannot place an order online easily.

Marketing: This factor was the base for every operations in online selling as the sellers told its significance is directly related to customer reach which in turn is increasing sales. More than 85% of sellers agreed on the same fact that this helped the sellers in showcasing their brand and ensuring customer loyalty. Various marketing tools, digital and social marketing, search engine optimization(SEO), all can be performed to attract customers and to make brand more visible. The more is the footprint, the more are the chances of conversion. The details on the page at marketplace can contain all the features and characteristics of products and the ratings of other customers that help in making buying decisions. Marketing also helped the sellers to observe the buying behaviour, choices, trends of the customers and advertise the products targeting the right audience.

Technical Knowledge: There was a 50-50 conflict in this factor. Some sellers told that it is difficult to learn the seller panel provided by the marketplace for operating the business, making it as an expense for the seller as for that training need to be conducted for the present staff which takes time and money or to hire someone having relevant technical knowledge. Other 50% told that they are or have someone from technical background and they can understand the panel very easily and can do the operations efficiently. To perform online activities, there need to be a clear understanding of the seller panel and the rules that need to be followed on marketplace so that the seller account does not get suspended.

Employment: More than 30% sellers had a proud feeling about giving employment to new employees. There are many operations that needed to be performed while doing business and as the online marketplace platform is new, so a new expertise can be employed to handle the online operations and learn new things to help the business grow. The sellers felt a responsibility to generate employment and contribute to the growth of the country. Rest of the sellers took employment as an expense as a new employee must be given salary and train them about the working and functioning of the company.

Low Investment: To start an E-Commerce Marketplace selling there was nil to some investment involved. There was 45% of sellers said that didn't invest separately to adopt the online selling. Rest had some costs involved in their business to make it adhere to the ways of doing online business. This entirely depended upon the nature and type of business conducted.

Warehousing: More than 65% of sellers had no inventory costs involved. They had a fixed product that needed no customization and for that reason they send the product to marketplace warehouses that was a facility provided by marketplace to deliver the products under their fulfilment. This helped the sellers in reducing the costs involved and managing inventory. The seller panel had everything to manage and automate the inventory counts. Also, extra expenses were eliminated.

These all were the factors that every seller agreed upon that were common to every one of them and affected the business performance directly. These were the common elements that matched every sellers interview and description.



Fig 2 : Common elements in every sellers' view

There were some more factors or to say challenges that were faced by the sellers in doing E-Commerce marketplace selling:

Barriers to Entry: There were certain approvals for the brand that needed to be passed to sell online. Also, there were classes required for selling in different categories of products. These were something that was not required when selling offline through stores.

Release of Money: The sellers are not paid immediately but in cycles of interval. The release of payments from marketplace vary from 1-2 weeks from the date of purchase.

Brutal competition: There is an algorithm in marketplaces that show or highlight the seller with the lowest price offer in front or first in search. This increases competitiveness in the market amongst sellers as the one purchasing in bulk for resale will be able to offer a more competitive price as compared to others. The whole fight in the online world is about how to get shown first in the relevant searches.

Difficulty in feedback: About 1 out of 30 customers leave a feedback for an order placed. Therefore, it takes time for the brand to get recognised and come to top and get recognition amongst customers.

Complicated and vast: In online marketplace selling, to understand the whole system for a non-technical person is a very difficult job. For a technical person too, it will take some time to learn the complete prep system.

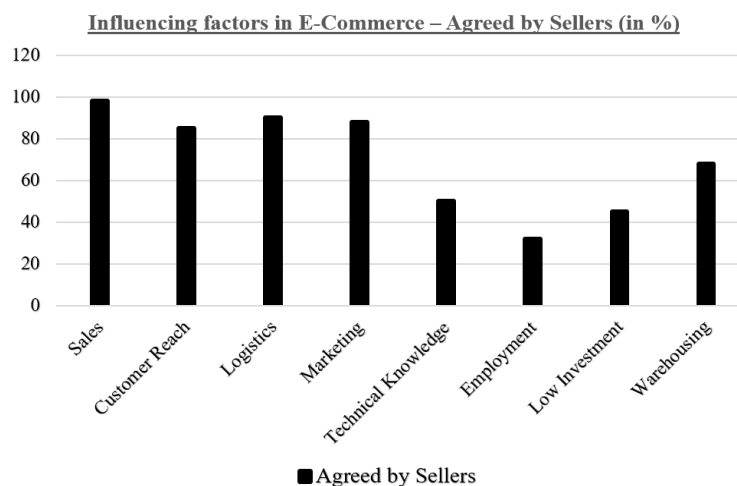
Storage issues: In marketplace fulfilment, the material or product in storage may be lost, damaged or misplaced which is responsibility of online companies.

Marketplace Fees: There is a fee charged by the marketplace on every purchase done by the customer. This fee can be as high as up to one-third of the total sale price of the product which doesn't include the monthly rental storage fees if the product is under marketplace fulfilment.

Table shows percentage of the factors that sellers agreed in common:

Influencing Factors in E-Commerce	Agreed by Sellers (in %)
Sales	98
Customer Reach	85
Logistics	90
Marketing	88
Technical Knowledge	50
Employment	32
Low Investment	45
Warehousing	68

Table 1: Influencing Factors in E-Commerce agreed by sellers



Graph 1: Influencing Factors in E-Commerce

V. Conclusion

The E-Commerce marketplace is a great platform to increase sales and to showcase our branding. It provides sellers with an additional channel to market and advertise their products. This helps them to increase their sales tremendously and increase their customer reach to different parts of the country and globally. Many sellers having struggling businesses were able to survive by selling online and make a living out of it. Online marketplace is a bit new to the Indian society but has a lot of potential to generate employment and to contribute to the growth of the country. There are employment opportunities for many people as it is a channel having various areas like developing, marketing, selling, delivering, servicing and paying for products and services. Sellers are embracing this platform and more sellers are in view to join the platform and start selling in future for better prospects.

The businesses are mostly in steep growth and once the pace catches online, there are huge gains. E-Commerce is growing constantly and with that so are the customers which adds variety of options and opportunities with sellers to target the right audience. Every business setup has some challenges that need to be faced to sustain but to overcome those issues and face the challenges effectively online marketplace is helping the sellers a lot so that they can expand efficiently. It is a win-win situation for both the marketplace and sellers.

VI. References

- [1]. Devaraj P. "A Hypothesis on Assorted E-Commerce Performances in Online Market" IJECS, Volume 5 Issue 12 Dec. 2016, Page No. 19316-19321
- [2]. Dahiya M. "Study on E-Commerce and it's Impacts on Market and Retailers in India" Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1495-1500
- [3]. Lui S. "An Empirical Study on E-commerce's effects on Economic Growth" ICETMS,2013
- [4]. Shahjee R. "the impact of electronic commerce on business organization" SJIF 2015:5.403 ISSN 2278-8808
- [5]. Kahai R. et al 2000 "Impact of e-commerce on entrepreneurs and small businesses: opportunities, challenges, and strategies" ISSN: 1087-8955
- [6]. <https://www.happiestminds.com/Insights/ecommerce-marketplace/> [Accessed on 10/04/2018]
- [7]. <https://www.export.gov/article?id=Types-of-eCommerce-Websites> [Accessed on 10/04/2018]
- [8]. <https://www.digitalcommerce360.com/2017/09/29/top-5-questions-ask-selling-marketplace/> [Accessed on 12/04/2018]
- [9]. <https://www.marketing91.com/advantage-and-disadvantages-of-selling-on-e-commerce-portals/> [Accessed on 10/04/2018]
- [10]. <http://www.freepatentsonline.com/article/Entrepreneurial-Executive/179817800.html> [Accessed on 08/02/2018]
- [11]. <https://www.tutor2u.net/business/blog/buss4-ecommerce-impact-of-ecommerce-on-functional-areas> [Accessed on 08/02/2018]
- [12]. <http://indianonlineseller.com/2017/08/gst-impact-in-ecommerce/> [Accessed on 05/04/2018]
- [13]. <https://www.digitalcommerce360.com/2017/11/10/amazon-discounts-marketplace-sellers-products/> [Accessed on 05/04/2018]
- [14]. <http://ventureburn.com/2016/03/what-are-the-benefits-of-ecommerce-for-your-business/> [Accessed on 12/06/2018]
- [15]. <https://economictimes.indiatimes.com/small-biz/startups/how-e-commerce-startups-are-helping-small-businesses-go-online/articleshow/51762973.cms> [Accessed on 08/02/2018]
- [16]. <https://www.ibef.org/industry/ecommerce.aspx> [Accessed on 12/06/2018]
- [17]. <https://blog.ipleaders.in/ecommerce-business/> [Accessed on 10/04/2018]<http://www.synaxiom.com/impact-of-ecommerce-in-todays-business-world/> [Accessed on 12/06/2018]



Analysis of Market Strategy of Smartphone Providers

Shreya Chawla¹, Dr. M. Mary Jacintha², Vivek Arya³, Shipra Sharma⁴
School of IT, B-30, Sector 62, CDAC, Noida
India

Abstract: This paper aims to do a detailed analysis of the market strategy adopted by leading smart phone providers in India. This includes analyzing the market strategy of leading smartphone providers for brand positioning, comparing the strategy of leading smartphone providers like with that adopted by Blackberry, analyzing the failure of the premium image of blackberry in India, doing the SWOT & TOWS analysis of blackberry. Blackberry did not succeed in India as the debut of many other competitors in the smartphone category with touch screen phones had changed the scenario. Blackberry was also assailed by launches from Samsung with cheaper versions. This indicates the failure of the market strategy of blackberry in India. Therefore, a need arises to analyze the possible reasons for the failure of premium image of blackberry in India thereby comparing its market strategies with those of the leading smart phone providers in India.

Keywords: smartphone, market strategy, brand positioning, SWOT, Premium Image

I. Introduction

The trend of smartphone market started from the year 2000. After the release of the android operating system Google had taken over the Android OS and released the 1st Android phone with alliance. Till that time market was ruled by different Mobile companies with different operating systems like Symbian, Blackberry & IOS.

The 1st Android Mobile was released in 2003, as an alliance of HTC and Google. After the release of android operating system Google had taken over the Android OS and the major impact started in the market which is making of Android OS announced by Google. As in the development cost of the Operating system is high, Android came into open source and smartphone cost is reduced to half of it. This gave a chance to the leading companies such as HTC, Samsung, Lava, Micromax and other companies to start releasing smartphones with smart applications.

Android is dominating the market from last decade. Its share in market is growing exponentially over the years.. BlackBerry's market share fell from 14.3% in Q4 2011 to 0.6% in Q4 2013. This can be seen in the table given below wherein the sales of the operating systems like Android & BlackBerry have been tracked over the years.

TABLE 1: Sales figures of the operating systems (in millions of units)

Year	Android (Google)	Blackberry OS
2007		11.77
2008		23.15
2009	6.8	34.35
2010	67.22	47.45
2011	219.52	51.54
2012	451.62	34.21
2013	758.72	18.61
2014 Q1	229	3.4
2014 Q2	255	2.1

Source: IDC Reports

2009, BlackBerry was named by Fortune magazine as the fastest growing company in the world, with earnings exploding by 84% a year. Today, BlackBerry has fallen to the back of the smartphone pack — with a minuscule 3% of the market — as Android is dominating the market . Blackberry has fallen leaps and bounds and it becomes the talk point of Industry, why Blackberry is not picked up the innovation bus in the ever growing technology driven market.

Black berry was little bit blurred with their strategy and vision. Firstly they very well in line with usage of corporate customers, but they failed to anticipate the demands of the general consumers. Moreover they slipped away during the time when app based market is booming up. BlackBerry also failed to keep up with the ever growing expectations of the customers in terms of full entertainment through phones, like they were not able to convert their full keypad phone to touch screens ,which ultimately led to their downfall. . Therefore, in this analysis, an attempt has been made to analyse the possible reasons as to why BlackBerry screwed up as well as

why its premium image failed in India. This analysis has been done by comparing the marketing strategies adopted by the various leading smartphone providers in India according to their market share in India, with the market strategy adopted by BlackBerry for its brand positioning.

Objective: The objective of the paper is to analyse the market strategy of various smart phone providers for brand positioning as well as to analyse the failure of the premium image of BlackBerry in India. The paper aims to compare various parameters of future prospects, manufacturing process advertising, export scenario, marketing strategies, technology adopted production policy and government policies for the various smart phone providers. The paper also seeks to analyse how a smart phone provider is tackling the aggressive marketing strategies adopt by its competitors.. This includes comparing the strategy of BlackBerry with those of the other leading smartphone providers & hence providing the SWOT & TOWS.

Scope of Work:

The scope of this analysis includes the following:

- Identifying the leading smart phone providers in India.
- Identifying the market share & Y-O-Y Growth of these leading smartphone providers.
- The market strategies adopted by these smartphone providers for their brand positioning.
- Performing the comparison between the strategies adopted by BlackBerry & these smartphone providers thereby analysing the reasons for the failure of the premium image of BlackBerry in India.
- The SWOT & TOWS analysis of BlackBerry.

II. Theoretical Framework:

Smartphone is a device which is having its primary functioning which includes calling facilities, receiving and sending text messages, surfing internet, communicating with the help of emails and different apps. It also features a camera, GPS Functionality, mobile payments etc.

By the end of forth quarter of 2017 , more than 408 million smartphones were shipped throughout the world. India had overtaken the US market in terms of 2nd largest smartphone market in the world.

As far as manufacturing of smartphones is concerned, China is leading the baggage with more than half of the shipments were rolled out through Chinese land.

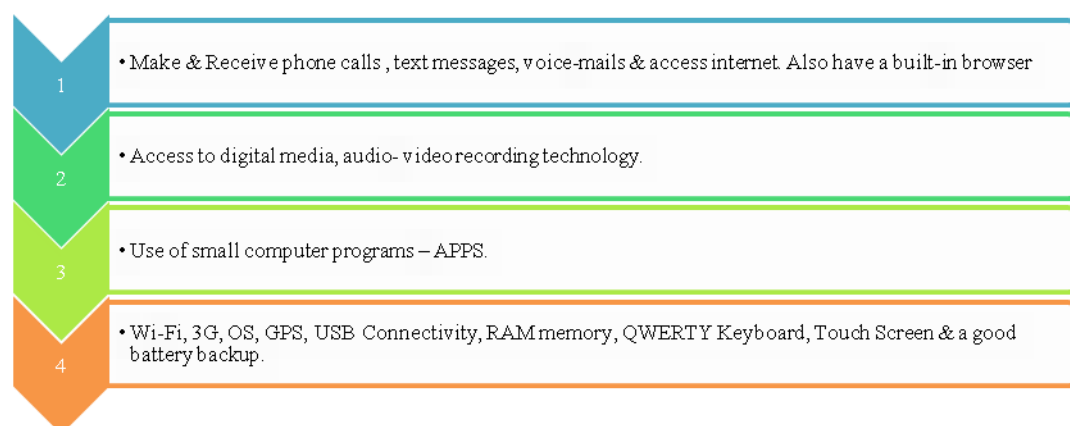
TABLE 2: Worldwide smartphone vendor market share

Source	Date	Samsung	Apple	Huawei	Xiaomi	Oppo	Others
Gartner Q4	2017	18.2	17.9	10.8	6.9	6.3	39.9

Source: Gartner

Features: In today's world smartphones are leading the roost in terms of providing real time communication, digital camera, running the businesses, running household chores or even securing your home just by tapping on apps available on smartphones.

FIGURE 1: Features of a Smartphone



Positioning Errors:

TABLE 3: Various positioning Errors:

Under Positioning	This is a scenario in which the customers have a blurred and unclear idea of the brand.
Over Positioning	This is a scenario in which the customers have too limited an awareness of the brand.
Confused Positioning	This is a scenario in which the customers have a confused opinion of the brand.
Double Positioning	This is a scenario in which customers do not accept the claims of a brand.

Source: Management Study Guide

Market Strategy:

FIGURE 2: 4 Principles of Marketing Strategy



Source: Forbes Magazine

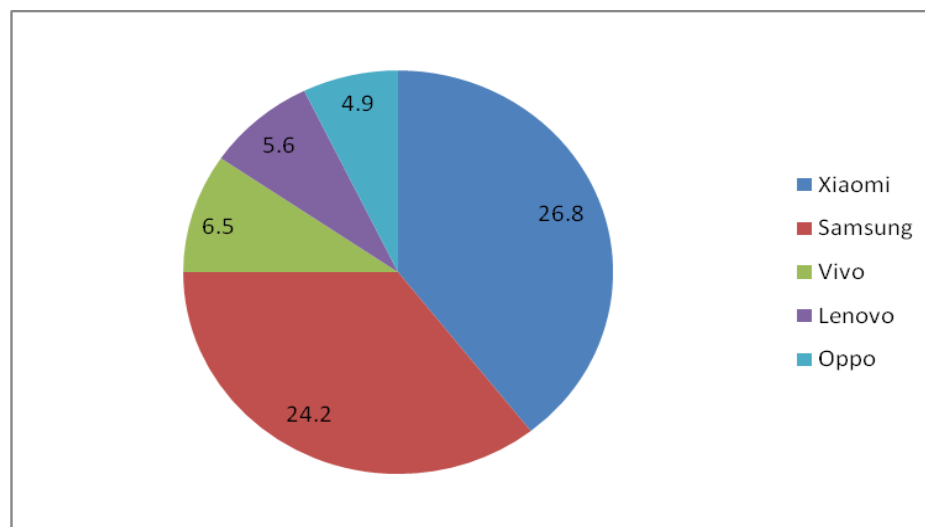
Specific Work Done:

Identifying leading smartphone providers in India

The top five smartphone players in India are as follows:

1. Xiaomi
2. Samsung
3. Vivo
4. Lenovo
5. Oppo

FIGURE 3: Market Share of leading Smartphone providers in India



Source: Q4 2017 report generated by Livemint

Above pie chart is depicting a startling figures as Samsung slipped away from its leader ship position and Xiaomi climbed up to the top spot in Q4, 2017. Other brands like Vivo, Lenove, Oppo are catching up to eat into the market share of its competitors.

Analysis of marketing strategy of smart phone providers

Market strategies adopted

SAMSUNG:

- Samsung launched its smartphone model by name of Galaxy SERIES in year 2010.
- Highest market share in the smartphone market.
- Provides good products with reasonable prices to its customers.
- Spends 5.2% of its revenue earned in research which is even more than the 2.4% spent by Apple..
- Target markets are high & medium income level communities.

- To create brand awareness in India Samsung signed cricket celebrities.
- Samsung also has organized the biggest ever national level school quiz Olympics for school children (2008)
- Recently, on July 23, 2014, Samsung joined hands with Google & Aircel to promote Internet adoption amongst women in India.

MICROMAX:

- Micromax entered into Indian market with its smartphones in 2009.
- Second highest market share in the smartphone market.
- Brand Ambassador is Akshay Kumar.
- Core Competency Strategy is delivering the right value at the right price.
- Has enough reach (1,30,000 retail outlets in 560 districts across India) and spends a lot on advertising.
- Understanding of the Indian Consumer (Price Value Equation: low on price & high on features).
- Target markets are Fashion Savvy Women, Youth & Professionals.
- On the occasion of mother's day, They announced the launch of its android phone - Bling 2.
- Invests heavily in the product development to cater to needs of its consumers.
- Distribution Strategies: B2C, B2B.
- A stunning global ad campaign in 2013 featuring Australian actor Hugh Jackman for Canvas turbo smartphone.
- Micromax simply buys what is available in the market rather than custom-building it like Samsung as it is always relatively cheaper.
- The company tripled its service centres from 436 last year to 1,250 by the year 2014.
- Recently, launched the 3D campaign in the online space across MSN and Yahoo homepages for canvas smartphone series.

KARBONN:

- Karbonn Mobiles came into existence in 2009. It is a joint venture company.
- Target Markets are rural & urban areas.
- It expanded its dealer network and service network very aggressively and they mainly concentrated on 2 tier and 3 tier cities.
- Brand Ambassadors are Gautam Gambhir, Manish Pandey, Virendra Sehwag & Virat Kohli.
- In 2010, Karbonn signed the most lucrative two year sponsorship deal with the Indian Premier League (IPL). Being associated with cricket helped them reach out to the consumer base and enlighten them about their offerings.
- In 2012, Karbonn had signed up as title sponsor of Champions League T20 that helped in creating the right brand visibility.

LAVA:

- Lava International Ltd., came into existence in 2009.
- Lava has its two R&D Centres one is located in India and other one in China
- Lava expanded its imprints in India very aggressively with more than 1000 distributors throughout the country.
- Target Audience is youth, particularly, entry level professionals as well as students.
- Target Market are tier 1 & tier 2 markets as well as rural markets for its feature phones.
- Lava recently has been recognized as one of India's most promising brands by World Consulting and Research Corporation (WCRC)..
- In 2013, LAVA International partnered with the 14th Annual International Indian Film Academy Awards (IIFA), as a 'Powered by' sponsor. LAVA is also the 'Associate Sponsor' of the most respected Indian reality game show Kaun Banega Crorepati (KBC) Season 7, hosted by Mr. Amitabh Bachchan.
- LAVA's premium series of mobile handsets, the IRIS Pro series comes with a promise of 'Art meets Smart' & competes in the niche category of high-end smartphones.
- Recently, Lava has become the official sponsors of English football club Liverpool in India.

MOTOROLA:

- Motorola was the first to enter the Indian market with its DYNATAC 8000X in 1983.

- Motorola was founded by GALVIN BROTHERS and its name was adopted in 1947.

Collaborations with cellular phone service providers:

- In 2004, Motorola optimized BPL India multi-vendor GSM network providing increased coverage, capacity and improved quality, greatly enhancing the end user experience. BPL utilized Motorola's wireless infrastructure multi-vendor GSM RAN optimization service to analyze & improve end-to-end performance of its network.
- In 2005, Motorola introduced a wireless handset that sends email (Motorola Q). It was thinner than the popular Motorola razor and featured Microsoft's Windows mobile software. It helped boost the sales of Motorola.
- Motorola made a come - back in the Indian market with its smartphone MOTO X which was available to the public by October 2013.
- By November 2013, Motorola mobility unveiled its low cost smartphone called MOTO G that saw overnight success in the Indian market because of its value-for-money proposition.
- MOTO G was launched exclusively with Flipkart in India. They decided to go with the only-online retail model to provide the best possible user experience (reaching a big part of India very quickly as well as offering a good service at a competitive price) rather than flooding the market with the product & aimed at taking one step at a time after making a comeback in the Indian market.
- Motorola has set up 130 walk-in service centres within months leading to their India launch to take care of consumer throughout the product life cycle.
- MOTO G has seen the largest sales ever in the history of Motorola. MOTO X has also been very well received.
- Motorola brought itself back in the Indian market through the following strategies:
 - Better design and performance at a lower price.
 - People love change – exploit it.

BLACKBERRY:

- BlackBerry is a line of wireless handheld device developed by the Canadian company, Research In Motion (RIM).
- Target market segment: Business Professionals, Innovators & high income segments.
- BlackBerry stormed the market by focusing on only enterprise sales & getting corporate to push the brand among their employees.

PROS OF BLACKBERRY:

- Super receptivity (no disturbance whatsoever in receiving or making calls).
- Higher level of security.
- BlackBerry Messenger (BBM).

III. Analysis and Interpretation

Comparative analysis

TABLE 4: Comparative Analysis of leading Smart phone providers

Parameters	Features	Samsung galaxy grand prime 	Micromax Canvas Knight A350 	Karbonn Titanium S9 
DISPLAY	PRICE	Rs. 14,024	Rs. 13,299	Rs. 13,766
	TOUCHSCREEN	YES	YES	YES
	TOUCHSCREEN TYPE	CAPACITIVE	CAPACITIVE	CAPACITIVE
	SCREEN SIZE	5.00 mch	5.00 mch	5.50 mch
GENERAL	RESOLUTION	540 X 960 Pixels	1080 X 1920 pixels	720 x 1280 pixels
	BATTERY CAPACITY	2600 mAh	2350 mAh	2600 mAh
	REMOVABLE BATTERY	YES	NO	YES
	WEIGHT	156 (g)	154 (g)	176.5 (g)
SOFTWARE	DIMENSIONS (mm)	144.80 x 72.10 x 8.60	145 x 72 x 9	143.5 x 76.8 x 9.6
	OS	Android 4.4	Android 4.4.2	Android 4.2
	APP MARKET	GOOGLE PLAYSTORE	GOOGLE PLAYSTORE	GOOGLE PLAYSTORE
	JAVA SUPPORT	YES	YES	NO
HARDWARE	RAM	1 GB	2 GB	1 GB
	INTERNAL STORAGE	8 GB	32 GB	16 GB
	EXPANDABLE STORAGE UPTO	32 GB	NO	32 GB
	PROCESSOR	1.2 GHz Quad core	1GHz Octa - core	1.2GHz Quad-core
CAMERA	REAR CAMERA	8 MP	16 MP	13 MP
	FRONT CAMERA	5 MP	8 MP	5 MP
CONNECTIVITY	SIM TYPE	MICRO - SIM	MICRO - SIM, DUAL	REGULAR, DUAL
	GSM / CDMA	GSM	GSM	GSM
	GPS	YES	YES	YES
	BLUETOOTH	YES	YES	YES
	WI-FI	YES	YES	YES
SENSORS	PROXIMITY SENSOR	YES	YES	YES
	ACCELEROMETER	YES	YES	YES
	AMBIENT LIGHT SENSOR	YES	YES	YES

Parameters	Features	Lava Iris Pro 30 	Motorola Moto G 2 nd Generation 	BlackBerry Z 10 
DISPLAY	PRICE	Rs. 12,429	Rs. 12,999	Rs. 14,900
	TOUCHSCREEN	YES	YES	YES
	TOUCHSCREEN TYPE	CAPACITIVE	CAPACITIVE	CAPACITIVE
	SCREEN SIZE	4.70 inch	5.00 inch	4.20 inch
GENERAL	RESOLUTION	720x1280 pixels	720x1280 pixels	768x1280 pixels
	BATTERY CAPACITY	2000 mAh	2070 mAh	1800 mAh
	REMOVABLE BATTERY	NO	NO	YES
	WEIGHT	114.00 (g)	149.00 (g)	136.00 (g)
SOFTWARE	DIMENSIONS(mm)	138 x 66.7 x 7.5	141.50 x 70.70 x 10.99	130.00 x 65.60 x 9.00
	OS	Android 4.2.1	Android 4.4.4	BlackBerry OS 10
	APP MARKET	GOOGLE PLAYSTORE	GOOGLE PLAYSTORE	BLACKBERRY APP WORLD
	JAVA SUPPORT	YES	YES	NO
HARDWARE	RAM	1 GB	1 GB	2 GB
	INTERNAL STORAGE	4 GB	16 GB	16 GB
	EXPANDABLE STORAGE UPTO	32 GB	32 GB	32 GB
	PROCESSOR	1.2GHz Quad-core	1.2GHz Quad-core	1.5GHz Dual-core
CAMERA	REAR CAMERA	8 MP	8 MP	8 MP
	FRONT CAMERA	3 MP	2 MP	2 MP
CONNECTIVITY	SIM TYPE	REGULAR, DUAL	MICRO-SIM, DUAL	REGULAR
	GSM/CDMA	GSM	GSM	GSM
	GPS	YES	YES	YES
	BLUETOOTH	YES	YES	YES
SENSORS	WI-FI	YES	YES	YES
	PROXIMITY SENSOR	YES	YES	YES
	ACCELEROMETER	YES	YES	YES
	AMBIENT LIGHT SENSOR	NO	YES	YES

Source: NDTV Gadgets & MySmartPrice

Interpretation:

It is evident from the comparative analysis that although blackberry is providing exceptionally good hardware functionalities in terms of storage space with 2GB of RAM and 16 GB of internal storage which can be further expandable upto 32 GB & a good connectivity, blackberry is still lagging behind the other smartphones in the following areas:

- Poor display in terms of the small screen size.
- Poor battery backup which provides a talk time of only 8hrs and 20 minutes which is 30mins behind the talk time provided by other smartphones.
- Complex UI and poor customization with no java support and poor application support.
- Slow processing leads to more execution time.
- Low quality camera with no features like bursts and auto focus.

Therefore, the reason as to why blackberry is screwed is the fact that it provides no value for money.

Figure 4. SWOT of Blackberry



IV. Conclusion

The conclusion of this analysis is as follows:

CONS OF BLACKBERRY

- Fixed monthly BlackBerry charges.
- Lack of multimedia capabilities.
- Low quality camera (front camera only 2 – megapixel).
- Poor application support.
- Poor battery back-up.

ANDROID V/S BLACKBERRY OS – FACE OFF!!

TABLE 5: Comparison between ANDROID & BLACKBERRY OS

ANDROID	BLACKBERRY OS
Largest number of devices to choose from	Limited number of devices to choose from- all from RIM
Large number of applications available (250,000)	Limited number of applications available (around 45,000)
Excellent UI	Complex UI
Free developer tools	Application development is somewhat more complex than other OSes

No restrictions on applications	Applications better targeted at corporate environment. Lack of personal apps (eg. games)
Customization is excellent	Customization is poor Multimedia capabilities
Lack of multimedia capabilities There is java support	No java support
Open architecture	Closed architecture
Frequently enhanced	No frequent enhancements
57% 0..of the apps are free	Only 26% of the apps are free. Applications tend to be more costly.
GPS is provided by the Google Maps Navigation	GPS is provided by the 3rd party Apps

Reason for the failure of Blackberry in India

In a market captured by Android's open architecture, it is very difficult for blackberry's closed architecture to survive. The BB OS offers poor customization and a complex UI. Also, the number of applications available on its App Store i.e. BlackBerry App World is limited to 45,000 as compared to Android's 2,80,000.

The biggest failure of blackberry in the Indian market is directly linked to its pathetic performance at the global level (lack of innovation).

The smartphones of blackberry are over-priced especially in comparison to the local smartphone providers like Micromax, Karbonn & Lava.

No value for money.

Lack of interest shown by ardent BlackBerry users in India since many of them had already started using Android.

No touch screen phone by BlackBerry and by the time Z10 entered the market it was already late (by 2013).

Indians get greatly influenced by advertisements and BB lacked in proper marketing and promotion.

The major reason for the failure of its premium image in India was that RIM refused to give security conformance to Indian Government. Every ruling government has the right to have due access to the encrypted data of its nation.

V. REFERENCES

- <http://www.pricedekho.com/mobiles.html>
- <http://forbesindia.com/article/boardroom/can-micromax-become-indias-leading-smartphone-maker/36577/0>
- <http://indianexpress.com/article/technology/mobile-tabs/interview-karbonn-executive-director-shashin-devsare/2/>
- <http://www.dnaindia.com/money/interview-karbonn-strategy-is-to-take-the-smart-phone- to-the-masses-1780359>
- <http://www.telecomlead.com/smart-phone/lava-drafts-smartphone-strategy-to-gain-market-share-in-india-64664-31947>
- <http://timesofindia.indiatimes.com/tech/tech-news/How-Lava-Intex-can-counter-Nokia- Samsung/articleshow/29099420.cms>
- <http://www.lavamobiles.com/aboutus>
- <http://www.slideshare.net/anilkr123/motorola-marketing-strategy>
- <http://www.bgr.in/manufacturers/motorola/heres-why-motorola-has-an-exclusively- online-retail-model-for-the-moto-g-in-india/>
- <http://thecorporatespirituality.com/four-things-learn-motorolas-success/>
- <http://www.tdktech.com/tech-talks/android-vs-ios-vs-blackberry-os>
- <http://www.in.techradar.com/news/phone-and-communications/mobile-phones/Android-vs-BlackBerry-which-is-the-best-smartphone-for-your- business/articleshow/38786087.cms>
- www.gizbot.com
- <http://managementstudyguide.com/brand-positioning.htm>
- <http://www.forbes.com/sites/gregsatell/2013/04/16/4-principles-of-marketing-strategy-in- the-digital-age/>
- <http://gizmodo.com/5812832/ten-reasons-why-blackberry-is-screwed>

ARTICLES/MAGAZINES

- International Data Corporation (IDC) Reports
- Gartner Reports
- Forbes Magazine
- Article on Android in Hindustan Times – by Rajiv Makhni

WEBSITES

- www.mysmartprice.com
- gadgets.ndtv.com
- www.phonearena.com