Proceeding for International Conference

on

Emerging Trends

in

Engineering and Technology

Bengaluru

18th October'15

Institute for Engineering Research and Publication

(A Unit of VVERT)

4A, Girija Apartment, MMDA

Arumbakkam, Chennai-600106, India

www.iferp.in

Publisher: IFERP Explore

Editorial:

We cordially invite you to attend the International Conference on Emerging Trends in Engineering and Technology (ICET-2015), which will be held in Hotel Pai Vaibhav, Bengaluru on October 18, 2015. The main objective of ICET-2015 is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Electrical, Electronics, Mechanical, Civil, Computer Science and Information Technology. This conference provides opportunities for the delegates to exchange new ideas and experience face to face, to establish business or research relations and to find global partners for future collaboration.

These proceedings collect the up-to-date, comprehensive and worldwide state-of-art knowledge on software engineering, computational sciences and computational science application. All accepted papers were subjected to strict peer-reviewing by 2-4 expert referees. The papers have been selected for these proceedings because of their quality and the relevance to the conference. We hope these proceedings will not only provide the readers a broad overview of the latest research results on Electrical, Electronics, Mechanical, Computer Science and Information Technology but also provide the readers a valuable summary and reference in these fields.

The conference is supported by many universities and research institutes. Many professors plaid an important role in the successful holding of the conference, so we would like to take this opportunity to express our sincere gratitude and highest respects to them. They have worked very hard in reviewing papers and making valuable suggestions for the authors to improve their work. We also would like to express our gratitude to the external reviewers, for providing extra help in the review process, and to the authors for contributing their research result to the conference.

Since January 2015, the Organizing Committees have received more than 120 manuscript papers, and the papers cover all the aspects in Electrical, Electronics, Computer Science and Information Technology. Finally, after review, about 10 papers were included to the proceedings of ICET-2015.

We would like to extend our appreciation to all participants in the conference for their great contribution to the success of International Conference 2015. We would like to thank the keynote and individual speakers and all participating authors for their hard work and time. We also sincerely appreciate the work by the technical program committee and all reviewers, whose contributions make this conference possible. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank to organizing committee for their hard work.

Editor-In-Chief
Dr. Nalini Chidambaram
Professor
Bharth University

Acknowledgement

IFERP is hosting the International Conference on Emerging Trends in Engineering and Technology this year in month of October. Technical advantage is the backbone of development and nanoelectronics has become the platform behind all the sustainable growth. International Conference on Emerging Trends in Engineering and Technology will provide a forum for students, professional engineers, academician, scientist engaged in research and development to convene and present their latest scholarly work and application in the industry. The primary goal of the conference is to promote research and developmental activities in Electrical, Electronics, Mechanical, Civil, Computer Science and Software, Information Technology and to promote scientific information interchange between researchers, developers, engineers, students, and practitioners working in and around the world. The aim of the Conference is to provide a platform to the researchers and practitioners from both academia as well as industry to meet the share cutting-edge development in the field.

I express my hearty gratitude to all my Colleagues, staffs, Professors, reviewers and members of organizing committee for their hearty and dedicated support to make this conference successful. I am also thankful to all our delegates for their pain staking effort to travel such a long distance to attain this conference .

Er. R. B. Satpathy
Secretary
Institute for Engineering Research and Publication(IFERP)

CONTENTS

ORGANIZATION COMMITTEE

MANAGING DIRECTOR
Dr. P. C. Srikanth
Head of the Department,
Department of ECE,
Malnad College of Engineering

PRESIDENT:
Dr. P A Vijaya
Professor, Department of ECE,
BNM Institute of Technology

ORGANIZING SECRETARY
Er. Rajesh Babu
Er. A. Siddharth

KEY NOTE SPEAKER
Prof.G P Ramesh
Head of the Department,Department of ECE,
st.Peter's University

PUBLICATIONS COMMITTEE
Dr. Shankar Narayanan
Dr.S. Sangeetha

PROGRAM CHAIR
Dr. Nalini Chidambaram
Professor
Bharth University

An FPGA Implementation of High Speed and Area Efficient Double-Precision Floating Point Multiplier Using Urdhva Tiryagbhyam Technique

[1]Y. Srinivasa Rao,[2]T. Subhashini,[3]Dr. M. Kamaraju
[1]M.Tech Student,Department of ECE, Gudlavalleru Engineering College, Gudlavalleru, India.
[2]Assistant Professor,Department of ECE, Gudlavalleru Engineering College, Gudlavalleru, India
[3]Professor& HoD, Department of ECE, Gudlavalleru Engineering College, Gudlavalleru, India

Abstract- **Floating-point arithmetic is ever-present in computer systems.Every computer language has supports a floating-point data types. Most of the computer compilers call upon floating-point algorithms from time to time for execution of the floating-point arithmetic operations and every operating system must respond virtually for floating-point exceptions such as underflow and overflow. The double-precision floating arithmetic is mainly used in the digital signal processing (filters, FFTs) applications, numerical applications and scientific applications. The double-precision floating arithmetic supports the addition, subtraction, multiplication, division, and square root. Among the all arithmetic operations, multiplication is widely used and most complex arithmetic operation. The double-precision (64-bit) floating point multiplier has a 1-bit sign bit, 11-bits exponent bits and 52-bits mantissa bits.The double-precision floating-point multiplier requires a large 52x52 mantissa multiplication to get the final result. The mantissa multiplication exits as a limit on both area and speed bounds of multiplication operation. The proposed work presents a novel way to reduce this large mantissa multiplication. The Urdhva Tiryagbhyam technique allows using less amount of multiplication hardware compared to the traditional method. In traditional method adding of the partial products are separately done and it takes more time in comparison with the proposed method. In proposed method the partial products are concurrently added with the multiplication operation and it can reduce the time delay. The double-precision floating multiplier is implemented using Verilog HDL with Xilinx ISE tools on Virtex-5 FPGA.**

Key words- **double-precision, floating point, multiplication, Vedic, Urdhva Tiryagbhyam, IEEE-754, FPGA, Virtex-5.**

I. INTRODUCTION

The real numbers[2] are represented in binary format are called as floating point numbers. In IEEE-754 standard, the floating point numbers are divided into two categories, they are binary interchange format and decimal interchange format. The floating point multipliers are very important in scientific computations, digital signal processing(likedigital filters, FFT, image processingetc…) and numerical computations. The IEEE-754 standard[8] defines the single precision (32-bit) and double-precision (64-bit)format.

The notation of a floating-point formats is in fig. 1 and fig. 2.
The fig.1 represents the format of the single-precision floating point number. The single-precision floating point number has 32-bits in length. The total 32-bits are divided into three components, sign (1-bit), exponent (8-bits) and mantissa (23-bits).

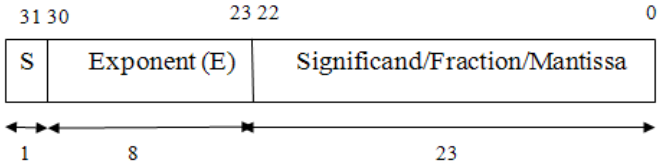


Fig. 1: Representation of single precision floating point number.

The fig. 2 represents the format of the double-precision floating point number. The double-precision floating point number has 64-bits in length. The total 64-bits are divided into three components, sign (1-bit), exponent (11-bits) and mantissa (52-bits).

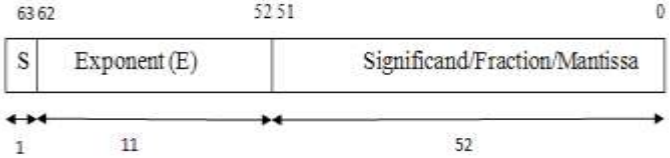


Fig. 2: Representation of Double Precision floating point number

Hardware implementations of IEEE-754 standard floating point arithmetic operations are very important in all

processors. In all the floating point arithmetic operations, the multiplication operation is very important and core operation. The area of the application is main constraint which should be minimum and provides a greater speed. The area efficient implementation of floating point arithmetic operation and thus the efficient implementation of floating point multiplier are of a major concern.

For the past few decades, a lot of work has been dedicated to improve the performance of floating point operations, both at algorithmic level and hardware level. Several works have also focused on designing the FPGA platforms.

FPGA (Field Programmable Gate Array)[7] are preferably used to compute the high performance operations.Theavailable speed, amount of logics and limited available on – board intellectual property (IP)[4] cores make them suitable for large set of applications. They are now used in tremendous applications like numerical and scientific computations, image processing, communications, cryptography computations and digital signal processing applications. In now days, super-computers are designed by using the FPGAs. As a result, this work is primarily aimed for improved implementation of double-precision floating point multiplication on FPGA platform.

The main important and crucial part of the double precision floating point multiplication is mantissa multiplication. This is the main bottleneck of the performance. The mantissa multiplication of double precision floating point numbers is 53-bits in length, and in general this would require a hardware implementation of 53x53 multipliers, which is very cost effective and expensive. In this work, the algorithm is proposed for the multiplication of double precision floating numbers and would be allows to using less amount of multiplication, achieving high speed at a relatively low hardware resources cost. Comparison of results has been done with Karatsuba algorithm[6] and Vedic algorithm.The implementation is mainly concerned only normalized numbers. The implementation has been carried out using Xilinx ISE synthesis tool, ISIM simulator and Virtex-5[4](xc5vlx110t-1ff1136) speed grade-1 FPGAplatform.

In this paper, the mantissa multiplication operations are carried out by using Urdhva Tiryagbhyam in binary. In section 2 represents proposed design approach, section 3 refers Implementation, section 4 refers results, section 5 refers conclusion, and section 6 refers future scope and references.

II. PROPOSED DESIGN APPROACH

The proposed design approach is taken from the Karatsuba MultiplicationTechnique. These technique is dividethe both 53-bits mantissas (including one hidden bit) into three parts is in fig. 3.

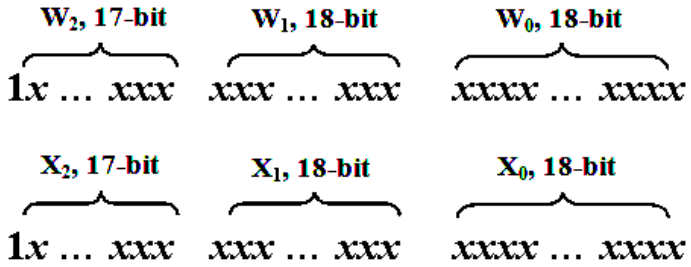


Fig. 3 53-bit mantissa is splitted into 17-bit and 18-bit fractions

The 53-bit mantissa[1] is splitted into 17-bit and 18-bit signed fractions. The terms w_0 and w_1 represents the 18-bit fraction and w_2 represents the 17-bit fraction of the one of the input operand. In the same way x_0 and $x_1$represents the 18-bit fractions and x_2 represents 17-bit fraction.

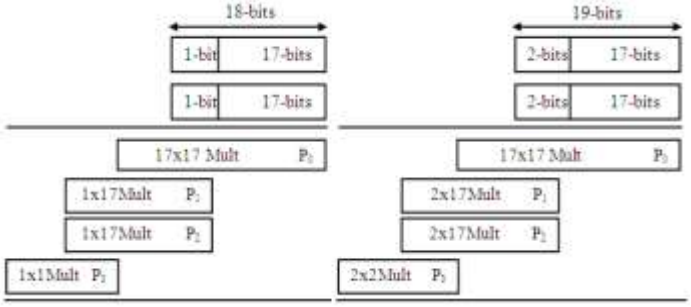


Fig.418-bit and 19-bit Multipliers

The 18-bit and 19-bit multipliers are used to design 53x53-bit mantissa multiplication. The fig. 4 represents the 18-bit and 19-bit multipliers. The 18-bit multiplier has 1-bit sign bit and 17-bits fractional bits, the first step is to multiply two 17-bits of the input operands and generate a partial product p_0 then generate a p_1,p_2, and p_3. In the same way 19-bit multiplier is also design.

Vedic multiplier is designed based on Vedic multiplication[5] formulae (sutra). The sutras are widely used for the multiplication of two large decimal numbers. This same sutra can be used to perform multiplication[5] on two large binary numbers and it is in fig. 5.. This proposed method is compatible to design digital hardware system. The Urdhva Tiryagbhyam technique[4] can take (2n-1) steps for designing the n-bit multiplier. In fig. 5 4-bit multiplier can take 7-steps for designing of the multiplier.

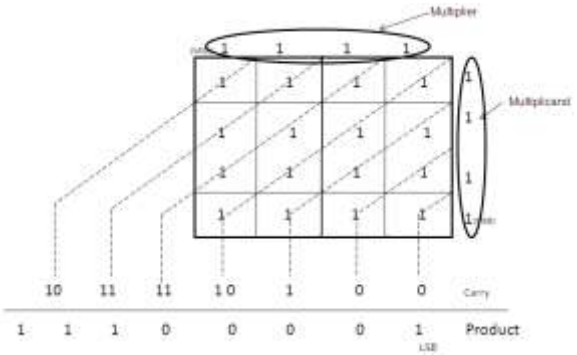


Fig.5Multiplication steps of 4-bit binary numbers using Urdhva Tiryagbhyam Sutra

The 4-bit binary multiplier using Urdhva Tiryagbhyam technique is in fig. 5. The 4-bit multiplier has 4-bit multiplier as one of the operand and 4-bit multiplicand is the another operand. The carry resents the previous state generated carry and it can be added to the current to get the final product.

III. IMPLEMENTATION

Designing of floating-point multiplication carries the operation on sign, exponent and mantissa separately.

A. Sign and Exponent operation

The sign and exponent operations are performed in a straightforward manner. The both operands of sign-bits are performed on logical XOR operation.

$$Sign\_out = Sign\_in1 \oplus Sign\_in2$$

The output exponent is addition of both input exponents and then subtracting the BASE i.e.

$$Exp\_out = Exp\_in1 + Exp\_in2 - 1023$$

For double-precision floating numbers, the BASE is equals to $1023(2^{11-1}-1)$.

The BASE of the any floating point number is determined by $(2^{\text{exponent bits}-1}-1)$.

B. Exceptional Case Handling

The IEEE standard as defined by the many exceptional cases like NaN (Not a Number), INFINITE, ZERO, UNDERFLOW, OVERFLOW. These are appearing for all the floating point arithmetic operations. Thus, the main operation has been also combined with the detection of all the exceptional cases, and determining the final output as per standard. All the exceptional cases executions are in parallel with the IEEE-754 standard.

The fig. 6 represents the flowchart of the handling of exceptional cases.

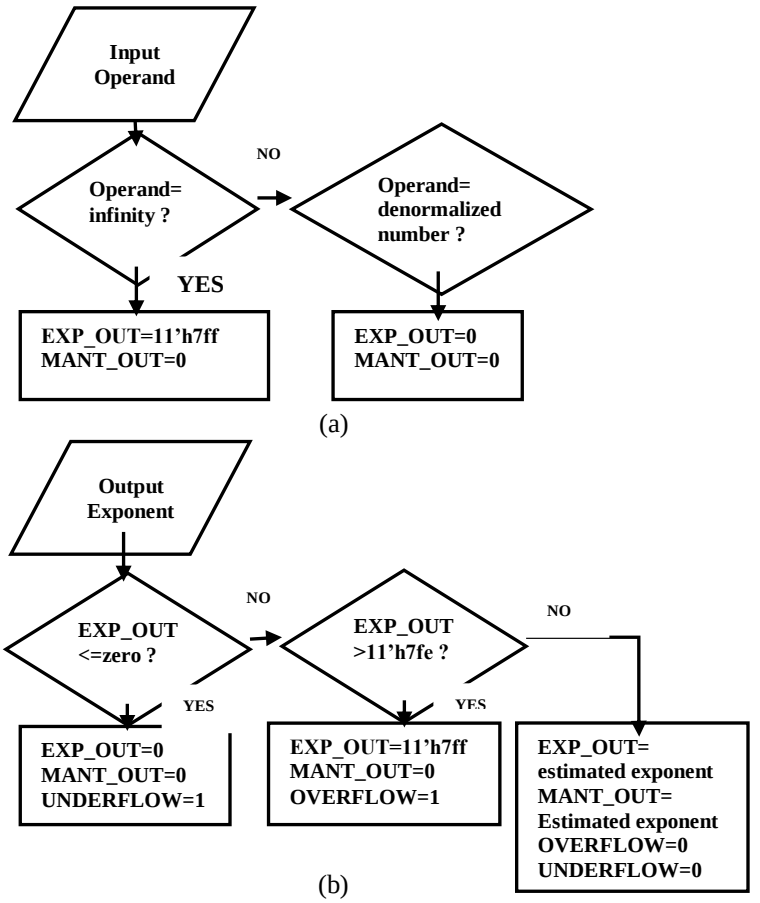


Fig. 6 Flowchart for handling of exceptional cases (a) infinite case (b) OVERFLOW and UNDERFLOW cases

If any/both of the operands are infinity; then produce a INFINITY as output (with computed sign-bit). If any one of the operand is denormalized number, then the output will be zero (w.r.to sign-bit). If the output exponent is zero or below zero, UNDERFLOW will show, and the exponent is above 11'h7fe (2046 in decimal), OVERFLOW will show.

C. Normalization and Rounding

The normalization is very important to get the final result in 64-bit format i.e. sign-exponent-mantissa. Often the mantissa multiplication, the resultant has one extra bit in the MSB before decimal point. Similarly, in sometimes the same situation will arise after rounding. These results need to be fixed to get the mandatory formatting of the result. So, whenever the extra carry bit is generated after multiplication/rounding, the product is right-shifted for necessary shifts and the corresponding changes shouldare made in exponent also to get the normalized result.

Rounding is necessary to get back the 106-bit mantissa multiplication result to 53-bit result only. In this paper we have only implementedRound to nearest rounding mode specified by the IEEE-754 standard. The remaining rounding modes also used depending on requirement.

IV. RESULTS

The FPGA implemenatation of double precision floating point multiplier using Urdhva Tiryagbhyam techniqueis divided into 18-bit multiplier, 19-bit multiplier and 53-bit mantissa multiplier. The 53-bit multiplier is designed by combining the 18-bit and 19-bit multipliers. The 18-bit multiplier, 19-bit multiplier and 53-bit mantissa multipliers are in followed figures 7-18.

Fig. 7 Simulation result of 18-bit Urdhva Tiryagbhyam multiplier

The simulation result of 18-bit multiplier using Urdhva Tiryagbhyam multiplication technique is in fig. 7.Here Ainand Bin are the two input signals.In this Ain is one of the operand which is 18-bits in length and Bin is other operand which is also an 18-bits in length. These two operands are performing a multiplication operation and produce a 36-bit of result.
INPUTS: Ain[17:0]=18'h00256; Bin[17:0]=18'h00100;
OUTPUT: Y[35:0]=36'h000025600;

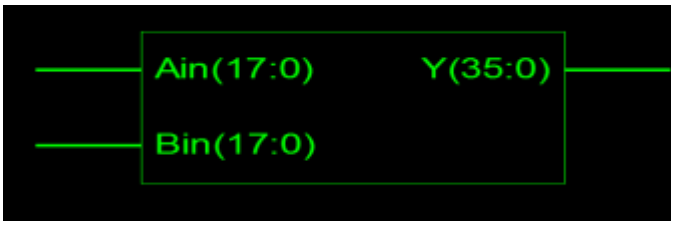


Fig. 8 RTL Schematic of 18-bit Urdhva Tiryagbhyam multiplier

The fig. 8 represents the RTL Schematic of 18-bit multiplier using Urdhva Tiryagbhyam multiplication technique. In this Ain and Bin are the input operand each one has the 18-bits in length and output Y is 36-bits in length.

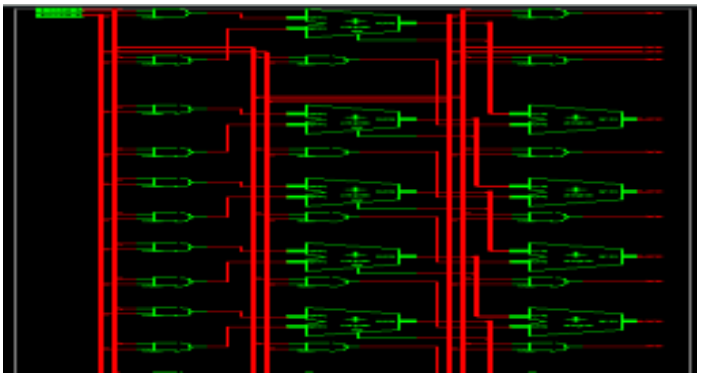


Fig. 9 Technology schematic of 18-bit Urdhava Tiryagbhyam multiplier

The technology schematic of 18-bit multiplier is in fig. 9 has represents the number of AND gates and several adders and these components are interconnecting to construct the 18-bit multiplier.

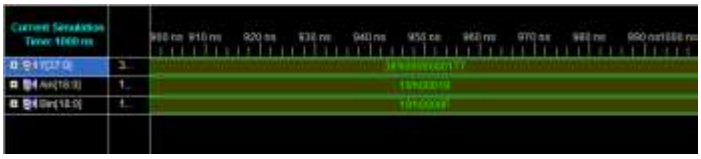


Fig. 10 Simulation result of 19-bit Urdhva Tiryagbhyam multiplier

The simulation results of 19-bit multiplier using Urdhva Tiryagbhyam multiplication technique is in fig. 10. Ain and Bin are the two input operand of the 19-bit multiplier and Y is the output produced by the multiplier.
INPUTS: Ain[18:0]=19'h00019; Bin[18:0]=19'h0000F;
OUTPUT: Y[37:0]=38'h0000000177;

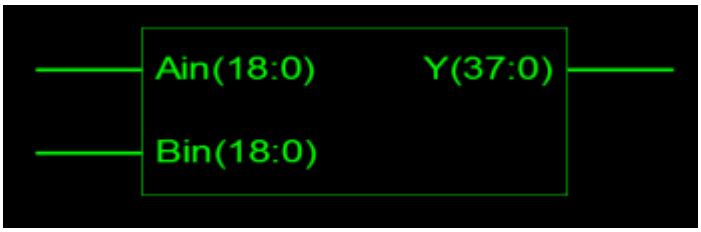


Fig. 11 RTL Schematic of 19-bit Urdhva Tiryagbhyam multiplier

The fig. 11 is the RTL Schematic of 19-bit multiplier using Urdhva Tiryagbhyam multiplication technique. Ain and Bin are two operands which are give to input to the multiplier and output of the multiplier displays as Y which is 38-bits in length.

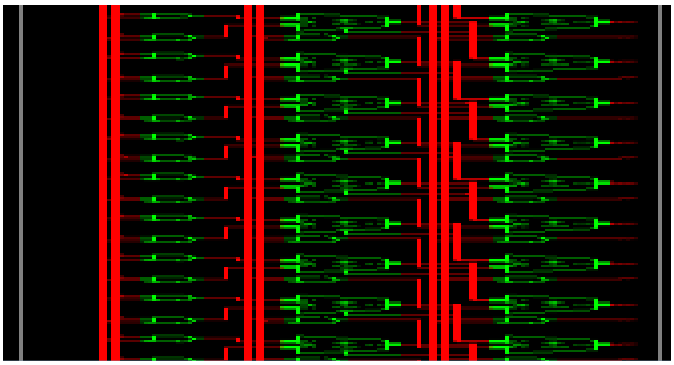


Fig. 12 Technology Schematic of 19-bit Urdhva Tiryagbhyam multiplier

The fig. 12 shows the technology schematic of the 19-bit multiplier using Urdhva Tiryagbhyam sutra. All the components are internally connected via the red color wire. This is also known as interconnection diagram of the 19-bit multiplier.

Fig. 13 Simulation result of 53-bits mantissa multiplier

Simulation result of 53x53-bit multiplier is in fig. 13. In this Ain and Bin are the inputs of the multiplier; each one has 53-bits in length and output Y has 106-bits in length.
INPUTS:Ain[52:0]=53'h01234589756478;
Bin[52:0]=53'h05689745689523;
OUTPUT:
Y[105:0]=106'h00000000000000000005B36392F6;

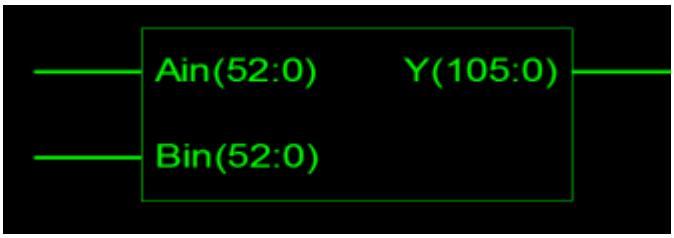


Fig. 14 RTL Schematic of 53-bit mantissa multiplier

The mantissa bits of the double-precision floating point multiplier have 53-bits. The 53-bit x 53-bit multiplier design is very difficult, the 53-bits are divided into 18-bits and 19-bits and design the multipliers. These multipliers are added to get the 53-bit mantissa multiplier.The fig.14 is the RTL Schematic of 53x53-bits multiplier using Urdhva Tiryagbhyam multiplication technique. Here Ain and Bin are the input operands of the 53-bit multiplier; Y is the output of the multiplier.

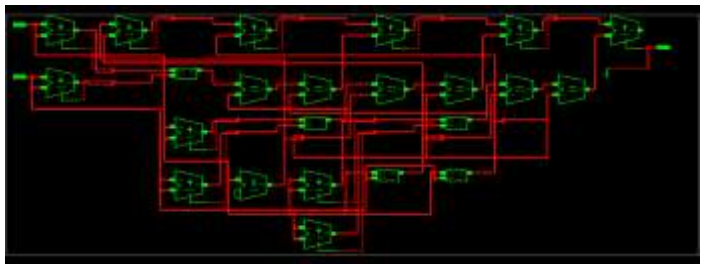


Fig. 15 Technology Schematic of 53-bit mantissa multiplier

The fig. 15 shows the technology schematic of the mantissa bits multiplier. In this the red color wires are interconnecting wiring network of the circuit. The technology schematic has adders, subtractors and multipliers. These components internally through wiring network to build the actual circuit.

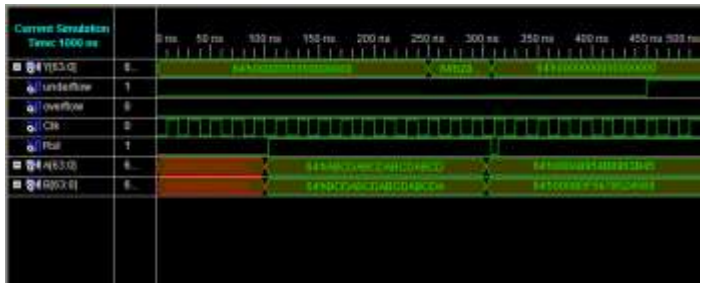


Fig. 16 Simulation result of double-precision Urdhva Tiryagbhyam multiplier

The resulting waveform of the double-precision floating point multiplier using Urdhva Tiryagbhyam sutra has two inputs and each one has 64-bits wide. The two 64-bit input operands are performed multiplication operation; the resultant has 64-bits length.The fig. 16 shows the simulation result waveform of double-precision floating point multiplier using Urdhva Tiryagbhyam sutra. In this the exceptional conditions like overflow and underflow are checked by giving the inputs are A[63:0]=000a8954b8963b45;B[63:0]=0008df5678524569; and the output will display as all zeros because the exceptional case is occur. The exceptional case is resultant exponent is less than or equal to zero.

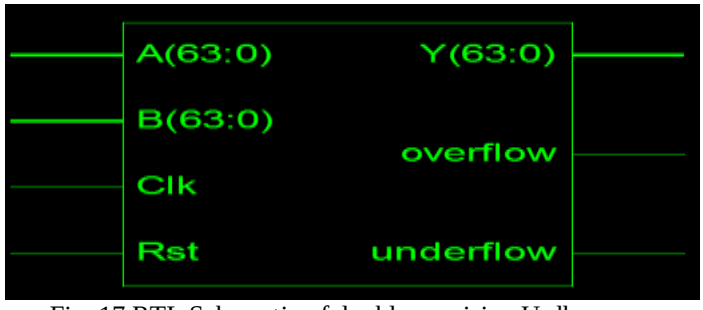


Fig. 17 RTL Schematic of double-precision Urdhva Tiryagbhyam multiplier

The RTL Schematic of double-precision Urdhva Tiryagbhyam multiplier is shown in fig. 17. The inputs are A[63:0], B[63:0], clock and reset and outputs areY[63:0],

underflow and overflow. A[63:0] has 64-bit of floating point operand and B[63:0] has 64-bit of another floating point operand. These two operands are applied to inputs of floating point multiplier. The Y[63:0] is 64-bit floating point multiplier output. Overflow and underflow are exceptional conditions occur during the floating point multiplication.

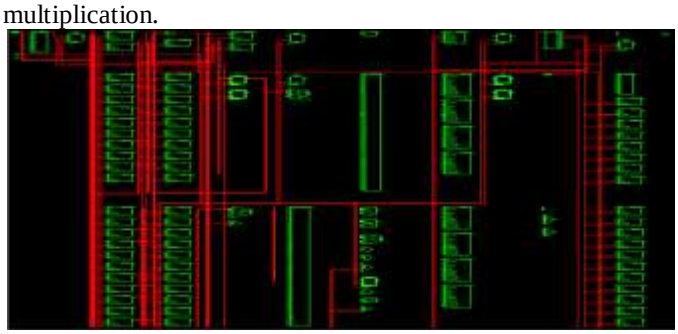


Fig. 18 Technology Schematic of double-precision Urdhva Tiryagbhyam multiplier

The Technology schematic is in fig. 18 represents the how the components are internally connected. It also represents the circuits are actually arranged in the FPGA.

Table 1: Device utilization summary of double-precision floating point multiplier using Urdhva Tiryagbhyam technique

Device	**No. of. Slice Registers**		**No. of. Slice LUTs**	
	Karatsuba Multiplier	Urdhva Tiryagbhyam Multiplier	Karatsuba Multiplier	Urdhva Tiryagbhyam Multiplier
Spartan-2 (xc2s15-6cs144)	1877	1970	3599	3607
Spartan-2 (xc2s200-6fg256)	1878	1970	3599	3607
Spartan-2e (xc2s600e-7fg676)	1933	1969	3599	3607
Spartan-3 (xc3s4000l-4fg900)	797	504	1384	900
Spartan-3a (xc3s1400a-5fg676)	798	526	1378	936
Spartan-3e (xc3s1600e-5fg484)	798	525	1378	936
Virtex-2 (xc2v6000-6ff1517)	796	506	1384	902
Virtex-2p (xc2vpx70-7ff1704)	796	506	1384	902
Virtex-4 (xc4vlx200-11ff1513)	798	441	1384	775
Virtex-5(xc5vlx110t-1ff1136)	390	373	1456	617

Table 1 represents the Device utilization summary of double-precision floating point multiplier using Urdhva Tiryagbhyam technique. In this the comparision parameters are number slice registers and number of slice LUTs. These two parameters are comparision with the Karatsuba multiplier and Urdhva Tiryagbhyam multiplier. The devices are changed from Spartan to Virtex the number slice registers used by the multipliers are reduced and also number of slice LUTs are reduced. In this the designinig of the Urdhva Tiryagbhyam multiplier requires minimum amount of logic resources to perform double-precision floating point multiplication operations.

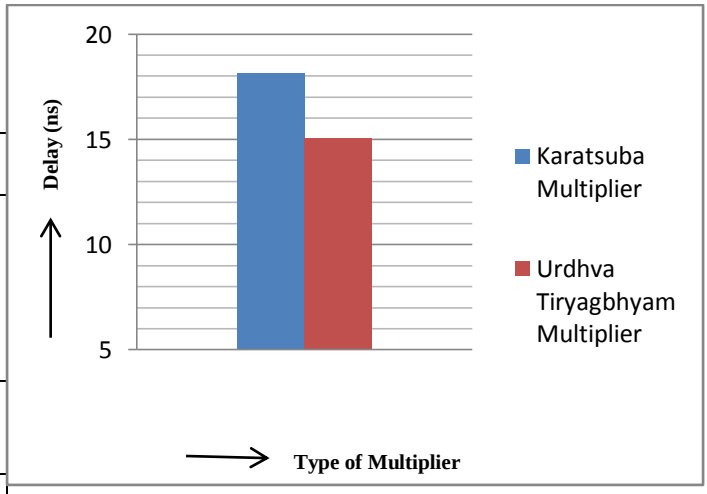


Fig. 19 Delay representation of Karatsuba and Urdhva Tiryagbhyam multipliers

The fig. 19 represents the delay representation of Karatsuba and Urdhva Tiryagbhyam multipliers. The delay of the both the multipliers are measured in nano seconds. The delay of the Karatsuba multiplier is 18.139ns and delay of the Urdhva Tiryagbhyam multiplier is 15.034ns. The delay of the Urdhva Tiryagbhyam multiplier is less compared to the Karatsuaba multiplier so Urdhva Tiryagbhyam multiplier is fast in comparison with the Karatsuba multiplier.

The fig. 20 represents the Power consumption of the Karatsuba multiplier and Urdhva Tiryagbhyam multipliers. In this X-axis takes frequency in MHzs and Y-axis takes power in terms of watts. The frequency is from 10MHz to 600MHz and note down the power at the

frequencies. In this fig. 20 observes the Karatsuba multiplier consumes more power in comparison with the Urdhva Tiryagbhyam multiplier.

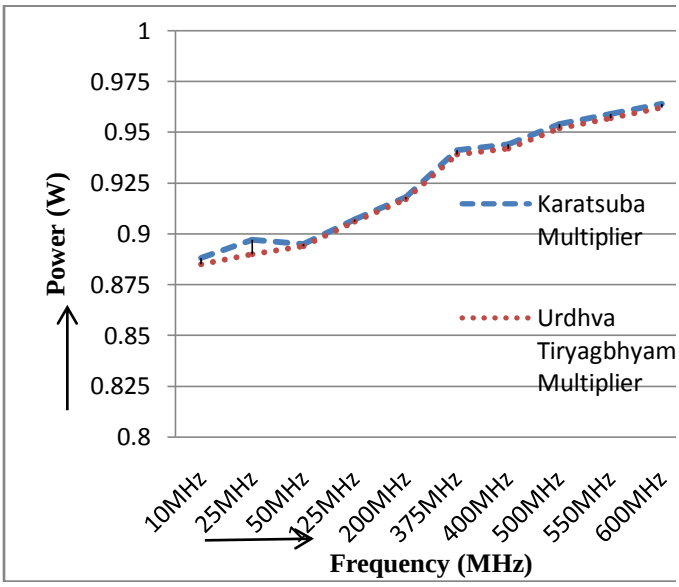


Fig. 20 Power consumption of Karatsuba and Urdhva Tiryagbhyam multipliers

CONCLUSION

The FPGA implementation of double-precision floating point multiplier using Urdhva Tiryagbhyam technique achieved the high speed and area efficient compared to the Karatsuba multiplier and it is fully compatible with the binary interchange format of IEEE-754 standards.The design handles the various exceptional conditions like OVERFLOW and UNDERFLOW .

FUTURE SCOPE

The double precision floating multiplier using Urdhva Tiryagbhyam technique can be implemented on ASIC flow to perform processor specified operations like square root, trigonometric functions.

REFERENCES

[1] Manish Kumar Jaiswal, Ray C.C. Cheung"VLSI Implementation of Double-Precision Floating Multiplier Using Karatsuba Technique",Circuits syst signal process, Springer science business media, vol. 32, pp. 15-27, July 2013.

[2] Purna Ramesh addanki, Venkata Nagaratna Tilak Alapati and Mallikarjuna Prasad Avana " An FPGA Based High Speed IEEE-754 Double Precision Floating Point Adder'/Subtractor and Multiplier Using Verilog" International Journal of Advanced Science and Technology, vol. 52, pp. 61-74, March 2013.

[3] M. al-Ashrafy, A. Salem, W. Anis, " An Efficient Implementation of Floating Point Multiplier", Saudi International Electronics, Communications and Photonics Conference(SIECPC), pp. 1-5, April 2011.

[4] Sandesh S. Saokar, R. M. Banakar, and Saroja Siddamal, "High Speed Signed Multiplier for Digital Signal Processing Applications", IEEE International Conference on Signal Processing, Computing and Control (ISPCC), pp. 1-6, March2012.

[5] Devika Jaina, Kabiraj Sethi, and Rutupama Pamda, "Vedic Mathematics Based Multiply Accumulate Unit ", International conference on computational intelligence and communication system, pp 754-757, 2011.

[6] Deena Dayalan, S.Deborah Priya, "High Speed Energy Efficient ALU Design using Vedic Multiplication Techniques", ACTEA IEEE pp 600-603, July 2009.

[7] P. Belanovic and M. Leeser, "A Library of Parameterized Floating-Point Modules and Their Use", in 12th International Conference on Field-Programmable Logic and Applications (FPL-02). London,UK: Springer Verilog, pp. 657-666, September 2002.

[8] Xilinx, Xilinx floating-point IP core. http://www.xilinx.com

Disaster Prediction using Back Propagation Algorithm Method

PrisillaJayanthi G

Abstract- **Every time a disaster occurs, development gains are lost. Disaster put development at risk and development can create and accumulate disaster risk. The Emergency Management Service evolves huge data transmission over the network.In this survey the datatransmission is carried over cloud computing networks.The clouds are hardware-based services which computes, network and storage capacity where hardware management is highly preoccupied and Infrastructure capacity is highly elastic.The cloud computing is a specialized form of grid and distributed computing which varies in terms of infrastructure, services, deployment and geographic dispersion.In a cloud, the security of the data is the biggest concern. The principle in back propagation algorithm is considered to fashion the minimum error in predicting the early warning of the floods. Prediction is a wonder of knowing what may happen to a process in the next coming eras.**

Key words- **Cloud Computing, Back Propagation Algorithm, Automatic Weather Station , Radio Sonde**

I. INTRODUCTION

Disaster Management aims to avoid lessens or transfers the adverse effects of hazards through activities and measures for prevention, mitigation and preparedness. Increased frequency in precipitation will increase the risk of floods and landslides, with consequence on agriculture, settlement and transport. Increased in overall temperature causes glaciers melting which may trigger lake outbursts and flooding in settlement. Disasters affect the countries in different ways, the effect of lower –middle income countries is much high when compared to the high income countries. The low income countries face huge death loss and financial loss whereas the high income countries have fewer death ratesbut financial loss is found to be more.

The cloud computing isn't network computing. In network computing, applications arehosted on a single enterprise's server and accessed over the enterprise's network. Cloud computing is a better-quality than that. It encompasses multiple companies, multiple servers, and multiple networks. Unlikenetwork computing, cloud services and storage are accessible from anywhere in the world over an Internetconnection; whereas with network computing, access is over the enterprise's network only. Both data in transit and at rest mustbe secure is a primary goal of any organization in Cloud computing.

The back tracking algorithm, the iterative process that begins with the last layer and moves backwards through the layers until the first layer is reached. Assume that for each layer, the error in the output of the layer is known. If the error of the output is known, then it is not hard to calculate changes for the weights, so as to reduce that error. The problem is that the error in the output of the very last layer only can be observed.

Back propagation helps to determine the error in the output of a prior layer by giving the output of a current layer as feedback. Then calculate the error in the output of the prior layer.Back propagation requires that the activation function used by the neurons be differentiable. With the increase number of hidden neurons, the number of independent variable of the error function also increases and the computing time also increases rapidly.

II. CLOUD COMPUTING

Cloud is an eco-friendly green technology as it reduces the purpose for the enterprises to maintain multiple servers which occupies lot of space and infrastructure. Cloud compingmakes use of shared services and resources as a form of distributed computing over the network. Cloud providers read fixed number of packets and perform timing calculation and store in databases placed in remote areas. In a cloud based computing, the resources are presentin someother network and accessed remotely by the cloud users. Multiple users can access a single server to retrieve and for data updatingfor different applications with cloud computing.Cloud computing and associated technologies like SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) have been initiating a lot of concern in changing technology. Lower costs and rapid scalability are few benefits provided by cloud computing by moving heterogeneous data into cloud

computing environment.The aim of using cloud based high performance computing is to deal with faster processing of huge amount of data and higher throughputs.

The Basic Code of Backpropagation algorithm

Step 1 :All the network inputs are assigned and generate output
Step 2: The weights are initialized with small random numbers, between -1 and 1
Step 3: Repeat until for all patterns in the network
// **the inputs are propagated forward through the network**
Each node in the layerin the network
1. Sum of weights of inputs to the nodeis computed
2. Add the threshold to the sum
3. Compute the activation function for the node
// **the errors are propagated backward through the network**
for every node in the output layer
calculate the error signal
Step 4: For all hidden layers
for each node in the layer
1. Compute the node's signal error
2. Update each node's weight in the network

Each input signal to the unit is associated with a weight. The weight vector $w = \{w_1, w_2, \ldots, w_n\}$, input vector $x = \{x_1, x_2, \ldots, x_n\}$, the input information is processed by calculating the weighted sum of the inputs. The neuron will fire if the summation exceeds a pre-set threshold value.The firing of any neuron is ruled by its activation function f(v).
The two learning methods used in neural network: **supervised** and **unsupervised** learning. In **supervised learning**, the concept of a teacher exists. The intent of the teacher is to intimate the network when its output is incorrect by showing it what the output should have been. In **unsupervised learning**, there is no teacher and therefore no rules by which the network can learn. The network must learn by inspecting the input data and discovers any essential properties that exist.

When the patternis fed to the input layer, the weighted sum of the input to the j^{th} node in the hidden layer is given by
$\mathbf{Net_j = \sum w_{i,j} x_j + \theta_j}$----------(1)This equation is used to calculate the aggregate input to the neuron. The $\theta_j$term is the weighted valuewith a thresholdθ (If the net input is greater than the threshold θ, then the output unit is turned on, otherwise it is turned off).

Back Propagation --Error Calculations and Weight Adjustments **1. Output Layer**
If the actual activation value of the output node, k, is O_k, and the expected target output for node k is T_k, the difference between the actual output and the expected output is given by:

$$\Delta_k = T_k - O_k$$

----------(2)The error signal for node k in the output layer can be calculated as

$$\delta_k = \Delta_k O_k (1 - O_k)$$ ----------(3)or $$\delta_k = (T_k - O_k) O_k (1 - O_k)$$------(4)where $O_k(1-O_k)$ is the derivative of the Sigmoid function. The change in the weight connecting input node j and output node k is proportional to the error at node k multiplied by the activation of node j. The formulas used to modify the weight, $w_{j,k}$, between the output node, k, and the node, j is:

$$\Delta w_{j,k} = l_r \delta_k x_k$$----------(5) $$w_{j,k} = w_{j,k} + w_{j,k}$$----------(6)

where $\Delta w_{j,k}$ is the change in the weight between nodes j and k, l_r is the **learning rate.**To updating the weights, a modification to equation (5) is made:

$$\Delta w_{j,k}^{n} = l_r \delta_k x_k + \Delta w_{j,k}^{(n-1)} \mu$$ ----------**(7)**

2. Hidden Layer
The error signal for node j in the hidden layer can be calculated as

$$\delta_k = (T_k - O_k) O_k \sum (w_{j,k} \delta_k)$$ ----------**(8)**where the sum term adds the weighted error signal for all nodes, k, in the output layer. The formula to adjust the weight, $w_{i,j}$, between the input node, i, and the node, j is:

$$\Delta w_{i,j}^{n} = l_r \delta_j x_j + \Delta w_{i,j}^{(n-1)} \mu$$

$$w_{i,j} = w_{i,j} + \Delta w_{i,j}$$

Observation Data Captured by weather Station The dataset is gathered from well -known meteorological department after the thorough survey from Pune [3]. The department takes the Global observational data such as wind, humidity, temperature at different levels from the Automatic Weather station, Radio sonde, Satellites etc. The figure 1 shows 'Nanauk' Cyclonic Storm over the Arabian Sea during 10th– 14th June 2014 taken by INSAT 3D satellite.

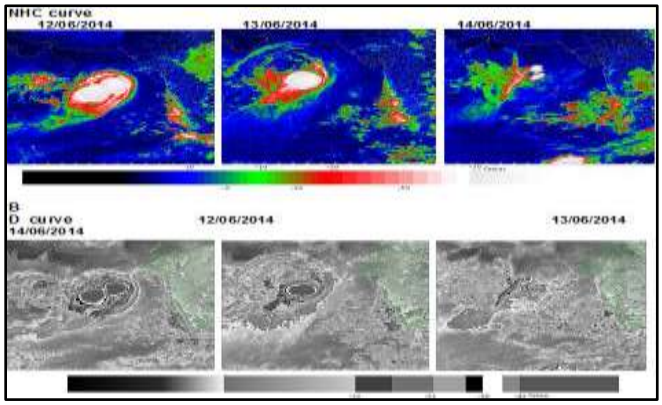


Fig. 1: INSAT-3D Enhancement curves at 0300 UTC of 12th, 13th and 14th June 2014.

INSAT-3D IR (4Km) resolution images for the CS 'Nanauk' (10^{th}–14^{th}June) which formed over east Arabian Sea of the monsoon current aided the advance of Arabian branch up to south Gujarat coast. The system was well monitored with satellite observations, supported by meteorological buoys and coastal and Island observations.

A **radiosonde**, a small weather station coupled with a radio transmitter; is attached to a helium or hydrogen-filled balloon, also known as a weather balloon, and the balloon lifts the radiosonde to altitudes exceeding 115,000 feet. It transmits data on temperature, pressure, and humidity to a sea or land-based receiving station whiles the radiosonde's ascents. Usually the position of the radiosonde is tracked through Global Position System or radar to provide data on the strength and direction of winds aloft. The radiosonde flight produces a vertical profile of weather parameters in the area above which it was launched.

An **Automatic weather station** (AWS) is an automated weather station, which enables measurements from remote areas orto save human labor. TheAWS have Thermometer for measuring temperature, Anemometer for measuring wind speed, Hygrometer for measuring humidity, Barometer for measuring pressure, rain gauge for measuring rainfall, ceilometers for measuring cloud height, present weather sensor or visibility sensor. AWS needed to be located where there was an availability of electricity and communication lines. In wireless technology, applications of the solar panel, wind turbine and mobile phone technology have made it possible to have wireless AWSs that are not connected to the electrical grid or telecommunications network.

'Nanauk' Cyclonic Storm over the Arabian Sea during 10th– 14th June 2014

Over the east-central Arabian Sea on 9th June, 2014 a Cyclonic Storm (CS) 'NANAUK' that developed from a low pressure area moved north-westwards, it intensified into a Cyclonic Storm on 11th June 2014. Continuingits movement north-north-westwards, it weakened into a low pressure area over northwest Arabian Sea in the morning of 14th June, 2014. The wind speed and the daily rainfall recorded by AWS in the state of Maharashtra is analyzed and shown in Fig.2 and Fig.3 respectively. The rise of wind speed is seen during 10-11 June 2014.

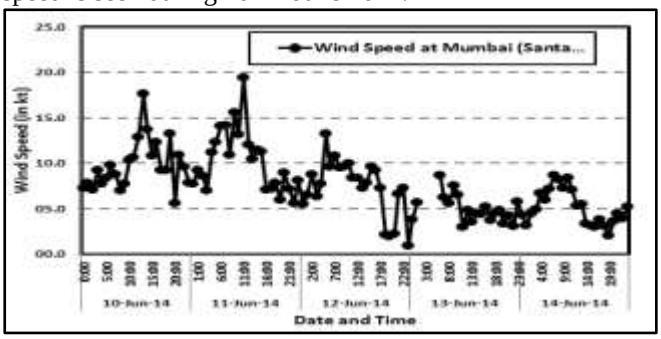


Fig 2 : Variation of Wind Speed at MumbaiSantaCruz station during 10-14,Jun 2014.

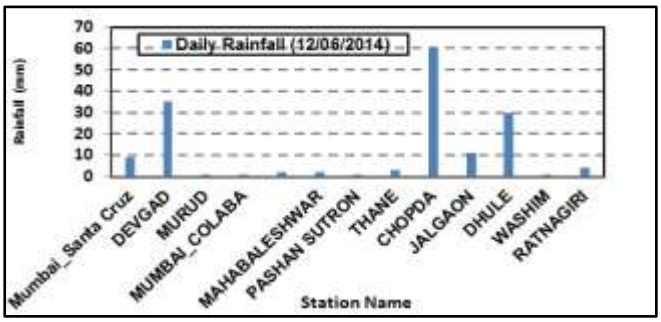


Fig 3: Daily Rainfall recorded at AWSs in Maharashtra under influence of NanaukCycloneon 12th June 2014

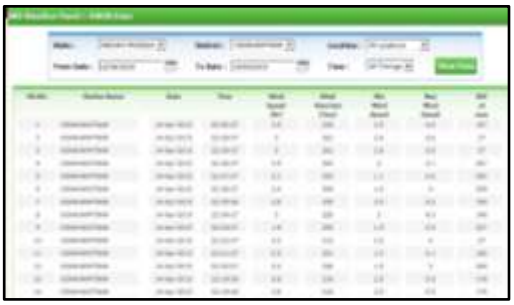


Fig 5: High Wind Speed Recorder data displays information related to wind speed, wind directionat a particular time.

Yet the forecast for the monsoon weather 2014 showed few errors in wind components hence thebias correction in the models can improve the range of predictability and increase the accuracy of the forecasts. By applying the bias correction in the models systematic errors can be removed or minimized. Bias correction has most impact at lower levels and at longer forecast periods.

CONCLUSIONS

The output received by AWS is passed to back propagation algorithm to reduce errors and hencepredict the disaster with the minimal errors and forecast the weather in timely manner and with accurate correctness. The coordination of the data transmission over the cloud computing can save the human live and economical loss. The work is in progress and shows possible directions for future research.

REFERENCES

1. Monjur Ahmed and Mohammad Ashraf Hossain (2014), "Cloud computing and security issues in cloud", International Journal of Network Security & Its Applications Vol 6(1), 25-36.
2. JiaqiDuan ,Parwiz Faker ,Alexander Fesak and Tim Stuart "Benefits and drawbacks of Cloud-based Versus traditional ERP systems", Proceedings of the 2012-13course on Advanced Resource PlanningW.J.H. van Groenendaal
3. Monsoon_report_2014 from meteorological department ,Pune
4. Shaminder Singh, Pankaj Bhambri and Jasmeen Gill (2011), "Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique", International Journal of Computer Science Issues, Vol. 8(5).
5. Michael Negnevitsky. Artificial intelligence: a guide to intelligent systems. Pearson Education Limited, 2005.
6. Raul Rojas. Neural Networks:A Systematic Introduction. Springer, 2005.z

Pseudo-Anonymization of Social Networks By Sequential Clustering and Classification

[1]S. Tulasi Krishna,[2]M. Vijaya Bharathi
GMR Institute of Technology
[1]Tulasi.sampangi@gmail.com,[2]vijayabharathi116@gmail.com

Abstract- **In these days every one placing the personal data like photos, bio data etc. in so many social sites like face book, Gmail, matrimony etc. but the data is not safe in all conditions this problem is called privacy preserving problem. It provides an anonymized view of the data through a unified network, without revealing information to any of the users. Finally we develop an algorithms which is based on sequential clustering that provides centralized settings. This algorithm works based on the SaNGreeA algorithm, because Campan and Truta which is the leading algorithm for achieving anonymity in networks by means of clustering. The disadvantage of SaNGreeA, it builds clusters gradually. It cannot use actual Information Loss. This information loss will be evaluated only when all of the clusters are defined. So it contains structural information loss.**

But in Sequential clustering algorithm overcome this problem. In this algorithm makes decisions based on the measure of real information loss. Finally it gives a framework to classify the data with less information loss.

Key words- **clustering, data mining, distributed computation, Information Loss, privacy preserving, Social networks.**

I. INTRODUCTION

Social networks are have been studied for last so many years in different streams like Biology, Economics, and Sociology etc. In recent days every one uploading the data in social networks such as MySpace, Facebook, Twitter. Even in the few online networks that are completely open so that the data will be shown in online. Most operators provide some privacy to their data. That data is like telephone numbers, emails, messages etc. Generally individual data sets can be stored in simpler, traditional and complex forms. The researchers doing work on these data models and providing different solution to improve the privacy of the data. Although most of the privacy work done in Healthcare data. In this paper manly concentrating on university data that contains student details, employees salary workers details etc. Online social interaction has most popular around the world and this technology drastically increases day by day, most of the researchers also accepted it. For example Facebook, in this site every one personal details, contact numbers, email ids, and photos are also available. So privacy in social sites is infancy. For reason number of techniques are proposed. Generally networks are modeled by a graph, where nodes of the graph are called entities and nodes connecting lines are called edges. These edges represents the relationship between them. Example it is a real social network. A financial network contains more nodes then it becomes more complex. These type of networks are called Asymmetric. Social networks are very curious to do research work because it contains number of departments like sociology, psychology, market research or studies related and epidemiology. However, this social data is published in online. This data contains some sensitive information. Therefore, it is needed to hide that information this is simply called Anonymization. This anonymization can be done in different areas like images, data, graph etc. but here data is anonymized. So the publishing data is anonymized and provided the security to that data in order to achieve the privacy preservation.

II. CONCOMITANT STUDY

Data anonymization characteristically trades off with effectiveness. Therefore, it is required to find a golden path in which the released anonymized data still holds enough utility and privacy preservation to some accepted degree on the other hand. In this paper we propose an anonymization technique based on clustering, the data is formed into clustering the nodes based on loss dynamically forms the clusters nodes likewise it forms big nodes known as super nodes in which each of size at least k, where k is called anonymity parameter means it has to form minimum k nodes. Before this the data is centralized manner so if

number of transactions are more then the information loss measure is increase. In order to reduce this problem the data is distributed between the nodes then it is called distributed network. The network data is split between several players.

III. MODELLING

GRPAH REPRESENTATION

The social network data is represented in graph formats. Those graphs are directed and undirected graph. Social network is a simple undirected graph. $A=(V_i, E_i)$, A is a graph, V_i is the vertices of the graph and E_i is the edges of the graph where i=0…n.

Figure1. Social network graph

In the above figure is Facebook graph. That doesn't representing any directions so it is undirected graph. The images are in circle shape is called vertices v_i and connecting between vertices is called edges E_i. for unique identification add some structural information to the edges. Each node is corresponding to an individual in that group it belongs to. Edge connects two nodes and describes a relationship between the two corresponding individuals. Each node has non identifying attributes, such as age, gender, zip code etc. the combination these attributes are called quasi attributes. These quasi identifiers are used for unique identification.

Anonymization by clustering

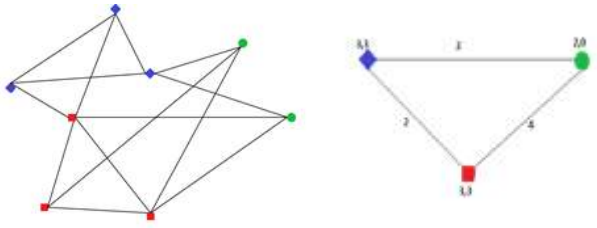


Figure 2. A Network and a Corresponding Clustering.

Definition 1:

The social network as a simple undirected graph G = (V, E), where $V = \{_{V1\ldots VN}\}$ is the set of nodes and $E \subseteq V^1$ is the set of edges. Each node corresponds to an individual in the underlying group, while an edge that connects two nodes describe a relationship between the two corresponding individuals. Where V^1 is the set of unordered pairs of elements from V. In every graph each node can be described by using the non identifying attributes. Those attributes are like zip code, age, sex, roll or employee numbers etc. These attributes are called quasi identifiers or unique identifiers. The combination of identifiers are uniquely identifies the nodes.

Let us a graph contains number of nodes $A_1 \ldots\ldots A_I$ quasi identifiers. Example A_1 is sex then A_1 contains Male and Female, then node V_i (i<N) is described by quasi identifier record as follows $R_i = (R_i(1) \ldots\ldots R_i(I)) \in A_1 \times \ldots\ldots \times A_I$. Where R is the particular record. Edge represents structural information where $E \subseteq (V_{2)}$. Clustering means grouping of similar objects. In this paper we are taking educational data set. So each record having the common row age. Based upon the age we can form the clusters $C_1 \ldots C_T$. let $C= \{ C_1 \ldots C_T\}$ be partition into disjoint subsets, or clusters. That is $V = \sum_{t=0}^{T} Ct$ and $C_t \cap C_s = \emptyset$

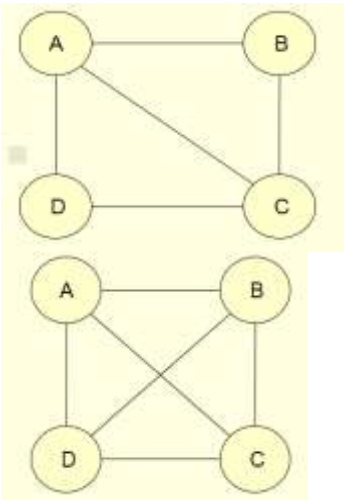
An example of a network of eight nodes, with two-dimensional quasi-identifier records and a corresponding clustered network with three super nodes. In the above figure having three colors in different shapes. Let us consider blue color nodes represent age, red color nodes represents location and blue color nodes gender then finally form clusters and finally calculates the information loss between the edges.

Edge Connectivity:

It is the minimum number of edges whose removal results in a disconnected graph. It is denoted by k(G). For a graph G, if k(G) = l then G is called an l-connected graph.

Example:

GRAPH 1 GRAPH 2



Figre3. Graphs with Edge connection

The edge connectivity for the GRAPH 1 is 2.
The edge connectivity for the GRAPH 2 is 3.

A cut in a graph is a set of edges whose removal disconnects the graph. A minimum cut is a cut with a minimum number of edges. It is denoted by S. For a non-trivial graph G iff |S| = k(G). The distance d(u,v) between vertices u and v in G is the minimum length of a path joining u and v. The length of a path is the number of edges in it.

Highly connected graph:

A graph G is k-connected if the removal of any collection of fewer than k vertices from G results in a connected graph with at least two vertices. Highly connected graphs represent robust networks that are resistant to multiple node failures. When a graph is not highly connected, it is useful to partition the vertices of the graph so that every part induces a highly connected subgraph. For example, designed a clustering algorithm where the vertices of a graph G are partitioned into highly connected induced subgraphs. For a graph with vertices n > 1 to be highly connected if its edge-connectivity k(G) > n/2.A highly connected subgraph (HCS) is an induced sub graph H in G such that H is highly connected. HCS algorithm identifies highly connected subgraphs as clusters. Properties: Diameter of every highly connected graph is at most two. That is any two vertices are either adjacent or share one or more common neighbors. This is a strong indication of homogeneity. Each cluster is at least half as dense as a clique which is another strong indication of homogeneity. Any non-trivial set split by the algorithm has diameter at least three. This is a strong indication of the separation property of the solution provided by the HCS algorithm.

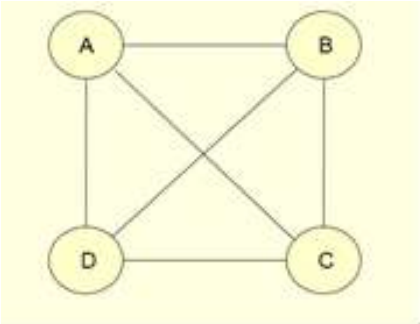


Figure4. Highly Connected Graph:

No. of nodes = 4 , Edge Connectivity = 3

The sequential clustering algorithm for k-anonymizing tables was presented in above. It was shown there to be a very efficient algorithm in terms of runtime as well as in terms of the utility of the output anonymization. We proceed to describe an adaptation of it for anonymizing social networks.

ANONYMIZATION

We categorize anonymization methods on graph formatted data into three main categories:

– Graph modification approaches: These methods anonymize a graph by modifying (adding and/or deleting) edges or nodes in a graph. There are two basic approaches:

- The simplest way alters the graph structure by removing and adding edges randomly. It is called randomization or random-based approach.
- Another way consists on edge addition and deletion to fulfil desired constraints, i.e. anonymization methods do not modify edges at random, they modify edges to meet some desired constraints. For example, k-anonymity-based approaches modify graph structure (by adding and removing edges) in order to get the k-anonymity value for the graph.

– Generalization approaches (also known as clustering-based approaches): These methods cluster nodes and edges into groups. Then, they anonymize each group into a super-node to publish the aggregate information about structural properties of the nodes. The details about individuals can be hidden properly, but the graph may be shrunk considerably after anonymization, which may not be desirable for analyzing local structures.

– Differentially private approaches: These methods refer to algorithms which guarantee that individuals are protected under the definition of differential privacy [11]. Differential privacy imposes a guarantee on the data release mechanism rather than on the data itself. The goal is to provide statistical information about the data while preserving the privacy of users.

Algorithm 1.

- Input: A social network $\mathcal{SN}$, an integer k.
- Output: A clustering of $\mathcal{SN}$ into clusters of size $\geq k$.

1) Choose a random partition $\mathcal{C} = \{C_1, \ldots, C_T\}$ of V into $T := \lfloor N/k_0 \rfloor$ clusters of sizes either k_0 or $k_0 + 1$.
2) For $n = 1, \ldots, N$ do:
   a) Let C_t be the cluster to which v_n currently belongs.
   b) For each of the other clusters, C_s, $s \neq t$, compute the difference in the information loss, $\Delta_{n:t\to s}$, if v_n would move from C_t to C_s.
   c) Let C_{s_0} be the cluster for which $\Delta_{n:t\to s}$ is minimal.
   d) If C_t is a singleton, move v_n from C_t to C_{s_0} and remove cluster C_t.
   e) Else, if $\Delta_{n:t\to s_0} < 0$, move v_n from C_t to C_{s_0}.

If there exist clusters of size greater than split each of them randomly into two equally sized clusters.
If at least one node was moved during the last loop, go to Step 2.
While there exist clusters of size smaller than k, select one of them and unify it with the cluster which is closest.
Output the resulting clustering.

The algorithm then starts its main loop (Steps 2-4). In that loop, the algorithm goes over the N nodes in a cyclic manner and for each node it checks whether that node may be moved from its current cluster to another one while decreasing the information loss of the induced anonymization. If such an improvement is possible, the node is transferred to the cluster where it currently fits best.

In the above algorithm the size of the cluster varies from [2, k1] where k1=βk and β is the fixed parameter. Based upon the main loop the cluster is formed, the data is removed and transferred to other based on best fit. On the other hand information loss (step sd). Likewise this process is continuous until the cluster becomes large means cluster size k. Then it partition the cluster randomly. This process is continuous until all the data should be moved to other clusters and the information loss is minimized. The stopping condition is repeatedly changed based on information loss. In this way we can minimize the information loss.

As a result number of clusters are formed but not all clusters having equal length, some of them having large size and some clusters are not having minimal size means atleast k. so have to apply agglomerative procedure to group all the small clusters into a single cluster. In this process also having some information loss we have to observe that loss also.

Information Loss: In [6], the proposed SaNGreeA algorithm uses a measure of structural information loss that differs from the measure I_S (.) that is given by (4)-(6). We proceed to define it. Let B be the N×N adjacency matrix of the graph G=(V,E) i.e $B(n,n^1) = 1$ if $\{V_n, V_{n1}\} \in E$ and $B(n,n^1) = 0$ otherwise then the hamming like distance is defined on V as follows

$$D(n,n^1) = \frac{|\{ l \neq n, n1 : B(n,l) \neq B(n1,l)\}|}{N-2} \quad \text{---------- (1)}$$

This definition of distance induces the following measure of structural information loss per cluster. Based upon the distance metric we can find the loss by using the equation

$$I'_s(C_t) = \frac{1}{\binom{|Ct|}{2}} \cdot \sum_{vn, vn' \in Ct}^{n} D(n, n') \quad \text{---------- (2)}$$

and overall structural information loss after forming the clusters

$$I'_s(C_t) = \frac{1}{N} \sum_{t=1}^{T} |Ct|.I's(Ct') = \sum_{t=1}^{T} x(Ct) \quad \text{---------- (3)}$$

Where $x(C_t) = \frac{2}{N(|Ct|-1)} \sum_{vn, vn' \in Ct}^{n} D(n, n')$ ---------- (4)

In other words, I'_s of a given cluster is the average distance between all pairs of nodes in that cluster, and I of the whole clustering is the corresponding weighted average of structural information losses over all clusters. The corresponding weighted measure of information loss is then

$$I'(c) = w.ID(c) + (1-w).I's \quad \text{---------- (5)}$$

Where w ϵ [0,1] and $I_D(c)$ ranges between zero and one as per the structural information loss.

The parameters ∝ and β changes the size of the clusters and find outs the structural information loss. Based on β and ∝ values the information loss is minimized. For higher β values would result lager size cluster and lower β values forms more number of clusters then finally we have follow the agglomerative phase. By this method information loss is high. So we have to take fixed β and ∝ values. But this is gives approximate values. So we have to apply the sequential clustering.

Distributed sequential clustering

Algorithm 2. Secure computation of sums
• Input: Each player m, 1 ≤ m ≤ M, has a private input vector am∈ Zd p.
• Output: a =∑M m=1am.
1) Player m selects M random share vectors am,ℓ ∈ Zd p, 1 ≤ ℓ ≤ M, such that∑M ℓ=1am,ℓ= ammod p.
2) Player m sends am,ℓto the ℓth player, for all 1 ≤ ℓ □= m ≤ M.
3) Player ℓ, 1 ≤ ℓ ≤ M, computes sℓ=∑M m=1am,ℓmod p.
4) Players ℓ, 2 ≤ ℓ ≤ M, send sℓto the player 1.
5) Player 1 computes a =∑M ℓ=1sℓmod p and broadcasts it.

In distributed sequential clustering the entire network data is split among different sites. Each site is connected to the

other sites. According to the above algorithm N is the network data and M is site. So each site M_i is connected to the all other sites with an edge called player. Where i=1…..n. so each player can be uniquely defined by all sites with losing any data. and each player has to protect all nodes under his control. As well as the existence and non existence of edges adjacent to his nodes.

Information loss:

The modified clusters having some information loss that is measured by using these formulas

Struct intra information loss = 2 e (1-(2 e/mod(c) * (mod (c) -1)))

Struct inter information loss = 2 Et,s (1-(Et,s / mod(Ct) * mod(Cs)))

IV. SECURITY

A secure network doesnot share any data to the other parties but it is not good always. Because some basic information is needed for all conditions and visibility of complete information is also not useful. By using distributed sequential clustering the sensitive information is hided in the entire network and all sites are linked to each other so that information loss is also minimized as well basic information will be published in web. Computing the sum of private integers has well known simple SMPs. The components of the vectors are rational numbers The denominators of those numbers are common and known to all, but their numerators depend on private integers, those are the private integers that appear in the numerator Hence, that problem reduces to computing sums of private vectors over the integers. Moreover, it is possible to compute upfront an upper bound p on the size of those integers and of their sum.

NETWORK DATA CLASSIFICATION

Network data is quite popular in Web and social networks applications in which a variety of different scenarios for node classification arise. In most of these scenarios, the class labels are associated with nodes in the underlying network. In many cases, the labels are known only for a subset of the nodes. It is desired to use the known subset of labels in order to make predictions about nodes for which the labels are unknown. This problem is also referred to as collective classification. In this problem, the key assumption is that of homophily. This implies that edges imply similarity relationships between nodes. It is assumed that the labels vary smoothly over neighboring nodes. A variety of methods such as Bayes methods and spectral methods have been generalized to the problem of collective classification. In cases where content information is available at the nodes, the effectiveness of classification can be improved even further. A different form of graph classification is one in which many small graphs exist, and labels are associated with individual graphs.

V. IMPLEMENTATION RESULTS

In this paper we taken educational data set which is created in sql language. A university contains all students, faculty and other workers information. Means the data set contains some sensitive information so that the data is represented in graph after that the data is clustered. The result is shown in figure5.

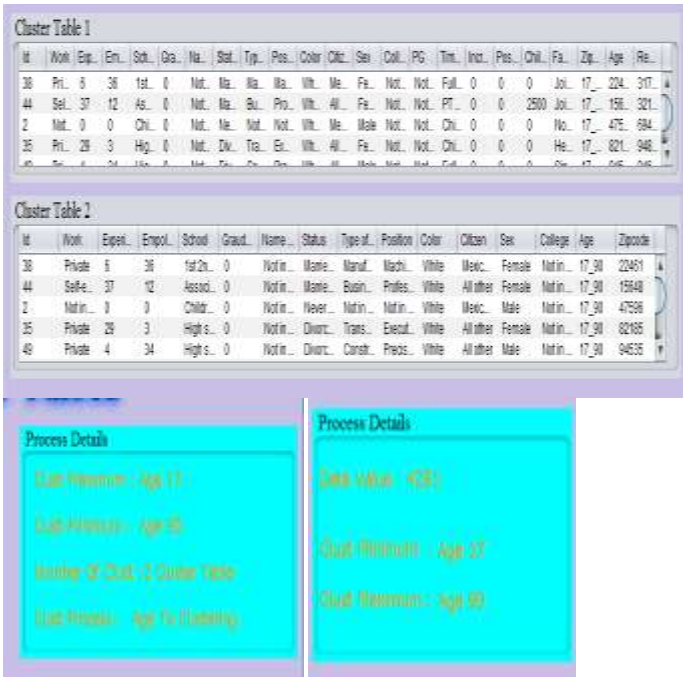
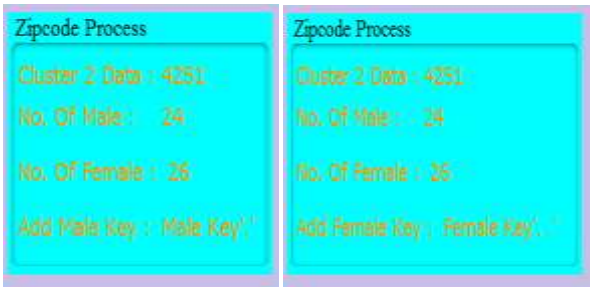


Fig5. Clustered data and Condition

Then we apply the anonymization condition on sensitive attributes of the data. so that we can hide the sensitive information. Here the data is clustered based upon the age and zipcode is anonymized and the result is shown in figure6.



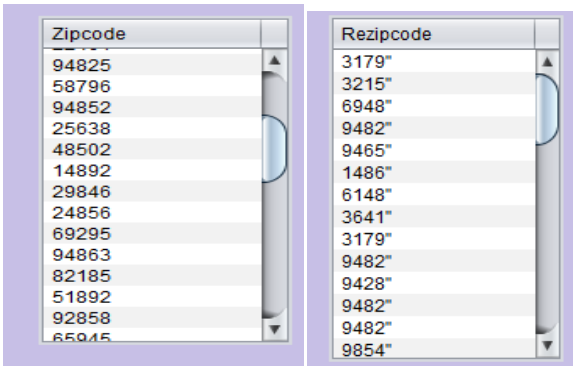


Fig6. Anonymization condition and its result

And finally calculates the information loss and these results are compared to SaNGreeA algorithm. The graphs are generated based on the loss of information and that graph explains how much effectively performs than compared to previous algorithms.

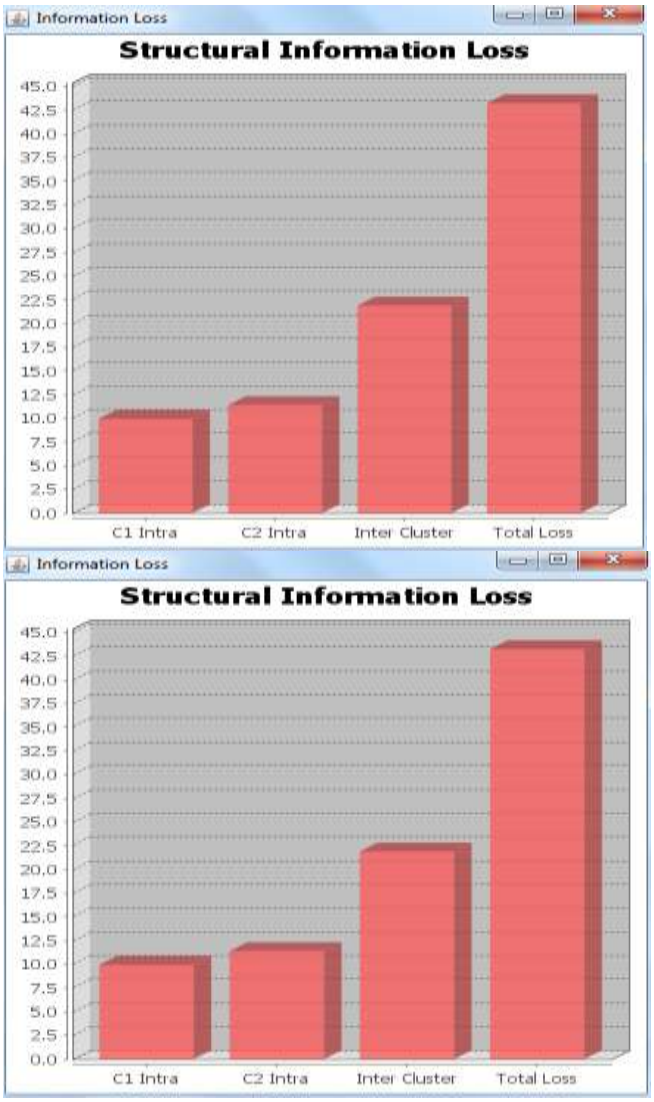


Fig7. Information Loss both Structural and General in the cluster

CONCLUSION

In this paper we represented university data that contains educational and other sensitive information. That information is distributed among the different sites so it has to be anonymized by one admin and here we used sequential clustering algorithms. The entire data is forms K clusters only so it is not safe all conditions it causes L-diversity problems. And finally I applied classification algorithm in reduce the number of computations every time. In classification if new record entered by using classifiers the record is sorted according to the condition. It is research direction so develop the number of algorithms which is having high security and less information loss.

REFERENCES

1. Anonymization of Centralized and Distributed Social Networks by Sequential Clustering Tamir Tassa and Dror J. Cohen
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 2, FEBRUARY 2013
2. De-anonymizing Social Networks Arvind Narayanan and Vitaly Shmatikov The University of Texas at Austin
NSF grants IIS-0534198, CNS-0716158, and CNS-0746888.
3. Data and Structural k-Anonymity in Social Networks Alina Campan and Traian Marius TrutaDepartment of Computer Science, Northern Kentucky University, Highland Heights, KY 41076, U.S.A. {campana1,trutat1}@nku.edu.
4. Data Classification Algorithms and Applications Edited by Charu C. Aggarwal IBM T. J. Watson Research Center Yorktown Heights, New York, USA.
5. A. Campan and T.M. Truta, "Data and Structural k-Anonymity in Social Networks," Proc. Second ACM SIGKDD Int'l Workshop Privacy, Security, and Trust in KDD (PinKDD), pp. 33-54, 2008.

● ● ●

Investigations on Vehicle Rollover Prevention Using Conventional PID Controller System

[1]Binda.M.B,[2]Dr.M.Rajaram
[1]Research Scholar,Dept of EEE, Sathyabama University, Chennai
[2]Vice Chancellor, Anna University,Chennai
[1]bindamb@gmail.com,[2]rajaramgct@rediffmail.com

Abstract- **In this paper, we present a novel methodology for preventing vehicle rollover model and control strategy which is suitable for preventing the untripped rollovers that was to be caused in vehicle. For vehicles that are deemed to be susceptible to wheel-lift off, various control strategies are implemented by earlier researchers. In this work, we propose a method for rollover prevention based on Z-N tuned PID controller. The validity of the proposed methodology is proved, and it is used to realize rollover prevention in the direction of the roll.**

Key words- **Rollover, Z-N Tuned PID Controller, PID Controller**

NOMENCLATURE

Parameters	Definition
v_x	Longitudinal velocity (body-fixed frame)
ω_r	Yaw rate (angular rate about vertical axis)
m	Vehicle mass
I_{zz}	Inertia about the vertical axis
a	front-axle-to-CG distance
b	rear-axle-to-CG distance
L	Track of vehicle
t	Width of vehicle
β	Slip angle of the vehicle body
k_1	Front cornering stiffness
k_2	Rear cornering stiffness
δ_f	Front steering angle

I. INTRODUCTION

The increased popularity of vehicles with high center of gravity (center of gravity is hereafter denoted CG),such as SUVs(Sport Utility Vehicles),makes the prevention of rollovers an important safety issue. Rollover occurs when a vehicle flips over, and there are two types: tripped and untripped. Tripped rollover is the most common type, and occurs when the vehicle has started to skid, hits an obstacle, and finally flips over. Untripped rollover is induced by the driver, either during extreme maneuvers, or in panic situations.

The main advantage of our proposed approach is: purely based on the safety Concerns and Understanding the Physics behind Rollover that are occurred in SUV vehicles.

The aim for this work is to find a control system that can prevent untripped rollovers, with minimum trajectory deviation. Vehicle control is an active research field, both in the academic world and in the car industry, and much effort is spent on finding better and better solutions.

This paper deals with standardized SAE vehicle axis system & three-degree-of-freedom vehicle model. Z-N tuned PID controller system is presented and also explains the need of this controller. The results concerning the aim in this paper is to suggest a solution to the problem of an initial investigation into models and control strategies suitable to prevent vehicle rollover due to untripped driving maneuvers.

II. RELATED WORKS

Rollover prevention is a topical area of research in the automotive industry (see, for example, the rollover section at http://www.safercar.gov/ for a good introduction to the problem) and several studies have recently been published. Relevant publications include that of Palkovics et al. [8], where they proposed the ROP (Roll-Over Prevention) system for use in commercial trucks making use of the wheel slip difference on the two sides of the axles to estimate the tire lift-off prior to rollover.

Wielenga [9] suggested the ARB (Anti Roll Braking) system utilizing braking of the individual front wheel outside the turn or the full front axle instead of the full braking action. The suggested control system is based on lateral acceleration thresholds and/or tire lift-off sensors in the form of simple contact switches.

Chen et al. [10] suggested using an estimated TTR (Time To Rollover) metric as an early indicator for the rollover threat. When TTR is less than a certain preset threshold value for the particular vehicle under interest, they utilized differential braking to prevent rollover. A number of metrics based on geometric principles have been developed for stability measurement. Researchers in mobile robotics have recognized that the location of the vehicle centre of gravity (c.g.) relative to the wheel–terrain contact points is critical to vehicle stability. The stability polygon (or support pattern) is defined as the convex hull of the polygon formed by wheel–terrain contact points projected onto a horizontal plane [1]. An early geo-metric measure defined stable vehicle configurations as those where the horizontal projection of the vehicle c.g. lies within this polygon [2].

A stability margin was then defined based on the shortest distance from the projected c.g. to a side of the polygon. Improvements to this measure were proposed [3, 4]. However, this metric ignores the effects of changes in c.g. height and the destabilizing influence of vehicle dynamic effects. It should be noted that a common geometric stability metric used in vehicle design is the static stability factor, which is computed as the ratio of vehicle width to c.g. height [5].

To help address this public safety issue, impending vehicle rollover needs to be detected and mitigated by activating chassis control actuators based on the vehicle rollover conditions [6].Various mathematical representations of the human driver have models been developed by earlier researchers for the prediction and avoidance of rollover [7]. In [13] several methods for detecting vehicle rollover were tested. The purpose of the tests was to find a method for detecting impending vehicle rollover that will activate active chassis subsystems when needed and minimize unnecessary activations of the chassis control systems. First, estimates obtained from the use of single sensors were examined. A lateral acceleration sensor, roll rate sensor, and suspension relative position sensor were all used to obtain estimates of the vehicle roll angle.

Ackermann et al. and Odenthal et al. [11], [12] proposed a robust active steering controller, as well as a combination of active steering and emergency braking controllers. They utilized a continuous-time active steering controller based on roll rate measurement. They also suggested the use of a static Load Transfer Ratio (LTRs) which is based on lateral acceleration measurement; this was utilized as a criterion to activate the emergency steering and braking controllers.

III. ROLLOVER MODEL

All numerical representations follow the standard SAE right-handed sign convention utilizing the simplified three-freedom vehicle model as depicted in Figure 1.

In Figure 1(a), x-y plane, F_{y1} and F_{y2} are the lateral forces over the front and rear tires respectively .The sideslip angle β and the roll angle ϕ are assumed to be negligible. The total forces and torques over the whole vehicle are given by

$$\sum F_y = F_{y1}\cos\delta + F_{y2} = ma_y$$
$$\sum M_z = aF_{y1}\cos\delta - bF_{y2} = I_z\dot{\omega}_r \quad (1)$$

Here a_y is the lateral acceleration, ω□$_r$ is the yaw acceleration, and I_z is the yaw moment of the vehicle inertia.

In addition, the lateral forces $F_{y1} = k_1\alpha_1$,$F_{y2} = k_2\alpha_2$,where k_1 and k_2 are the cornering stiffness of front and rear tires, respectively, α_1 and α_2 are the slip angles of front and rear tires, $\alpha_1 = (\beta+a\omega_r/v_x)-\delta$, $\alpha_2 = (\beta-b\omega_r/v_x)$, v_x is the longitudinal velocity, and ω_r is the yaw rate. Consequently, the following equation is derived:

$$a_y = \frac{(k_1+k_2)\beta}{m} + \frac{(ak_1-bk_2)\omega_r}{mv_x} - \frac{k_1\delta}{m}$$
$$\dot{\omega}_r = \frac{(ak_1-bk_2)\beta}{I_z} + \frac{(a^2k_1+b^2k_2)\omega_r}{I_zv_x} - \frac{ak_1\delta}{I_z} \quad (2)$$

In Figure 1(b), y-z plane, the total forces and torques over the whole vehicle are given by

$$\sum F_z = F_{n1} + F_{n2} - mg = 0$$
$$\sum M_x = mgh\sin\phi + ma_y(h_0+\Delta h-h) +$$
$$F_{n2}(\frac{T}{2}+\Delta T) - F_{n1}(\frac{T}{2}-\Delta T) = I_x\ddot{\phi} \quad (3)$$

Where $\ddot{\phi}$ is the body roll acceleration, h_0 is the height of the vehicle's center of gravity (CG) standing above the ground level, h is the distance between the vehicle CG and the assumed roll axis, T is the width of the vehicle track, and Δh and ΔT are the deformations of suspensions and tires. F_{n1} and F_{n2} are normal forces over the left and right wheels, and I_x is the roll moment of vehicle inertia.

The restoring force F_a shown in Figure 1(c) is related with the roll movement, $F_a = (k_\Phi\phi + c_\Phi\dot{\phi})/h$, where k_Φ and c_Φ are the total roll stiffness and roll damping of suspensions and $\dot{\phi}$ is the roll rate.

The total torques over the body with respect to the roll axis are given by

$$\sum M_{xs} = m_sa_yh + m_sgh\sin\phi - F_ah = I_{xs}\ddot{\phi} \quad (4)$$

where m_s is the sprung mass that represents the total vehicle mass m excluding the suspension and tire and I_{xs} is the roll moment for the inertia of the sprung mass.

Compared with the sprung mass, the unsprung mass is negligible. Assume that $m_s = m$ and $I_{xs} = I_x$, and establish (5) as follows:

$$\ddot{\phi} = \frac{mh(a_y + g\sin\phi)}{I_x} - \frac{k_\phi \phi}{I_x} - \frac{c_\phi \dot{\phi}}{I_x},$$

$$F_{n1} - F_{n2} = \frac{2ma_y(h_0 + \Delta h - h)}{T} + \frac{2mg\Delta T}{T} - \frac{2(k_\phi \phi + c_\phi \dot{\phi})}{T}, \quad (5)$$

$$F_{n1} + F_{n2} = mg,$$

$$LTR = (\frac{F_{n1} - F_{n2}}{F_{n1} + F_{n2}})$$

where LTR stands for the load transfer ratio, an important criteria to evaluate the vehicle status.

Since $a_y = \dot{v}_y + v_x \omega_r$, β = v_y /v_x , after introducing the state vector $x = [v_y \quad \omega_r \quad \dot{\phi} \quad \phi]^T$, the motions of this model can be described by

$$\dot{x} = Ax + B\delta \quad (6)$$

Where

$$A = \begin{bmatrix} \frac{k_1 + k_2}{mv_x} & \frac{ak_1 - bk_2}{mv_x} - v_x & 0 & 0 \\ \frac{ak_1 - bk_2}{I_z v_x} & \frac{a^2 k_1 + b^2 k_2}{I_z v_x} & 0 & 0 \\ \frac{k_1 + k_2}{I_x v_x} & \frac{ak_1 - bk_2}{I_x v_x} & \frac{-c_\phi}{I_x} & \frac{mgh - k_\phi}{I_x} \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (7)$$

$$B = \begin{bmatrix} \frac{-k_1}{m} & \frac{-ak_1}{I_z} & \frac{-k_1}{I_z} & 0 \end{bmatrix}^T$$

IV. ROLLOVER PREVENTION METHOD BY MITIGATION CONTROL

The running vehicle intends to overshoot in its response to a suddenly applied lateral acceleration. Therefore, the inner wheels may have lifted off during the steering when the applied acceleration is lower than the nominal permit value that is in compliance with the steady-state laws .

From (5), we can see

$$\ddot{\phi} = \frac{mh(a_y + g\sin\phi)}{I_x} - \frac{k_\phi \phi}{I_x} - \frac{c_\phi \dot{\phi}}{I_x} \quad (8)$$

where the roll angle ϕ is small enough to get $\sin\phi \approx \phi$, and after applying Laplace law to the aforementioned differential equation, we established the following new equation:

$$\frac{\phi(s)}{a_y(s)} = \frac{(k_\phi - mgh)/I_x}{s^2 + (c_\phi / I_x)s + (k_\phi - mgh)/I_x} \cdot \frac{mh/I_x}{(k_\phi - mgh)/I_x} \quad (9)$$

where the angular frequency is

$$\omega_n = \left[(k_\phi - mgh)/I_x \right]^{1/2} \quad (10)$$

and the damping ratio is

$$\xi = \frac{c_\phi}{2\left[(k_\phi - mgh) I_x \right]^{1/2}} \quad (11)$$

When the input a_y is set to be the unit step function, the overshoot is

$$M_P = \exp(-\frac{\xi\pi}{\sqrt{(1-\xi^2)}}) \quad (12)$$

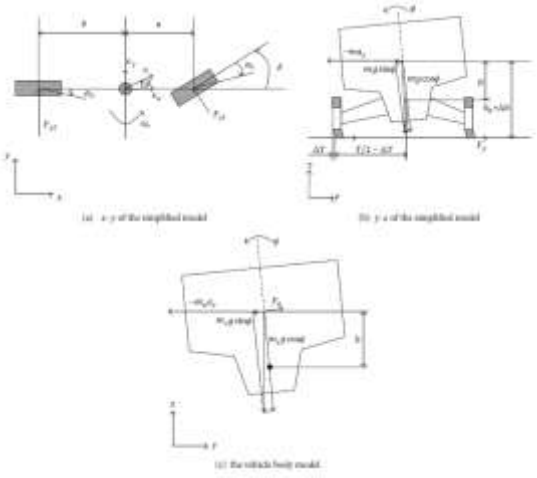


Figure.1. Simplified vehicle rollover model.

At this point, we found that vehicle dynamics is determined by the first part of (9), whereas the second part defines the vehicle steady state. These findings form the base of the mitigation control referred to in the rest of this section.

A specific brand of midsize SUV was used as the test vehicle in this study, of which the relevant parameters are shown in Appendix.

With the rollover model and vehicle parameters, we achieved the damping ratio of the original vehicle, ξ = 0.71, and the overshoot value of the roll angle at 38.18% as shown in Figure 5. Thus, the damping ratio needs to be raised in order to reduce the overshoot value. In other words, we can improve the restoring force F_a by increasing the roll stiffness k_Φ or the roll damping c_Φ . However, it needs to bear in mind that the roll stiffness k_Φ and the roll damping c_Φ must be increased simultaneously. Otherwise, the damping ratio ξ may be decreased.

With the rollover model and vehicle parameters, we improve the damping ratio of the vehicle, ξ = 0.95, and the overshoot value of the roll angle at 0.71% as shown in Figure 6.

V. Rollover Prevention Method by Z-N Based PID Controller

PID CONTROLLER

A PID is the most commonly used feedback controller. A PID controller calculates error value as the difference between a measured process variable and a desired set point. The controller attempts to minimize the error by adjusting the process control inputs. The PID controller calculation (algorithm) involves three term control: the proportional, integral and derivative values denoted by P, I and D. Where P depends on the present error, I on the accumulation of past errors, and D is prediction of future errors.

A PID controller will be called a PI, PD, P or I controller in the absence of the respective control actions. PI controllers are fairly common, eliminates the steady state error. A PID controller has proportional, integral and derivative terms that can be represented in transfer function form as

$K(s)=K_p+K_i/s+K_ds$ (13)

K_p: proportional gain, a tuning parameter,

K_i: Integral gain, a tuning parameter,

K_d: Derivative gain, a tuning parameter

By tuning these PID controller gains, the controller can provide control action designed for specific process requirements. The proportional term drives a change to the output that is proportional to the current error. This proportional term is concerned with the current state of the process variable. The integral term (K_i) is proportional to both the magnitude of the error and the duration of the error. It (when added to the proportional term) accelerates the movement of the process towards the set point and often eliminates the residual steady-state error that may occur with a proportional only controller.

- Optimizing Of PID Controller

For the system under study, Zieger-Nichols tuning algorithm based on critical gain K_{er} and critical period P_{er} will be used. Here, the integral time T_i will be set to infinity and the derivative time T_d to zero. This is used to get the initial PID setting of the system [15].

In this method, only the proportional control action will be used. The K_p will be increase to a critical value K_{er} at which the system output will exhibit sustained oscillations. In this method, if the system output does not exhibit the sustained oscillations hence this method cannot be used.

- Designing PID Parameters

From the response given below in Figure 2, the system under study is indeed oscillatory and hence the Z-N tuning rule based on critical gain K_{er} and critical period P_{er} can be applied.

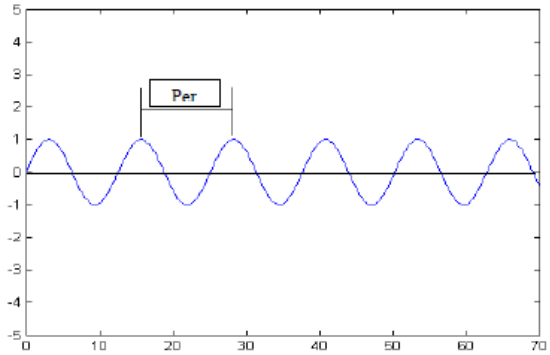


Figure 2. Illustration of Sustained Oscillation with Period P_{er}

The transfer function of the PID controller is

$G_c(s) = K_p(1+ 1/sT_i + T_d s)$ (14)

The objective is to achieve a unit-step response curve of the designed system that exhibits a maximum overshoot of 25 %. If the maximum overshoot is excessive says about greater than 40%, fine tuning should be done to reduce it to less than 25%.

Since the $T_i = \infty$ and $T_d = 0$, this can be reduced to the transfer function of

$$\frac{C(s)}{R(s)} = \frac{127.33K_p}{s^3+16.02s^2+127.33s+127.33K_p} \quad (15)$$

The value of K_p that makes the system marginally stable so that sustained oscillation occurs can be obtained by using the Routh's stability criterion. Since the characteristic equation for the closed-loop system is

$$s^3+16.02s^2+127.33s+127.33K_p = 0 \quad (16)$$

From the Routh's Stability Criterion, the value of K_p that makes the system marginally stable can be determined.

The sustained oscillation will occur if K_p =16.02.Hence the critical gain K_{er} is K_{er}=16.02.Thus with K_p set equal to K_{er}, the characteristic equation becomes

$$s^3+16.02s^2+127.33s+2039.83 = 0 \quad (17)$$

The frequency of the sustained oscillation can be determined by substituting the s terms with $j\omega$ term. Hence the new equation becomes

$(j\omega)^3 + 16.02(j\omega)^2 + 127.33(j\omega) + 2039.83 = 0$ (18)

From the above simplification, the sustained oscillation can be reduced to

$\omega^2 = 127.33$

or

$\omega = 11.28$ rad/sec

The period of the sustained oscillation can be calculated as

$P_{er} = 2\pi/11.28 = 0.557$

From Ziegler-Nichols frequency method of the second method ,the table suggested tuning rule according to the formula shown. From these we are able to estimate the parameters of K_p, T_i and T_d .

Type of Controller	K_p	T_i	T_d
P	$0.5K_{er}$	∞	0
PI	$0.45K_{er}$	$(1/1.2)P_{er}$	0
PID	$0.6K_{er}$	$0.5P_{er}$	$0.125P_{er}$

Table 1: Recommended PID value setting

Hence from the above Table 1, the values of the PID parameters K_p, T_i and T_d will be $Kp = 0.6K_{er} = 9.612, T_i = 0.5 \text{ X } 0.557 = 0.2785, T_d = 0.125 \text{ X } 0.557 = 0.0.06925$.

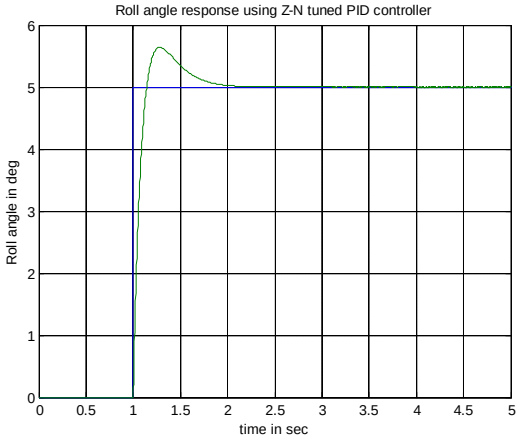


Figure 3: ϕ versus a_y error transient response for Z-N tuned PID controller

SIMULATION RESULTS

In order to verify the proposed mitigation control, simulation tests are conducted according to NHTSA's Fishhook 1a test description, and the steering maneuver patterns are shown in Figure 7 and the simulink diagram is shown in Figure 4 and the simulation result is shown in Figure 3. Table 3 gives the time domain specifications obtained from this simulation.

Note that the wild fluctuations occur when the hand wheel angle rate changes rapidly, especially within the time ranges of [2s, 3s] and [4s, 6s].

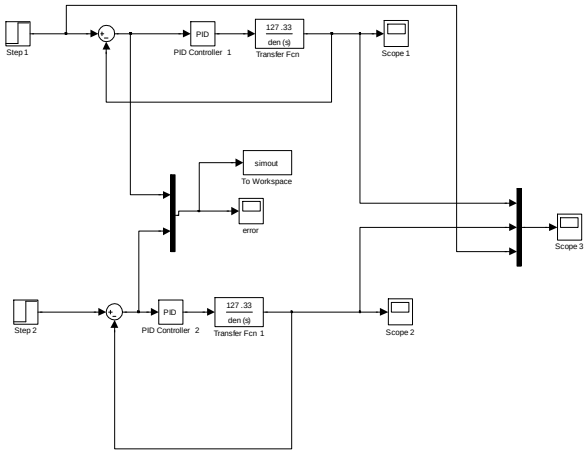


Figure 4:Simulation Diagram for Roll Model

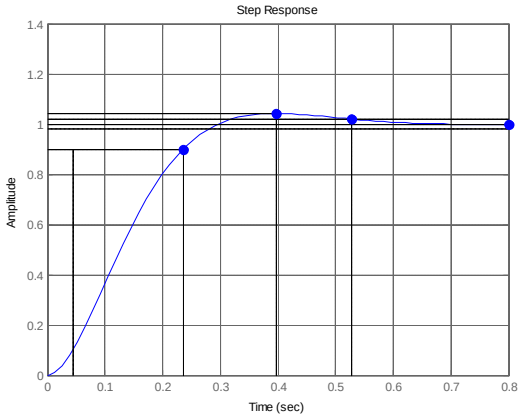


Figure 5: ϕ versus a_y transient response for ξ=0.71

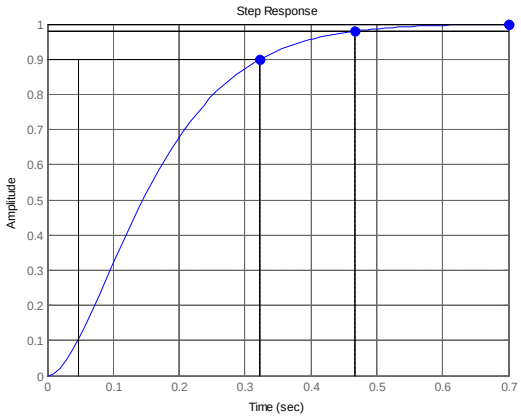


Figure 6: ϕ versus a_y transient response for ξ=0.95

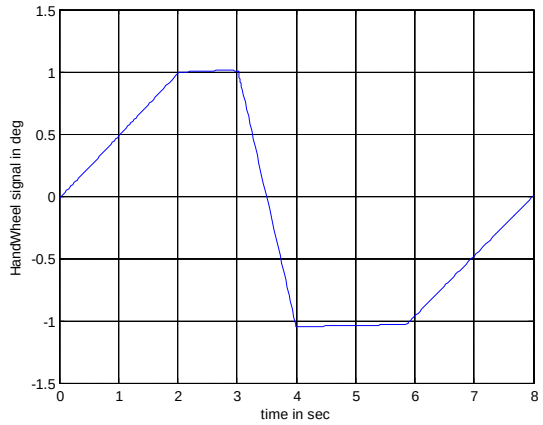


Figure 7: Fishhook 1a maneuver

Measuring factors	Z-N tuned PID controller
Rise Time in sec	1.1
Maximum Overshoot in %	12.5
Settling Time in sec	2.2

Table 3. Results Of Z-N tuned PID controller for a 3 DOF vehicle model

CONCLUSION

The main purpose of this research work is to propose a novel methodology for preventing vehicle wheel-lift prior to sliding. Additionally, multiple control strategies were presented in order to mitigate the vehicle rollover. To investigate the vehicle transient and steady states, an improved rollover model using Z-N tuned PID controller was established in this study. Simulation results indicated a decreased overshoot of the roll angle and a better confined steady value.

The adjusting method proposed in this paper may help the design of both passive and active suspension controls to increase vehicle stability. We are confident that, with additional studies, the proposed model will be applicable for a real vehicle in the near future.

REFERENCES

[1] R.B. McGhee and A.A. Frank, On the stability properties of quadruped creeping gait, Math.Biosci. 3(2) (1968), pp. 331–351.

[2] R.B. McGhee and G.I. Iswandhi, Adaptive locomotion of a multilegged robot over rough terrain, IEEE Trans. On Syst. Man Cybernet. SMC-9(4) (1979), pp. 176–182.

[3] S. Sugano, Q. Huang, and I. Kato, Stability criteria in controlling mobile robotic systems, IEEE/RSJ International Workshop on Intelligent Robots and Systems, Yokohama, Japan, July 1993, pp. 832–838.

[4] J.K. Davidson and G. Schweitzer, A mechanics based computer algorithm for displaying the margin of static stability in four-legged vehicles, Trans. ASME J. Mech. Design 112
(1990), pp. 480–487.

[5] J.P. Chrstos and D.A. Guenther, The measurement of static rollover metrics, SAE Transactions no. 920582, 1992.

[6] D. W. Shuttlewood, D. A. Crolla, and R. S. Sharp, "Active roll control for passenger cars," *Vehicle Syst. Dyn.*, vol. 22, no. 5/6, pp. 383–396, 1993.

[7] V. Cherian, R. Shenoy, A. Stothert, J. Shriver, J.Ghidella, and T. D.Gillespie, "Model-Based Besign of a SUV anti-rollover control system," SAE Paper No. 2008-01-0579,
2008.

[8] Palkovics L., Semsey A` and Gerum E., "Rollover prevention system for commercial vehicles-additional sensorless function of the electronic brake system", Vehicle System Dynamics, Vol. 4, pp. 285-297, 1999.

[9] Wielenga T.J., "A method for reducing on-road rollovers: anti-rollover braking", SAE Paper No. 1999-01-0123, 1999.

[10] Chen B. and Peng H., "Differential-brakingbased rollover prevention for sport utility vehicles with humanin-the-loop
evaluations", Vehicle System Dynamics, Vol.36, pp. 359-389, 2001.

[11] Ackermann J. and Odenthal D., "Robust steering control for active rollover avoidance of vehicles with elevated center of gravity", Proceedings of International Conference on
Advances in Vehicle Control and Safety,Amiens, France, 1998.

[12] Odenthal D., B¨unte T. and Ackermann J., "Nonlinear steering and braking control for vehicle rollover avoidance", Proceedings of European Control Conference, Karlsruhe, Germany, 1999.

[13] Aleksander Hac, T.B., and John Martens, Detection of Vehicle Rollover. SAE,2004(SAE Paper No. 2004-01-1757).

[14] Binda.M.B, Dr.M.Rajaram, "Intelligent Prediction And Prevention Of Vehicle Rollover Using FLQG Regulator", in Journal of Theoretical and Applied Information Technology,(JATIT),20th November 2014, Volume. 69, No.2, (ISSN: 1992-8645) .

[15] Binda.M.B, Dr.M.Rajaram, "Investigations on vehicle Rollover Prevention using LQG Regulator ", in Advances in Mathematical Physics, Hindawi Publishing Corporation, Volume 2014, Article ID 308285, 11 pages,(ISSN:1687-9120), December 2014 .

[16] Binda.M.B, Dr.M.Rajaram, "Intelligent Prediction And Prevention Of Vehicle Rollover Using NNLQG Regulator",

in International Journal of Electrical & Electronic Engg &Telecommunications, Volume. 4, No. 1, (ISSN: 2319-2518), January 2015.

Privacy enrichment in personalized Web search

[1]Anju Sukumar,[2]S Gayathri
[1]M.E (Computer Science and Engineering) in Vins Christian College of Engineering, Anna University.
[2]Information Technology department,Vins Christian College of Engineering, Anna University
[1]anju123s@gmail.com

Abstract- **Web search engines are widely used to find certain data among a huge amount of information in a minimal amount of time. Personalized web search is a process of presenting right information to a specific person at right time. In order to provide personalized search results for the users, Web search engines store all the queries which have been submitted by the users and the search results which they have selected in previous searches. Personalized web search have many advantages over normal search, in improving quality of search result. Web search engines profile their users by storing and analyzing their past searches. However, these useful tools also pose a privacy threat to the users. They have reluctance to share their private information, hence wide usage and benefits of personalized web searches not proliferated. The proposed system introduces a framework to generalize user profile based on their queries while respecting their requested privacy requirements.**

Key words- **personalized web search, profile, privacy**

I. INTRODUCTION

Now a days, the Internet has million users and million web pages. It has become the largest source of information world- wide. Web search engines (WSEs) - e.g., Google, yahoo etc. are useful tool for finding specific data in minimal amount of time from huge amount of information. It is a software system that is designed to search for information on the World Wide Web. However, when the same query is submitted by different users, most search engines return the same results regardless of who submits the query or what is the right intention of the query. In general, each user has different information needs for his/her query. For example, for the query "apple "some users may be interested in documents dealing with fruit, "apple," while other users may want documents related to computer company "apple." Therefore, Web search results should adapt search results according to users different infor- mation needs. Personalized web searches (PWS) are used to adapt search result by analysing user's interests.

Web pages are personalized based on the characteristics (interests, social category, and context) of an individual. Personalization is based on implicit data analyzed from items purchased or viewed links. Customization can be used instead of personalization when the site only uses explicit data such as ratings or preferences. Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. General approaches to personalizing search results are, one involving modifying the user's query and the other re-ranking search results.

Click-log based and profile-based techniques are used for PWS. Click-log based is straight forward methods they bias through user's query history. Profile-based methods profile its users implicitly from query history [2],[3],[4], browsing history[5],[6] etc. A profile-based method fits for all sorts of queries, but under some circumstances it is reported as unstable [1]. Although there are advantage and disadvantage over both types of PWS methods, the profile-based PWS has more effectiveness in improving search quality. Profiles created may contains private information of user, user hesitate to share private information. In fact, privacy issue limits wide usage of PWS services.

1.1 Motivations

Privacy concerns have a long history in web search. They already existed in information retrieval from public databases. When a user submits a query, he/she is also exposing her interests to the database operator. Several proposals are used to hide this personal information. Those proposals can be classified according to the level of privacy offered to users:

1. Provide perfect privacy but no personalization.
2. Protect the privacy of users and provide a certain degree of personalization.

To protect user privacy in profile-based PWS, have to consider two contradicting effects. On the one side, to improve search quality with personalization utility of the user profile. On the other side,to hide privacy content in user

profile. When using a Personalized Search service, how is the user and the search provider to ensure the privacy and protection of a users identity and information that is supplied to the service? If a user does not trust the search provider, then the user is not going to either provide sufficient personal information in order to optimize his/her search results and will not use the personalized search feature at all.

Users may have concerns regarding the security surrounding the storage of their data. After a user's personal information has been given away, then the search provider have to ensure that the information remains private and does not fall into the hands of people or organizations with malicious intentions for that data. Previous work of privacy preserving PWS is not that much best. The problems with the existing methods are following.

1. *Runtime profiling is not done in existing profile-based* PWS. Generalization of user profile is done only offline, and same profile is used for all queries submitted by the user. One profile fits all strategy has certain drawbacks in dealing variety of queries. Profile –based personalization may not help to improve search quality of ad hoc queries[1], though exposing may result in a risk of user's privacy. A better way is to make decision online, whether to personalize the query. Also want to make decision what all data in the user profile should expose to get better result. No previous work has supported such features.

2. *Customization of privacy requirements are not done in existing systems.* For privacy customization sensitive topics are identified using an absolute metric called *surprisal* based on information theory .Sensitive topic are topics whose disclosure may creates privacy problems to user. Sensitive topic are identified implicitly and explicitly.

3. *Less user interactions to obtain personalized search results.* In existing system personalization are made by using higher user interactions. Here predictive metrics are used to measure the search quality and risk after personalization, without iterative user interactions

1.2 Contributions

Project proposes a user-oriented personalized web search framework. This framework generalize profiles for each queryaccording to user specified privacy requirements. Generalization is based on two conflicting metrics, personalization utility and privacy risk. Two generalization algorithms are proposed GreedyIL (Informaation loss) and GreedyDP (Disriminating Power). An effective generalization algorithm GreedyIL, to support runtime profiling attempts to minimize the information loss (IL). GreedyIL out performs compared to existing algorithm GreedyDP. Provide an inexpensive mechanism for the client to decide whether to personalize a query in the framework. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. Online generalization on user profiles to protect the personal privacy without compromising the search quality. Propose privacy preserving personalized web search framework .Generalize profiles for each query according to user-specified privacy requirements. The search results are personalized with the profile and delivered back to the query proxy in client machine. Finally, the proxy presents the raw results to the user, or re-ranks them with the user profile.

As illustrated in Fig. 1, frameworks consist of non-trusty server and a number of clients (users). Each clients request service from the server. The framework works in two phase online and offline. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified requirements.

During the online phase queries are handled as follows.

1. When a user issues a query the client side proxy generates a user profile in runtime in light of query terms. The output is a generalized user profile satisfying the privacy requirements.
2. The query and the generalized user profile are sent together to the PWS server for personalized search.
3. The search results are personalized with the profile and delivered back to the query proxy.
4. The proxy presents the raw results to the user, or re-ranks them with the complete user profile.

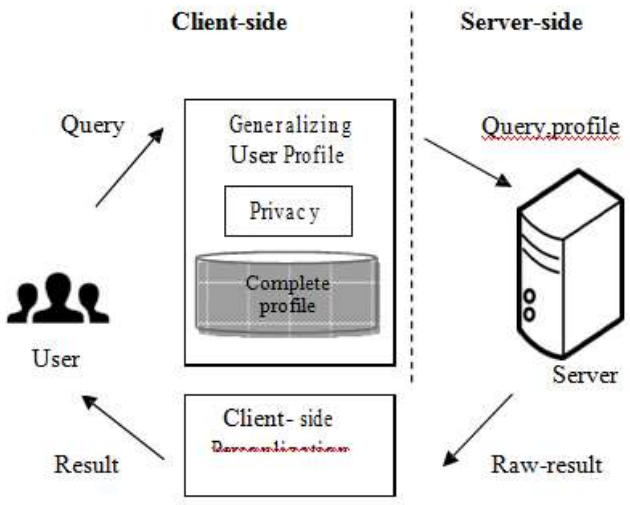


Fig. 1.Personalized search System Framework

Client-side personalized search strategy avoids the privacy risk of storing personal user information on search providers severs by letting the client maintain and be responsible for their own 'set' of personal information. With this information, the client transports it to the search provider whenever they perform a search by respecting user specified privacy. The search provider will then take the received information along with the search query and then perform a personalized search for the client. By allowing the user to give their privacy requirements, their own personal data are protected in client machine. Thus, the search provider will not have to store a
copy of the data on their severs. When a query term such as "Golf" is used for search. In one scenario, the user may be looking for information relating to the sport of "Golf" while another user may be inquiring about the automotive offering of a Volkswagen "Golf". This will lead to situations where users are both unhappy with their search results and the contextual ads, which search providers heavily rely on for revenue , will be ineffective. One way of handling this ambiguity is to supply more personal information about the client to the search provider. This personal information should be protected in client side and can enhance search result

II. RELATED WORKS

This section focus on the literature of profile based PWS and privacy in PWS system.

2.1 Profile-based PWS

Previous works on profile-based PWS focus on improving search utility. Two main aspects are considered in Profile- based PWS, first is the representation of user profile and second is the measure of personalization. Many profile representations are used such as list/vectors [5] or bag of words [2] etc. Most effective way to represent profile is hierarchical structure. Hierarchical representations are used in ODP [1], Wikipedia and so on. The user profile is incrementally developed over time and it is stored for use in later sessions. The information exploited for constructing the profile usually comes from various sources, so it relies on different aspects of the user. Any personal documents such as browsing history and email on a user's computer could be the data source for user profiles. Our hypothesis is that terms that frequently appear in such documents represent topics that interest users. This focus on frequent terms limits the dimensionality of the document set, which further provides a clear description of users' interest. Here hierarchical user profile constructed based on frequent terms. Terms with higher frequency are placed at higher levels and terms with lower frequency are placed at lower levels of the profile [8].

2.2 Privacy in PWS

In Previous works [9], four levels of privacy protection are defined and analyzed: pseudo identity, group identity, no identity and no personal information. In the pseudo identity level, the user identity is replaced by a pseudo-identity which contains less identifiable information. Nevertheless, this does not prevent the WSE from creating a profile associated to the pseudo-identity. This profile contains sensitive information that can be used to identify the real user who hides behind the pseudo-identity. When AOL released its search engine query log , they replaced each IP addresses with pseudoidentities [7]. The second level of privacy corresponds to the group identity. In this case, a group of users share a single identity. Therefore, the WSE is only able to build a group profile. It cannot identify single users and cannot profile the user .This mechanism enhance the privacy level achieved by query submitters. By using the group profile in trems of individual profile reduces the effectiveness of the service. Third level is no identity in this case user identity is not available to the user, search engine cannot know the identity of individual user. Last level is no personal information neither the user identity nor the description of user information is available to the search engine. For achieving the fourth level privacy is that a search engine would be required by law to guarantee that it does not store any user information. Hence, the search engine will have no memory for activity of user.

III. PRELIMINARIES and PROBLEM DEFINITION

This section focus on structure of user profile, customized privacy requirement customization in user profile, attack model in PWS

3.1 User Profile

In proposed framework a hierarchical user profile is constructed based on the availability of a public taxonomy. The repository consists of huge collection of topic covering entire human knowledge. The repository is regarded as publicly available and can used by anyone with its background
knowledge. Examples of such repository are ODP [1].
Profile generator creates user profiles representing the user preferences. User profile is in the form of topic hierarchy that reveals user interests. Matching the profile to descriptions of the topic repository. User profiles are defined using a simple approach. Profiles are created automatically as the user dohis/her searches [8].
For example, a user profile could look like this:
/Arts/Music/Artists
/Arts/Architecture/Famous
The above profile indicates that the user visited articles or websites related to artists under the subtopic music and which is under the main topic arts. Each leaf topic in user profile has a support factor indicating preference on respective topic.

Support factor for other topics in the user profile is calculated recursively. Repository and user profile are in the form of tree like structure. User profile is obtained on the basis of repository.

User profile is a rooted sub tree of repository. Each topic in public repository has a support value indicating how often the topic is visited by people. The user support can be recursively aggregated from those specified on the leaf topics.

SupH(t) =Σt'€C(t,H)SupH(t') (1)

The user support is different from the repository support as the first describes the user's preference on topic, while the second shows the importance of topic in the entire human knowledge. User profiles are adaptive to changing user interests.

3.2Privacy Requirement Customization

Nodes in user profile consist of sensitive and non-sensitive node.

Sensitive nodes

Sensitive nodes are those whose disclosure (to server) causes privacy risk to user. Given a user profile H, the sensitive nodes are a set of user specified sensitive topics S subset of H, whose sub trees are non overlapping. Example of sensitive nodes S = {Adults, Privacy, Harmonica, Figure (Skating)}. Here user is not willing to share details that he searched about Adults, Privacy, Harmonica, Figure (Skating) .

For example user selected figure (Skating) only to avoid advertisements. User's privacy concern differs from one sensitive topic to another. Selection of sensitive nodes depends on users interest. Studies revealed that user might still tolerate the exposure of such interests to trade for better personalization utility and to improve search result.

Sensitivity Value

A sensitivity value indicates the severity of privacy leakage caused by disclosing corresponding sensitive node. User specifies sensitivity values for each sensitive node. Sensitivity of rest of the topic is found recursively. Customized privacy requirements specified by number of sensitive topics by user in the profile. Generally there are two main privacy protection problems for personalized search. One which cause privacy risk as the identification of an individual, the other includes identification of sensitive data. In proposed system sensitive nodes in hierarchical profile represents the sensitive information/topic in the profile. A person can specify her/his sensitive values to the sensitive topics to show severity of disclosure of the topic to server.

3.3 Attack in PWS

This work aims at providing protection against attacks such as eavesdropping and man-in the middle attack. Eve intercepts the communication between user and the server via some measures. When user issues a query the entire copy of query together with a profile will be captured by Eve. Based on the profile eve will attempt to identify the sensitive nodes of user by recovering the topic hidden from the original complete user profile. Eve knows the background knowledge of the taxonomy repository. The profile and privacy are defined based on repository; hence it is easy to identify sensitive topics. The eavesdropping will be started and ended within a single of query session. Privacy attacks on the web are undertaken by some automatic programs by sending spam advertisements to PWS-users. These programs act as a real person that collects p information of a victim for a long time.

3.4 Generalizing Profile

The generalization technique can seemingly be conducted during offline without user involvement. It is impractical to perform offline generalization because of two reasons: The output from offline generalization may contain many topic which are irrelevant to a query. Second problem is forbidding which may remove all the sensitive topic in the tree. A solution to this problem is online generalization, which depends on the queries. Online generalization avoids unnecessary privacy disclosure and also removes topics that are irrelevant to the current query.

Utility Metrics

Utility metric is to predict the search quality of the query on a generalized profile. Main reason that search quality not measured directly is because search quality depends on the implementation of personalized search engine, which is difficult to predict. Also, it is too expensive to include user feedback on search results. Here the utility prediction problem is used to the estimation of the discriminating power of a given query on a profile. Discriminating power shows how two queries are discriminated each other.

Privacy Metrics

Privacy metrics deal with the privacy factor that affected during the disclosure of the profile. The privacy risk when exposing user generalized profile is defined as the total sensitivity of the topic contained in the profile. When the complete user profile is exposed risk of privacy is too high. Which means exposing all sensitive nodes reaches its maximum value. if a sensitive node is pruned and rest of the nodes are retained during the generalization privacy risk is reduced. The generalization process is done by considering two metrics, utility metrics and the privacy metrics; both are defined for user profiles. Generalized user profile should satisfy these two metrics and should be under a threshold value□. G is the generalized user profile.

G*=argmax (util (q, G)), risk (q, G) <□ (2)

IV. PROCEDURES

This section, present the procedures used in the PWS system.

Two execution phases are there in the procedure to obtain personalized result.

1. Offline phase.
2. Online Phase.

The offline phase consists of profile creation and privacy customization. The online phase consists of query-topic mapping and generalization.

Profile Creation

The first step of the offline is to build the complete user profile that reveals user interests and activity. Here user's preferences are represented in a set of plain text documents.

Step to construct the profile

1. Detect the topic in repository for every document that is visited by user.
2. Construct the profile using the topic identified as a path to leaf node
3. Give user support for each topic and recursively find Support

value of rest of the topic

Privacy Customization.

First step is the user to specify a sensitive-node and the respective sensitivity value (positive value) . Sensitive values

of inner topics are identified by recursively calculating the sensitive value of subtopics. Sensitive value associated with cost of disclosure of the topic.

Query-topic Mapping

The aim of query-topic mapping is to compute a rooted sub tree of profile called as seed profile. Seed profile is the output

of the query-topic mapping which includes topic relevant to the query. These seed profile enhance the search quality and decrease the privacy risk of exposure of complete user profile.

Mapping the query topic with the user profile is done mainly in this phase. Based on the sensitivity of topics in the user profile prune operation is done and seed profile of less private content is obtained

Profile Generalization

This procedure generalizes the seed profile in a cost-based iterative manner based on the privacy and utility metrics. In addition, this procedure computes the discriminating power for online decision on whether personalization should be employed.

V. EXPERIMENTAL RESULTS

This section contains experimental setup needed to implement system and result obtained on the basis of analysis of metrics.

5.1Experimental Setup

This framework is implemented on a PC with minimum requirements. Such as a Pentium Dual-Core 2.50-GHz CPU and 2-GB RAM memory, running Microsoft Windows XP. All implementations can be done in Java.

The ODP web Directory is used to create the user profle and topic repository. The click logs are downloaded from the online AOL query log. The AOL query log data contain 20 million of queries and 30 million clicks of 650k users. User profiles are created from the real query log that are previously searched by user. Queries are classified as Distinct, Medium, and Ambiguous on the basis of discrimination power.

5.2 Scalability of Profile Generalization

The scalability of the proposed system analyzed on the basis of the seed profile size and number of queries. For each possible seed profile size randomly choose 100 queries from The query log, and take their corresponding seed profiles using the user profile. All leaf nodes in a same seed profile are given equal user preference .These queries are then processed using the two algorithms GreedyDP and GreedyIL algorithms.

CONCLUSION AND FUTURE WORK

A client-side privacy protection framework for personalized search is presented in this paper. As the system works on client side it already ensure minimum amount of privacy to its users. It can be adopted by any personalized web search that captures user profiles in a hierarchical manner. Users are allowed to specify their privacy requirements. They can also limit their privacy, and they can specify sensitive value for their sensitivity topic. This system generalize user profile online on user profiles to protect the personal privacy without compromising the search result. Proposed greedy algorithm for the online generalization experimental results revealed that framework could give better quality search results while respecting privacy requirements specified by user. The results also confirmed the effectiveness and efficiency. The profile-based personalization contributes little or even reduces the search quality, while exposing the profile to a server would for sure risk the user's privacy. To address this problem, we develop an online mechanism to decide whether to personalize a query. The basic idea is straightforward. If a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile. The future work aimed to overcome the adversaries with broader background knowledge. It mainly includes richer relationship among topics or capability to capture a series of queries from the user. Also seek better methods to implement the user profile.

REFERENCES

[1] Z.Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluvation Evaluvation
and Analysis of Personalized Search Strategies, "Proc.
Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.
[2] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search
via Automated Analysis of Interests and Activities,"
Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development
in Information Retrieval (SIGIR), pp. 449-456,
2005.
[3] M. Spertta and S. Gach, "Personalizing Search Based on U
Ser Search Histories," Proc. IEEE/WIC/ACM Int'l
Conf.Web Intelligence(WI), 2005.
[4] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search
History to Improve Search Accuracy," Proc. ACM
SIGKDD Int'l Conf. Knowledge Discovery and Data
Mining (KDD), 2006.
[5] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive W
-eb Search Based on User Profile Constructed without any
Effort from Users," Proc. 13th Int'l Conf. World WideWeb
(WWW), 2004.
[6] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for
Personalized Search," Proc. 14th ACM Int'l Conf. information
and Knowledge Management (CIKM), 2005.
[7] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information
Retrieval Using Implicit Feedback," Proc. 28th Ann.
Int'l ACM SIGIR Conf. Research and Development
Information Retrieval (SIGIR), 2005.
[8] J. Castelli´-Roca, A. Viejo, and J. Herrera-Joancomarti´,
"Preserving User's Privacy in Web Search Engines," Computer
Comm., vol. 32,no. 13/14, pp. 1541-1551, 2009.
[9] K. Hafner, Researchers Yearn to Use AOL Logs, but They
Hesitate, New York Times, Aug. 2006.

A Query Recommender System for Interactive Database Exploration

[1]Diancy Denny,[2]Ramya P
[1][2]M.Tech (Information Technology) in Vins Christian College of Engineering, Anna University.
[1]diancydenny7@gmail.com,[2]ramgees@gmail.com

Abstract- **Recommender systems plays an important role in the information and e-commerce ecosystem. It represents a powerful method for enabling users to filter through large information and product spaces. The collaborative filtering systems predict a person's affinity for items or information by connecting that person's recorded interests with the recorded interests of a community of people and sharing ratings between like-minded persons. In this paper, a query recommender framework is instantiated to assist non-expert users of the scientific database by providing personalized query recommendations. For this, the fragment-based approach is proposed in which each user session is represented by the query fragments identified in the respective queries. In order to provide recommendation, the item-based collaborative filtering is used which is based on pair-wise similarity among the items involved in the recorded user sessions. The fragment-based approach provides better precision and recall than previous tuple-based approach.**

Key words- **Collaborative filtering, recommender systems, relational databases, interactive exploration.**

I. INTRODUCTION

The recommendation system is a particular type of information filtering technique that attempts to present information items which are likely of interest to the user. Nowadays, it is of great importance for the success of e-commerce and IT industry and gradually gains popularity in various applications. Intuitively, a recommendation system builds up a user's profile based on his/her previous records, and compares it with some reference characteristics, and seeks to predict the `rating' that a user would give to an item he/she had not yet evaluated. Ratings may be gathered through explicit means, implicit means, or both. Explicit ratings are those where a user is asked to provide an opinion on an item. Implicit ratings are those inferred from a user's actions. For example, a user who visits a product page perhaps has some interest in that product while a user who subsequently purchases the product may have a much stronger interest in that product. Mostly, the recommendation system corresponds to a large-scale data mining problem.

Collaborative Filtering is the process of filtering or evaluating items using the opinions of other users. Usually users may not know which parts of the database hold useful information, so the task of knowledge discovery is a big challenge. An exploration of the whole database is very time-consuming because the database size is continuously growing. To address the problem of assisting users when exploring a database, the QueRIE (Query Recommendations for Interactive data Exploration) framework is designed. QueRIE assists users by generating dynamic, personalized query recommendations which is in ad-hoc (without any previous planning) or form-based query environments. The idea is to provide a set of SQL queries to the users that are expected to be relevant to their information needs. The user will be able to directly submit or further edit these queries, instead of having to compose new ones.

This framework is built on a simple idea that is inspired by Web recommender systems: If users U1 and U2 have similar querying behavior, then they are likely interested in retrieving the same data. Hence, the queries of user U2 can be made as a reference for user A. Collaborative filtering is a method of making automatic predictions about the interest of a user by collecting preferences from other users. This ides has been used in web recommender systems, but when transfer this method into database context raises several challenges. First, SQL is a declarative language, so syntactically different queries may produce the same data. Consider the queries Q1 and Q2, Q1: SELECT NAME FROM TABLE1 WHERE ID=4; and Q2: SELECT TABLE1.NAME FROM TABLE1 JOIN TABLE2 ON (TABLE1.RID=TABLE2.RID) WHERE TABLE1.ID=4;. HERE, TABLE1 and TABLE2 have a key/foreign key on attribute RID, then both queries retrieve same results. In case of web recommender system, the similarity between two users can be analyzed through the similarity between items they visit/rate/purchase, but the SQL queries cannot be compared similarly. A second important challenge is the absence of explicit rating system for the queries posted by the user-how to know which queries are necessary for the computation of user similarity. Finally, the recommended SQL queries need to be simple, so that the user can

understand and refine it necessary. Too complex queries may even more confuse the users. The QueRIE addresses these challenges defining by close the loop approach. It can be done by first decomposing each SQL queries into basic elements in order to compute similarities between users and make predictions. Then the recommended elements are mapped backto meaningful and simple SQL queries that users can understand or refine.

In [1], [2], the authors propose the QueRIE framework and the application of user-based collaborative filtering using witness tuples to represent user queries. In this paper the QueRIE framework is presented, including the overview of item-based approach that uses query fragments to represent the user queries in [3]. Each recorded fragments are used to identify similar query fragments in the previously recorded sessions, which are then "assembled" as interesting queries for the active user. A scalable design is proposed which includes generalization and parsing of queries to compute fragment similarity for query recommendation

II. RELATED WORK

Some work has been done in the area of personalized recommendations for keyword or free-form query interfaces. In this scenario, a user queries a database using keywords or free-form text and the personalization system recommends items of interest. This approach is different from the scenario because it aims to assist users who query relational databases using either ad-hoc or form-based queries. Also, our framework recommends queries instead of "items" from the database. Finally, the query framework does not require from the users to explicitly declare their preferences before in order to generate recommendations.

The problem of generating personalized recommendations has been broadly addressed in the Web context [4], but only a handful of related works exist in the database context. In this paper, a user queries a database using keywords or free-form text and the personalization system recommends items of interest. But our approach is different from this scenario because it aims to assist users who query relational database. Also, our framework recommends queries instead of "items" from the database. Finally, the query framework does not require from the users to explicitly declare their preferences beforehand in order to generate recommendations.

The necessity of a query recommendation framework is emphasized in [5], where the authors outline the architecture of a collaborative query management system targeted at large-scale, shared-data environments. As part of this architecture, they suggest that data mining techniques can be applied to the query logs in order to generate query suggestions. The authors present a general out-line of a framework for query recommendations pointing out that this is a challenging process. But, they do not provide any technical details on how such a recommendation system could be implemented.

Two very recent works propose frameworks for query recommendations using the information recorded in the query logs [6, 7]. In [6], the authors propose a query recommender system that represents the past queries using the most frequently appearing tuple values. Then, after predicting which new tuples might be of interest to the end user, they reconstruct the query that retrieves them. Contrary to this work, the approach is tuple-based. More-over, the proposed scheme works better with relations that have discrete attribute values, contrary to scientific databases, where most attributes are numeric. The authors also propose a global ranking of the queries, based on the statistics of the database and not the query logs. Both approaches are evaluated in a preliminary empirical study, yet no discussion on scalability issues is provided.

In [7], the authors propose a framework that recommends join queries. They use the data recorded in the query logs and reconstruct queries, however they assume that the end user should pro-vide the system with some tables to be used as input and other tables to be used as output, along with the respective selection conditions. This approach clearly differs from ours in that they do not take the current user's session into consideration, neither they per-form recommendations in the traditional "personalized" form (i.e. finding similarities among users or items).

III. PRELIMINARIES

A relational database can be explored through a sequence of SQL queries posted by the user. The goal of this exploration is to identify interesting information and hence based on this goal the queries are formulated. The queries posted by a user during one session to the database are correlated, and the user formulates the next query in the sequence only after having examined the results of previous queries. Given a user i, let Qi denote the set of SQL queries that the user has posed.

A session summary is the subset of the database covered by the queries of each user. This summary captures the parts of the database accessed by the user and gives a measure of importance for each part. A summary may contain the names of the relations that appear in the queries, and the importance of each relation can be measured by the number of queries that reference it. A detailed summary may contain the actual results given by the user, along with an explicit rating of each result tuple. Considering that the choice of the summary is fixed for all users. The Si can be used to denote the summary for user i. In order to generate recommendations, the framework computes a "predicted" summary Spred.

$$Spred = f(\alpha, S0, S1, \ldots, Sx)) \quad (1)$$

The predicted summary Spred is a function that combines information from both active user's summary S0 and the summaries of S1….Sx past users. This summary shows the predicted degree of interest of the active user S0 and this can serve as the "seed" for the generation of recommendations. In other words, the predicted summary

depends on both the active user S0 and the summaries S1,….,Sx of past users.

Depending on the choice of reference characteristics, a recommendation system can be content-based collaborative filtering, user-based or item-based collaborative filtering and both. As their names indicate, content-based approach is based on the matching of user profile and some specific characteristics of an item (e.g. the occurrence of specific term in a document) while user or item-based approach is a process of filtering information or pattern based on the collaboration of users, or the similarity between items.

A "mixing factor" $\alpha \varepsilon [0, 1]$ is introduced which allows us to include or exclude the fragments of the active user session in the recommendation process. Here α is a parameter of the QueRIE framework, it determines the importance of active user session with respect to the past users session. When $\alpha = 0$, the item-based collaborative filtering approach is used, which means only the past users fragments are taken into account. When $\alpha = 1$, the content-based approach is used, in that only the fragments included in the active user's queries are taken into consideration. Neither of these condition is a good setting for all cases. By considering predicted summary Spred, the framework models queries that cover the subset of the database with the highest predicted importance. These queries are presented to the user as recommendations.

IV. PROPOSED SYSTEM

A.THE QUERY FRAMEWORK

The proposed Query framework is depicted in Fig.2. The active user formulates his/her queries through query interface. The SQL syntax analyzer in query interface will check the correctness of SQL syntax. Then the queries are forwarded to both the DBMS and Recommendation Engine. The DBMS processes each queries and produces the result sets. Simultaneously these queries are stored in Query Log. The Recommendation Engine analyzes the active user's query with the past user's query as recorded in query log and generates a set of recommended queries to the user.

In this framework users can explore a relational database through a sequence of SQL queries. The goal of this exploration is to identify interesting information to the user. The queries posted by an user during one session to the database are correlated and the user formulated next query in the sequence after having inspected the results of previous queries. For example, a real user session is shown in Table 1.

Query 1	SELECT count(*) FROM table1 WHERE name like 'system';
Query 2	SELECT count(category) FROM table1 WHERE name like 'system';
Query 3	SELECT id, count(*) FROM table1 WHERE name like 'system' GROUPBY id;
Query 4	SELECT id, count(*) FROM table1 WHERE name like 'system' GROUPBY id having count(*)>1;

Table 1: User session queries

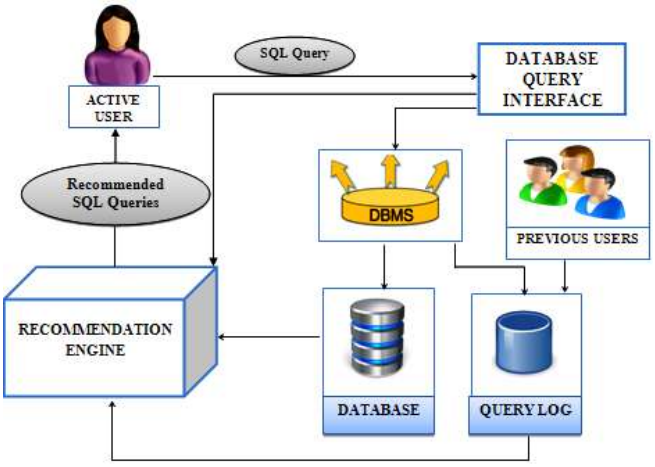


Fig.1: QueRIE framework

This query pattern clearly shows an interactive exploration of the database: the user starts by counting the number of tuples satisfying a predicate, then counts the different objects corresponds to these tuples and then retrieves the object that occur more than once. The user could save a lot of time if the system could recommend the appropriate query (Query 4) after the first attempt (Query 1). This is possible if a similar session is already present in the query logs to generate appropriate recommendations for the current user.

The query recommendations presented to the user can be performed in several ways. One approach is to analyze queries using the characteristics present in Spred. But this approach is not optimal because it is important to recommend meaningful and intuitive according queries. In order to address these issues, an alternative approach is used: the queries that are included in the query log of the DBMS are re-used. This framework consists of mainly four components: (a) a model for generating session summary Si for each user i, (b) the computation of session summary S0 ,S1,….Sx (c) the computation of predicted summary in Spred and (d) generation of query recommendation based on Spred. The framework forms a close loop with the components (a), (b) and (d), going from SQL queries to session summaries and back. Through the declarative query language, all user interaction with a relational database is occurred.

B.FRAGMENT BASED QUERY RECOMMENDATION

An item-based collaborative filtering approach is used in order to generate recommendations to the active user. This approach is based on pair-wise similarity among items included in the recorded user sessions. The main advantage of this approach is that all similarity calculations are done during training process and hence very small overhead during recommendation generation phase [4]. It is considered that the items that co-appear in many sessions are similar to each other and these similarities are used in order to generate recommendation for the active user session. The user-based

collaborative filtering requires real-time similarity calculations, since the active user's profile gets updated. So this slows down the real-time generation of recommendation.

The main aim of this technique is to recognize fragments that co-appear in various queries posted by different user, and use them in recommendation process. These fragments may or may not include in user's active session depending on the value of mixing factor. Thus the pair-wise similarities of all query fragments recorded in the query log are initially calculated. These similarities are used for ranking each fragment with regards to active user session and then highest ranked query fragments are used as recommendations. The session summary Si for a user i consists of all the query fragments of the user's past query. Two different weighing schemes are used to compute the fragment weights in Si.

Binary Scheme: In this approach, all participating fragments have the same weight of importance, regardless of whether they appear in many queries in the session or only one.

Si = □ Q ε Qi SQ (2)

Weighted Scheme: In this scheme, the fragments that appear more than once in a user session will receive higher weight than others.

Si = ΣQ ε Qi SQ (3)

The fragment-fragment matrix is constructed using the session summary of the past users and vector similarity metric. According to item-based collaborative filtering approach, if a session includes both the fragments it is more similar. The similarity metric employed depends on weighting scheme. The Jaccard's coefficient is applied for binary scheme and Cosine similarity for weighting scheme. The predicted summary Spred shows the estimated importance of each query fragment with respect to the active user's behavior. As shown in Equation 1, a mixing factor, α is used to include or exclude the fragments of the active user session in the recommendation process. The recommendation set will contain queries that have been previously recorded in the query logs. It can be ensured that the recommended queries are simple and understandable, since they were authored by humans. Once the predicted summary Spred has been calculated, the top N fragments (the fragments that having higher weight) are selected and given as recommendations to the user.

C.METHODOLOGY

Each query submitted by the user is decomposed into basic elements that capture the essence of that query logic. These elements are used to calculate similarity between users and hence compute the predicted summary. By evaluating the predicted summary, top N recommendations are generated. The Query framework consists of four sections :(a) query generalization, (b) query parsing, (c) similarity calculation, and (d) query recommendation. In order to find similarity between different user sessions, it is necessary to relax the queries in the

query log. This will increase the cardinality and thus can avoid dissimilar queries in the query log.

During the relaxation process, all WHERE clauses are relaxed by converting numerical data and string literals to generic string representations. For example, all strings are replaced by STR, "=" by EQU and "<, ≤, >, ≥" by COMPARE. The generalized query of Query 4 is SELECT ID, COUNT(*) FROM TABLE1 WHERE NAME LIKE STR GROUPBY ID HAVING COUNT(*) COMPARE NUM;. After the generalization process, these queries are converted into fragments. Each fragment is separated using regular expressions. Table 2 shows the start and end keywords that is used to identify the fragments. The fragments of Query 4 are: COUNT(*), TABLE1, TABLE1.NAME PATMATCH, COUNT(*) COMPARE NUM.

Fragment name	Start keyword	End keyword
Attribute String	Select	From
Relation String	From	Where, Group By, Order By, End Of Query
Where String	Where	Order By, Having, End Of Query
Group By String	Group By	Order By, Having, End Of Query
Having String	Having	Order By, End Of Query

Table 2: Parsing keywords

The advantage of generalization is that the fragments in the WHERE clause are not differentiated based on their actual values, rather based on the attributes used for filtering. For example, value=10 and value=100 will be represented by same fragment (value EQU NUM). The fragment-based approach is used since no online calculations are required. Then the fragment-to-fragment similarity is calculated. This real–time update of fragment similarities resolves the cold start problem of recommender systems, because new fragments will be considered immediately for subsequent recommendations.

V. EXPERIMENTAL RESULTS

A. Evaluation of the Top-N parameter: The Top-N fragments of previous queries are used to generate recommendations for the active user. Figs. 2 and 3 show the average precision and F-Score for various Top-N values. It can be noted that the accuracy increases with the value of N. But when N>10, the accuracy again start decreasing. The Query framework achieves higher precision when N=5 and N=10(0.75 and 0.8 respectively) whereas the F-score is almost same for N=5 and N=10(0.75 and 0.76 respectively).

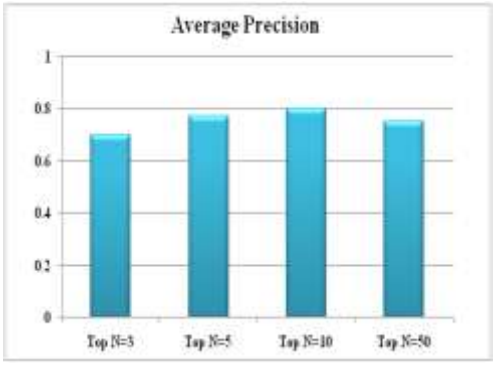


Figure 2: Average precision for various Top-N values

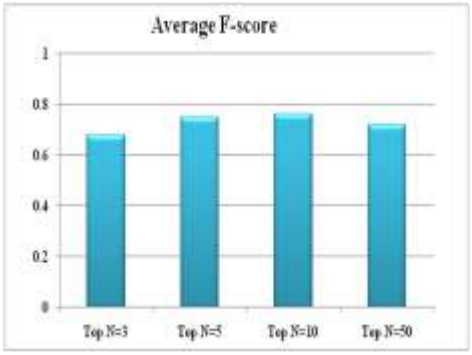
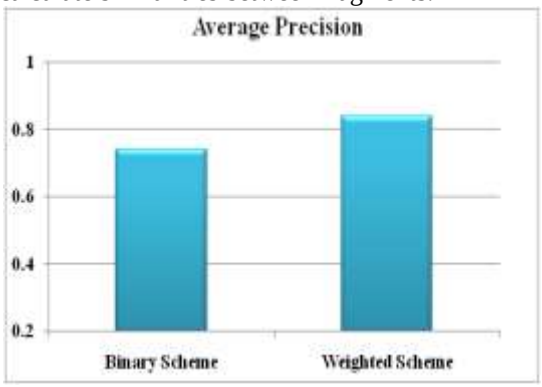
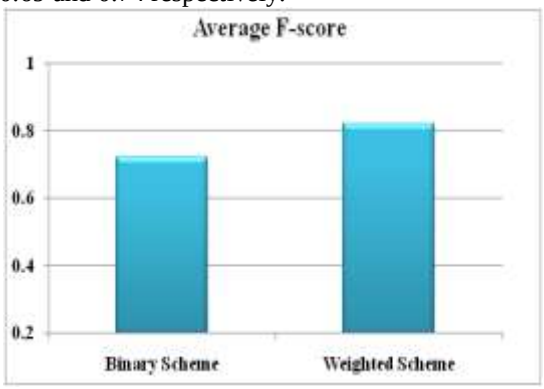


Fig. 3: Average F-score for various Top-N values

B. Evaluation of the weighting scheme: The Query framework uses binary and weighted schemes to represent the query and session vectors and to calculate similarities between fragments.



The Jaccard's coefficient and Cosine similarity metric is employed respectively for weighting schemes. Figs 4 and 5 show the average precision and F-score of binary and weighted schemes. The average precisions of binary and weighted schemes are 0.69 and 0.74 respectively, and an F-score of 0.69 and 0.74 respectively.



CONCLUSION

In this paper, a fragment-based approach is instantiated for the query framework to generate useful SQL recommendations to the users of relational database. An item-based collaborative filtering method is used in order to generate query recommendations to the active user. Each user query is converted into fragments to represent user sessions. Overall, it is showed that the precision of the recommendations is close to 80% when the active user's session is included in the prediction process. The weighted scheme is selected as default weighting scheme and top-N recommendation is best for N=10. This shows that the query framework is very effective in generating useful recommendations to the end users of relational database systems. The current implantation of query framework supports SPJ(Select Project Join) queries, whereas if a query includes sub-query it will be dropped. There are many interesting directions that can be explored in future. Exploring on sub-queries in the user session will be an added advantage.

REFERENCES

[1] G. Chatzopoulou, M. Eirinaki, and N. Polyzotis, "Collaborative filtering for interactive database exploration," in Proc. 21st Int. Conf. SSDBM, New Orleans, LA, USA, 2009, pp. 3–18.

[2] S. Mittal, J. S. V. Varman, G. Chatzopoulou, M. Eirinaki, and N. Polyzotis, "QueRIE: A recommender system supporting interactive database exploration," in Proc. IEEE ICDM, Sydney, NSW, Australia, 2010.

[3] J. Akbarnejad et al., "SQL QueRIE recommendations," PVLDB, vol. 3, no. 2, pp. 1597–1600, 2010.

[4]G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76–80, Jan./Feb. 2003.

[5]N. Khoussainova, M. Balazinska, W. Gatterbauer, Y. Kwon, and D. Suciu. A case for a collaborative query management system. In Proc. of the 4th Biennal Conference on Innovative Data Systems Research (CIDR 2009), 2009.
[6]K. Stefanidis, M. Drosou, and E. Pitoura. "You May Also Like" Results in Relational Databases. In 3rd International Workshop on Personalized Access, Profile Management, and Context Awareness in Databases (PersDB 2009), 2009.
[7] X. Yang, C. M. Procopiuc, and D. Srivastava. Recommending join queries via query log analysis. In 25th International Conference on Data Engineering (ICDE 2009), pages 964–975, 2009.

Hostel Management System

[1]Shubham,[2]Shweta,[3]Vinuta,[4]Vishal,[5]Anagh
Department of CSE, B. V. BhoomaraddiCollege of Engineering & Technology, Hubli
[1]shub406@gmail.com,[2]shweta9448@gmail.com,[3]vinuta.t@gmail.com,
[4]vishalchauhancr7@gmail.com,[5]anaghkalburgi@gmail.com

Abstract- **Today almost all colleges have a very good infrastructure which includes library facility, hostel facility etc. Normally each college has hostel for boys and girls. The staff in the hostel should maintain each student's details like attendance, mess bill, room details etc. All these activities are carried out manually and record books are maintained for each activity. But this indeed is a laborious task which requires a large amount of patience and time.**

Key words- **Attendance, Bill calculation, Gallery, Login, Mess menu.**

I. INTRODUCTION

Every educational institute has a hostel for the accommodation of its student. The number of educational institutes is increasing rapidly, thereby increasing the number of hostels. Hostel management by manual way is a tedious process, since it involves managing student details, room details, mess bill calculation, room allocation, complaint registry and hostel attendance.
To reduce the work load and time consumption, we came up with the idea of an automated system, which is more user friendly and smart way for carrying out all these activities.
The proposed system is more user friendly, reduces redundancy and also reduces the mistakes occurred due to manual entries and thus it is an easier and smart way to manage hostel records.

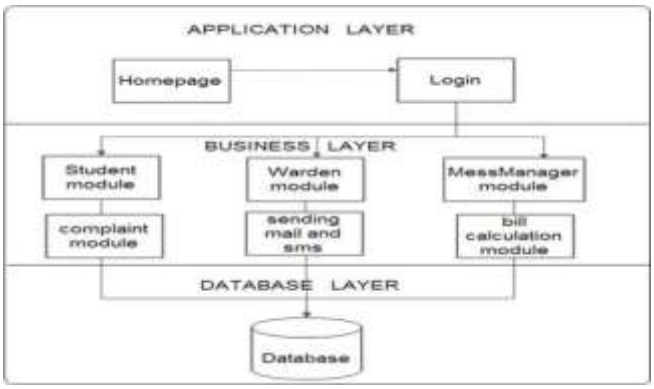
II. EXISTING METHODOLOGIES

GSHK is one of the software designed for hostel management. It provides the functionalities of registration, bill calculation, fee collection etc. It lacks the functionalities of sending mail, attendance registry and complaint registry.

III. PROPOSED METHODOLOGY

In the proposed system, we can easily manage the student details, room details, hostel and mess bill calculation, student complaint registry and sending mail. It also provides an easier way of managing hostel attendance and updating mess bill according to guest information provided by the student.

IV. ARCHITECTURAL DESIGN



It is a 2 tiered architecture diagram. It describes the 3 layers of the hostel management system. The first layer consists of the application layer which has the login page and the home page. The next which is the business layer consists of the student page, warden page and the mess manager page. The third layer is the Database layer which has the details of the students and workers.

V. USER INTERFACE DESIGN

Figure 1.Main Page

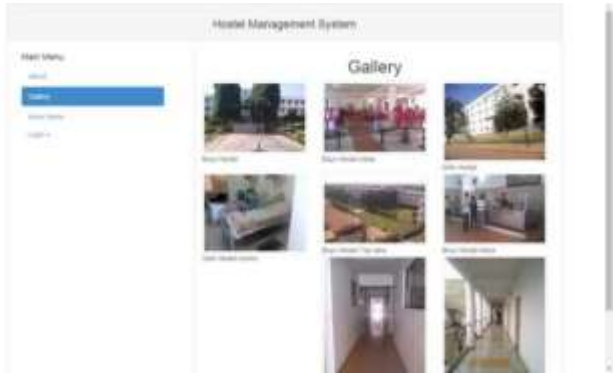


Figure 2.Gallery

VI. MODULES

A. LOGIN
User has the option of login as student, warden or mess manager.
B. COMPLAINT

Student can register his complaints anonymously regarding hostel or mess which will be directly fetched by warden or mess manager.
C. ATTENDANCE

Warden and Mess manager can update hostel and mess attendance.
D. ADDING DETAILS

Warden and Mess manager can add student details, hostel and mess bill details according to which bills will be calculated.
E. BILL CALCULATION

Warden and Mess manager can calculate and update hostel and mess bill respectively.
F. SENDING MAIL

Warden and Mess manager can send mail to the parent regarding hostel or mess payment due.

VII. WORKING OF MODULES

A. LOGIN PAGE

The user can login either as warden, student or mess manager. The user enters his login id and password. If he is an authentic user then he is granted permission to the system or an error message is generated.

Figure 3 Warden Login page

B. COMPLAINTS

The student enters his account using his login id and password. He then has the option of registering complaints regarding hostel or mess which can be fetched by the warden or mess manager.

Figure 4 complaint registry

C. ATTENDANCE

Warden and Mess manager can check and update the attendance of the students on daily basis. The mess attendance is further used by the mess manager to calculate the mess bill.

Figure 5 Attendance

D. ADDING DETAILS

Warden and Mess manager can add several details like opening balance, payment in bill, etc which will be input to calculate hostel and mess bill.

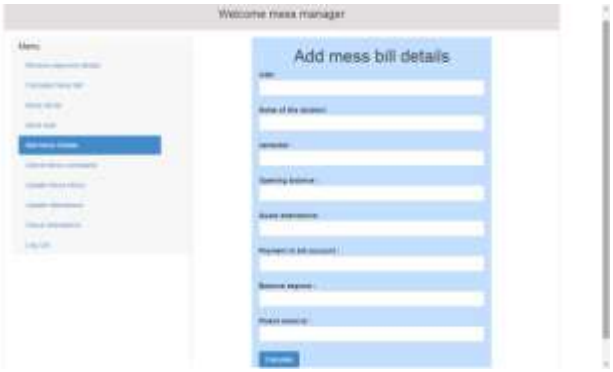


Figure 6 Adding details

E. BILL CALCULATION

Bill details of the student are used to calculate hostel and mess bill. Warden or Mess manager enters ID of the student. After processing this query, name of the student and balance amount is calculated and displayed.

Figure 7 Bill calculation

F. SENDING MAIL

Warden and Mess manager log in using their login id and password. They check the bill details. If any balance is found in either the mess or the hostel bill, they have an option of sending an appropriate mail to the parent.

REFERENCES

[1] James Rumbaugh ,MichaelBlaha , Object Oriented Modelling and Design, edition 2.

[2]"GSHK" http://www.freestudentprojects.com

[3]"HMSReport" https://www.scribd.com/doc/26623621/Hostel-Management-System

Printer Aggregator Model

[1]Aashish,[2]Ankita,[3]Chirag,[4]Prafulla,[5]Manjula Pawar
[1][2][3][4]Student,[5]Lecturer Department of CSE, B. V. B College of Engineering & Technology, Hubli
[1]aashishparbat@gmail.com,[2]akejriwal6@gmail.com,[3]chiragkamat.kamat@gmail.com,[4]prafulladiweseh@gmail.com,[5]manjulap@bvb.edu

Abstract- **Our paper proposes a web service which will provide a single platform to aggregate all the printers in the market. It will give the opportunity to the designers to choose their favorable service providers with the specification of their choice for the printing of their photo books.**

Key words- **aggregator, photobook, printer, professional.**

I. INTRODUCTION

In today's scenario, about 40 percent of world's population uses internet for myriad purposes like news, entertainment, communication etc. [source: Internet World Stats]. Despite, more and more people using internet for selling and purchasing of different kinds of goods and services there is no such platform for the photo book designers to get their photo books printed by their choice of service providers and all the materials.

Photo printing industry is one of the huge industries present in the market. But the industry is scattered. Nowadays there is increase in the number of professional photographers in the market. Everyone wants to print his photos with good quality and cheap cost. PRINTER AGGREGATOR MODEL is a web service that will aggregate all the printers that are available in the market. A photographer can search or print his photos anywhere in the world using this service.

There are almost 11000 printers available in India. But all are scattered. There is a demand of a system which will connect all these printers. There are many aggregator systems in the market such as flipkart for online shopping, make my trip for travel and hotel booking, redbus for online bus booking. But there is no such system for photo printing industry.

Professional photographers who want to print a photobook for their customers will be the main users of this system. Photographers are the professionals who know about the exact size, type and quality of the photobooks they want to print. Photographers can also make the photobook using any editing software they know and get it printed by choosing from the available options provided on the web service. This will give the photographers the freedom to choose the type of the materials and designs of their choice with the minimum cost. They will be able to select the service providers on the basis of the photo book size which will be the basic criteria for the search. They will be able to filter their search results on the basis of different criteria's predefined in the printing industry such as material or tech.

We will be providing an isolated platform for the admin who will be responsible for all the activities happening on the web service and will provide the service providers with the option of getting registered with the web service and use this platform for their business purpose.

Each and every service provider have to register them self first then only they will be able the use the service. The service providers will be free to choose from the predefined specifications for their products and add as many products as they want. They will be provided with a unique id and password which they can modify later. They will be free to change anything about their product at any point of time.

II. REQUIREMENTS

Functional Requirements

1) From the users prospective he should be able to search all the service providers based on the preference of the photo book size selected by him. After getting the results he should be able to filter the results depending on several other parameters such as material etc. He should be able to specify additional requirements like a metal box or a jute

bag to cover his photo book. Before the payment part we are providing the users with the option of paying the bill as a guest user. So he should be able to pay the bill online. Also the user should be able to share his file using any of the provided methods in the web service.

2) From the service providers prospective they should be able to register themselves on the web service. After the registration they should be able to add as many products as they have and be able to change the specifications whenever necessary.

3) From the admins prospective he should be able to provide the new service providers with the password using which they can login for the first time. The admin should also be able to manage everything on the web service.

Non-Functional Requirements

1) Performance: Many users should be able to use the web service simultaneously without any lagging.

2) Safety: As the users are going to use our web service using the browsers and these browsers provides with their own security measures, there is no need for providing extra safety.

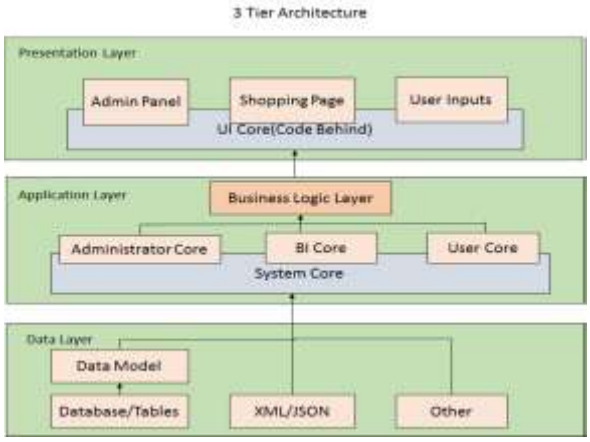
3) Security: The web service should be able to restrict any intruder from altering information about the service providers. Only the authenticated persons should be able to change or modify the attributes about their products. For this a unique id is provided to each registered service provider using which he can log in.

Quality: The user should be able to get the search results instantly. The user interface of the web service should be user friendly. The users should be provided with user manuals and online support. They should be able to send mails for queries.

III. PROPOSED METHODOLOGY

In the proposed system we are providing a single platform from searching a product till printing the invoice. Using our system any printer can easily register with us and can add their products. The can change the details whenever they want and can add new products. The users can easily search for the service providers to get their photobook printed. They will be provided with searching options based on the photobook size and several other criteria. The can easily send the photos through courier to be printed.

IV. ARCHITECTURAL DESIGN

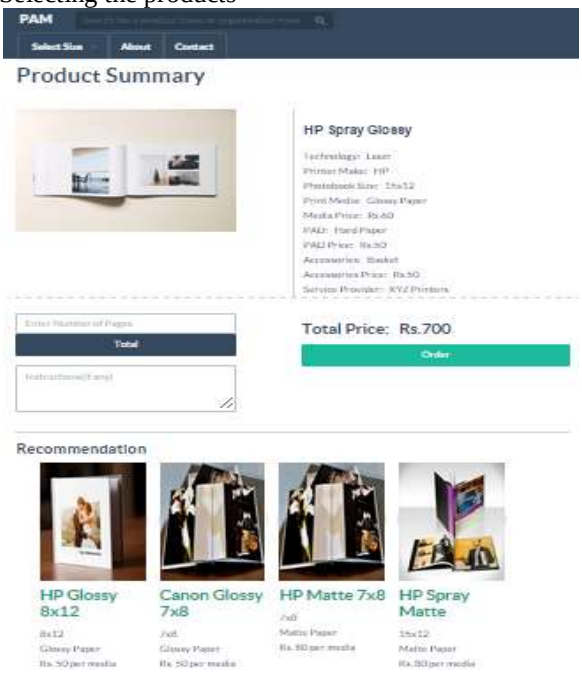


It is a 3 tiered architecture design where there will be a server running all the time and clients who will be accessing the server with a database provided for the storage. The first layer is the presentation layer where we have the admin panel for the admin of the system and the user interface from where the user will be interacting with the system. The second layer consists of the application layer which has the business logic as to how the system is going to work. The third layer is the Database layer which has the all the information stored in it.

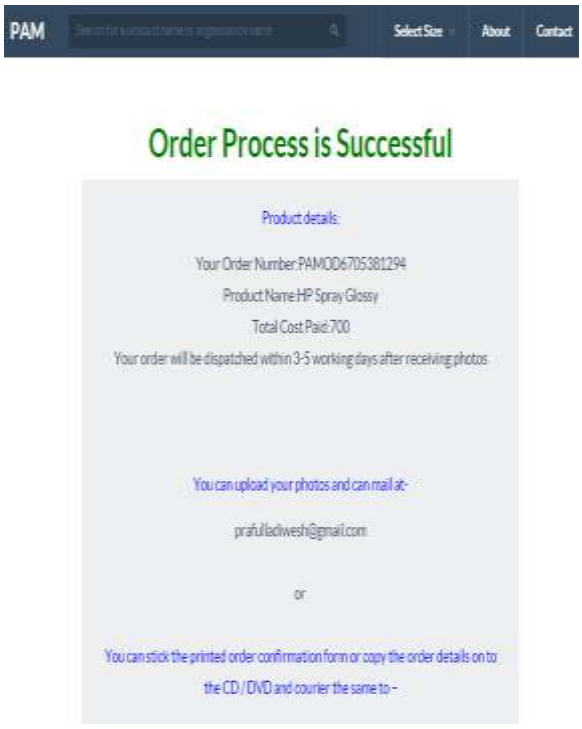
V. USER INTERFACE DESIGN

First page

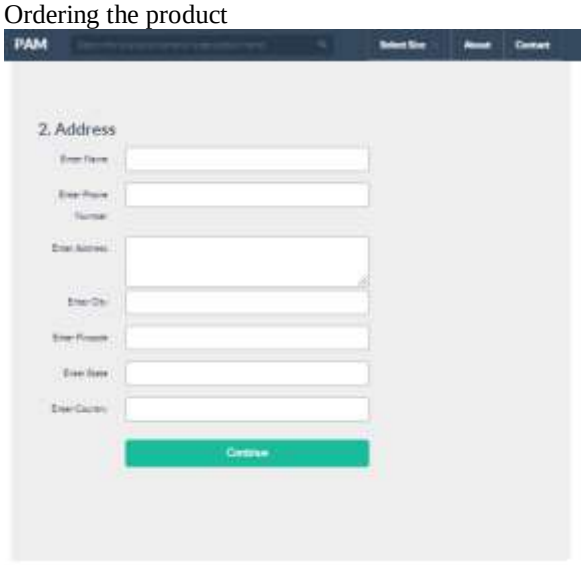
Selecting the products



Providing the total no of pages to be printed



Ordering the product



Filling the Delivery Address

VI. ADVANTAGES

The point that a photographer gets to choose a service provider based upon his choice makes it very convenient. By using our web service the photographer will be able to compare the prices charged by different service providers and also he can get the time that it will take to deliver the product.

Our web service will be running 24*7 which makes it relevant for the users around the world. It is providing the companies with an online platform for selling their services which is much easier and convenient for both the customers

and the seller. The customer need not go to the service provider. Only using a laptop and internet connection he can select where and by whom he wants to get his photobook printed.

VII. constraints

- The system works on e-commerce and cloud for which a constant internet connection is required. Any drop in the internet connection will affect the usage and performance of the system.
- For the transaction part we are including a predefined payment gateway.
- We have to purchase the licensed software if used like oracle for the DB creation.

VIII. References

[1] E-Store Project by Marvel and Home Entertainment version 4.0
www.utdallas.edu/
chung/RE/Team1=Documents=SRS4:0:doc

[2] A Shopping Store Online Database System
http://ieeexplore.ieee.org/xpl/login.jsp?tp=arnumber=6245788url=http

[3] Workflow model of online shopping system
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5622312

[4] FlowDiagram.html for definition of Data Flow Diagram
http://www.agilemodeling.com/artifacts/data
.

[5] http://www.redbus.in

[6] http://www.albumizer.in

[7] http://www.uml-diagrams.org/examples/online-shopping-uml-diagram-example.html

[8] For information on Transactions in ASP.NET
http://samples.gotdotnet.com/quickstart/aspplus/doc/mtstransactions.aspx

Interpretation of Telugu Language through Visual Discriminating Features

[1]G.Shulamithi, [2]M.V.Srikanth
[1]PG Student, [2]Asst.professor
[1]Department Of ECE, [2]Department Of ECE
[1]Gudlavalleru Engg college, [2]Gudlavalleru Engg college
[1]shulamithi.g@gmail.com, [2]sree.02476@gmail.com

Abstract: **In a multi language speaking country like India, a document generally consist of multiple languages. For multiple language text language, multilingual Optical Character Recognition is required to read the multilingual texts. The OCR is to identify existing languages in the document. The main objective is to intend visual clues based practice to identify Telugu text portions of the multiple language text image and separation of machine printed Telugu language with the support of script identification from multiple language script image**

Keywords: **Feature Extraction, V/H Projections, Segmentation, Zonalization**

I. INTRODUCTION

Language detection is a main topic in pattern detection and image processing based automatic document scrutiny and recognition. Humans can identify languages easily by looking at script but not the machines. The purpose of text identification or recognition is to translate human detectable text to machine detectable codes for processing the text to identification of language. Now a days the necessity of electronic libraries are become very essential and presenting a text in any language as automatic translated text into other languages. Detection of the particular language in a document is a primary importance for selection of a specific Optical Character Recognition system for processing multilingual documents.

In India, public required documents contain text words are more than one language. In such instance multi lingual OCR system is essential to read the multilingual text from documents. To arrange a multilingual OCR system function successful, separation of different language portions of the document is necessary before feeding to individual OCR systems.

The existence of multilingual documents in India leads to increase in demand for automatic processing the documents to identification of text in all states in India. Most of the states adopted printing documents in three languages especially like reservation forms, application forms and examination forms. The three languages are state official language, national language and English. We are considering Andhra Pradesh state regional language Telugu for separation from the multilingual image document. As mentioned earlier to arrange a multilingual-OCR system function successful, the multilingual OCR system shall work in two stages. First to Identification and separation of different language texts of the document and secondly feeding of individual language texts derived from first stage to appropriate OCR system. The first stage of the multilingual OCR system and procedures for identification and separation of Telugu text from trilingual input text image is concentrated in this paper. Telugu, Hindi and English text of the image are commonly used in Andhra Pradesh state. Since three languages belong to different scripts and separation of Telugu text based on identification script may be treated as language identification.

II. VISUAL DISCRIMINATING FEATURES OF TELUGU LANGUAGE

Telugu is the official language of Andhra Pradesh and popular language in south India. Telugu is a Dravidian language and is the only language other than Hindi, English and Bengali that is predominantly spoken in more than one Indian state, being the primary language in Andhra Pradesh state and Yanam where Telugu is an official language. Most of the Telugu letters have tick mark shaped structures at the top portion of their letters and also observed that majority of Telugu letters have upward shaped curves are present at the bottom portion of their letters. These properties of Telugu letters are useful in separating

them from other languages.

2.1 Zonalization Text Lines

Text lines of some Indian languages are classified into three zones. In this paper the zonalization procedure proposed by Pal and Choudhuri [2] is used for zonalization is useful for feature extraction. A sample text line consists of English languages is classified into three zones. An imaginary line where the first lowermost black pixels of letters of a text line lies is called as lower line. An imaginary line, where the maximum number of uppermost black pixels of letters of a text line lies is called a mean line.

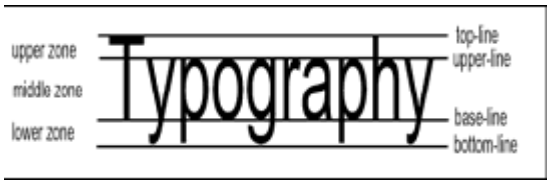


Fig. 1 Imaginary Line

An imaginary line, where maximum number of lowermost black pixels of characters of a text line lies, is called a base line. The upper zone covers the portion between the upper line and the mean line of a text line. The middle zone covers the portion between the mean line and the base line of a text line. The lower zone is the portion between the baseline and the lower line. The height of the text line is the normal distance between the upper line and the lower line. The X-height is the distance between the mean line and the base line.

2.2 Model

The proposed model is on the four visual features such as

i. Horizontal lines
ii. Vertical lines
iii. Variable sized blocks
iv. Number of components present in each block. Two assumptions are made in our proposed model

a) Input image is a machine printed image with standard font for Telugu, Hindi and English text lines.
b) Every text line must have at least four words and text line may have different font sizes but the font and font size within a text line is same.

2.3 Supportive Knowledge Base

Knowledge base plays an important role in recognition of any pattern and knowledge base is a repository of derived information. A supportive knowledge base is constructed for each specific class of patterns, which further helps during decision making to arrive at a conclusion. The percentage of the presence of the four features for each text of the three languages Telugu, Hindi and English, are practically computed using sufficient data set. Based on the experimental results, a supportive knowledge base is constructed to store the percentage of the presence of the four visual features. The technique of obtaining the four visual features from the input image through experimentation is explained below.

2.4 Feature 1-Horizontal lines

In the binary image of all text lines, if there are continuous one's in a row greater than the horizontal threshold value (Horizontal threshold value is calculated for all text lines. Horizontal threshold value = 75% of the X- height of that text line), then such continuous one's are retained resulting in a horizontal line and if there are no continuous one's greater than the horizontal threshold value, then such one's are changed to zeroes. A component has a horizontal line-like structure, if a black run length (sequence of continuous one's) of that component is greater than the horizontal threshold value of that text line.

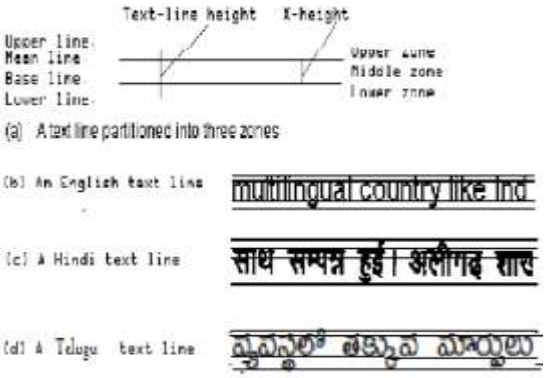
2.5 Vertical lines

In the binary image of all text lines, if there are continuous one's in a column greater than the vertical threshold value (Vertical threshold value is computed for all text lines. Vertical threshold value = X-height of that text line) then such continuous one's are retained resulting in a vertical line and if there are no continuous one's greater than the vertical threshold value, then such one's are changed to zeroes. A component has a vertical line-like structure, if a black run length (sequence of continuous one's) of that component is greater than vertical threshold of the text line.

2.6 Zonalization process

A text line can be considered as being collected of three zones. The zones are upper zone, middle zone and lower zone as shown in Fig 1. These zones are delimited by four virtual lines. The virtual lines are top-line, upper-line, and base-line and bottom-line. All text lines have at least a middle zone. The upper zone depends on capital letters and letters with ascenders, like h and k. The

lower zone depends on letters with descended, like g, p and y. This structure allows the definition of four kinds of text line.



III. Line wise Identification Model

In this model, a line level identification /detection model shown in Fig 2 will accurately detect and separate different language text portions of Telugu, Hindi and English text lines of the input image and group the portion of the Telugu text in the image. If the input image consists of other than these three languages are treated as separate class as OTHERS.

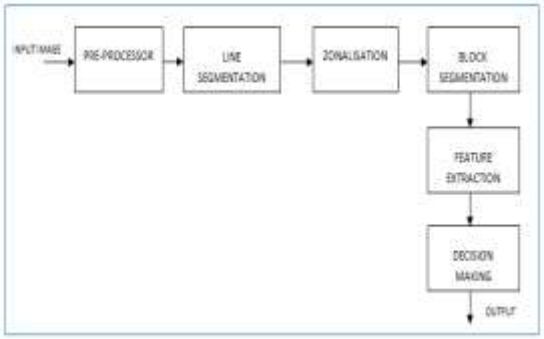


Fig 3. Line wise Identification

The discriminating features horizontal lines, vertical lines, variable sized blocks and blocks with more than one component are extracted from processed image and compared with the values that are stored in the knowledge base, to arrive at a decision regarding type of the text language. The different steps consists in the proposed model are as follows.

3.1 Preprocessing

The input image document is preprocessed i.e., noise removed, smoothing done, skew compensated and binarized.

3.2 Line segmentation

To segment the image to a number of text lines, the valley of the horizontal projection computed by a row-wise sum of black pixels is used. The position between two successive horizontal projections where the histogram height is least denotes one boundary line. by these boundary lines, image is separated into several text lines.

3.3 Zonalization

Every text line in image is classified into three zones - upper zone, middle zone and lower zone as shown in Fig 1.

3.4 Block segmentation

The zonalized text line, upper line and lower line is from zonalization is used as two boundary lines for every text line. Then every text line is scanned vertically from its upper line to reach its lower line without touching any black pixels to get a boundary line. Such characters within each boundary lines lead to a stream of blocks.

3.5 Feature extraction

Horizontal line detection: From the input image, the horizontal lines are obtained. Then the percentage of the presence of these horizontal lines for each text line is computed and compared with the stored values in the knowledge base.

Vertical line detection: From the input image, the vertical lines are obtained. Then the percentage of the presence of these vertical lines for each text line is computed and compared with the stored values in the knowledge base.

Variable Sized blocks: The size of the blocks of each text line is calculated by taking the ratio of width to height of each block. Then the percentage of equal and unequal sized blocks of each text line is calculated.

Blocks with more than one component:

The percentage of the number of components present in each block of every text line is computed.

3.6 Decision making

Condition-1: If 90% of the horizontal lines on the mean line is greater than two times the X-height of the corresponding text line; if there are 80% of vertical lines in the middle zone and also if 70% of the blocks have width greater than two times the X-height, then such portion of the document is recognized as Hindi language.

Condition-2: If 65% of the horizontal lines on the mean line is greater than half of the X-height of the corresponding text line and if there are 40% of

unequal sized blocks in a text line, then such portion of the document is recognized as Telugu language.

Condition-3: If there are 80% of vertical lines in the middle zone greater than half of the text line height and if 80% of the blocks are equal in size, then such portion of the document is recognized as English language.

Condition-4: If the output image does not belong to any of the above three classes, then such portion of the document is grouped into a separate class as OTHERS.

3.7 Limitations

One of the limitations of this algorithm is that if a text line contains words in more than one language, then such a text line is classified as the language type of the majority of words in that text line. Another limitation of this method is that if a text line contains some numerical Figs in addition to the text words, then the entire text line is grouped into the language type of the remaining words, without identifying and grouping numerical Figures into a separate category.

IV. WORD-WISE IDENTIFICATION MODEL

In the previous section, we have suggested text language identification at the line level, with the assumption that the input document contains text lines in one and only one language. In this section, we suggest another model of language identification at word level, to overcome the limitations of the previous method. In a more practical scenario as found in majority of business applications, most of the documents such as language translation books, electric bills, application forms, marks cards and reservation forms, are comprised of several text lines, in which a text line itself could contain words in two or more languages.

To identify the type of the language at word level using some visible discriminating features, which are relatively strong enough to identify the text language, without reading the complete content of text line. An attempt is made to develop algorithm that should accurately identify and separate text words of Telugu, Hindi and English language portions of the document. different stages involved in the word level identification model shown in Fig3 are explained as follows.

4.1 Preprocessing

The input image document is preprocessed i.e., noise removed, smoothing done, skew compensated and binarized and followed by line segmentation.

4.2 Word Partitioning

Each text word is partitioned into three zones - upper zone, middle zone and lower zone as explained in zonalization to get upper line and lower line as two boundary lines for every text word.

4.3 Word segmentation

All text lines are segmented into words by detecting the valleys of the vertical projection. If the width of the valley is greater than the threshold value (Threshold value = two times the inter character gap), then a vertical line is drawn at the middle of the columns with zero values (Inter word gap). Using these vertical lines, words within a text line are segmented.

4.4 Block segmentation

From the partitioned text word, upper and lower lines are used as two boundary lines for every text word. Then, every text word is scanned verticallyfrom upper line to lower line without touching any black pixels to get a stream of blocks. Thus a block is defined as a rectangular section of the text words that has one or more characters with more than one component.

4.5 Block size evaluation

The size of each block of every text word is calculated by taking the ratio of width to height of each text word. Then the percentage of equal and unequal sized blocks of each text word is calculated.

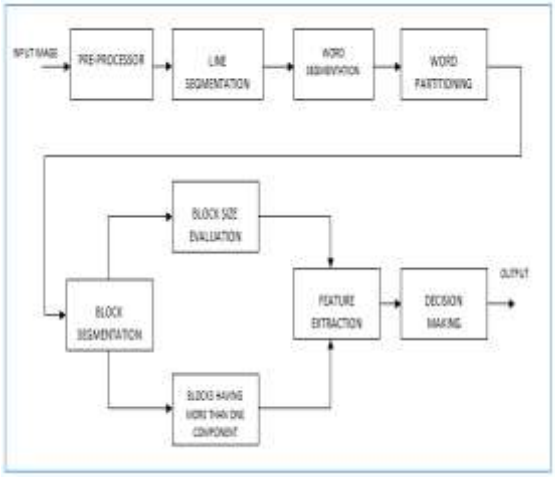


Fig 4. Word wise Identification

4.6 Blocks having more than one component

The number of components (a connected component is one in which the pixels are aggregated by an 8-connected points analysis) present within each block is calculated using 8-neighbour connectivity. Then the percentage of the occurrence of blocks having more than one component is calculated.

4.7 Decision making

Level-1: If the length of the horizontal line on the mean line is greater than two times the X-height of the corresponding text word, if there are vertical lines in the middle zone, if the block has width greater than two times the X-height and also if the word/block contains only one component, then that text word is identified as a text word in Hindi language.

Level-2: If the length of the horizontal line on the mean line is equal to the x-height of the corresponding text word; if there are 70% of unequal sized blocks in the output image and also if 30% of the blocks contain more than one component, then that text word is recognized as a text word in Telugu language.

Level-3: If there are vertical lines in the middle zone greater than half of the text word height; if a text word contains 70% equal blocks in size and also if a text word contains 90% of the blocks having only one component, then that text word is identified as a text word in English language.

Level-4: If a text word does not belong to any of the above three levels, then such a text word does not belong to the above categories and hence it is grouped into a separate class as OTHERS.

4.8 Limitations

One of the limitations of this method is that, if a text word contains alphanumeric values, then such text word is misclassified into any one of the four groups depending on the majority features present in that word

V. RESULT

Input Image

Following is the input image as shown in Fig 5 for processing to identification and separation telugu language text from trilingual image

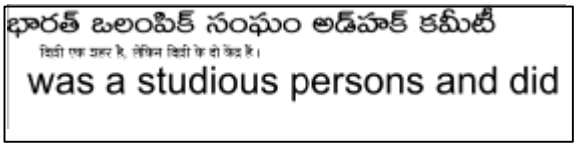


Fig 5. Input Image

Line Segmentation

The above Fig 6 shows the line segmentation of first line which is segmented from the input image.

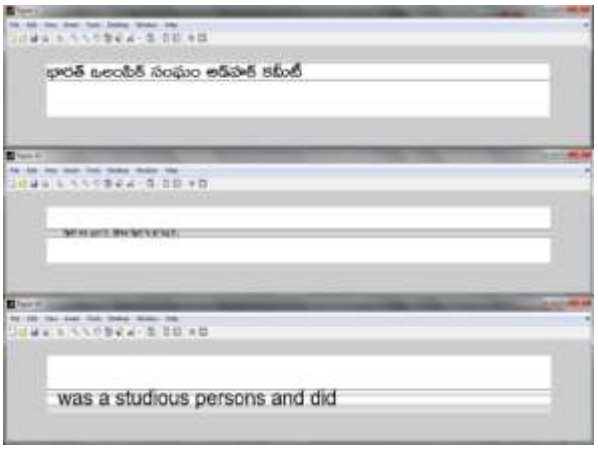


Fig 6. Line Segmentation of Multilingual language

Feature Extraction

Fig.7 shows the result of feature extraction of first line, second line and third line where for the first line the up component, bottom component and middle component is shown respectively. It is achieved by zonalization. The content in the upper zone is called the up component, the content in the lower zone is called the bottom component, the content in the middle zone is called the middle component these components are useful in extracting the features from the input.

Fig 7. Feature Extraction

Horizontal Projection

The above Fig 8 shows the horizontal projections of all the text lines in image which are horizontally projected from right to left.

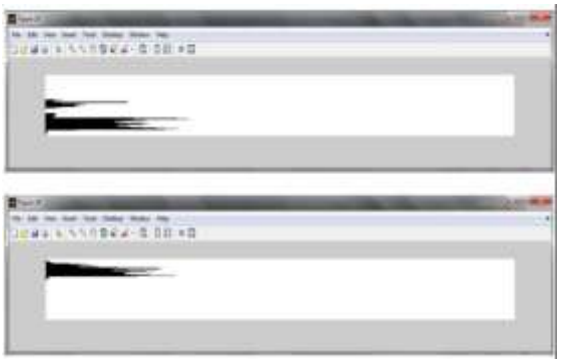


Fig 8. Horizontal Projection

Vertical Projection

The above Fig 9 shows the horizontal projections of all the text lines in image which are horizontally projected from right to left.

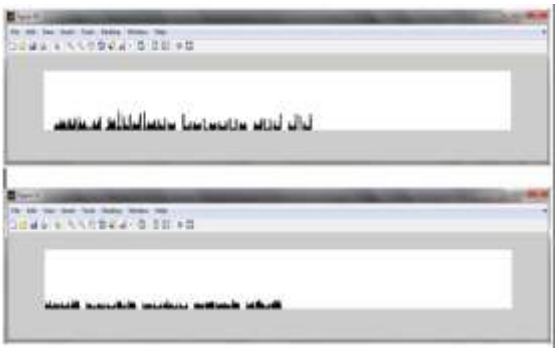


Fig 9. Vertical Projection

Word Segmentation

The Fig 10 shows the result of word segment for the first line. The Segmented line is again segmented into words and these segmented words will be used in separation of Telugu language from trilingual language text image.

Fig 10. Word Segmentation

OUTPUT

The below Fig 11 is the output image. It contains only telugu language text which is identified and separated from the trilingual language input image

Fig 11. Separated Telugu Text

CONCLUSION

The presented line-wise and word wise identification models to identify and separate Telugu language text words from Indian multilingual machine printed documents. The simulated results indicates the two methods are good enough to identify and separate the Telugu text from multilingual image document, which further helps to feed individual language regions to specific OCR system. Proposed model fails in the case of blurred image when given as input.

ACKNOWLEDGEMENT

We would like to thank to Dr. M.KamaRaju for the opportunity on working on such a beneficial project and special thanks to Mr. Dr. Tilak for his guidance, advice, and continuous support for the duration of this project.

REFERENCES

[1] S.Basvaraj Patil, N.V.Subba Reddy, "Character script class identification system using probabilistic neural network for multi-script multi lingual document processing", Proc. National Conference on Document Analysis and Recognition, Mandya, Karnataka, 1-8, (2001).

[2] U.Pal B.B.Choudhuri, "Automatic Separation of Words in Multi Lingual multi Script Indian Documents", Proc. 4th International Conference on Document Analysis and Recognition, 576-579, (1997).

[3] S.Chanda, U.Pal, "English, Devanagari and Urdu Text Identification", Proc. International Conference on Document Analysis and Recognition, 538-545, (2005).

[4] U.Pal, S.Sinha, B.B.Choudhuri, "Word-wise script identification from a document containing English, Devanagari and Telugu text", Proc. 2nd National

Conference on Document Analysis and Recognition, Karnataka, India, 213-220, (2003).

[5] M.C.Padma, P.Nagabhushan, "Identification and separation of text words of Kannada, Hindi and English and languages through discriminating features", Proc. 2 National Conference on Document Analysis and Recognition, Mandya, Karnataka, 252-260, (2003)

[6] U.Pal, B.B.Choudhuri, "Automatic Identification of English, Chinese, Arabic, Devanagari and Bangla Script Line", Proc. 6th International Conference on Document Analysis and Recognition, 790-794, (2001).

[7] R.C.Gonzalez, R.E.Woods, Digital Image Processing Pearson Education Publications, India, 2002.

[8] U.Pal, S.Sinha, B.B.Choudhuri, "Multi-Script Line Identification from Indian Documents", Proc. 7th International Conference on Document Analysis and Recognition (ICDAR 2003) vol. 2, 880-884, 2003.

Improved Mechanical Properties and Microstructure of Mechanical Vibration Moldduring Solidification

Kalla Girija
M.Tech, Machine Design, Mechanical Engineering Department, Chaitanya college of engineering, JNTUK, Visakhapatnam Andhra Pradesh India.
kallagirija@gmail.com

*Abstract:***Production is increased now a day'sdemands error free castings. Many researchers are following different methods to improve castings. The mechanical properties of a cast depends on the pouring, solidification of the cast. In the current paper the mechanical properties of aluminum alloy are investigated with attachment of mechanical vibration to the mold during pouring .The properties are then compared with vibration and without vibration. The experimental results shows significantly improvement in mechanical properties with vibration when compared to without vibration.**

Keywords: **AluminumAlloys, Mechanical Vibration, Mechanical Properties and Microstructure**

I. INTRODUCTION

Aluminum –Copper-Silicon based alloys are the most widely used nonferrous alloys. The properties of this alloy are heat treatable, high strength at room temperature**[1-4].** The alloy whose microstructure have been subjected to grain refinement possesses many advantages such as good mechanical properties thermal stability and high strain rate. The grain refinement can be attained either when external forces are applied to increase the fluid flow during solidification mechanical stirring or electromagnetic stirring of melt. The use of mechanical vibration during pouring and solidification is one of the technique to improve grain refinement[5-13]. Therefore this paper have been investigated by imparting mechanical vibration to the mold and mechanical properties are compared with and without vibration

II. EXPERIMENTAL PROCEDURE

2.1. Experimental Method

The device which converts electrical signals to mechanical vibration with small amplitude is called mechanical vibration table.

The mechanical vibration table consists of mold holding plate, A.C motor, electric mass, four springs, power source, A.C Varying transformer, base plate, sensor, vibration measuring instrument, display unit. the mold is attached tomold holding plate of the vibration table .when the power is on the motor which to attached to the mold holding plate rotates with eccentric mass so that vibration is created which transfers to the mold, then vibration is measured by using vibration measuring instrument. The use of springs is to hold the vibration and reflect it to the mold vibration. The frequency of the vibration is varied by using A.C. varying transformer by cutting the voltage of source or motor.

2.2 Material and Procedure

The chemical composition of the Aluminum Copper alloy was selected as shown in **table 1**. The mold was prepared by using green sand and preheated to 200□C.The melt was produced in an open hearth furnace by heating around 650^0C-800^0C and the slag was removed by adding NaCl. Themoldplaced on the vibration table. The vibration table is switched ON before the pouring and switched OFF after 50 seconds of pouring. The specimens are made under different frequency such as 0HZ,10HZ,20HZ,30HZ,40HZ,and 50HZ. The cylindrical rods of 20mm of diameter is casted for each vibration frequency.

Table1: Chemical Composition of AlCuSi Alloy in Wt.%

Al	Cu	Si	Fe	Mg	Zn	Ti	Ni	Sn	Pb	Mn
91.56	2.39	5	0.39	0.29	0.18	0.04	0.02	0.01	0.03	0.03

 ISBN: 9788192958093

2.3. Mechanical Testing's

After casting with and without vibration, the specimens were made according to the ASTM Standards. Three different tests were done for each specimen are Tensile test, Hardness test and Impact test on Universal Testing Machine, Brinell Hardness Testing Machine and Charpy Testing Machine respectively. The results has been tabulated in the **Table 2**.

Table 2:Experimental Values For Various Tests

Frequency (Hz)	Ultimate Tensile Strength (N/mm2)	Brinell Hardness Number (BHN)	Impact Strength (J/mm2)
0	130.0699	27.6608	5.4
10	141.1134	32.0356	6
20	148.6202	35.4125	6.2
30	154.1274	39.3356	6.3
40	151.3214	41.2564	6.4
50	159.5301	42.4832	6.6

2.4. Metallography Test

The five specimens under different frequency are first grinded on belt grinding machine. Then the specimens are polished with silicon carbide emery paper under different grades. After polishing, disc polishing is done and cleaned with water and fine polishing was done by surgical cotton.The sample was etched with Keller's reagent (H2O: 95 ml, HNO3: 2.5 ml, HCl: 1.5 ml, HF (48%): 1ml) for 10 seconds than etched samples were cleaned in acetone and dried in hot air. Then these samples were undergone microstructure testing's on optical microscope.

III. RESULTS AND DISCUSSION

3.1. Microstructure Analysis

Microstructure of an Al-Cu-Si alloy is shown in **fig.1**. By observing the microstructure of specimen without vibration in fig.1 (f). Having coarse grain due to the presence of silicon flakes and intermetallic compounds of copper and aluminum. But while observing the microstructure of the specimens with vibration in fig.1 (a-e). Having fine grains due to the refinement of silicon flakes and due to the decrease of intermetallic compounds of aluminum and copper.

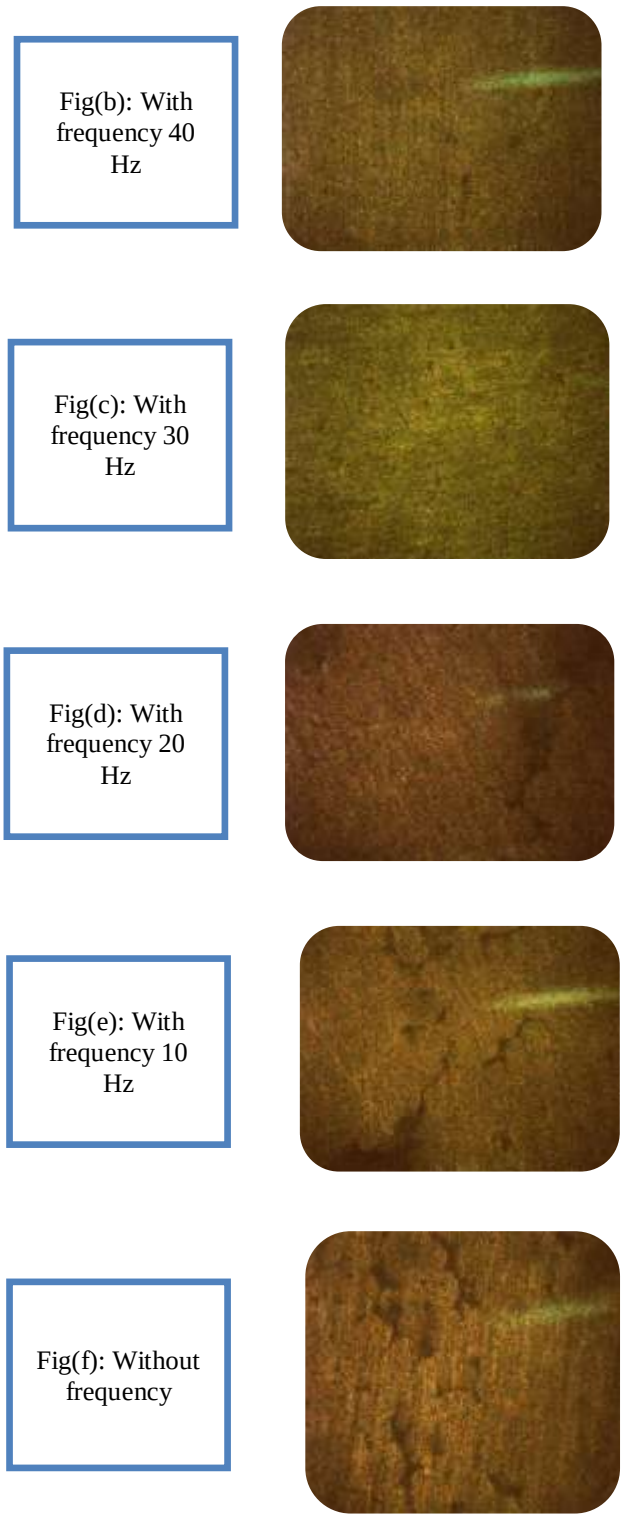


Fig.1. (a, b, c, d, e, f) Optical Microstructure with 300X magnification.

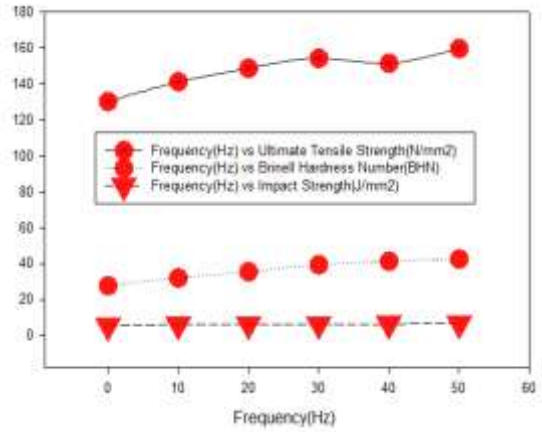


Fig.2. plotting of variation of properties with and without frequency.

3.2. Analysis of Mechanical Properties

The values obtained in **table 2**was plotted Graphs versus frequency and without frequency shown in **fig.2.** By observing the curve of frequency vs ultimate tensile strength, the ultimate tensile strength is increased with mechanical vibration when compared to ultimate tensile strength without mechanical vibration and increases with increase in frequency of mechanical vibration.

Similarly, the impact strength and hardness also improved with mechanical vibration.

CONCLUSIONS

The mechanical properties and microstructure with mechanical vibration and without mechanical vibration during pouring has been investigated and the major conclusions are as follows:

1. The microstructure of the specimen without mechanical vibration looks like coarse grain structure such that there is no grain refinement and hence the ultimate tensile strength, hardness and impact strength values are less.
2. After imposing mechanical vibration to the mold, the specimens having fine microstructure and hence there is a grain refinement. Therefore the properties are improved when compared with specimens without vibration.
3. The final conclusion is that if the casting is done with mechanical vibration will have improved mechanical properties and high strength.
4. Establishment of mechanical vibration is very cheap and easy to install in foundry shops and easy to handle when compared to other methods.

ACKNOWLEDGMENTS

This work was fully and completely supported by Dr. Satynarayana, B.E, M.E, PGDAS, FIE, FIIP, MISTE, C(ENGG), born in west Godavari district andhrapradesh, INDIA. He is also a council member of IEI. He has 35 years of industrial experience and 11 years teaching experience as professor in mechanical engineering department in Chaitanya engineering college, Visakhapatnam.

REFERENCES

[1] W. Khalifa, Y. Tsunekawa and M. Okamiya, J Mater Process Technol 210 (2010) 2178.

[2] S. Janudom, T. Rattanochaikul, R. Burapa, S. Wisutmethangoon and J. Wannasin, *Trans NonferrousMet Soc China* **20** (2010) 1756.

[3] Z.Y. Wang, M.L. Hu and H.Y. Xu Mater Charact 62 (2011) 925.

[4] Y.G. Li, Y.Y. Wu, Z. Qian and X.F. Liu, *Mater Sci Eng A* **527** (2009) 146

[5] A. Ohno: Kinzoku Gyouko-gaku, (Chijin Shokan, Tokyo, 1973) pp.60–64. [in Japanese]

[6] T. P. Fisher: Brit. Foundryman 66 (1973) 71–84.

[7] N. Abu-Dheir, M. Khraisheh, K. Saito and A. Male: Mater. Sci. Eng .A 393 (2005) 109–117.

[8] Y. Osawa and A. Sato: J. JFS 72 (2000) 733–738. [in Japanese]

[9] O. V. Abramov: Ultrasonics 25 (1987) 73–82.

[10] J. Dong, J. Cui, X. Zeng and W. Ding: Mater. Lett. 59 (2005) 1502–1506.

[11] B. Zhang, J. Cui and G. Lu: Mater. Sci. Eng. A 355 (2003) 325–330.

[12] Y. Mizutani, S. Kawai, K. Miwa, K. Yasue, T. Tamura and Y. Sakaguchi: Mater. Trans. 45 (2004) 1939–1943.

[13] Y. Mizutani, Y. Ohura, K. Miwa, K. Yasue, T. Tamura and Y. Sakaguchi: Mater. Trans. 45 (2004) 1944–1948.