7-1-2012

Proceedings of International Conference on Advancement in Engineering & Technology

Dr. Sushanta Panigrahi

Proceedings

of

International Conference

On

Advancement in Engineering & Technology

(ICAET-2012)

1st JULY, 2012

BANGALORE, India

Organized by:

ASTAR, INDIA

About *ICAET*-2012

International Conference on Advancement in Engineering and Technology (ICAET). ICAET aims to bring together researchers, scientists, engineers, and scholar students to exchange and share their experiences, new ideas, and research results about all aspects of Computer Science, Electronics and Communication engineering, and discuss the practical challenges encountered and the solutions adopted. The conference will be held every year to make it an ideal platform for people to share views and experiences in Computer Science, Electronics and Communication Engineering and related areas.

ICAET invites original contributions on topics covered under broad areas such as (but not limited to):

- Scientific and engineering computing
- Problem-solving environments
- Advanced numerical computation and optimization
- Complex systems: modeling and simulation
- Signal and System theory, Digital signal processing
- Architectures and computation models, compiler, hardware and OS issues
- Memory system, I/O, tools, programming environment and language supports
- Microwave theory and techniques, Radar, Sonar
- Artificial intelligence
- Visualizations and virtual reality
- Hardware/software co-design and VLSI support
- Cluster, grid and web-based computing and simulation
- Education in computational science and engineering
- Related applications

The conference designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these disciplines of engineering. We received a great response from all parts of country and abroad getting more than 100 papers and peer review 27 papers got selected for the presentation and publication in the proceeding of the conference. I sincerely thank all the authors for their invaluable contribution to this conference. I am indebted towards the reviewers and Board of Editors for their generous gifts of time, energy and effort.

TABLE OF CONTENTS

Concept of Adaptive Front Light Leveling Algorithm

Aryalekshmi. B. N

Sree Buddha College of Engineering, Pattoor P.O, Alappuzha Dist, Kerala, India
E-mail : arya9ohm@yahoo.co.in

Abstract – Adaptive front light system/ advanced front light system (AFLS) is the development trend of lighting system in motor vehicles. AFLS can swivel and level the headlamp beam pattern based on sensor signal inputs including steering angle, vehicle speed, chassis height, etc. As a result, the driver can get optimal road illumination effects. Automatic headlamp leveling system work to keep light parallel to the road surface regardless of the vehicle's tilt. A vehicle may tilt as a result of a standstill event, such as boarding passengers, loading the trunk or even while driving, the vehicle tilt changes during breaking or acceleration. In all these cases, the headlamps must be maintained level, with the road way. Headlamp leveling systems correlate their adjustment angles based on a variety of sensor data in particular suspension compression data from the front and rear axles. This paper presents the concept of leveling algorithm for adaptive front lighting system. MATLAB-SIMULINK software will be used for the development of leveling algorithm. An auto-code will be generated from this SIMULINK model and it will be integrated into AUTOSAR application interface. AUTOSAR (Automotive Open System Architecture) is a worldwide development cooperation of car manufacturers, suppliers and other companies from the electronics, semiconductor and software industry. Since 2003 they have been working on the development and introduction of open, standardized software architecture for the automotive industry.

Keywords - *AFLS, Leveling algorithm, AUTOSAR*

I. INTRODUCTION

One of the most important topics in the world is the research of safety driving of vehicle. The poor illumination during driving at night is the main cause for the high rate of traffic accidents. The research in literatures [1-2] indicates that more than 80 percent of all road traffic accidents occur in darkness and bad weather, it also indicates that a driver of 50 years old needs 3 times of illumination than a driver of 20 years old. About 80-90% of the information that humans perceive is obtained by vision, therefore for safe driving good vision is very essential. Car headlight is one of the important parts of motor vehicles, and it serves as a guarantee for safety driving. One of the most important tasks of modern automotive lighting technology is to support human perception at night and under adverse weather condition. In typical traffic situations the contrast sensitivity and visual acuity as well as the speed of perception and danger recognition are strongly dependent on the ambient light distribution. False or missing adaptation of driving speed to visibility is one of the main parameters that cause failure or accidents. In today's headlights, the driver can choose between dipped beams, main beams and fog lights based on the traffic situations and environmental conditions. Since these conditions change continually, the driver's attention is required for adapting the light distribution. This should be done automatically by an ideal lighting system by taking into account information on the present traffic and environmental situation as well as information on the driver's preference.

Adaptive front lighting system is an electro-mechanical system that control and adjust the headlights of a vehicle to suit different road and driving situations. It helps to improve visibility for the driver and so as to achieve a significant increase in road safety and driving comfort. The headlights mainly used in an automobile are to provide enough road illumination at night travel. The essential factors for safety driving are the brightness and the illumination direction of headlight [3]. Most vehicles apply fixed front light system to provide illumination for the front road, but it doesn't allow vehicles to project front-light beam onto road properly when making sharp acceleration or deceleration or cruising up and down hills, and resulting in increased incidence of accidents.

Automotive embedded system in the automotive domain means, automotive electronic system that consists of ECUs that are interconnected by a communication network. AUTOSAR [14] (AUTomotive Open System Architecture) defines embedded software as ECU software that specifies standards of software architecture, methodology and meta model for ECU software. AUTOSAR software is implemented in the form of reusable and modularized components that can be widely interchanged and deployed on a large variety of hardware platforms. Thus the overall efficiency of the development process is improved, and the maintenance of software is simplified throughout the whole software life cycle.

As the implementation and validation methods of application software are not mentioned in the AUTOSAR, model based development and tests are required for improving the efficiencies in the application software validation and verification process.

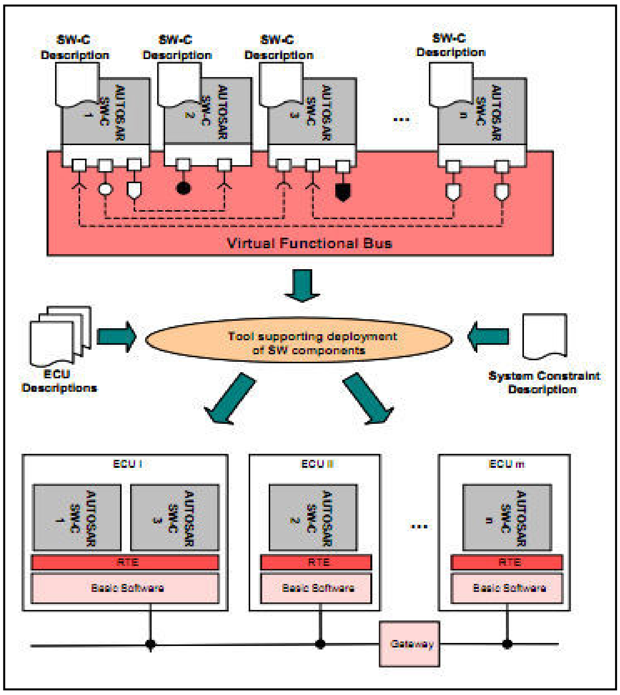


Fig 1: AUTOSAR Technical Overview [13]

The technical overview [13] of AUTOSAR version 3.1 is shown in figure 1.

- AUTOSAR SW-Cs: encapsulate the application running on to the AUTOSAR background.
- SW-C description: the necessary information for the SW-Cs integration from interfaces and other components.
- VFB-Virtual Function Bus: Enables virtual integration of SW-Cs and sums of the communication mechanisms.
- System Constraint and ECU descriptions are independent of SW-C descriptions which are used to integrate SW-Cs into an ECUs Network.
- Mapping all descriptions together including configuration and generation of RTE and basic software.
- RTE- Runtime Environment: VFB functionality implementation on a specific ECU.
- Basic software: providing the functional infrastructure on an ECU.

This paper presents the concept of leveling algorithm for adaptive front lighting systems and also how its software component could integrate into AUTOSAR environment. The organization of the rest of the paper is as follows. In section 2 basics of Adaptive Front Lighting System is explained. In section 3 concept of leveling algorithm is given. In section 4 integration of leveling algorithm for AFLS into AUTOSAR is explained. The paper ends with conclusion.

II. ADAPTIVE FRONT LIGHTING SYSTEM

Adaptive headlamp control is a key factor for enhancing visibility at night. AFLS contains adaptive control strategies which adapts to the different changes under different speed, road and weather conditions. At night, reduced visibility and driver fatigue affect the driving ability and are often the cause for accidents. Optimal lighting is crucial to increase traffic safety and to avoid accidents. The architecture of Adaptive Front Lighting System (AFLS) is shown in figure 2. It shows that the Electronic Control Unit (ECU) processes inputs and decides what is the best bending angle, beam level and beam pattern for that particular driving condition. The ECU sends commands to the Driver Board via LIN messages. Driver Board processes the command and drives the relevant steppers as per the command, thereby achieving the required leveling, swiveling and beam pattern.

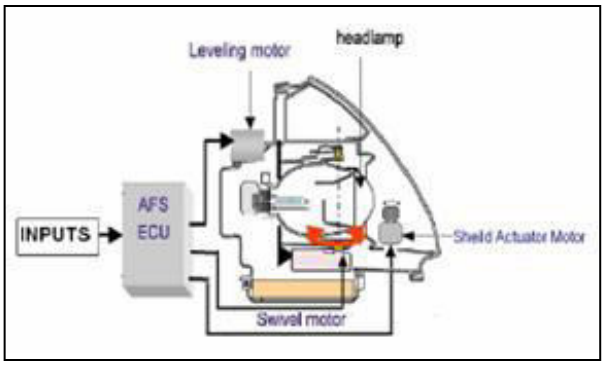


Fig. 2 : Architecture of AFLS

Dynamic bending lights help to optimize vehicle lighting on winding roads by turning the headlights horizontally in order to ensure that the headlights follow the curvature of the road ahead. But vertical motion of the vehicle chassis also affects the illumination of the road ahead.

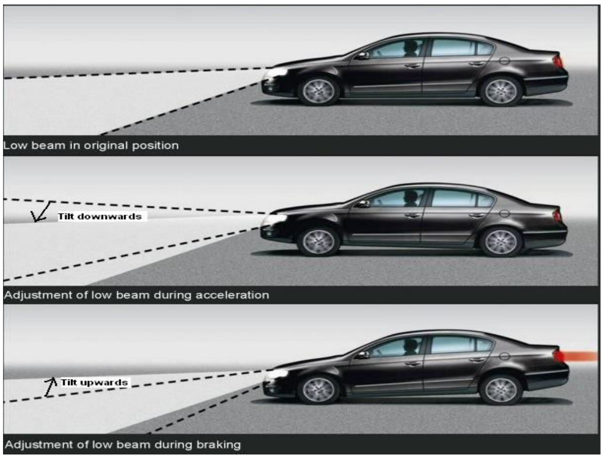

Low beam in original position
Tilt downwards
Adjustment of low beam during acceleration
Tilt upwards
Adjustment of low beam during braking

Fig 3: (a) Low beam in original position
(b) Adjustment of low beam during acceleration
(c) Adjustment of low beam during braking

Possible causes for vertical motion are as follows:

- Adding/removing passengers.
- Placing/removing heavy load in the trunk.
- Hard acceleration/deceleration.
- Bumps or potholes in the road.

When the car accelerates rapidly and the vehicle rear axle center of gravity drops, front axle lifts, to cause the brightness which are projected onto the ground is obviously inadequate; When the car decelerates hardly and the vehicle rear axle center of gravity lifts, and front axle drops, causing the brightness which are projected onto the ground is near, so that drivers can not see the front traffic. The vehicle headlight needs to be adjusted according to corresponding strategy to maintain the level with roadway [4,5].

There has been wide range of studies in the field of headlamp leveling [6-12]. A lot of improvement has been introduced in this headlamp leveling field.

III. LEVELING ALGORITHM

The Head Lamp Leveling unit may be broadly divided into two categories based on the leveling device used for the Headlamps. They are Static leveling and Dynamic leveling systems. The Static and Dynamic leveling systems adjust vehicle light beam's vertical inclination automatically, for the change in vehicle load conditions. In static leveling actuators with D.C. motors (L/D) are used to adjust the headlamps. Intelligent Stepper Motors (ISM) is used to adjust the headlamps in dynamic leveling.

Static Leveling operates in the following conditions:

- When the vehicle is standing
- When the vehicle is moving with constant speed.

Dynamic Leveling System operates in the following conditions:

- When the vehicle is standing.
- When moving with constant speed.
- When the vehicle got acceleration or deceleration.

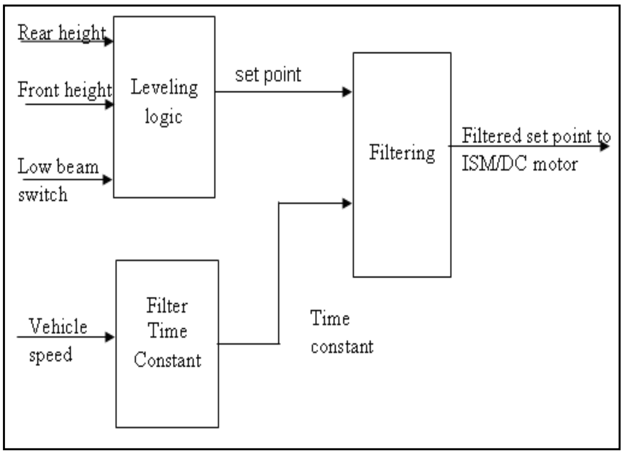


Fig. 4: Basic Block diagram for Static and Dynamic Leveling

The headlamp leveling functionality will be enabled if both AFLS and Low beam switches are ON and also when the vehicle speed is greater than a minimum value. The headlamp leveling module will first checks for the failure of its inputs and outputs. If a failure is detected it will invoke the headlamp leveling fail-safe sub module. The normal operation will continue by invoking the headlamp leveling normal module if there is no failure. There will be two sub modules for normal operation namely, Static Leveling and Dynamic Leveling. Each of these systems is further divided into Leveling Logic, Filter Time Constant and Filtering. The basic block diagram of both Static Leveling and Dynamic Leveling modules are shown in figure 4.

A. Leveling logic

This sub module receives the front and rear height values from two height sensors which are offset in the longitudinal direction of the vehicle body for measuring vehicle body pitch angle in the form of a level difference. The target value or set point will be calculated from this level difference and will be given to the filter. In Static leveling only Rear vehicle height will be the input. This rear vehicle height is converted to set point using linear interpolation table and that value will be given to the filter block. The Dynamic headlamp leveling also reacts to inclination changes due to acceleration and brake actions. Here both front and rear vehicle height will be taken and the effective vehicle height difference will be calculated. This value will be converted to set point using linear interpolation table and will be fed to the filter block.

B. *Filter Time Constant*

This sub module receives vehicle speed as the input and calculates the filter time constant.

In static leveling the following cases are handled:

- Vehicle is at rest: Time constant set to time constant under rest.
- Vehicle is moving with constant speed: Time constant set to time constant under constant speed.

The algorithm for filter time constant calculation in Static leveling system is as follows:

a) Initialize the system.

b) Check whether the vehicle speed is zero or not.

c) If the vehicle speed is zero then the time constant will be taken as time constant under rest condition.

d) If the vehicle speed is non zero then check if the speed is greater than or equal to the minimum speed.

e) If the condition is false then the system will be stopped.

f) If the condition is true then acceleration will be calculated.

g) Next step is to check if acceleration is less than minimum acceleration.

h) If the condition is true then time constant will be taken as time constant under constant speed.

i) If condition is false then the system will be stopped.

Similarly in dynamic leveling the following cases are considered for time constant calculation:

- Vehicle is at rest: Time constant set to time constant under rest condition.
- Vehicle is moving with constant speed: Time constant is set to time constant under constant speed.
- Vehicle is moving with constant acceleration/ deceleration.
- Vehicle is moving with varying acceleration/ deceleration.

C. *Filtering*

Filtering is used to smoothen the process of leveling. This module receives the filter time constant and target set point as the inputs. It will calculate the filtered set point to be given to the ISM or DC motor. The filtering module will always check if it has achieved the target set point. If it has not achieved the target set point then it will check whether there is any change in set point. If there is any change then the current position will be taken as the reference and filtered set point will be calculated. If there is no change then the filtered set point will be calculated and will be given to ISM/ DC motor.

A Simulink model for this leveling algorithm based on the above concept will be developed. From this Simulink model an auto code will be generated with the help of auto-code generation tools such as TargetLink from dSPACE and Realtime Workshop Embedded Coder from the MathWorks.

IV. INTEGRATING INTO AUTOSAR ENVIRONMENT

The next step is to integrate this auto-code to the AUTOSAR application interface. Figure 5 shows how an executable SW-C will fit into AUTOSAR environment. When the driver change the knob/presses the switch, the signal will be sent to the digital input/output (DIO). In the AUTOSAR environment, it will be sending the signal to the *SwitchEvent()* module for checking the signal. SwitchEvent will send signal to LightRequest (event), including the information like the type of signal and mode.

Through the communication network of AUTOSAR such as controller area network (CAN), these signals will come through AUTOSAR interfaces which are used to define exchange information and act as software information ports between software components via a network or local. Software components also exchange information with each other through the AUTOSAR Runtime Environment (RTE) of the system [15]. Thus the headlamp module will get the input signal as well as the type and mode of light and puts them in the system memories. The headlamp module will then set the level of light.

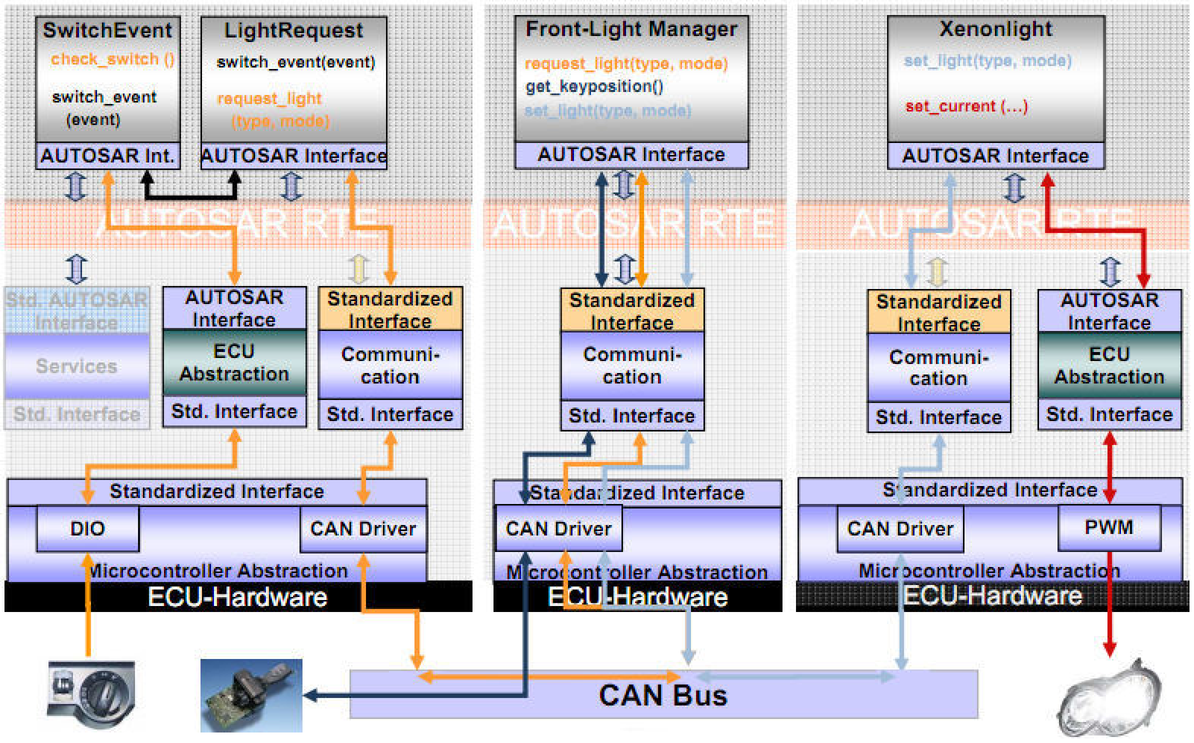


Fig 5: Integrating into AUTOSAR environment [15]

V. CONCLUSIONS

The concept of Leveling algorithm for Adaptive Front Lighting System has been presented, which helps to keep light parallel to the road surface regardless of vehicle's tilt. The Leveling algorithm will be developed using the simulation software Simulink. In this paper it is also shown how the leveling algorithm for Adaptive Front Lighting System will be integrated into an AUTOSAR environment.

REFERENCES

[1] Hamm M (2002).“Adaptive lighting functions history and future–performance investigations and field test for user’s acceptance,” Advanced Lighting Technology for Vehicles, SAE SP-1668, Paper # 2002-01-0526.

[2] Burkard Wordenweber, Jorg Wallaschek, Peter Boyce, Donald D.Hoffmsn, “Automotive Lighting and Human Vision”, Springer International Edition, 2007.

[3] T. Hacibekir, S. Karaman, E. Kural, E.S. Ozturk, M.Decmirci and B.A Guvenc, “Adaptive Headlight System Design Using Hardware-In-The-Loop Simulation”, IEEE International Conference Control Applications, 2006

[4] Qing Wu, Ling Lei, Jianlin Chen and Weifeng Wang, “Research on Hardware-in-the-Loop Simulation for Advanced Front-lighting System”, IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 2008.

[5] Hui Rong, Jinfeng Gong and Wulin Wang, “Kinematics Model and Control Strategy of Adaptive Front Lighting System”, Second International Conference on Intelligent Computation Technology and Automation, 2009.

[6] Frank Bilz, “Automatic Headlight Leveling System for Motor Vehicles”, U.S.Patent 6 480 806 B1, Nov 12,2002.

[7] Atsushi Toda, Hideaki Takeuchi, “Automatic Automotive Headlamp Leveling Device”, U.S.Patent 6 663 268B1, Dec 16, 2003.

[8] Atsushi Toda, “Auto Leveling Apparatus of Vehicle Headlamp”, U.S.Patent 6 693 380B2, Feb 17, 2004.

[9] Atsushi Toda, “Auto Leveling System for Vehicle Headlamp”, U.S.Patent 7 014 345B2, March 21,2006.

[10] Hiroki Kitajima, “Automatic Leveling Device and Automatic Leveling Method”, U.S.Patent 2010/0 091 505A1, Apr 15, 2010.

[11] Masashi YAMAZAKI Kazuo GOTO, Toshikazu TOMONO, “Control Device for Vehicle Lamp, and Methods of Controlling Vehicle Lamp”, U.S.Patent 2012/0 002 430A1, Jan 5, 2012.

[12] Akinobu Todani, Toshinori Origane, “Headlight Optical Axis Adjustment Device”, U.S.Patent 2012/0 014 122A1, Jan 19, 2012.

[13] AUTOSAR. Technical Overview, V2.2.2, R3.1, Rev 001. http://www.autosar.org/, 2008.

[14] AUTOSAR, AUTOSAR documents and specifications. http://www.autosar.org/, 2008.

[15] AUTOSAR, AUTOSAR tutorial. http://www.autosar.org/, 2008.

◈◈◈

Design, Development and Validation of Diagnostic Event Manager Based on AUTOSAR 3.2

Namya P

Sree Buddha College of Engineering, Pattoor P.O., Alappuzha Dist., Kerala, India
E-mail : namyap@gmail.com

Abstract – To handle the growing complexity of software in modern vehicles, automotive OEM's are increasingly developing their electronic system based on AUTOSAR. AUTOSAR (Automotive Open System Architecture) is an open and standardized software architecture for the automotive industry. AUTOSAR uses a layered software architecture where software components can be developed independent from hardware and low-level services. The main layers include Application layer, Real Time Environment (RTE) Layer, Basic Software layer and Microcontroller layer. With the increase in the complex automotive applications, the number of Electronic Control Units (ECUs) used for building a motor vehicle has exponentially increased during the past few years. Nowadays, it is really a challenge with these vehicles regarding the diagnosis and error detection. Therefore, these systems need to have some form of self diagnosis and means to communicate to the outside world. Diagnostics is a module of ECU software which is responsible for error tracking in automobiles. Diagnostic event management shall be established as basic software module known as Diagnostic Event Manager (DEM). The DEM is responsible for processing and storing of diagnostic events (errors) and associated freezeframe data occurring in the automobiles. The DEM module comes under the Services Layer of AUTOSAR. In this paper the design, development and validation of DEM module based on AUTOSAR3.2 is being done.

***Keywords* -** *AUTOSAR; Diagnostics; DEM; errors.*

I. INTRODUCTION

Driven by market demands the automotive industry has changed rapidly during the past few years [1]. A lot of new hardware and software technologies are now increasingly being used in automobiles. The hardware handles the basic functionality in a car whereas the multiple extra functionalities are solved by software. This leads to an increase in the complexity. Nowadays, we cannot imagine a modern vehicle without microcontrollers controlling the safety and comfort systems like Anti Lock Braking System, Airbags, Electric Power Steering (EPS), Parking aid, Automatic Gear job, Adaptive Cruise Control, Adaptive Front Lighting System (AFLS) etc. Upto 90 Electronic Control units (ECUs) are used in a high end car [2]. Tomorrow's car will be a complex embedded electronic system with an embedded network containing dozens of Electronic Control Units (ECUs).This leads to an increase in the software complexity and cost because software itself is not reusable.

AUTomotive Open System Architecture (AUTOSAR) emerged as a solution to address the issue of software complexity of ECUs [3]. AUTOSAR is a standardized automotive software architecture jointly developed by automotive manufacturers, suppliers and tool developers.

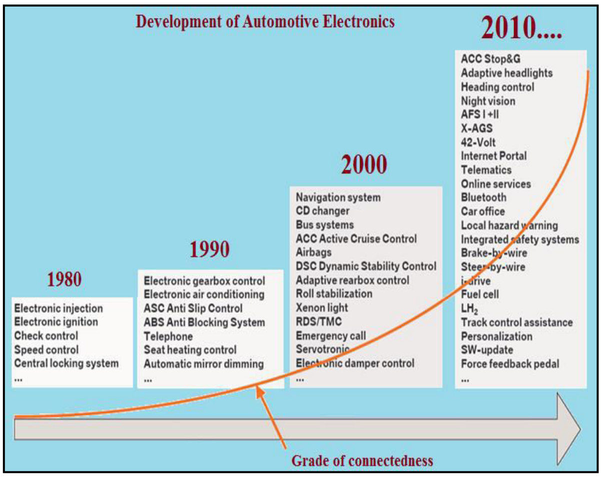


Fig. 1 : Development of automotive electronics [2]

The AUTOSAR partnership comprises of 9 core partners, 50 premium members and development members. More than 160 members play a significant role in this partnership.

As vehicles became more complex with a number of ECUs, there emerged a lot of problems regarding the diagnosis and error detection. It became a challenge how to extract diagnostic data, how to evaluate it to expose possible problems and how to fix these problems. So these systems need to have some form of self diagnosis and means to communicate to the outside world. The Diagnostics module is responsible for error detection and tracking in automobiles. The Diagnostic Event Manager (DEM) is a module of Diagnostics stack of ECU software. This paper presents the design, development and testing of one such module in diagnostics - Diagnostic Event Manager.

II. AUTOSAR

AUTOSAR is a standardized automotive software architecture designed to handle the software complexity of ECUs. The AUTOSAR provides a framework that completely separates the application from its infrastructure [4]. AUTOSAR facilitates the integration and reuse of software components. The motivation for the development of AUTOSAR is to manage the increasing E/E complexity, provide flexibility for product modification, scalability of solutions and to improve the quality and reliability of E/E systems. The major objectives of AUTOSAR are standardization of basic systems functions, scalability to different vehicles and platform variants, transferability of functions throughout the network, integration from multiple suppliers, maintainability throughout the entire product life cycle, software updates and upgrades over the vehicle's lifetime [5].

AUTOSAR achieves the technical goals like modularity, scalability, transferability and re-usability of functions by using layered software architecture. Layered software architecture [6] of AUTOSAR uses 3 layers namely BSW layer, RTE layer and Application layer.

The Basic Software (BSW) Layer [6] is standardized software layer situated below the AUTOSAR Run Time Environment. BSW doesn't have any functionality but it provides services to the AUTOSAR software components. This is achieved through the use of APIs. It contains standardized and ECU specific components. BSW layer comprises of Services Layer, Microcontoller Abstraction Layer, ECU Abstraction Layer. The Services layer contains the modules responsible for diagnostics. The Diagnostic Event Manager (DEM) is situated in the services layer.

The Run time Environment (RTE) [6] acts as a communication environment for inter-and intra- ECU information exchange. The RTE decouples the application software components from the hardware as well as the application software components from themselves. The application layer [6] is the only layer that is composed of standardized software. The actual functionality is located inside the application layer. The application layer consists of application software components as well as sensor/actuator software components.

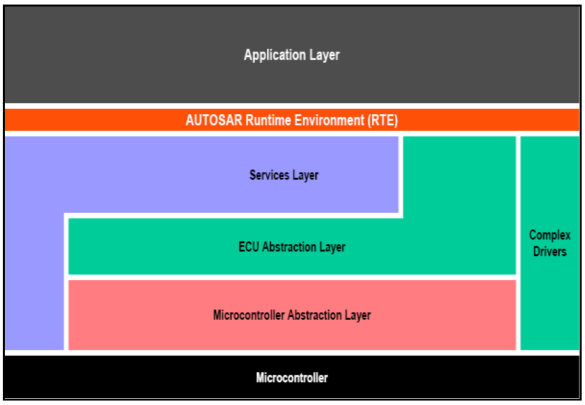


Fig. 2 : AUTOSAR Layered Software Architecture [6]

III. SYSTEM DESCRIPTION

Diagnostic Event Manager (DEM) [7] is a module of diagnostics stack of ECU software. DEM is responsible for processing and storing of diagnostic events (errors) and associated freeezeframe data occurring in the automobiles [8]. For each detected fault DEM stores the corresponding Diagnostic Trouble Code (DTC) to its memory. These DTCs are unique codes which direct the mechanic when and where the fault has occurred instead of dismantling all parts. The DTCs contains information about the fault, its time of occurrence, location etc. The DTCs can be accessed with the help of diagnostic equipment called scan tool which is plugged inside the ECU.

A. Modules assosciated with DEM

The Function Inhibition Manager is responsible for controlling function entities according to the status updates from the DEM module. It can inhibit specific monitor functions depending on the status changes. The DEM informs and updates the Function Inhibition Manager upon changes of event status in order to stop or release function entities according to assigned dependencies.

The Diagnostic Communication Manager is responsible for the communication mechanisms and processing and execution of diagnostic service from an external tester or onboard test system [8].

SW-Components can provide and retrieve data from the DEM. The Monitor Function is a sub-component of a SW-Component [8].

DEM stores the status information regarding the events in permanent memory blocks called NVRAM (Non Volatile Random Access Memory) blocks via mechanisms provided by the NVRAM Manager. The NVRAM blocks are configurable in size.

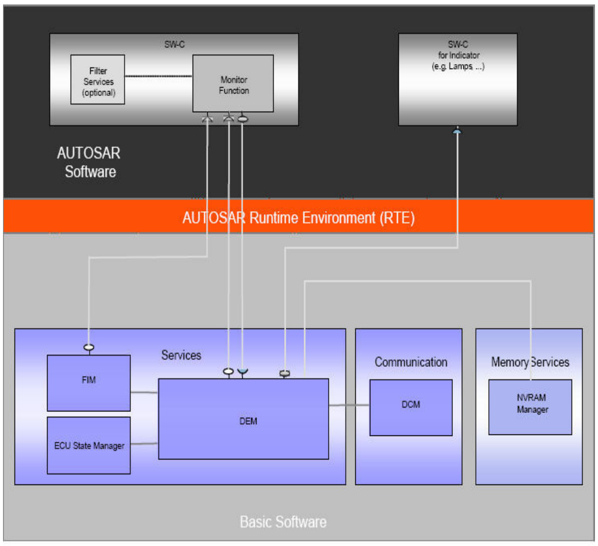


Fig. 3 : Dependencies of DEM to other software modules [8]

The ECU State Manager is in charge of the basic initialization and de-initialization of basic SW components including DEM [8].

B. *Functional Specification of DEM*

The DEM handles and store the events detected by the SW-Components using a monitor function above the RTE. The stored event information is available via an interface to other basic Software modules and Software Components [8]. Each diagnostic event handled by the DEM module represents the result of a monitor function. Each diagnostic event has a unique EventId and the related EventName [7,8].

The Event memory is a set of event records located in a dedicated memory block [7,8]. The DEM uses the NVRAM Manager for storing to non-volatile memory. The DEM module has several memories like primary memory, secondary memory and permanent memory. The parameter DtcOrigin is used to distinguish between memory areas.

Event Memory Management is defined as the process of adding, updating and removing event records in and out of the DEM module. A qualified event has TestFailed bit either “1” or “0”. If “1” it is an active event and passive event if “0”. The DEM module shall store a failed event (TestFailed = 1) and can neglect a passed event (TestFailed = 0). For storage it should check for storage conditions and memory availabilities.

In case a qualified event is passed to the DEM module, the DEM module shall perform the event transition immediately for the bits being relevant for fault reactions.

Event Status Management is referred to as the ability of the DEM module to record and retain events, event status and associated data. Dem_SetEventStatus is used by the DEM module to report the status of an event. The DEM module uses the APIs Dem_ResetEventStatus, Dem_GetEventStatus, Dem_GetEventFailed, Dem_GetEventTested etc in connection with status management. Dem_ClearDTC provides the provision for deleting a single DTC or a group of DTCs from the event memory.

Event combination allows several monitor results to be mapped to one specific DTC. Individual events can be combined or compressed to an additional event.

Event related data are additional data that are stored in case of an event. DEM supports 2 types of event related data- snapshot data(freezeframe) and extended data. The DEM module may support the prestorage of a freezeframe via the API function Dem_PrestoreFreezeFrame. The processing of extended data records is accomplished with the API function Dem_GetExtendedDataRecords.

C. *BSW Error Handling*

BSW-C can also detect errors. These errors are detected before DEM is fully initialized, information is buffered until DEM is fully available. For BSW error handling, the DEM module has the interface Dem_ReportErrorStatus.

D. *Debouncing of events*

DEM supports 2 types of debouncing- signal debouncing and event debouncing [8]. Signal debouncing is done in hardware whereas event debouncing can be done by the monitor function or by the DEM. The DEM module may provide standard algorithms for debouncing of events like counter based, time based and combination of these.

In case of counter based algorithm, the counter will increase with count in step size at every call of Dem_SetEventStatus/ Dem_ReportErrorStatus with status PREFAILED. The debounce counter will decrease by the count out step size with PREPASSED as the status.

In time based debouncing, the signal is unqualified until the first call of Dem_SetEventStatus/ Dem_ReportErrorStatus.

With the call with status PREFAILED/ PREPASSED the debounce time out is started. If the status toggles, the time is restarted and the direction will change.

In frequency based debouncing, an event is unqualified until Dem_SetEventStatus is called. With a PreFailed/ PrePassed event a time window is opened. Different counters are used here. When one of the counters reaches the threshold, and the time window is still open the event is qualified. Reporting of the next event reopens the time window and a new qualification process starts. If neither threshold reaches within the time window, the event is disqualified.

E. API Specification

There are around 48 APIs specified for DEM. Dem_PreInit should be called before initialization. All APIs except Dem_GetVersionInfo and Dem_ReportErrorStatus will work only if DEM is initialzed. There are different APIs to interface DEM with BSW Components, DCM, ECU State Manager, SW_C, NVRAM Manager etc.

IV. SYSTEM DESIGN

The design of DEM comprises of both Low Level Design (LLD) and High Level Design (HLD). HLD phase includes the identification of the APIs needed for the implementation whereas LLD phase includes the detailed design of the APIs in the HLD phase. The design consists of the following steps.

- Extract the functional requirements
- Partition the functional requirements
- Identify all the inter module interfaces
- Detailed design of each module

HLD and LLD is done using the Enterprise Architect tool version 7.5. EZ 7.5 is UML based design tool for designing and constructing software systems. An example of Low Level Design done using Enterprise Architect is shown in figure 4 below. The figure shows the design of the API Dem_ReportErrorStatus.

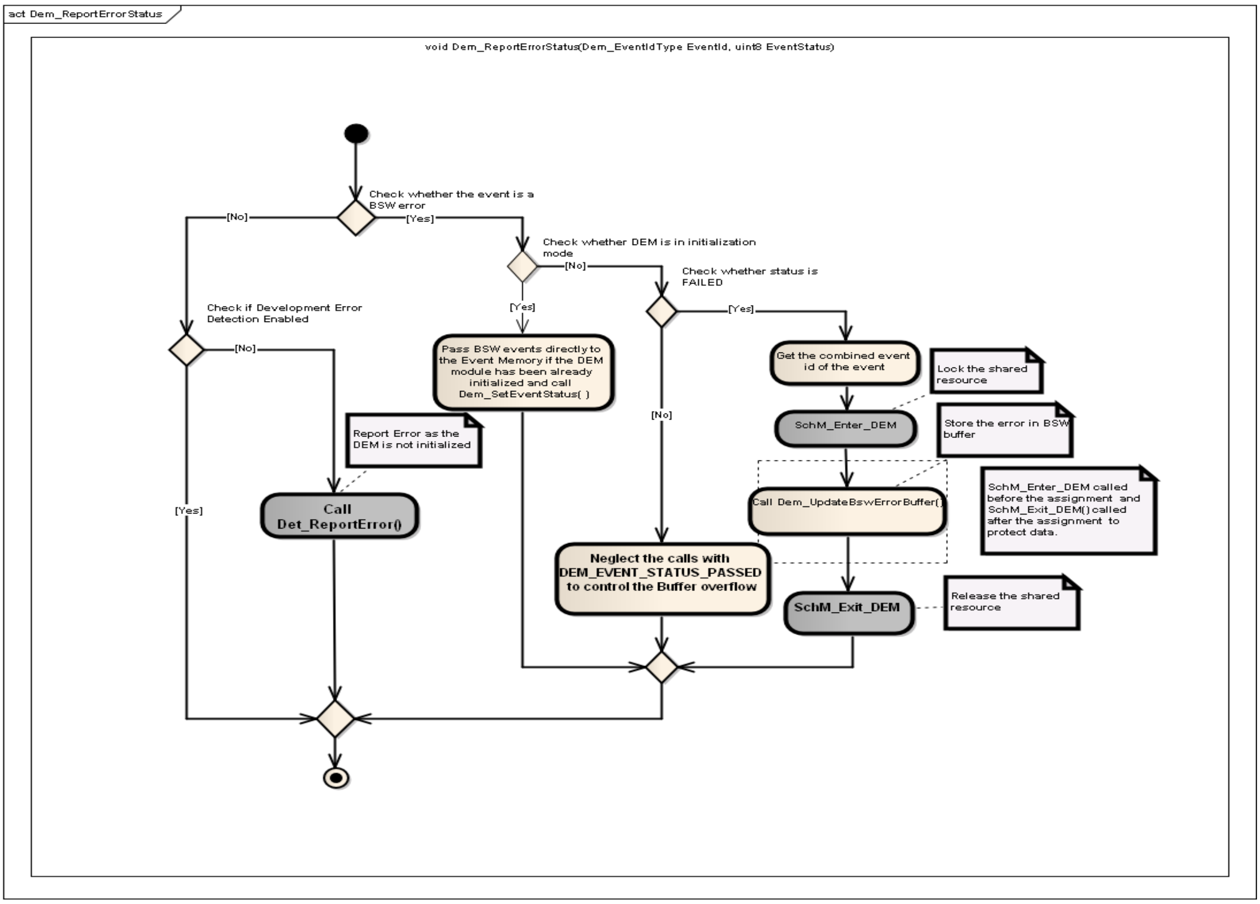

act Dem_ReportErrorStatus
void Dem_ReportErrorStatus(Dem_EventIdType EventId, uint8 EventStatus)
Check whether the event is a BSW error
[No]
[Yes]
Check whether DEM is in initialization mode
[No]
Check whether status is FAILED
[Yes]
Check if Development Error Detection Enabled
[No]
[Yes]
Pass BSW events directly to the Event Memory if the DEM module has been already initialized and call Dem_SetEventStatus( )
Get the combined event id of the event
Lock the shared resource
[No]
SchM_Enter_DEM
Store the error in BSW buffer
Report Error as the DEM is not initialized
[Yes]
Call Det_ReportError()
Call Dem_UpdateBswErrorBuffer()
SchM_Enter_DEM called before the assignment and SchM_Exit_DEM() called after the assignment to protect data.
Neglect the calls with DEM_EVENT_STATUS_PASSED to control the Buffer overflow
SchM_Exit_DEM
Release the shared resource

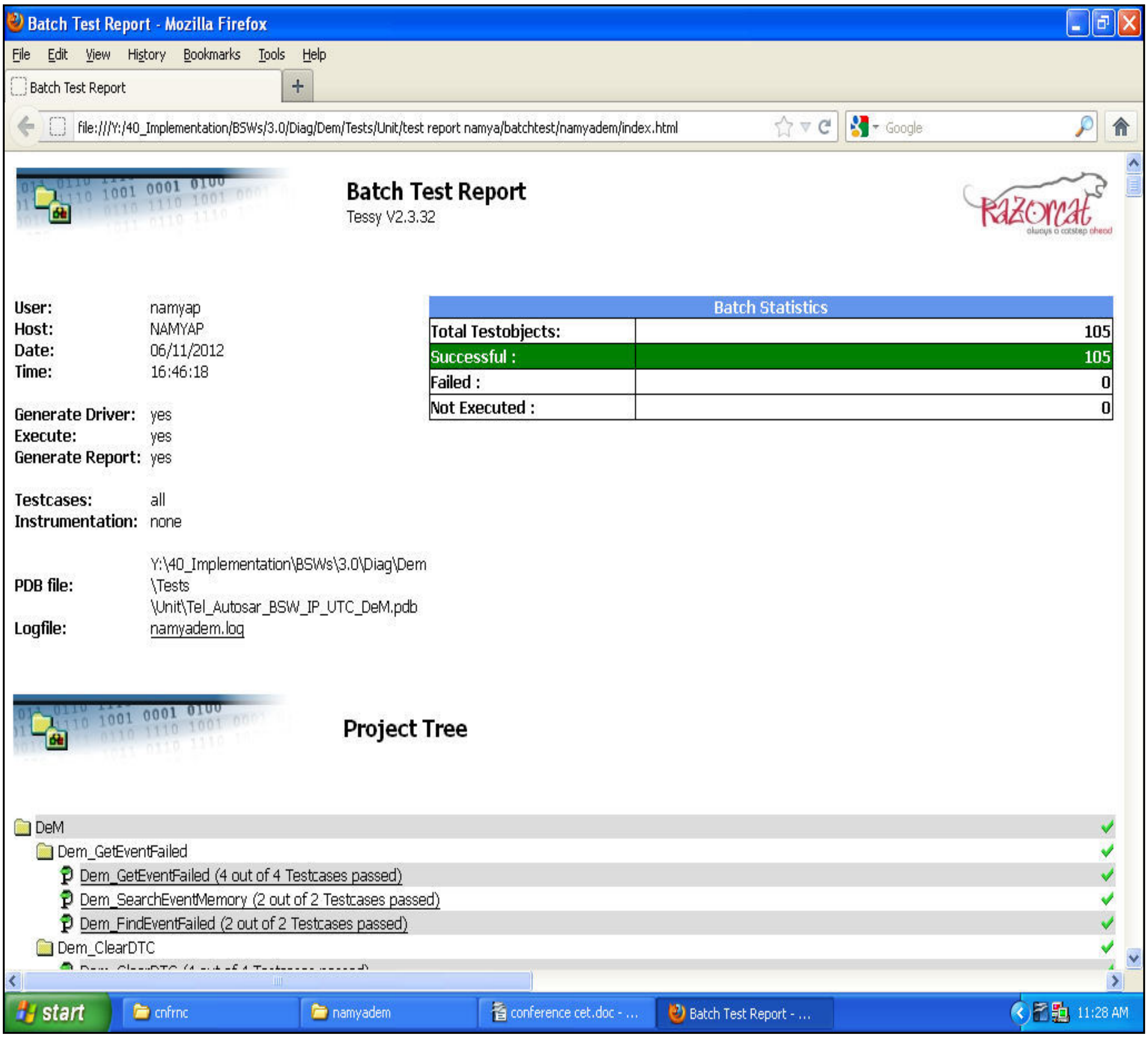
Fig. 4: Low Level Design (LLD) of Dem_ReportErrorStatus

V. CONFIGURATION OF DEM

Configuration parameters define the variability of the generic parts of an implementation of a module. The configuration can be achieved at different times of the implementation process viz before compile time, before link time or after build time.

- Pre compile time: This configuration is for enabling or disabling optional functionality
- Link time: Constant data outside the module; the data can be configured after the module has been compiled. Used for the configuration of modules that is only available as object code.
- Post build time: Loadable constant data outside the module; the data is located in a specific memory segment that allows reloading

The configuration tool used is EzyConfig version 2.1. The configuration design is done using EzyConfig by creating definition project and description project. The necessary Cfg and Pbcfg files are generated.


Batch Test Report - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Batch Test Report
file:///Y:/40_Implementation/BSWs/3.0/Diag/Dem/Tests/Unit/test report namya/batchtest/namyadem/index.html
Google
Razorcat
always a catstep ahead
start
cnfrnc
namyadem
conference cet.doc - ...
Batch Test Report - ...
11:28 AM

Batch Test Report
Tessy V2.3.32

User:	namyap
Host:	NAMYAP
Date:	06/11/2012
Time:	16:46:18
Generate Driver:	yes
Execute:	yes
Generate Report:	yes
Testcases:	all
Instrumentation:	none
PDB file:	Y:\40_Implementation\BSWs\3.0\Diag\Dem \Tests \Unit\Tel_Autosar_BSW_IP_UTC_DeM.pdb
Logfile:	namyadem.log

Batch Statistics	
Total Testobjects:	105
Successful :	105
Failed :	0
Not Executed :	0

Project Tree

- DeM ✓
  - Dem_GetEventFailed ✓
    - Dem_GetEventFailed (4 out of 4 Testcases passed) ✓
    - Dem_SearchEventMemory (2 out of 2 Testcases passed) ✓
    - Dem_FindEventFailed (2 out of 2 Testcases passed) ✓
  - Dem_ClearDTC ✓
    - Dem_ClearDTC (4 out of 4 Testcases passed)

Fig. 5 : Screenshot of batch test report in Tessy

The coding is done in Embedded C. The codes for all APIs are compiled using Visual C++. Then the unit testing is done using Tessy Version 2.3.32. The testing process is completely automated in Tessy thus significantly reducing the testing time and costs. Tessy offers an integrated graphic user interface that guides us comfortably through the testing process. Unit testing in Tessy consists of the following phases.

- Test case determination
- Test data and expected value determination
- Test execution
- Test evaluation and test documentation

Each phase is separately executed for every test object. Test objects are the export functions of a module.

VI. RESULTS

The design of the DEM module is done based on the functional requirements. The configuration design of DEM is done using EzyConfig tool. The coding is done in Visual C++ and the codes have been compiled successfully. The unit testing of DEM module is carried out in Tessy. The batch testing has been done and the test report is generated. The batch test report as shown in Figure 5 consists of 105 test objects.

VII. CONCLUSION & FUTURE WORK

The Diagnostic Event Manager responsible for error detection in automobiles has been developed. The integration of the DEM module is to be done. The integration is accomplished using MPC5668G Evaluation board. MPC5668G is built using Dual core power architecture technology. MPC5668G board connects all the currently available automotive communication protocols. It is widely being used in automotive gateways and central body controllers. After integrating the DEM module to the board the functional unit testing is done using vector CANoe tool.

REFERENCES

[1] Dietmar Schreiner, Karl M. Goschka and Jens Knoop, “Component based Communication Middleware for AUTOSAR”, in Proc. 2008 Intl. Conf. Mechatronic and Embedded Systems and Applications (MESA 08).

[2] Jonas Sandberg, “Autosar today – A roadmap to an AUTOSAR implementation”, Master's Thesis, Information Engineering Program, Chalmers University of Technology, Goteborg, pp. 1- 4, 2006.

[3] Daehyun Kum, Gwang – Min Park, Seonghum Lee and Wooyung Jung, “ AUTOSAR Migration from existing automotive software” presented at the International Conference on Control Automation and Systems (ICCAS 2008) on 14-17 Oct. pp. 558 - 562, 2008.

[4] Gareth Leppla, “Mapping Requirements to AUTOSAR Software Components”, Master's Thesis, Waterford Institute of Technology, pp. 2 – 4, June 2008.

[5] AUTOSAR GbR: Technical Overview. Version 2.2.2 Online (2012) http://www.autosar.org/download/R3.2/AUTOSAR_TechnicalOverview.pdf.

[6] AUTOSAR GbR: Layered Software Architecture Version 2.2.2 Online (2012) http://www.autosar.org/download/R3.2/AUTOSAR_Layered_Software_Architecture.pdf.

[7] AUTOSAR GbR: Requirements on Diagnostic Version 2.1.1 Online (2012) http://www.autosar.org/download/R3.2/AUTOSAR_SRS_ Diagnostic.pdf.

[8] AUTOSAR GbR: Specification of Diagnostic Event Manager Version 3.2.0 Online (2012)

http://www.autosar.org/download/R3.2/AUTOSAR_SWS_ DEM.pdf

◈◈◈

Design, Development and Testing of Protocol Data Unit Router Module Based on AUTOSAR 3.2

Shilpa Das

Sree Buddha College of Engineering, Alappuzha Dist., Kerala, India
E-mail : shilpagiresh@gmail.com

Abstract – AUTomotive Open System ARchitecture (AUTOSAR) is the new standard software architecture for automobiles. The increased usage of Electronic Control Units (ECU) in automobiles is contributing to higher software complexity in vehicles. Here is the significance of AUTOSAR. The AUTOSAR is a new framework used to reduce the ECU complexity in new generation automobiles. Its core frame work consists of four layers; the top layer is the Application layer and the bottom layer is the Microcontroller layer. Run Time Environment and Basic Software Layer are the two layers sandwiched between these two layers. The Basic Software Layer provides an infrastructural functionality to an ECU. The scope of this paper is limited to a module within the Basic Software Layer named Protocol Data Unit (PDU) Router module. PDU Router module must be instantiated in every ECU. The main task of this module is the routing of Interaction Layer Protocol Data Units (I-PDUs) between the communication services and the hardware abstraction layer modules. It also provides the gateway functionality which means the routing of I-PDUs between the same or different networks. This paper focuses the design, development and testing of AUTOSAR 3.2 based PDU Router module.

Keywords **-** *AUTOSAR, ECU, PDU Router, I-PDU.*

I. INTRODUCTION

The use of embedded software in automotive industry is growing by each day. In an automotive vehicle of modern day, most of the functionalities such as air conditioning, engine management, anti lock breaking system etc are under computer control.

So, off late, modern vehicles are characterized by the presence of high number of ECUs (Electronic Control Units). Luxury cars can easily have up to 70 ECUs. The high number of ECUs will increase the software complexity of the vehicles. In order to handle this enormous complexity in automotive software systems, a new innovative standard was introduced called AUTOSAR (AUTomotive Open System Architecture). AUTOSAR was founded as a development partnership in 2003 by a group of companies.

AUTOSAR is a standardized architecture for Electric and Electronic (E&E) systems. The re-use of software components between different vehicle platforms, like OEMs (Original Equipment Manufacturer) is one of the major goals of AUTOSAR [1]. There are a number of technical factors that have motivated the development of AUTOSAR. The most significant of those are listed below:

- Improving scalability
- Improving quality and reliability
- Improving flexibility
- Enabling the early detection of errors during project's design phase.
- Consideration of availability and safety requirements
- Software updates and upgrades over vehicle life time
- Redundancy activation
- Reduction of costs for software development and service in the long term [2].

AUTOSAR supports the modular development of ECU software. The functions inside the ECU network are easier to integrate and replace. It helps the vehicle manufactures to be more flexible in the use of ECU software [8].

II. SOFTWARE ARCHITECTURE

The AUTOSAR software architecture is a layered architecture. That has the goal to abstract the software from hardware components [3]. The software architecture of AUTOSAR shown in figure 1 consists of four main layers. The lowermost layer is the microcontroller layer which contains the ECU hardware. The next two layers above the microcontroller layer are Basic Software and the Run Time Environment layer. The uppermost layer is the application layer; which contains all the application specific hardware components [4].

The layers in the AUTOSAR have different functionalities. The Basic Software and the Run Time Environment layers are responsible for the abstraction between the hardware and the application software. The Basic Software is the standardized software; providing necessary services for running the functional part of the software. The Run Time Environment layer connects the basic software with the application software. It enables inter component communication as well as communication from software components to basic software modules.

The Basic Software layer includes three layers called the Service layer, the ECU abstraction layer and the microcontroller layer. It may also contain the Complex Device Driver. It breaks the layered architecture where the direct access of hardware is needed.

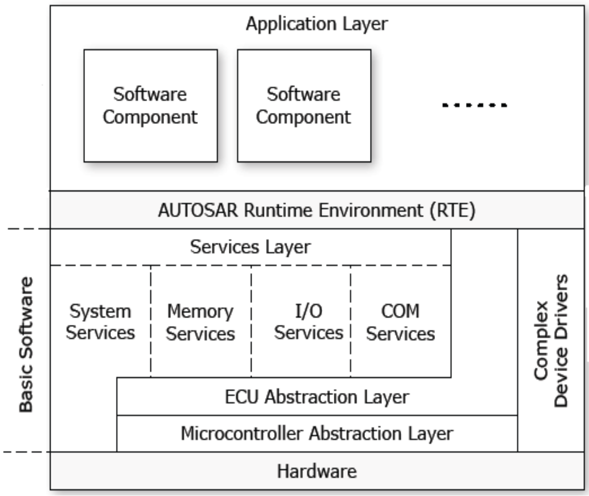


Fig. 1: AUTOSAR software architecture layers [4].

The MCAL (Microcontroller Abstraction Layer) is the abstraction of the hardware; to avoid the direct access of microcontroller's register. Abstraction layer provides software interface to a specific ECU. Service layer is the highest layer of the Basic Software and it provides basic services to the application layer, the basic services includes memory services, diagnostic services, communication services, system services etc [3].

Protocol Data Unit (PDU) Router module is a part of AUTOSAR service layer. PDU Router module provides communication services to other modules [6].

The Operating System of AUTOSAR is based on standard OSEK based operating system. This operating system is widely used by the automotive industry. AUTOSAR OS supports fixed priority based scheduling, facilities for handling interrupts and it also provides inter task communication [5].

III. PROTOCOL DATA UNIT ROUTER MODULE

PDU is an abbreviation of Protocol Data Unit. Each PDU contains an SDU and a PCI. SDU is Service Data Unit which is the data passed by an upper layer with the request to transmit the data. PCI is an abbreviation of Protocol Control Information; it contains the source and target information. It is added on transmission side and removed on receiving side.

The signals from the AUTOSAR application layer come to the RTE layer; the RTE performs the transformation from the AUTOSAR signal to the signal in COM. AUTOSAR signals are packed and unpacked by using the COM module. The signals are packed into I-PDUs. I-PDU group is a collection of I-PDUs in COM and contains more I-PDUs [1].

The PDU Router module is an important part of AUTOSAR ECU. The main task of Protocol Data Unit Router module is to route the I-PDUs (Interaction Layer Protocol Data Units) between:

- Communication interface modules such as LINIF,CANIF and FlexRayIf
- Transport protocol modules such as CAN TP,FlexRay TP
- AUTOSAR Diagnostic Communication Manager (DCM) and Transport Protocol modules such as CAN TP,FlexRay TP
- AUTOSAR COM and communication interface modules such as LINIF, CANIF or FlexRayIf or I-PDU multiplexer.
- I-PDU multiplexer and communication interface modules such as LINIF, CANIF or FlexRayIf.

PDU Router module can be adapted for gateway operations and for internal routing purposes. It transfers I-PDUs from one module to another module [6].

Each PDU has a static PDU ID; which is used to identify the PDUs. Each PDU Router module finds out the destination path by using the static configuration table and the PDU ID. I-PDUs has a maximum length of 8 bytes.

PDU Router module provides APIs for modules below and above the PDU Router. The modules below the PDU Router module are communication interface modules and the transport protocol modules. The modules above the PDU Router module are COM and DCM (Diagnostic Communication Manager). More over PDU Router module provides an interface to I-PDU Multiplexer (I-PduM) which is a module that is beside the PDU Router.

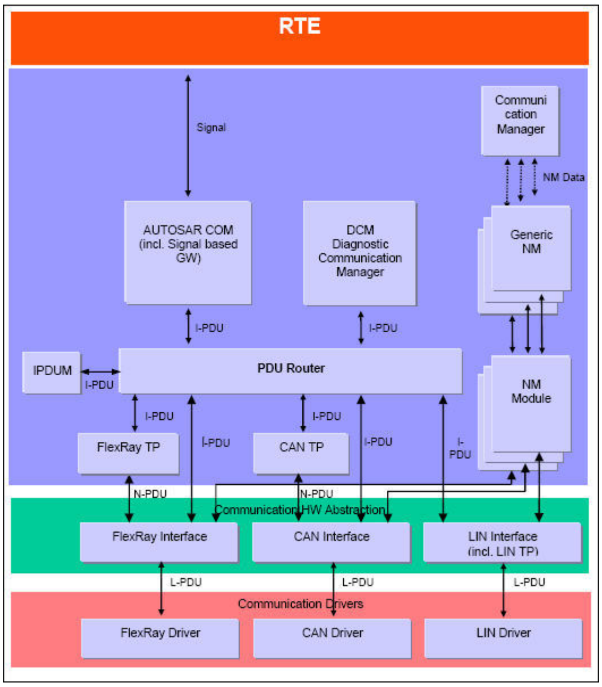


Fig. 2: Shows the AUTOSAR communication structure [6]

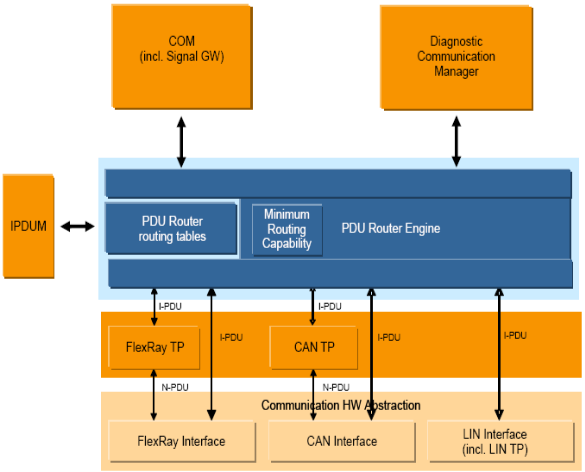

COM (incl. Signal GW)
Diagnostic Communication Manager
IPDUM
PDU Router routing tables
Minimum Routing Capability
PDU Router Engine
I-PDU
I-PDU
I-PDU
I-PDU
I-PDU
FlexRay TP
CAN TP
N-PDU
N-PDU
Communication HW Abstraction
FlexRay Interface
CAN Interface
LIN Interface (incl. LIN TP)

Fig. 3: The detailed PDU structure [6]

The PDU Router module consists of PDU Router routing tables and the PDU routing engine. The PDU routing tables can be updated during the post build time in the programming state of the ECU,and it describes the routing attributes for each PDU shall be routed.PDU routing engine describes the actual code performing the routing actions is based on the PDU Router routing tables. The detailed PDU structure is shown in figure 3.

The PDU Router engine provides a minimum routing capability to route specific PDUs without using the routing tables. This supports accessing of the DCM for activation of the ECU bootloader even when the post build time routing table is modified.

A. Functional Specification of PDU Router

The PDU Router module can perform three different classes of operations. They are PDU reception, receive the I-PDU and send to the upper layer, PDU transmission; transmit I-PDUs on request of upper layer modules and PDU gateway, receive the I-PDU from an lower layer module and transmit this I-PDUs to the same or another lower layer module [1].

The PDU Router module will refresh the routing tables only when they are not in use. The behavior of PDU Router module functions should be synchronous; although the overall action of a function might be asynchronous.

The PDU Router module provides a special type of gateway functionality known as routing-on-the-fly, which means that the forwarding of data between two communication modules is started before all data have been received.

PDU Router module transfers the I-PDU to one destination module (single cast transmission) or more than one destination modules (multicast transmission).The modules above the PDU Router module such as COM and DCM, can request the cancellation of an I-PDU transmitted from the lower layer modules, if that I-PDU is not needed ,it can be done by using the API PduR_<Lo>CancelTransmit.

Cancellation of an I-PDU will only be confirmed by the destination protocol module. Like cancel transmission the modules can request for cancel reception also, for example if the upper layer module want to cancel the reception of I-PDUs from the lower layer modules then the upper layer modules will use the API PduR_<Up>CancelReceive..

It is also possible for the upper layer modules can change the transport protocol parameters for a definite Rx I-PDU.For this upper layer modules call the API PduR_<Up>ChangeParameter [6].

B. PDU Router State Management

The PDU Router module consists of three states. They are PDUR_UNINIT, PDUR_REDUCED, and PDUR_ONLINE. After power up the PDU Router module shall be in the state PDUR_UNINIT. The PDU Router module shall change to the state PDUR_ONLINE, when the PDU Router has successfully initialized via the function PduR_Init. If the initialization did not succeed the PDU Router module shall change the state to PDUR_REDUCED.PDU Router states is shown in figure 4.

PDU Router module shall perform routing when it is in online state (PDUR_ONLINE), PDU Router shall perform minimum routing and routing when it is in the reduced state (PDUR_REDUCED) and the PDUR shall perform no routing when it is in the uninitialized state (PDUR_UNINIT) [6].

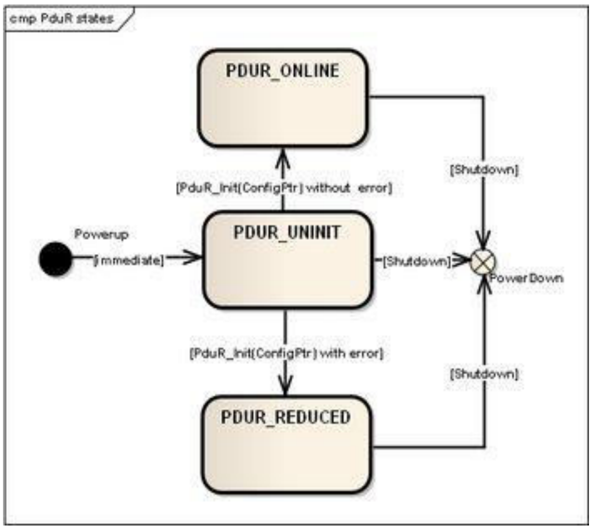


Fig. 4 : PDUR States

C. API Specification of PDU Router module

In order to initialize the PDU Router module PduR_Init API is used. If this API is successfully initialized then the PDU Router module shall go to the state PDUR_ONLINE.A set of routing paths can be enabled and disabled by using the APIs PduR_EnableRouting and PduR_DisableRouting respectively [6].

D. Functional Definitions for lower layer modules

The modules below the PDU Router modules are CAN,LIN,and FlexRay.Lower interface layer modules calls the function PduR_<Lo>IfRxIndiaction, after an L-PDU has been received. L-PDU is a Data Link Layer Protocol Data Unit which is assembled or disassembled in AUTOSAR hardware abstraction layer. In AUTOSAR; Data Link layer is correspondent to the Microcontroller Abstraction Layer. The maximum length of CAN and LIN L-PDU is 8 bytes and the maximum length for FlexRay L-PDU is 254 bytes.

PduR_<Lo>IfTxConfirmation and PduR_<Lo>TpTxConfirmation functions are called by the lower layer interface and Transport protocol modules after the PDU has been transmitted. PduR_<Lo>TpProvideTxBuffer , API is used to provide the Tx buffer for the lower layer modules.

NM (Network Management) interface is a new added feature of AUTOSAR 3.2 version. NM module shall act as a bus independent adaptation layer between the bus specific Network Management modules such as CanNm and FrNm. PduR_<Can,Fr>NmRxIndication is used, after an L-PDU has been received. PduR_<Can,Fr>NmTxConfirmation is called after the PDU has been transmitted. In order to trigger the transmission of an NM message PduR_<Can,Fr>NmTriggerTransmit API is used.

E. Functional Definitions for Upper layer modules

The modules above the PDU Router modules are COM and DCM. PduR_<Up>Transmit is called by the upper layer modules to request transmission .PduR_DcmCancelReceive API is used to terminate the currently ongoing data reception. PduR_DcmCancelTransmit is used to terminate the currently ongoing TP data transmission. PDUR shall provide a communication interface API for the CDD. PduR_CddTransmit is called by complex device driver to request a transmission. PduR_CddTpTransmit is called by an upper layer Cdd to request transmission using a transport protocol.

The IpduM (I-PDU Multiplexer) is placed next to the PDU Router module. PduR_IpduMTransmit requests the transmission for the IpduM. IpduM module is acting as an lower layer module. PduR_IpduMTriggerTransmit is called by the IpduM to request I-PDU contents from upper layer module [6].

IV. DESIGN OF PDU ROUTER MODULE

Design of PDU Router module consists, of both High Level Design (HLD) and Low Level Design (LLD).HLD specifies, the identification of APIs needed for the implementation of PDUR and the LLD point out the detailed design of APIs.

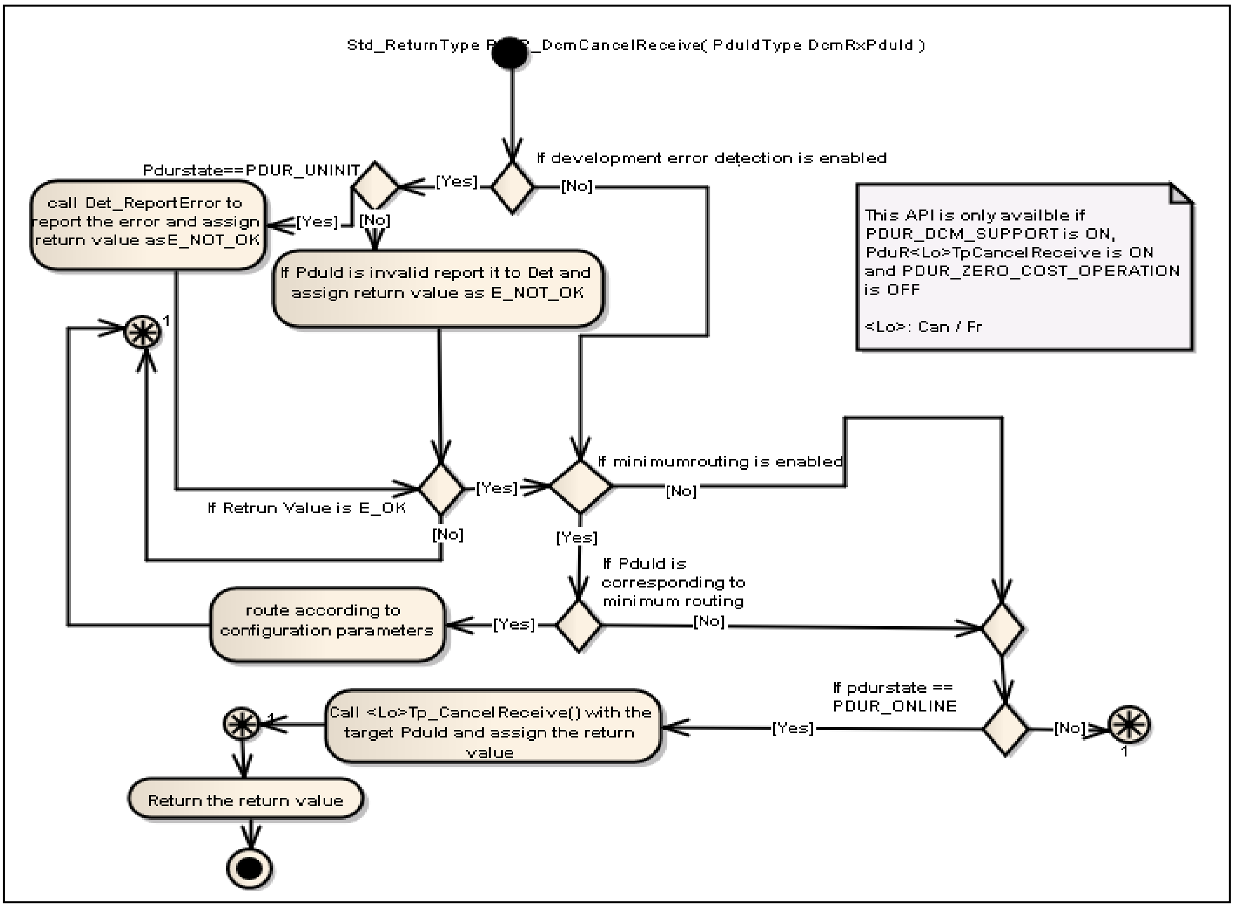

Std_ReturnType P R_DcmCancelReceive(PduIdType DcmRxPduId)
If development error detection is enabled
[Yes]
[No]
Pdurstate==PDUR_UNINIT
call Det_ReportError to report the error and assign return value as E_NOT_OK
[Yes]
[No]
If PduId is invalid report it to Det and assign return value as E_NOT_OK
This API is only availble if PDUR_DCM_SUPPORT is ON, PduR<Lo>TpCancelReceive is ON and PDUR_ZERO_COST_OPERATION is OFF
<Lo>: Can / Fr
1
If minimumrouting is enabled
[Yes]
[No]
If Retrun Value is E_OK
[No]
[Yes]
If PduId is corresponding to minimum routing
route according to configuration parameters
[Yes]
[No]
If pdurstate == PDUR_ONLINE
Call <Lo>Tp_CancelReceive() with the target PduId and assign the return value
[Yes]
[No]
1
1
Return the return value

Fig. 5: Low Level Design of PduR_DcmCancelReceive

Figure 5 shows the Low Level Design of PduR_DcmCancelReceive. This function is called by DCM to request a transmission. High level and low level design was done by using Enterprise Architect version 7.5 software.

V. CONFIGURATION

Configuration parameters for the coding were obtained form eZyConfig.This tool was developed by Tata Elxsi Limited. This tool is used for generating the configuration codes for BSW modules. The code file structure of the PDU Router module shall include the following files;PduR_Cfg.c PduR_Lcfg.c and PduR_PBcfg.c.This files represents the pre compile time ,link time and the post build time configuration parameters [7].

VI. DEVELOPMENT AND TESTING OF PDUR MODULE

The development of PDU Router module was done by implementing all the APIs intended for it. The file structure of the PDU Router module consists of all the header files and the code files. Coding was done in C and by using Microsoft Visual C++ 2008 express edition. It is successfully compiled and builds without error.

Testing of PDU Router module consist of Unit testing and Integration testing. Unit testing refers to the process of testing modules against the detailed design. Integration testing mainly deals with testing for the correctness of the interface.

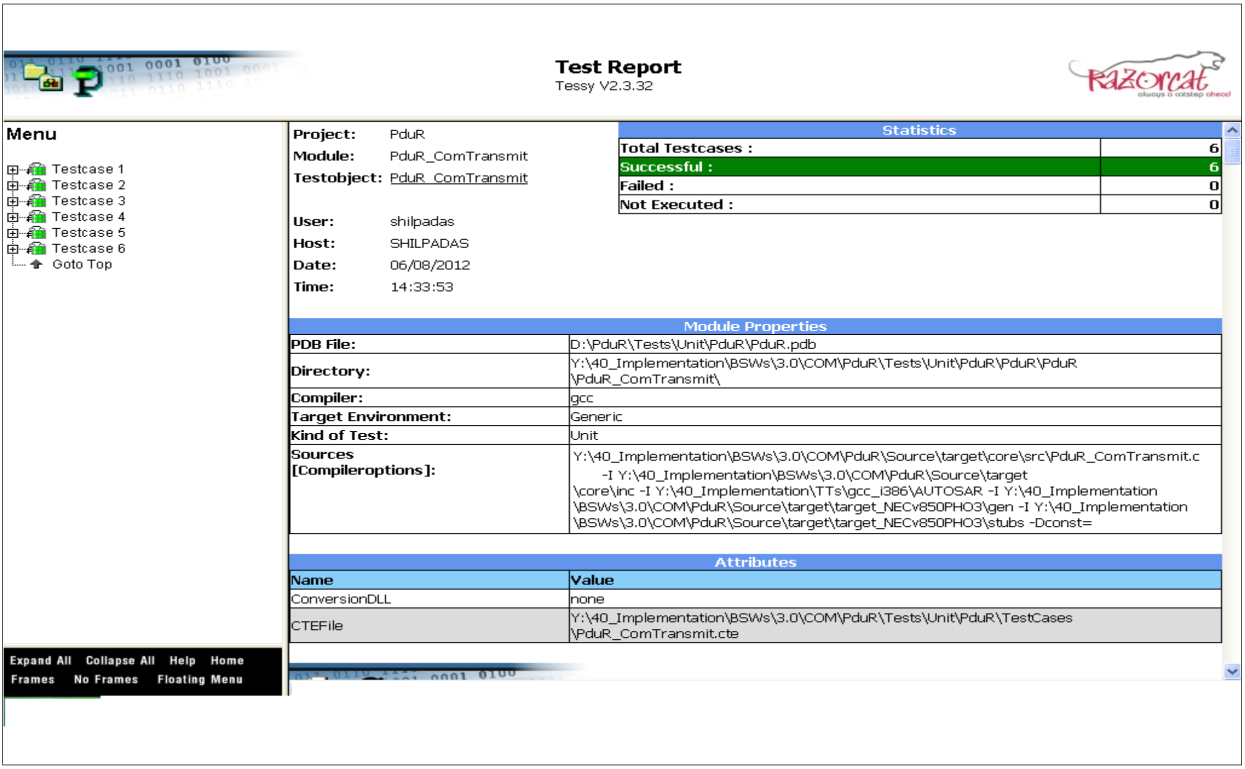
Test Systems 2.3 (TESSY) is used for unit testing of PDU Router. Validation platform used for integration testing is MPC5668G, a 32 bit MCU built in dual core power architecture technology.

VII. RESULTS

After executing the test in Tessy,if the expected and the actual values are same the test case will pass by indicating it in green colour. Test report will be generated in XML or HTML format. The test report for the function PduR_ComTransmit is shown figure6.

VIII. FUTURE WORK

After the unit testing, the integration testing has to be done. Integration testing mainly deals with the testing for the correctness of the interface. Integration testing will verify the basic functionality of the BSW modules after this module is integrated. Validation platform for integration testing is MPC5668G evaluation board. For Integration testing, source files for all the modules has to be build by using Diab 5.7.0.0 and Lauterbach debugger. Vector CANoe software can be used for integration testing. After comprehensive testing and making sure that module is fool proof, it can be integrated with AUTOSAR 3.2 for the production environment.



Test Report
Tessy V2.3.32

Project:	PduR
Module:	PduR_ComTransmit
Testobject:	PduR_ComTransmit
User:	shilpadas
Host:	SHILPADAS
Date:	06/08/2012
Time:	14:33:53

Statistics	
Total Testcases :	6
Successful :	6
Failed :	0
Not Executed :	0

Module Properties	
PDB File:	D:\PduR\Tests\Unit\PduR\PduR.pdb
Directory:	Y:\40_Implementation\BSWs\3.0\COM\PduR\Tests\Unit\PduR\PduR\PduR \PduR_ComTransmit\
Compiler:	gcc
Target Environment:	Generic
Kind of Test:	Unit
Sources [Compileroptions]:	Y:\40_Implementation\BSWs\3.0\COM\PduR\Source\target\core\src\PduR_ComTransmit.c -I Y:\40_Implementation\BSWs\3.0\COM\PduR\Source\target \core\inc -I Y:\40_Implementation\TTs\gcc_i386\AUTOSAR -I Y:\40_Implementation \BSWs\3.0\COM\PduR\Source\target\target_NECv850PHO3\gen -I Y:\40_Implementation \BSWs\3.0\COM\PduR\Source\target\target_NECv850PHO3\stubs -Dconst=

Attributes

Name	**Value**
ConversionDLL	none
CTEFile	Y:\40_Implementation\BSWs\3.0\COM\PduR\Tests\Unit\PduR\TestCases \PduR_ComTransmit.cte

Fig. 6: Tessy Report for the function PduR_ComTransmit

REFERENCES

[1] Gosda,Johannes.”AUTOSAR Communication Stack” .Technical report,Hasso-Plattner Institute fur Softwaresystemtechnik ;pp.13- 23;2009

[2] AUTOSAR Partnership,”AUTOSAR_Technical Overview V2.2.2R3.2Rev1”. [Online].Available.http://www.autosar.org/download/R3.2/AUTOSAR_ TechnicalOverview.pdf 2011

[3] Warschofsky, Robert,”AUTOSAR Software Architecture”,Technical report,Hasso-Plattner Institute fur Softwaresystemtechnik ;2009

[4] Schreiner Dietmar “Component based middleware for AUTOSAR”Master's Thesis,Vienna University of Technology,pp.42-45,2009

[5] Leppla,Gareth”Mapping Requirements to AUTOSAR software components”;Master's Thesis;Waterford Institute of Technology Awards Council;pp.51-58;June 2008

[6] AUTOSAR GbR:”Specification of PDU Router”V2.4.0R3.2Rev1 [Online]Available:http://www.autosar.org/download/R3.2/AUTOSAR_ SWS_PDU_Router.pdf 2011

[7] eZyConfig manual,Tata Elxsi Limited,November 2009.

[8] Han Wu;”Vehicle Diagnostic with AUTOSAR”Master's Thesis;KTH Computer Science and Communication ;pp.11-14;2008

◈◈◈

Design of Neural Network as Data Flow Model for Image Compression

Daneshwari I. Hatti, Savitri Raju & Mahendra M. Dixit

Dept of ECE, S.D.M College of Engineering & Technology, Dharwad, India
E-mail : daneshwari_hatti@yahoo.co.in, savitriraju@sdmcet.ac.in, mmdixitmm@yahoo.co.in

Abstract – In digital communication bandwidth is essential parameter to be considered. Transmission and storage of images requires lot of memory in order to use bandwidth efficiently neural network and Discrete cosine transform together are used in this paper to compress images. Artificial neural network gives fixed compression ratio for any images results in fixed usage of memory and bandwidth. In this paper multi-layer feedforward neural network has been employed to achieve image compression. The proposed technique divides the original image in to several blocks and applies Discrete Cosine Transform (DCT) to these blocks as a pre-process technique. Quality of image is noticed with change in training algorithms, convergence time to attain desired mean square error. Compression ratio and PSNR in dB is calculated by varying hidden neurons. The proposed work is designed using MATLAB 7.10. and synthesized by mapping on Vertex 5 in Xilinx ISE for understanding hardware complexity.

Keywords - *backpropagation, Discrete cosine transform, image compression, mean square error, neural network, PSNR.*

I. INTRODUCTION

Importance of image compression increases with advancing communication technology. The amount of data associated with visual information is so large that its storage requires enormous storage capacity. The storage and transmission of such data require large capacity and bandwidth, which could be very expensive. Image data compression techniques are concerned with reduction of the number of bits required to store or transmit images without any appreciable loss of information. Compression makes it possible for creating file sizes of manageable, storable and transmittable dimensions. Compression is achieved by exploiting the redundancy [1]. DCT represents image as a sum of sinusoids of varying magnitudes and frequencies. The DCT has a property that, for a typical image, most of the visually significant information about the image is concentrated in just a few coefficients of the DCT. Hence DCT is often used for image compression. The neural network has good performance in non-linear capacity [2]. Neural network is used for further compression in further reducing size. It has proved that the multi-layer neural network can accurately approximate to any linear or nonlinear function. Through learning the incomplete data, the neural network can achieve an accurate prediction of the whole data with its generalization capability. Neural network can be utilized to approximate to the continuous quantity, because any continuous quantity can be express by the combinations of linear and nonlinear function. Hence the neural network is designed to store the information of the continuous quantity.

Back-propagation algorithm [3] is a widely used learning algorithm in Artificial Neural Networks. In this paper Backpropagation neural network is designed and trained using different learning algorithms. The input data is entered into the network via the input layer. Each neuron in the network processes the input data with the resultant values steadily "percolating" through the network, layer by layer, until a result is generated by the output layer. The actual output of the network is compared to expected output for that particular input. This results in an error value. The connection weights in the network are gradually adjusted, working backwards from the output layer, through the hidden layer, and to the input layer, until the correct output is produced. Fine tuning the weights in this way has the effect of teaching the network how to produce the correct output for a particular input, i.e. the network learns. The general parameters deciding the performance of Backpropagation Neural Network Algorithm includes the mode of learning, information content, activation function, target values, input normalization, initialization, learning rate and momentum factors. [4], [5], [6].The neural network structure can be illustrated in Figure1. Three layers, one input layer, one output layer and one hidden layer. Both of input layer and output layer are fully connected to hidden layer.

Compression is achieved by designing the value of the number of neurons at the hidden layer, less than that of neurons at both input and output layers.

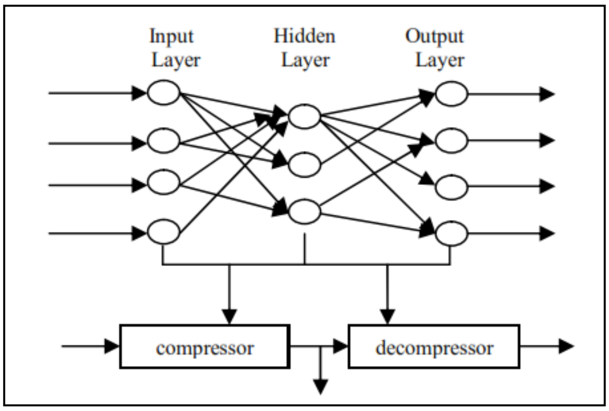


Fig. 1: The basic architecture for image compression

II. METHODOLOGY

Image compression is done using neural network by varying number of hidden neurons and made less than input layer neurons. Three layered neural network is designed and trained using backpropagation algorithm and applied with different learning algorithms such as one step secant (trainoss), BFGS quasi Newton (trainbfg) and gradient descent backpropagation with adaptive learning rate algorithm (traingda).

DCT2 is used for images. It is represented as shown in eqn1

$$C(u,v)=\alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{\Pi(2x+1)u}{2N}\right]\cos\left[\frac{\Pi(2y+1)v}{2N}\right] \quad (1)$$

for u,v = 0,1,2,…,N −1 and α(u) and α(v)

$\alpha(u)=\sqrt{\frac{1}{N}}$ for u = 0

$\sqrt{\frac{2}{N}}$ for u $\neq$ 0

The output is calculated at each node using eqn 2

$$output = f\left(\sum_i a_i w_i\right) \quad (2)$$

f is activation function represented in eqn3

$$f(x)=\frac{2}{1+e^{-2x}}-1 \quad (3)$$

Backpropagation refers to back propagation of error that is

$$E = (\text{target} - \text{output})^2 \quad (4)$$

In this paper tangent sigmoid is used in hidden layer shown in eqn3 and pure linear is used at output layer as activation function. The error obtained is back propagated to adjust the weights until error is minimized using delta rule represented in eqn3

$$w_{new} = w_{old} + \alpha \partial E / \partial w_{old} \quad (5)$$

where α is the learning rate.

Algorithm1

1. First read the input image, here the cameraman image is used for processing. The image is of 256x256 and it has to be segmented in to 8x8 blocks or apply DCT for 8x8 blocks and rearrange each block in to column vector and form it as 64x1024 size image matrix.

2. Apply this input image to neural network and consider that image as target image for training neural network.

3. After training the neural network the output obtained at the output layer is rearranged to 256x256. Hidden layer output is the compressed image it consist of hidden neurons hi<ni, hi is number of neurons in hidden layer and ni is number of neurons in input layer. It is of size 16x1024 rearrange to a matrix of 128x128 it is compressed image.

4. Repeat step2 to step 4 for learning algorithms namely trainbfg, traingda and trainoss.

III. RESULTS

A. *Image compression using backpropagation neural network and different training algorithms*

i. using trainbfg

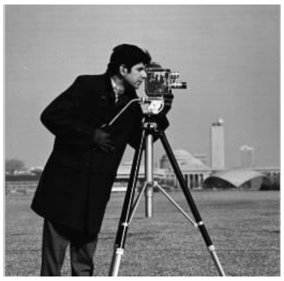


Fig. 2a: Original image of size 256x256

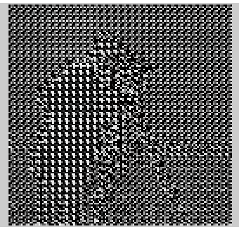


Fig. 2b: Compressed image of size 128x128

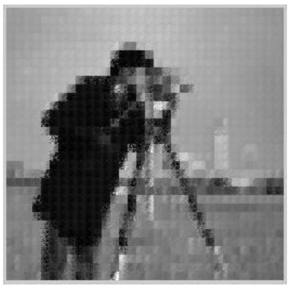


Fig. 2c : Reconstructed image of size 256x256

The network is trained using trainbfg for 1000 epochs and error is decreased and network is trained using trainbfg algorithm. Fig 2a, Fig 2b & Fig 2c shows the input image and compressed image which is output of hidden layer and reconstructed image obtained at output layer respectively. It can be easily shown that reconstructed image needs more iteration to converge and hence goal is set as 0.01.

ii. Using traingda

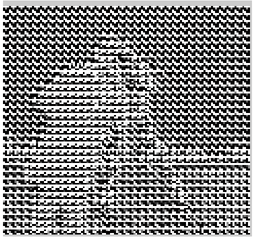


Fig. 4a: Compressed image of size 128x128

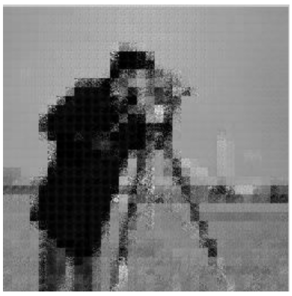


Fig. 4b: Reconstructed image of size 256x256

iii. Using trainoss

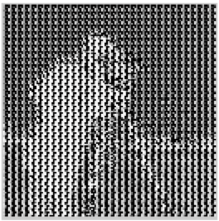


Fig. 5a: Compressed image of size 128x128

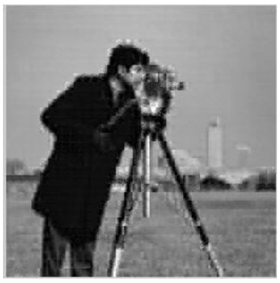


Fig. 5b: Reconstructed image of size 256x256

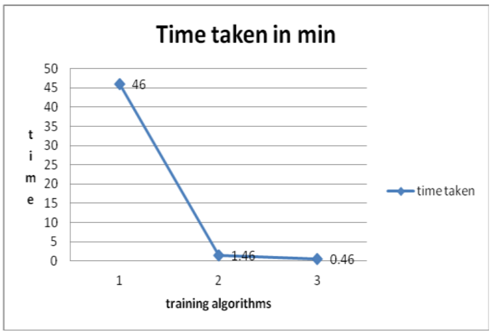


Fig. 6 : Time taken for training algorithms

Figure 6 shows th evariation of convergence time for trainbfg(1), traingda(2) & trainoss(3) in min.training algorithm takes less time for assigned MSE resulting in goood PSNR.

B. Reconstructed image for untrained network

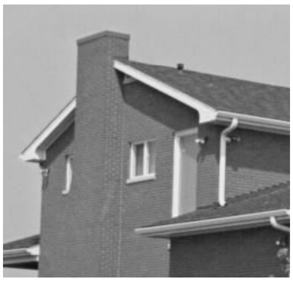


Fig. 7a: Input image

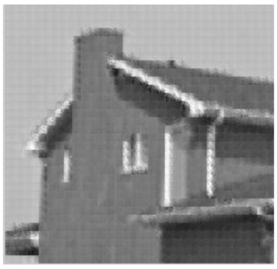


Fig.7b: Reconstructed image

Figure 7b shows the decompressed image obtained from the network which was trained for different image.

C. DCT as preprocessing technique

DCT is applied for 8x8 block and arranged in to column vector and hence given as input image for neural network. The network is trained using trainoss algorithm. The results obtained for considering all DCT coefficients is shown in Figure 8a and Figure 8b.

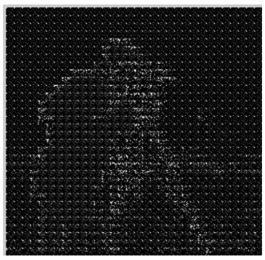


Fig. 8a: DCT image

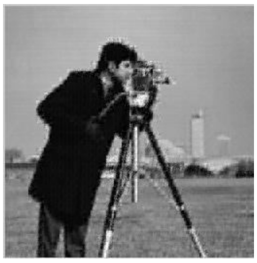


Fig. 8b: Reconstructed image

Figure 8b shows the decompressed image obtained from trained neural network.

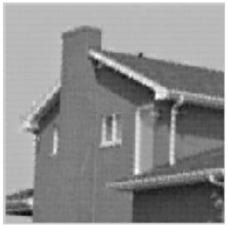


Fig. 8c: Reconstructed image for untrained neural network

Figure 8c shows the decompressed image obtained from untrained neural network.

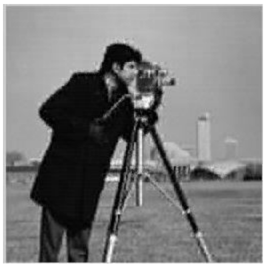


Fig. 8d: Reconstructed image for trained neural network

Figure 8d represents the decompressed image obtained from neural network when considering low frequency DCT coefficients that is 4x4 from each 8x8 block of whole image 256x256. Hence only 16 coefficients are considered out of 64 coefficients which results to compression.

D. Comparison of PSNR in dB obtained using trainoss algorithm

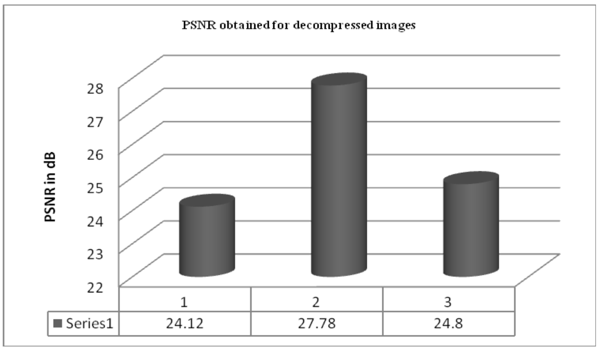


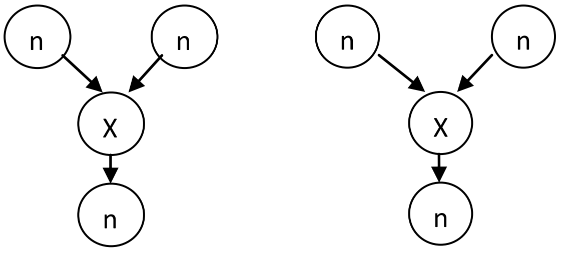
Fig. 9: PSNR for trained network

Figure 9 describes varying PSNR in dB obtained from trained network for traditional method in figure it is given as (1),DCT as preprocessing method considering all coefficients (2) & considering only 4x4 coefficients from each 8x8 block(3). PSNR obtained for test image given to this trained network is 23.02 for another approach is 24.08.

E. Neural network as data flow model

The design proposed consists of matrix multiplication of two matrices, one is the input image samples, and the second is the weight matrix obtained after training. This multiplied output is passed through the nonlinear transfer function to obtain the compressed output that gets transmitted or stored in compressed format. On the decompression side, the compressed data in matrix form is multiplied with the weight matrix to get back the original image.

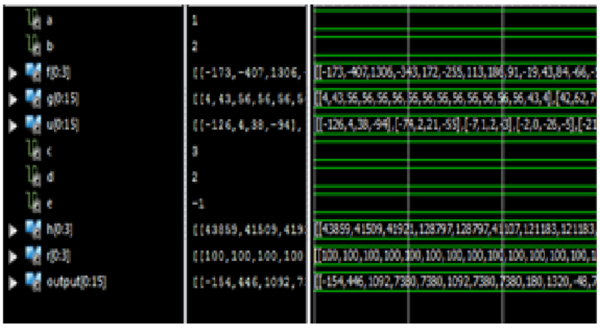
Neural network is modeled as data flow model. Input matrix is saved in one node n1 and weight matrix in another node n2, multiplication is performed between this two matrixes the output obtained is sent through the node n3, it behaves as activation function. Hence compressed image is stored in node n4.The weight matrix between hidden and output layer is stored in node n5. The output obtained from n4 and matrix stored in node n5 is multiplied results in to reconstructed image.



(a) (b)

Fig. 10(a) Compressor (b) Decompressor

Figure 10a & Figure 10 b shows data flow graph model for compression and decompression of image designed using HDL.



```
a            1                      
b            2                      
f[0:3]       [[-173,-407,1306,      [[-173,-407,1306,-343,172,-255,113,186,91,-19,43,84,-66,-
g[0:15]      [[4,43,56,56,56,5      [[4,43,56,56,56,56,56,56,56,56,56,56,56,56,43,4],[42,62,7
u[0:15]      [[-126,4,38,-94],      [[-126,4,38,-94],[-78,2,21,-55],[-7,1,2,-3],[-2,0,-26,-5],[-21
c            3                      
d            2                      
e            -1                     
h[0:3]       [[43859,41509,419      [[43859,41509,41921,128797,128797,41107,121183,121183,
r[0:3]       [[100,100,100,100      [[100,100,100,100,100,100,100,100,100,100,100,100,100,1
output[0:15] [[-154,446,1092,7      [[-154,446,1092,7380,7380,1092,7380,7380,180,1320,-48,7
```

Fig. 11: Simulation results of compressed and reconstructed image

Figure 11 shows the results for proposed work, here input image matrix is in integer and weight matrix is also rounded to integer value. In graph h[0:3] represents 4x16 output obtained for multiplication of input image and weight matrix, r[0:3] is the compressed image matrix of size 4x16 and output[0:15] is reconstructed image matrix of size 16x16.

IV. CONCLUSION

This paper has introduced a comparison among backpropagation training algorithms in image compression namely One step secant (trainoss), BFGS quasi Newton (trainbfg) and gradient descent backpropagation with adaptive learning rate algorithm (traingda).

The employed algorithms were tested for against parameters, like the number of epochs, and goal. With respect to time of execution, it is important to choose algorithms of acceptable time of execution. Finally, Image compression could be achieved with high accuracy using feedforward back propagation. The best reconstructed image is obtained for trainoss learning algorithm with less time. Using DCT as preprocessing technique gives better result compared to traditional method. Compression ratio is fixed for 1:2:1 neural network that is 1 hidden layer and 1 output layer having 64 input neurons and 16 hidden neurons are resulting 4:1 compression ratio.DCT using as preprocessing techniques looks that compression achieved is more that is after applying DCT less number of bits are required to represent coefficients then feeding to network still compresses and better reconstruction is also achieved through neural network. Hence this kind of method can be used as it compresses in same ratio for any images and hence bandwidth will be fixed which can be used for different purpose. The proposed neural network designed as data flow model in VHDL and simulated.

Integer DCT can be performed as preprocessing technique then it is to be fed as input image for this trained network. The output obtained is to be compared with results obtained in this paper.

ACKNOWLEDGMENT

The authors wish to thank the authorities at AICTE for granting RPS to the department of ECE, SDM college of Engineering and Technology Dharwad. This work is towards partial fulfillment of the said project. The authors are also grateful to the authorities of SDMCET Dharwad, for the facility extended to carry out the present work. The authors also would like to thank the unknown reviewers for their invaluable comments which help in improving the paper.

REFERENCES

[1] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing, 1993.

[2] Zhe Xue Guo, Zhi Qiang Shun, The Neural Network and The Realization of The Matlab 2007 ,Publishing House Of Electronic Industry, Beijing, China, 2007,9.

[3] Simon Haykin, “Neural Networks – A Comprehensive foundation”, 2nd Ed., Pearson Education, 2004.

[4] Bogdan M.Wilamowski, Serdar Iplikci, Okyay Kaynak, M. Onder Efe “An Algorithm for Fast Convergence in Training Neural Networks”.

[5] Fethi Belkhouche, Ibrahim Gokcen, U.Qidwai, “Chaotic gray-level Image transformation, Journal of Electronic Imaging -- October - December 2005 -- Volume 14, Issue 4, 043001 (Received 18 February 2004; accepted 9 May 2005; published online 8 November 2005.

[6] Hahn-Ming Lee, Chih-Ming Cheb, Tzong-Ching Huang, “Learning improvement of back propagation algorithm by error saturation prevention method”, Neurocomputing, November 2001.

❑❑❑

Visualization of Network Traffic

Sowmya C. L, C. D.Guruprakash & M.Siddappa

Department of Computer Science and Engineering, Sri Siddhartha Institute of Technology, Tumkur
E-mail : sowmyalnp@gmail.com

Abstract – Network security in today's world is critical. System administrators must quickly perceive the security state of their networks, but they often have only text-based tools to work with. These tools often provide no overview that would help users grasp the big-picture.

Network traffic visualization tools have successfully enabled security analysts to understand the nature of traffic present in a network. However these tools rely mainly on human expertise to discover anomalies in traffic and attack patterns.

The visualization capability provided by Visualization tool allows an operator to assess the state of a large and complex network given an overall view of the entire network and filter/drill-down features with a friendly user interface that allows users to request more detailed information of interest such as specific protocol traffic flows.

Visualization tool allows operators to detect and investigate anomalous internal and external network traffic. Visualization tool shows network events graphically. We model the network as a graph with hosts being nodes and traffic being flows on edges. We present a detailed description of Visualization tool functionality and demonstrate its application to traffic dynamics in order to monitor, discover, and investigate security-relevant events.

Keywords - *Visualization, NetFlows, Parallel axes representation, Link based visualization, Security visualization, Network traffic visualization.*

I. INTRODUCTION

Networks are becoming increasingly complex. Complexity is the enemy of security. Networks have become more complex in terms of size, topology, and especially traffic flows. Traffic flows are faster and the number of different applications generating flows grows continuously. While network traffic is an ideal place to monitor for security events since all security events leave some form of network trace, there is a major problem in that security events can also be concealed among the vast amount of legitimate traffic. It is often difficult to just to capture and store network traffic, so analyzing and detecting attacks in near-real-time with current command line driven text-based tools can be especially challenging for non-experts.

However, humans excel at visual processing and identifying abnormal visual patterns. Visualization is the act or process of interpreting in visual terms or of putting into visual form. A new definition is a tool or method for interpreting image data fed into a computer and for generating images from complex multi-dimensional data sets. Visualization tools can translate the myriads of network logs into animations that capture the patterns of network traffic in a sufficient way, thus enabling users to quickly identify abnormal patterns that warrant closer examination. Such visualization tools enable network administrators to sift through gigabytes of daily network traffic more effectively than searching text-based logs.

It is challenging to visualize all the information relevant to network and security administrators. We must be able to provide near real-time analysis. To deal with the storage and processing limitations, special data structures must often be created. Analysis is also complicated since not only must we show what is happening on each host, but we must visualize the relationships and interactions between all the hosts.

Visualization tool has these distinguishing features: (1) it uses animations to visualize network traffic dynamics and (2) it provides multi-level views of network traffic including an overview and drill-down views that allow users to query for details on-demand, and (3) it provides filtering capabilities to remove known legitimate traffic so as to focus on potential security events.

Although techniques to collect packets are important, the purpose and scope of collection have to be clearly described. As packets are collected in a single location, the location must satisfy the purpose and scope. Since network packets are digital signals transmitted by hosts, servers, network devices and others, it is important to understand the whole network structure. If the network is connected to the Internet, properties, such as the IP bandwidth of the collection location, type of the service, availability of VPN, rules of the firewall and network speed, have to be considered.

As packets are collected over a time period, details of the packets are analyzed by monitoring the time flow. For example, types of services used are achieved from the port information. The concentration and distribution of services can be analyzed by investigating specific IP bandwidths. Therefore, packet data including IP or port data are analyzed in a binary or hexadecimal format. Due to the large amount of packets, wide IP bandwidths, and various services, acquiring specialized knowledge under limited time leads to serious analysis problems.

Extract IP addresses and ports from network packet files are used as Ethereal and Sniffer systems. After storing relevant traffic data, classification by session is performed and finally the amount of traffic that flows in is measured. Port scanning is the process that identifies listening ports of target systems. Attackers check if target systems are on and then look for open ports and find weak points to attack. Since the port scanning is highly likely to be a preliminary step of attacking, it is important to identify such attacks.

System administrator can make use of the online network visualizing process to trace suspicious network activities or potential attacks to the network system. By applying visualization technique, the network analyst can examine the entire network traffics and discover the characteristics of traffic.

Most of the visualization techniques exist in generating the captured network data into informative graphical 2D or 3D view and to improve decision making and organization management performance.

Visualization tool is its ability to display a large amount of data and highlight unusual behaviors in two dimensional views that are easily readable. Obviously, three dimensional views would provide additional information compared to those that are two-dimensional (hereafter respectively called 3-D and 2-D views). However, to observe such 3-D views we have to project them down onto a 2-D visual aid (e.g., a screen or paper). Two main issues are raised by this dimensionality reduction, namely disorientation and occlusion. Disorientation means that the position of the plotted data is not clear and the values corresponding to the plots are difficult to retrieve. Occlusion occurs when plots hide one another, so information is omitted from view. These two problems are well-known in the field of computer vision, and a common solution is to display several 2-D projections instead of a single 3-D view.

1.1 Objective

Visualization is a technique to graphically represent sets of data. When data is large or abstract, visualization can help make the data easier to read or understand.

The objective of this project is to develop a technique for visualization of network traffic. The network traffic visualizing method is the main goal to be achieved. It visualizes network traffic flow using NetFlow source data. It detects and investigates anomalous internal and external network traffic. The tool provides multi level views of network traffic. Tool process network log records in real time.

Although visualization makes the analytical process more convenient, complex graphics makes the process difficult to conduct. By using simplified graphics network traffic can be easier to conduct because it is more understandable. Through logging network traffic will provide back tracking an incident of network to determine the source. Therefore the network traffic should be logged periodically and on demand for this purpose.

1.2 Motivation

Network attacks are a serious problem that network attacks cost businesses an estimated $666 million in 2003. There are a variety of network traffic visualization tools available, each focusing on a different level of network abstraction. Previous work focuses on representing activities occurring on machines while tool focuses on network flows between machines.

This project is motivated by concern of network security by system administrator and visualizing capability of network state. The text based tool, which perceive security state of network are not useful for users to grasp the big picture.

II. LITERATURE SURVEY

Linkages between different hosts and events in a computer network contain important information for traffic analysis and intrusion detection.

“Home-Centric Visualization of Network Traffic for Security Administration” [5] and “Combining Visual and Automated Data Mining for Near-Real-Time Anomaly Detection and Analysis in BGP” [10] focus on visualizing linkages in a network.

“Home-Centric Visualization of Network Traffic for Security Administration” presents a VISUAL(Visual Information Security Utility for Administration Live), visualization tool. It focuses on visualizing linkages in a network. Rapidly perceive the security state of their network. Use a matrix visualization technique combined with a simple graph drawing technique. The prototype whose design and testing is discussed in this paper is called Visual Information Security Utility for Administration Live (VISUAL). VISUAL aids network administrators by showing a graphical, home-centric overview of their network. Aside from seeing abnormal traffic, VISUAL allows network administrators to develop an accurate mental model of what is normal on their own network so that they can diagnose problems better. VISUAL's design follows "Overview first, zoom and filter, then details on demand." We wish to display an overview of network traffic for a small to mid-size network. We show a home-centric, internal vs. external perspective of the network.

"Elisha: a Visual-based Anomaly Detection System" [12] the authors present a visualization of network routing information that can be used to detect inter-domain routing attacks and routing misconfigurations.

In "A Visual Exploration Process for the Analysis of Internet Routing Data",[13] they go further and propose different ways of visualizing routing data in order to detect intrusions.

"NVisionIP", [9] a security visualization tool that shows network traffic flows from a host-centric view. The visualization approach behind NVisionIP is to present an overview first, zoom and filter capabilities next, and then details on demand. In the overview/galaxy view, NVisionIP represents an entire class B IP address space 64,000 hosts on one screen with attributes highlighted such as traffic volume, number of connections and port or protocol activity-color and shape representations can be chosen by users on demand. Drill-down views to a subnet small-multiple view and an individual machine view allow operators to make analytical observations of flow activity.

III. PROBLEM STATEMENT

Network is becoming complex in terms of size, topology and network traffic flow. So it is important to maintain the network secure. System administrator must maintain their system secure from network attacks.

Previously text based tools are used to monitor the network but it is difficult for non experts to analyze and detect attacks.

Thus visualization technique is used to analyze and detect attacks, which makes users to identify attacks easily by seeing patterns.

The title of the project is "Visualization of Network Traffic". The visualization technique monitors, analyzes the network traffic for attacks. Where visualization technique captures the network traffic and projects the data into graphical representation. Graphical representation can be 2D or 3D. Visualization technique enables user to understand the network anomalies and attack easily by using the graphical representation. By using visualization we can identify any anomalies in the network by seeing patterns in the graphical representation.

IV. METHODOLOGY

In visualization of network traffic, it captures the network traffic and analyzes the captured network traffic. The traffic is captured in the form of packets. Thus packets information like source, destination address and port number, packet size, time of packet capture, protocol type used and some information about the packet are captured as network traffic flow. Then captured packets are analyzed. These analyzed packets are then sent to visualizer, which creates new image after some interval of time from the captured packet data. The visualization of network traffic flow can be done in real time also, i.e by capturing live network traffic and then analyzing the captured data and then viewing the captured data through visualizer.

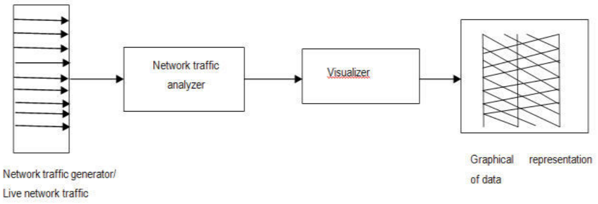

Network traffic analyzer
Visualizer
Network traffic generator/ Live network traffic
Graphical representation of data

Fig. 1: General architecture of visualization

Firstly, it requires generation of network traffic flow then capturing the network traffic flow. Later analyzing the network traffic and then filtering the network traffic based on some parameters like source, destination address, port number, protocol and etc. Then filtered or analyzed packets can be visualized through visualizer, which presents the data in user understandable format, i.e in graphical representation. Store the captured data for future reference. Identify any anomalous traffic flow by examining the flow with stored data.

Denial of Service Attack

In this case, an unfamiliar domain is generating a significant amount of traffic to multiple hosts on the local network. For example consider while traffic is capturing, at that time filtering is applied. i.e filter out all traffic except that on port 80. This reveals that a single domain is generating a large amount of traffic to many internal hosts on this port, suggesting that this external domain is contacting many local web servers. Thus many machines on the external network are heavily accessing a set of hosts on the local network. This type of pattern strongly suggests the occurrence of an attempted denial of service attack in progress on local web servers.

Visualization approach is capable of revealing a variety of network traffic patterns relevant to security. Here by using visualization we identify some variety of traffic patterns like Dos attack.

V. CONCLUSION

Visualization tool for visualizing traffic flows on a network using NetFlow source data. This visualization framework is extensible to other domains beyond IP networks. Visualization tool can visualize traffic in near realtime for security monitoring purposes.

Visualizing network traffic increases the ability to detect network anomalies or attacks.

Visualization tool enhances the situational awareness of a security administrator by providing an intuitive view of IP flows using link analysis. The primary graphical user interface to Visualization tool is a parallel axes view which is used to represent the source and destination of network traffic flows. A highlevel overview is provided first with the user able to drill down to lower level views for more resolution and details on demand. Filtering mechanisms are provided to facilitate either extracting known legitimate traffic patterns or highlighting flows of special interest.

REFERENCES

[1] R. Ball, G. A. Fink, C. North, “Home-Centric Visualization of Network Traffic for Security Administration”, CCS Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC), 2004.

[2] Kiran Lakkaraju William, Yurcik Ratna Bearavolu, Adam J. Lee, “NVisionIP: An Interactive Network Flow Visualization Tool for Security”, National Center for Supercomputing Applications (NCSA) University of Illinois, Urbana-Champaign.

[3] Meghan Allen, Peter McLachlan, “Network Analysis Visualization”, University of British Columbia Department of Computer Science.

[4] Sven Krasser, Member, IEEE; Gregory Conti, Member, IEEE; Julian Grizzard, Member, IEEE; Jeff Gribschaw, Member, IEEE; Henry Owen, Senior Member, IEEE, “Real-Time and Forensic Network Data Analysis Using Animated and Coordinated Visualization”, Proceedings of the 2005 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.

[5] K. Lakkaraju, W. Yurcik, A. Lee, R. Bearavolu, Y. Li, and X. Yin. NVisionIP: NetFlow Visualizations of System State for Security Situational Awareness. In ACM CCS Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC) held in conjunction with the 11th ACM Conference on Computer and Communications Security, 2004.

◈◈◈

Porting Nymble System into Mobiles

B. Muni Lavanya

CSE Dept., JNTUA College of Engineering, Anantapur, Andhra Pradesh, India
E-mail : munilavanya45@gmail.com

Abstract – Anonymizing networks route the traffic through independent nodes in separate administrative domains to hide a client's IP address. Users employ this anonymity for abusive purposes, so that these networks are limited. For blocking the misbehaved users Tsang et.al [1] proposed and built a comprehensive credential system called Nymble, which can be used to add a layer of accountability to any publicly known anonymizing network. This paper extends the nymble system to mobiles by porting it through Android OS. By porting the nymble system in mobiles the anonymizing networks services can be accessed and users can verify their blacklist status using mobiles.

Keywords - *anonymous blacklisting, revocation.*

I. INTRODUCTION

Anonymizing networks such as Tor [3] allow the users to access Internet services privately by using a series of routers to hide the client's IP address from the server. Users employ this anonymity for abusive purposes such as defacing popular websites, so that the success of such networks has been limited. For disabling access to misbehaving users,the administrators of website aredepend on IP-address blocking, but blocking IP addresses is not practical if the misbehaved user routes through an anonymizing network. For denying anonymous access to misbehaving and behaving users alike, the administrators block all known exit nodes of anonymizing networks.

There are several solutions, to address this problem, each providing some degree of accountability.In pseudonymous credential systems [4], [5], users log into websites using pseudonyms.These pseudonyms are added to a Blacklist if a user misbehaves. This approach results in pseudonymity for all users. And anonymizing networks provide anonymity, it is weaken by this system. In Anonymous credential systems [6], [7] group signatures are employed. To cancel a misbehaving user's anonymity basic group signatures allow servers by complaining to a group manager. For every authentication, servers must query the group managerthrough this it lacks in scalability. With dynamic accumulators, a revocation operation results in a new accumulator and public parameters for the group, and all other existing users' credentials must be updated, making it impractical. Verifier-local revocation (VLR), fixes this shortcoming by requiring the server ("verifier") to perform only local updates during revocation. VLR requires heavy computation at the server that is linear in the size of the blacklist.

For this, a system called Nymble, in which servers can "blacklist" misbehaving users, thereby blocking users without compromising their anonymity.

The remainder of this paper is organized as follows. In Section 2, we describe the nymble system. Section 3 discusseshow to port the nymble system into mobile. Section 4 concludes the paper.

II. NYMBLE SYSTEM

In this section, we describe nymble system, which is the solution of blocking misbehaving users in anonymizing networks, and also addresses the Sybil attack [8] to make its deployment practical.

The Nymble system provides the properties: anonymous authentication, backward unlinkability, subjective blacklisting, fast authentication speeds, rate-limited anonymous connections, revocation auditability (where users can verify whether they have been blacklisted).

In Nymble system, users connect to websites by acquiring an ordered collection of nymbles.These nymbles are special type of pseudonyms. These nymbles are computationally hard to link without additional information. Users access the services using the stream of nymbles. Websites obtain a seed for a particular nymble to blacklist users.Seed is used to link future nymbles from the same user, those used before the complaint remains unlinkable. Servers blacklist anonymous users without knowledge of their IP addresses while allowing behaving users to connect anonymously. The nymble system ensures that users are aware of their blacklist status before they present a nymble, and disconnect immediately if they are blacklisted.

The Nymble system architecture:

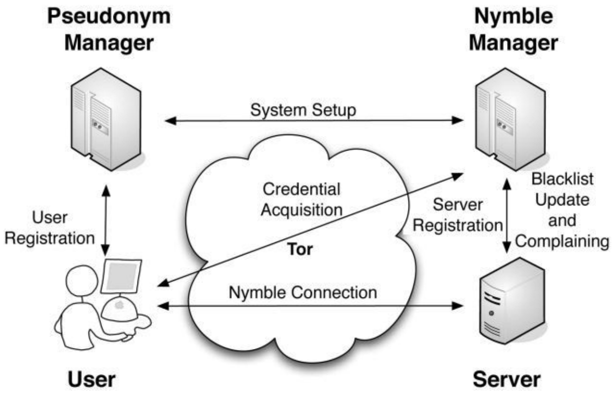


Fig. 1: Nymble system [1]

In the Nymble system architecture, it shows the various modes of interaction. Users interact with the NM to get the nymbles and interact with servers to access the services though the anonymizing network.

2.1 RESORCE BASED BLOCKING

To limit the number of identities a user can obtain (calledthe Sybil attack), the Nymble system binds nymblesto resources that are sufficiently difficult to obtain ingreat numbers. IP addresses usedas the resource in our implementation, but our scheme generalizes to other resources such as email addresses,identity certificates, and trusted hardware.

2.2 PSEUDONYM MANAGER

The user must first contact the Pseudonym Manager and demonstrate control over a resource. Pseudonyms are deterministically chosen based on the controlled resource, ensuring that the same pseudonym is always issued for the same resource. PM's duties are limited to mapping IP addresses to pseudonyms. The user contacts the PM only once per linkability window. PM knows only the user identity-pseudonym pair.

2.3 NYMBLE MANAGER

The user connects to the Nymble Manager (NM) through the anonymizing network, and requests nymbles for access to a particular server. nymbles are generated using the user's pseudonym and the server's identity. These nymbles are thus specific to a particular user-server pair. Nymbles are encapsulated into tickets to provide the requisite cryptographic protection and security properties.

2.4 TIME

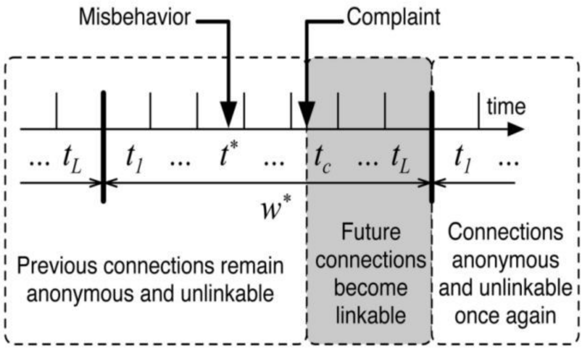

Misbehavior
Complaint
time
... t_L
t_1 ... t^* ...
t_c ... t_L
t_1 ...
w^*
Previous connections remain anonymous and unlinkable
Future connections become linkable
Connections anonymous and unlinkable once again

Fig 2: The life cycle of a misbehaving user[1].

The above figure shows the life cycle of a misbehaving user. If the server complains in time periodt_c about a user's connection in$t$*, the user becomes linkable starting in t_c. The complaintin $t_c$can include nymble tickets from only t_{c-1} and earlier.

Time is divided into linkabilitywindows of duration W, each of which is split into Ltime periods of duration T (i.e., *W = L*T).* The user access the server with in a time period is tied to a single nymbleticket.Different nymble tickets are used across timeperiods, it grants the user anonymity between time periods. With in linkability window resources such as IP addresses can get reassigned i.e. it allows dynamismand it is undesirable to blacklist such resourcesindefinitely, and it ensures forgiveness of misbehaviorafter a certain period of time.

2.5 NOTIFYING THE USER OF BLACKLIST STATUS

Users expect their connections to be anonymous in anonymizing networks. Users be notified of their blacklist status before theypresent a nymble ticket to a server. In our system, theuser can download the server's blacklist and verify herstatus. If blacklisted, the user disconnects immediately.Since the blacklist is cryptographically signed by theNM, the authenticity of the blacklist is easily verifiedif the blacklist was updated in the current time period(only one update to the blacklist per time period isallowed).

2.6 BLACKLISTING A USER

A user connects and misbehaves at a server within linkability window. The server later detects this misbehavior and complains to the NM of the same linkability window.The server presents the nymble ticket of the misbehaving user and obtains the corresponding seed from the NM. The server is then able to link future connections by the user of the same linkability window to the complaint. once the server has complained about a user, that user is blacklisted for the linkability window.

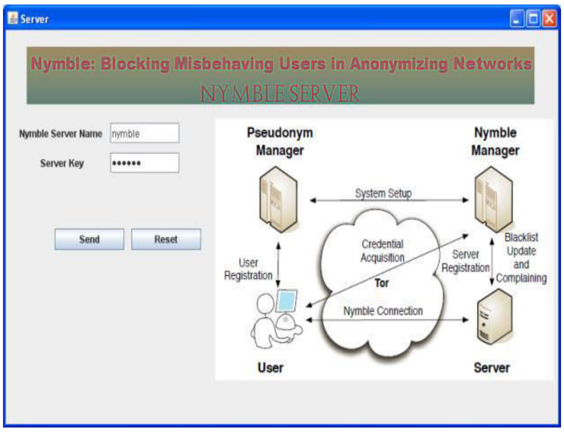
III. PORTING NYMBLE INTO MOBILE

Anonymizing networks allow users to access Internet services privately by using a series of routers to hide the client's IP address from the server. To hide the client's IP address and to provide the security nymble system is proposed. This paper ports the nymble system into mobiles, so that we can access the services of anonymizing networks in mobiles and with this the user can know the blacklist status in mobiles.The user can nymble connect or disconnect to the server for the accessing the services through mobiles.

With the help of Android OS the nymble system is ported into mobiles, by creating the class that extends the class activity and by importing the packages of Android and calling the nymble system in the class with this we can access the services of nymble system in mobiles.

Pseudonym code for porting nymble into mobile:

```
packagecom.exercise.AndroidInternetTxt;
importjava.io.BufferedReader;
import java.io.IOException;
importjava.io.InputStreamReader;
importjava.net.MalformedURLException;
import java.net.URL;
importandroid.app.Activity;
importandroid.os.Bundle;
importandroid.widget.TextView;

public class AndroidInternetTxt extends Activity
{
finalString textSource =
"http://192.168.1.55:8080/nymble_source/hackers.txt";

   // to access the nymble home page in mobiles

 /** Called when the activity is first created. */

  @Override
public void onCreate(Bundle savedInstanceState)
 {
try
 {
      textUrl = new URL(textSource);
//create url

while ((StringBuffer = bufferReader.readLine()) != null)
        {
            stringText += StringBuffer;
  //for user entered passwords
       }
     }
catch (MalformedURLException e)
     {
      // TODO Auto-generated catch block
           //to display misbehaved users
}
catch (IOException e)
     {
// TODO Auto-generated catch block
     }

   }

}

packagecom.exercise.AndroidInternetTxt;

public class EBCDIC
{
      public static String encrypt(String str)
            {
            //a method to encrypt the string
      }

            public static String decrypt(String str)
            {
            // a method to decrypt the string
            }
}
```



Server

Nymble: Blocking Misbehaving Users in Anonymizing Networks

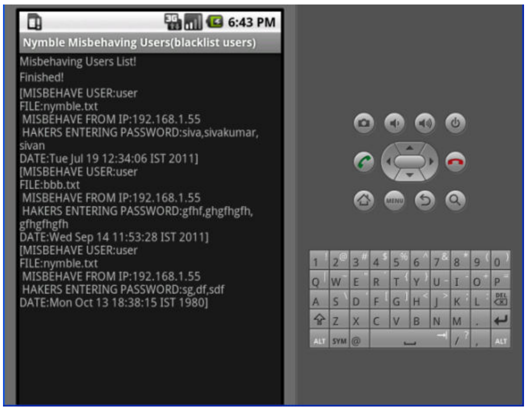
NYMBLE SERVER

Nymble Server Name: nymble

Server Key: ••••••

Fig 3: Nymble server webpage

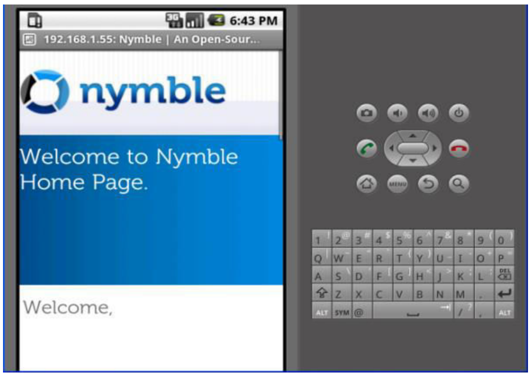
In the above nymble server webpage it shows the server login and send the reply to users. And also it shows the nymble system in anonymizing networks.



```
Nymble Misbehaving Users(blacklist users)
Misbehaving Users List!
Finished!
[MISBEHAVE USER:user
FILE:nymble.txt
 MISBEHAVE FROM IP:192.168.1.55
 HAKERS ENTERING PASSWORD:siva,sivakumar,
sivan
DATE:Tue Jul 19 12:34:06 IST 2011]
[MISBEHAVE USER:user
FILE:bbb.txt
 MISBEHAVE FROM IP:192.168.1.55
 HAKERS ENTERING PASSWORD:gfhf,ghgfhgfh,
gfhgfhgfh
DATE:Wed Sep 14 11:53:28 IST 2011]
[MISBEHAVE USER:user
FILE:nymble.txt
 MISBEHAVE FROM IP:192.168.1.55
 HAKERS ENTERING PASSWORD:sg,df,sdf
DATE:Mon Oct 13 18:38:15 IST 1980]
```

Fig . 4: misbehaving users information in mobiles

In the above figure, mobile displays the misbehaving user's information when the nymble system ported into mobiles. In that information it displays misbehaved IP address, when the user entered and passwords used for entering.



192.168.1.55: Nymble | An Open-Sour...

nymble

Welcome to Nymble Home Page.

Welcome,

Fig. 5: Nymble home page in mobiles

In the above figure it shows the nymble home page on mobiles when the nymble system is ported into mobiles through Android OS. By porting nymble system into mobiles, services of anonymizing networks are accessed in mobiles.

IV. CONCLUSIONS

The nymble system for blocking the misbehaving users in anonymizing networks, which can be used to add a layer of accountability to any publicly known anonymizing network as proposed by Tsang et. al.[1]is discussed. Nymble system provides anonymous authentication, backward unlinkability, subjective blacklisting, fast authentication speeds, rate-limited anonymous connections andrevocation auditability. These features of Nymble system paved a way to port it into mobile applications. The users can access the services of anonymizing networks and verify their blacklist status using their mobiles, once ported.

REFERENCES

[1] Patrick P. Tsang, ApuKapadia, Cory Cornelius, and S. W. Smith. Nymble:Blocking Misbehaving Users in Anonymizing Networks. IEEE transactions on Dependable and Secure Computing, vol 8, no. 2, Mar-Apr 2011.

[2] P. C. Johnson, A. Kapadia, P. P. Tsang, and S. W. Smith. Nymble:Anonymous IP-Address Blocking. In Privacy Enhancing Technologies,LNCS 4776, pages 113–133. Springer, 2007.

[3] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. In Usenix Security Symposium, pages303–320, Aug. 2004.

[4] D. Chaum. Showing Credentials without Identification TransfeeringSignatures between Unconditionally Unlinkable Pseudonyms.In AUSCRYPT, LNCS 453, pages 246–264. Springer, 1990.

[5] I. Damgard. Payment Systems and Credential Mechanisms withProvable Security Against Abuse by Individuals. In CRYPTO,LNCS 403, pages 328–335. Springer, 1988

[6] J. Camenisch and A. Lysyanskaya.An Efficient System for NontransferableAnonymous Credentials with Optional AnonymityRevocation. In EUROCRYPT, LNCS 2045, pages 93–118. Springer,2001.

[7] J. Camenisch and A. Lysyanskaya.Signature Schemes and AnonymousCredentials from Bilinear Maps. In CRYPTO, LNCS 3152,pages 56–72. Springer, 2004.

[8] J. R. Douceur. The Sybil Attack. In IPTPS, LNCS 2429, pages 251–260.Springer, 2002.

◈◈◈

Development of BPEL Processes Adopting Dynamic Business Strategies

V. Siva Madhuri

JNTUA College of Engineering, Anantapur, Andhra Pradesh, India
E-mail : shivamadhuri5204@gmail.com

Abstract – A template-based method is proposed to map business level process to BPEL process in order to support just-in-time reconfiguration of business process. This helps in mapping relationships between business level process and BPEL process, which helps users to make changes in the business level process with simplified operations, and automatically maps those changes to BPEL. This paper proposes a new approach for improving web service adaptability through a semantics modification in the invocation primitive. Hence programmers can define dynamic web service compositions without making any changes in the source code. A loan approval system was applied in order to illustrate this approach.

Keywords - *business process execution language (BPEL), template, adaptability.*

I. INTRODUCTION

BPEL (Business Process Execution Language, [1]) is the most widely used solution for work flow-based cooperation amongst web services. In enterprises, business processes play an important role in enhancing the competitiveness and to keep the sustainable development of enterprises. With the application of business process, an efficient method to organize services is the process modeling, in order to meet the needs of the business collaboration. Because the original rigid modeling method could not adapt to the rapid changes in the business environment, so there is an urgent need for a flexible business process construction method.

WS-BPEL(Web service business process execution language) (BPEL in short) is a language for specifying business process behavior based on Web services, which could generate a new Web service by composing acquired Web services. To allow enterprise collaboration, web service composition has been emerged as an important strategy [2]. In order to describe complex business process, BPEL defines various activities including service calling, data operating, troubleshooting and so on. It also defines a series of structured activities, such as sequence, flow. All these factors cumulatively contribute towards making BPEL a popular process modeling language. However, it is difficult for business staff to construct BPEL process, because BPEL is a language for the specification of executable business process which means it is too complex for business staff to understand. Therefore, the transformation from the easily-understandable business process modeling methods to the executable BPEL Process is attracting more attentions. It is an urgent need to realize the automatic or semi-automatic mapping from business process to executable BPEL process. Specifications based on WS-BPEL use XML to describe aspects of the process: partners in the message exchange, available services, the orchestration [3], message's format, and so on.

Our proposal addresses template based runtime adaptability in web service composition, specifically in the level of exchanging partners and web services, considering neither additional tasks to programmers nor changes in the syntax of standard specifications. We propose a semantic extension to a WS-BPEL primitive (without changing its syntax) in such way that it can be adaptive, making the composition adaptive. As a consequence, adaptability becomes transparent to programmers. This paper is organized as follows: Section II briefly introduces basic concepts of Service Level Agreements. Section III presents the proposed approach in details. Section IV illustrates a practical use of the proposed approach. Finally, last section presents an evaluation of the research.

II. SERVICE LEVEL AGREEMENTS

Service level agreement means the part of a contract between the client and provider of an internet service that describes the quality attributes that the service is required to possess. Service Level Agreements can be defined by using SLAng, which is based on XML language.

Increasingly, distributed systems primitives are being used to build mission-critical applications that cross autonomous organizations. Examples include the use of web services in, for example, supply chain management, or computing on demand using distributed component or grid technology as offered by IBM. Because the usefulness, and sometimes even the functioning of a business, depend not only on the functionality but also the quality of these services, for example performance and reliability, and because these qualities not only depend on the behavior of the service but on that of the client, contracts between the provider and client of a service must contain terms governing their individual and mutual responsibilities with respect to these qualities.

This paper describes the approach taken to produce a rigorous yet understandable specification of the semantics of SLAng, and the use of the semantics in reasoning about service composition.

SLAng is an evolving language. The related work selection highlights desirable features of previous SLA languages that SLAng could benefit from incorporating. These include: Payments related to service violations; reuse features, including SLA templates, clause reuse and generalization hierarchies; the specification of management actions, performed in response to QOS state changes or SLA violations; the ability to define new parameter types in terms of existing types; the ability to distribute clauses to third parties without exposing sensitive details of the client and server; and the ability to define dynamic relationships between service levels, effectively defining the permissible states for a system capable of adapting its QOS behavior. Using SLAng [4] it is possible to produce Service Level Agreements between two parties at (application level, as well as at application services and system resources level), through expressing quality parameters pertaining the particular working level and type of agreement. The major contribution of SLAng is to fill the gap of QoS, by producing a formal language, with a well-defined syntax and semantics, for describing service level specifications.

We have presented the approach taken to defining the semantics of SLAng. The advantage of the semantic is that they are helpful to reduce the ambiguity of the language, improve its consistency and to support SLA composition. The model forms the basis for a common understanding of SLA terms between parties to an SLA, reducing the risk of outsourcing functionality to other organizations, and providing an unambiguous reference for negotiation and arbitration of agreements. These facilities are vital to the emerging Internet service business model.

III. PROPOSED APPROACH

This section presents the proposed approach, their related elements and implementation.

1. General Overview

As mentioned before, this work provides adaptability both in a transparent way and preserving standards (more specifically, WS-BPEL).Two main points must be considered in order to carry out this task: different implementations of a service must have the same interface, i.e., they must have the same WSDL specification; and the composition's designer must be able to make available existing new web services to the engine.

2. Adaptability

The interest in adaptive systems has been increased. Due to the dissemination of ubiquitous and autonomic computing, several issues related to adaptability have been widely studied. In the context of web services, adaptability comes from the fact that new web services are made available in a very frequent way. New web services may be considered in an automatic way, without stopping the whole system. Additionally, the Internet does not provide a reliable communication environment and servers where the web service is hosted may encounter problems. In this context, the importance of using adaptability is apparent in the web services world [5].

Adaptations can be static or dynamic, manual or automatic, and proactive or reactive [6]. The static adaptation is carried out through modifications in the source code, while the dynamic one modifies software runtime characteristics. Manual adaptation means direct intervention in the system, whilst automatic one can be performed by the system itself. Finally, proactive adaptation occurs before a specific event and reactive adaptation happens after it. For example, the system may realize that one service is not available, and take actions before the execution of the composition (proactive adaptation). If the system takes action after the unsuccessful call, a reactive adaptation will be made.

3. ActiveBPEL

The adoption of ActiveBPEL was based on some criteria. Firstly, despite the fact that the proposed approach is independent from any particular engine, one had to be chosen. Secondly, most of the engines are not open-source. As a consequence, their source code cannot be modified to incorporate the proposed approach. Thirdly, depending on the engine, different languages and platforms are used, which imposes further difficulties to do a generic implementation. Finally, ActiveBPEL is widely used in academy and industry, written in Java and has a good documentation.

WS-BPEL specifies many primitives to allow various forms of composition, e.g., sequence (<sequence>), flow (<flow>) and while (<while>). Input and output operations are specified with the use of the primitives receive (<receive>) and reply (<reply>), respectively. In particular, our approach concentrates on the primitive invoke (<invoke>), which performs invocations on participating web services of the composition.

IV. USING THE ACTIVEBPEL: THE LOAN APPROVAL SYSTEM

In order to illustrate the use of the ActiveBPEL extension, we present a case study that benefit from the adaptation.

1. Overview of the Loan Approval System

As a case study, a traditional application was chosen, the Loan Approval System. This application was originally presented in [7] and has also been adopted in [8]. A company may be interested on verifying whether its clients can make loans. So a system was implemented in order to automate this task. Two web services, named “Assessor” and “Approver”, compose this business process.

Depending on both the value required by the customer and the return of assessor service (defining whether the loan has a low or a high risk), it is defined whether the intervention of the approver service is necessary or not. The reason to use an Assessor instead of having only an Approver is not overload it with a large number of requests. If only one of them is necessary to evaluate the request, a choice composition is executed. Otherwise, a sequence composition is carried out.

2. Tools

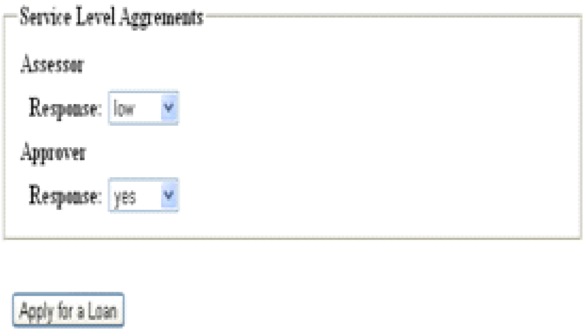
All experiments have been executed using the Apache-Tomcat web server. Two servers were configured to evaluate the composition of distributed web services. ActiveBPEL is a web based tool used for executing BPEL processes.

3. Execution

The composition consists of two web services (Assessor and Approver) in a sequential/choice composition, depending on the loan amount and service level agreements. When the administrator login into the home page if the credentials are authorized, the administrator can go to the next page otherwise he has to enter the authorized details. Then he can access the java server page, it will be taken by Runtime parameter java class and setting these values to the xml file.XML file is based on the SLA agreements. Administrator can enter his values with two Service Level Agreements called “Assessor with value YES” and “Approver with value NO”. These values will give effect at XML file called “bpel_example_config.xml”. This is the XML file called “bpel_example_config.xml” where its values get updated by the values entered by the user with corresponding SLAs (Service Level Agreements) given through template page.

If an Assessor value is “high” then BPEL Engine internally invokes Web Service by using the URL http://localhost:8080/active-bpel/services/AssessorWebService and if Approver value is “No” then BPEL Engine internally invokes Web Service by using following URL http://localhost:8080/active-bpel/services/ApproverWebService. Similarly remaining service level agreements will be invoked by the BPEL Engine.

Generally SLAs in distributed applications comprising web services are written in a text file using some language. However, the work done in this project to simplify this is to have a user-friendly template through which administrator can easily provide SLAs. The following is the template part of SLAs input in the web application that has been built.



Service Level Agreements

Assessor

Response: low

Approver

Response: yes

Fig. 1: shows template for SLAs

4. Performance Evaluation:

Expected BPEL process answer; depends upon risk-level and accept. Here, amount 0 represents any value less than 10,000 and 20000 represents any value greater than or equal to 10,000.

AMOUNT	RISK LEVEL	ACCEPT	EXPECTED RESPONSE
0 – 19999	High	Yes	Yes
0 – 19999	High	No	No
0 – 19999	High	FAULT	FAULT
0 – 19999	Low	Ignored	Approved
0 – 19999	FAULT	Ignored	FAULT
>20000	Ignored	Yes	Yes
>20000	Ignored	No	No
>20000	Ignored	FAULT	FAULT

Table 1: Experimental Results

Risk level and accept values are the service level agreements used in this paper.

V. CONCLUSION

We present the template-based mapping method to map business level process to BPEL process.

Our investigation showed that these templates can attract the users with a limited background, but with knowledge of business domain.

REFERENCES

[1] Luciano Baresi and Sam Guinea,"Self-Supervising BPEL Processes", IEEE Trans. Software Eng., Vol 37, no. 2,pp.247 to 263,March/April 2011.

[2] B. Benatallah et al, "Towards Patterns of Web Services Composition",in: Patterns and Skeletons for Parallel and Distributed Computing.

[3] M. Juric, Business Process Execution Language for Web Services. Packt Publishing, Mumbai, India. 2006.

[4] D. D. Lamanna, J. Skene, and W. Emmerich. Specification language for service level agreements. Technical Report D2,University College London, March 2003.

[5] N. Milanovic and M. Malek, "Current Solutions for Web ServiceComposition", in IEEE Internet Computing, vol.8, n.6, pp.51-59. 2004.

[6] C. Courbis and A. Finkelstein, "Towards Aspect WeavingApplications", in 27th International Con-ference on SoftwareEngineering – ICSE. St. Louis, Missouri, USA. 2005.

[7] LLC ActiveBPEL, "The Open Source BPEL Engine".http://www.activebpel.org (last visit: 2th Sep-tember 2006). 2006.

[8] F. Lins et al, "Avaliação de Desempenho da Composição de WebServices usando Redes de Petri", in Brazilian Symposium on ComputerNetworks (SBRC 2006). Curitiba, Paraná, Brazil. 2006.

◈◈◈

Top-*k* Query Processing in Integrated Uncertain Databases Using TPRQ

S Azad Razvi & S Vikram Phaneendra

Dept. of Computer Science and Engineering, Madanapalle Institute of Technology & Science, Madanapalle, India
E-mail : azadrazvis@gmail.com,

***Abstract* –** Querying integrated uncertain databases using Threshold Probabilistic Ranked Query (TPRQ) is different from Top-*k* processing. Top-*k* processing provided results based on scores, membership probabilities, traditional top-*k* semantics and possible world semantics. An integrated uncertain database contains data from different data sources and also consists of multiple tuples relating to same object. As the number of duplicate tuples related to query increases, the number of possible worlds increases significantly and this will take very long time to produce results so to overcome this problem this paper propose methods that will cut off processing time as well as response time. This paper introduce new framework to speed up the TPRQ with scoring functions over the integrated uncertain database and effective duplicate tuples elimination methods to remove the duplicate tuples and to reduce the TPRQ search space. The results will be returned to the user with greater than or equal to minimum probability provided by the user. Our formulations are based on traditional top-*k* semantics and possible world semantics. By using these techniques there is a reduction in processing time and materialized search states.

***Keywords* -** *duplicate tuples; framework; ranked query; TPRQ; uncertain database;*

I. INTRODUCTION

At present processing of uncertain data is getting important in different areas like sensor networks, moving object environments and data cleaning. Some examples related to uncertain databases include sensor networks which have large number of sensor devices are used in different areas experimentally. Different applications from different areas need to manage this sensor data which are collected from sensor devices. Some applications are environmental monitoring, surveillance and traffic monitoring. There are some limited capacities of sensors like limited bandwidth, low ranges and high losses. Sensor data collected from its surroundings have disturbance and sometimes unavailable [1]. There are many advances in wireless sensor networks and it can collect data from all types of environments [2]. Systems that observe continuous movement of objects that receive updated data as the object move in space. Due to limited battery power and bandwidth of moving objects precise value will not be stored in the database. Due to transmission delay the value which is present in the database may deviate from the actual sensed value so exact position of object may not be available in the database at the query time. The internal property of applications that monitor this data is to update the positions of objects periodically and we cannot know the precise position until the next update. If the value is not updated the uncertainty of the exact position of object increases. Due to these problems only approximate values present in the database rather than precise values [3].

To keep valid data in a database we need to impose integrity constraints on data. But sometimes constraints are violated due to integration of data from different sources which have different constraints and after integration duplication of constraints may occur. Two sources designed independently with different constraints exchange data may not satisfy the destination constraints. Violation of a constraint means data which does not satisfy a constraint [4]. After integration of records from different data sources there is a chance of existence of duplicate records. Duplication can also occur from data entry errors. Some of the techniques used for duplicate detection are clustering and classification. Duplication is occurring due to recording information about the same object by two or more different data sources. Duplicate tuples is alternative representation for same real world object. If duplicate tuples exist in the database then we cannot know exactly whether the answer is exact so there is some uncertainty exists in the answer. Some integrated databases collect information from different sources may contain inconsistent information about the same customer so some data integration tools use conflict resolution rules which support merging of tuples belong to same entity. Dirty databases mean containing conflicting tuples. Dirty database can violate a set of key constraints [5].

We need a model to deal with integrated uncertain data from different databases and there are two ways to deal with uncertainty: probabilistic data models and fuzzy data models. Many of them use probabilistic data models to deal with uncertainty. So this paper concentrates on probabilistic data models. Some analysis extends the relational model to represent uncertainty using probability calculus. There are two types of uncertainty attribute uncertainty and tuple uncertainty [6]. In the previous frameworks [7] when we query uncertain database it returns tuples which also consists of duplicate tuples and possible world construction takes more time than expected and also increase in processing time so we need an approach to remove duplicate tuples with same probability and unwanted tuples with low probability using mutual exclusion [7].

With threshold based approach we can query the database with minimum threshold probability so that we can reduce number of tuples for processing. Let us assume a location with different sensors which are used for detecting temperature automatically when there is a change in temperature. Due to the limitations of equipment like delay or loss in data transfer and low battery power cannot detect properly so instead of precise value some unknown values are stored or may not return value or may return some wrong values due to interference. So for this reason if we deploy two or more sensors at the same location there can be at most one sensor with precise value. Multiple sensors are deployed to increase the detection quality. Some examples which shows different tables with existence of duplicate values.

TID	Time	Sensor	Location	Temp	Conf
T1	11:30	L1	A1	25°C	0.3
T2	11:30	L2	B1	21°C	0.4
T3	11:45	L3	B1	13°C	0.5
T4	11:45	L4	A1	12°C	1.0
T5	11:55	L5	C1	17°C	0.5
T6	12:00	L6	C1	17°C	0.5

Table 1 : Wireless Sensor Uncertain Database with duplicate values

ID	Time	Radar Loc	Car Model	Plate No	Speed	Conf
T1	11:30	S1	Honda	X-12	130	0.3
T2	11:35	S2	Toyota	Y-24	120	0.4
T3	11:20	S3	Toyota	Y-24	80	0.5
T4	11:45	S4	Mazda	O-54	90	1.0
T5	11:50	S5	Mazda	O-54	110	0.5
T6	11:55	S6	Chevy	L-10	105	0.5

Table 2 : Integrated Uncertain Database with duplicate values

This table consists of duplicate tuples which are eliminated with mutual exclusion principle.

TID	Time	Sensor	Location	Temp	Conf
T1	11:30	L1	A1	25°C	0.4
T2	11:30	L2	B1	21°C	0.7
T3	11:45	L3	B1	13°C	0.3
T4	11:45	L4	C1	12°C	0.4
T5	11:55	L5	C1	17°C	0.6
T6	12:00	L6	D1	11°C	1.0

Rules: (t2⊕t3) , (t4⊕t5)

Table 3. Uncertain Database with temperature values

Possible world	Probability
PW1={T1,T2,T6,T4}	0.112
PW2={T1,T2,T5,T6}	0.168
PW3={T1,T6,T4,T3}	0.048
PW4={T1,T5,T6,T3}	0.072
PW5={T2,T6,T4}	0.168
PW6={T2,T5,T6}	0.252
PW7={T6,T4,T3}	0.072
PW8={T5,T6,T3}	0.108

Table 4. The possible worlds of Table 3

Example 1: Let us consider an above example with snapshot of sensor database in the last half an hour. six different sensors which are deployed at three different locations A1,B1 and C1. T1 to T6 are tuples and L1 to L6 are sensors. Sometimes sensors which are deployed at same location may detect different temperature values. Each tuple is associated with a confidence field “Conf” indicating its membership probability. At particular time sensor with highest probability will be considered.

When data is integrated from different data sources in to particular database there is a chance of occurrence of duplication of tuples so we need to eliminate such kind of duplicate tuples before we sending these extracted tuples according to query for processing. The Figure 1 carries the possible world semantics [7]. The possible worlds are validated by the generation rules. T1 and T2 are cannot be true at the same time can be given by the generation rule t1⊕t2.

1.1 Threshold Probabilistic Top-*k* Ranked Queries

A set of tuples in table T can be specified using generation rules as $R : t_{r_1} \oplus \cdots \oplus t_{r_m}$ where $t_{r_1} \in T$ (1≤i≤m) and $\sum_{i=1}^{m} Pr\ (t_{r_1}) \leq 1$. The constraint given by rule R is only one tuple can be selected among all tuples which are involved in the rule. The selected tuple will be appear in the possible world [8].

From [7][8] let us assume that each tuple is involved in atmost one generation rule. For independent tuples we can give a little rule R_t : t. so uncertain table has a set of generation rules such that only one rule from generation rule in $\mathcal{R}_T$ can be applied to each tuple. If tuple is involved in rule R then we can say that t ∈ R . The probability of a rule is given by sum of membership probabilities of all tuples involved in the rule and it is denoted by

$$Pr(R) = \sum_{t \in R} Pr(t)$$

A generation rule R is a singleton rule if $|R| = 1$. A tuple is said to be dependent if it is involved in multi-tuple rule such that $|R| > 1$ otherwise it is said to be independent. Given the generation rule R and subset of tuples $S \subseteq T$. The tuples involved in R and also appeared in S is denoted as

$R \cap S$. We denote W as a possible world and $\mathcal{W}$ as a set of possible worlds.

The uncertain table T consists of tuples with membership probability. There are number of tuples $t \in T$ such that $Pr(T)$>0. A Threshold Probabilistic Ranked Queries (TPRQ) on an uncertain table T consists of a top-k query Q and a threshold probability p ($0<p\leq 1$). For each possible world W, Q is applied and a set of k tuples $Q^k(W)$ is returned. For a tuple $t \in T$, the top-k probability of t is the probability that t is in $Q^k(W)$ in all $W \in \mathcal{W}$, that is,

$$Pr^k_{Q,T}(t) = \sum_{w \in \mathcal{W}, t \in Q^k(W)} Pr(W)$$

When Q and T are clear from context, we often write $Pr^k_{Q,T}(t)$ as $Pr^k(t)$ for the interest of simplicity. The answer set to a TPRQ is the set of all tuples whose top-k probability values are at least p, that is

$$Answer(Q,p,T) = \{t|\ t \in T, Pr^k_{Q,T}(t) \geq p\}.$$

The new method to answer the TPRQ is to determine the number of possible worlds and query is applied to each possible world. Then, compute the probability of each tuple in the possible world according to top-k and select the tuples according to the user specified threshold [8]. We construct a novel framework which extracts results from integrated uncertain database in sorting order according to scoring function which is applied on a predicate in a relation.

1.2 Understanding of possible world semantics

Possible world is modeled with membership uncertainty where each tuple in the database have confidence and generation rules that determine the valid possible world. For example mutual exclusion rule which does not consider two or more identical tuples belong to same possible world. In Table-4 PW5 is computed by assuming the existence of T2,T6 and T4 and the absence of T1,T3 and T5. Using generation rule constraint [7] T3 and T5 has been excluded and absence probability of T1 is given as 1-Pr(T1). Therefore $Pr(PW5)$=0.7 × 1.0 × 0.4 × 0.6 = 0.168. Therefore 0.7, 1.0, 0.4 are existence probabilities of T2, T6, T4 and 0.6 is the absence probability of T1.

1.3 Understanding novel Framework

Our framework shown in Fig. 1 consists of a integrated uncertain database and several working modules. First user issues a query to the database with some minimum probability, user defined threshold access module retrieve tuples based on query and probability with atleast p. duplicate tuple eliminator module merges two or more tuples with same probability in to one and sorted score access module sorts tuples based on scoring function and some attribute. Space navigation module will materialize the search states and to compute the probability of each state, state formulation module formulates each state and computes state probability by contacting rule engine.

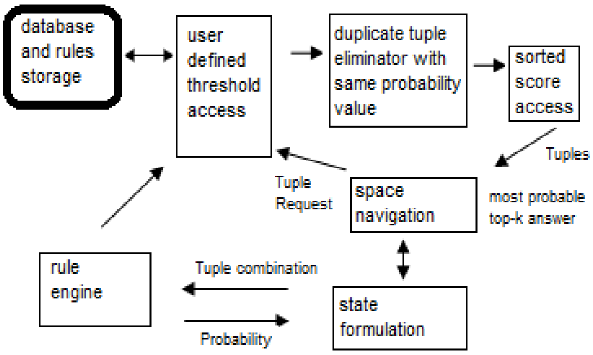

database and rules storage
user defined threshold access
duplicate tuple eliminator with same probability value
sorted score access
Tuples
Tuple Request
space navigation
most probable top-k answer
rule engine
Tuple combination
Probability
state formulation

Fig. 1 : Processing Framework.

In this paper section-1 gives introduction about uncertain databases, possible world semantics and our novel framework. Section-2 compares previous work and the current work and explains our motivation. Section-3 gives algorithms related to our framework and section-4 concludes our paper and explains about future work.

II. RELATED WORK

The motivation of this work very similar to recent work of Mohamed A. Soliman [7]. The focus of [7] is to produce Top-k results using possible world symantics. Their formulations are based on state space search so as the number of accessed tuples increases, processing time as well as response time also increases and their techniques can be integrated with very small databases with very low storage capabilities. There is a need to develop much more effective framework and the goal of this work is to derive such effective framework with minimal processing time. We make use duplicate removal techniques and user defined threshold probability. Hence motivations are similar but our ideas and techniques are different that those [7]. Our work also has same motivations related to a probabilistic threshold approach [9].

There are a number of uncertainty database systems with different methods and different goals. For example Top-k processing in uncertain databases [7] is a framework attempting to separate the storage layer and processing layer. The storage layer is used for tuple retrieval, indexing and traditional query processing. Uncertain data and generation rules are present in relational database and provide different data access methods for processing layer to retrieve tuples. In processing layer different states are constructed with retrieved tuples and from that valid possible worlds are constructed and methods that are provided are navigate through this possible worlds to find top-k answers. Data Cleansing and The Merge/Purge Problem [8] is an integrated database and provides methods for removal of common information about same entity, to solve Merge/Purge problem and data cleansing.

Our work mainly focus on reduction on access tuples from database using using user defined threshold probability and duplication removal. Our ideas for user defined threshold probability is similar to [9] but we have presented modified version of exact algorithm. For duplicate detection, incremental merge/purge algorithm is used which is [8].

III. ALGORITHMS

3.1 Threshold based top-k processing

$Q^k_{P,f}$ is the top-k query, T is an uncertain table, $P(T)$ is set of tuples satisfying query predicate. Each tuple in $P(T)$ has same membership probability as T. Each tuple in the uncertain table belongs to only one generation rule in $\mathcal{R}_T$. One or more tuples can belong to rule R but two or more rules cannot belong to one tuple. The probability of the rule is given as sum of membership probabilities of all tuples involved in the rule.

$$Pr(R) = \sum\nolimits_{t \in R} Pr(t)$$

The set of tuples satisfying the query predicate is given as $P(T) = \{t|\ t \in T \wedge P(t) = true\}$. The generation rule which was applied to T was applied to $P(T)$ by taking tuples which belong to $P(T)$. TPRQ results are calculated by finding those tuples which passes given probability threshold in Top-k. P is called probability threshold. A top-k query $Q^k_{P,f}$ which contains predicate P, ranking function f, and an integer k>0. When query Q is applied on tuples, tuples are ranked according to the scoring function f according to the predicate p and top-k tuples are returned. For a tuple t ϵT, the top-k probability of tuple t means the probability with which tuple t is in possible world W.

$$Pr^k_{Q,T}(t) = \sum_{w \in \mathcal{W}, t \in Q^k(W)} Pr(W)$$

Input: an uncertain table T, a set of generation rules $\mathcal{R}_T$, a top-k query $Q^k_{P,f}$, and a probability threshold p;

Output: Answer(Q,p,T);

Method:

```
1:  retrieve tuples in $P(T)$ in the ranking order in the ranking order one by one
    For each $t_i \in P(T)$ do
2:  compute $T(t_i)$ by rule-tuple compression.
3:  compute subset probability values and $Pr^k(t_i)$.
4:  if $Pr^k(t_i) \geq p$ then output $t_i$.
5:  if all remaining tuples in $P(T)$ fail the probability threshold then exit.
    end for
```

3.2 Duplication removal algorithm

3.2.1 The Basic Sorted-Neighborhood Method

Consider integrated uncertain database and apply sorted-neighborhood method. We are dividing the sorted-neighborhood method in to three phases

1) Create Keys: the key is computed by extracting distinct fields rather than partial fields.

2) Sort Data: Sort the records in the data list using the key created before.

3) Merge: move the fixed size window through the sequential list of records. If the window size is w records then every new record entering the window is compared with the previous w-1 records to find "matching" records.

3.2.2 Incremental Merge/Purge

Definitions:

R_0: The initial relation

$\triangle_i$: The i-th increment relation

c_i : A relation of only "prime representatives" of the clusters

Identified by the Merge/Purge procedure.

Initially:

$\triangle_o \leftarrow R_0$

$c_o \leftarrow \emptyset$

$i \leftarrow 0$

Incremental Algorithm:

For every $\triangle_i$ do:

Begin

1. Let $I_i \leftarrow CONCATENATE(c_i, \triangle_i)$.
2. Apply any Merge/Purge procedure to I_i. The result is a cluster

   Assignment for every record in I_i.
3. Separate each record in I_i into the clusters assigned by the previous step.
4. For every cluster of records, if necessary, select one or more records as prime representatives

   For the cluster. Call the relation formed of all selected prime representatives, c_{i+1}.

   end.

The merge/purge procedure presented above assumes a single data set. When new data set arrives, it concatenates to the previously processed data set and merge/purge procedure runs again on the whole data set. The incremental merge/purge algorithm [8] removes this constraint by using information gathered from previous executions. This incremental algorithm reduces time compared to the merge/purge algorithm.

3.3 Top-*k* Processing algorithm

Require:

Source: Score-ranked tuple stream

k: Answer length

Ensure: Top-*k* queryanswer

```
Q ← empty priority queue for states ordered on probabilities
e ← $s_{0,0}$ where $\mathcal{P}(e) = 1$ {init empty state}
d ← 0 {scan depth in source}
insert e into Q
while ( source is not exhausted AND Q is not empty) do
    $s_{l,i}$ ← dequeue (Q)
    if (l=k) then
        return $s_{l,i}$
    else
        if(i=d) then
            t ← next tuple from source
            d-d+1
        else
            t ← tuple at pos i + 1 from seen tuples
        end if
        Extend $s_{l,i}$ using t into $s_{l,i+1}$, $s_{l+1,i+1}$
        Insert $s_{l,i+1}$, $s_{l+1,i+1}$ into Q
    end if
end while
return dequeue(Q)
```

Top-*k* processing algorithm buffers the ranked tuples from the sorted score access module. Let d be the number of tuples retrieved from buffer. When next tuple is drawn from buffer, the algorithm extends the state which has the highest probability. Let s_l be a state and $s_{l,i}$ be the extended state where *i* is the position of last seen tuple in score ranked tuple stream. Let e be the empty state with length zero. Let us initialize e with $s_{0,0}$ where $\mathcal{P}(s_{0,0})$=1. Q is the priority queue with states ordered according to probabilities and initialize Q with e. let d be the seen tuples from sorted score access module. The top-*k* processing algorithm iteratively retrieves the state from Q and extends the state $s_{l,i}$ with two possible states [7] and reloads in the queue according to the order of probability. Retrieving new tuple from buffer occurs due to extending the state. When top state in Q becomes complete then top-*k* processing algorithm will be terminated.

IV. CONCLUSION AND FUTURE WORK

We have defined the top-*k* query processing using possible world semantics and modified the previous framework to process TPRQ. Previous frameworks was taking more time to produce possible worlds due to processing of more number of tuples related to query and presence of duplicate tuples in the obtained query results from database so we need a process to remove duplicates and also to produce top-*k* results based on greater than or equal to user defined probability so that we can increase the storage capacity of the database and also we can reduce the response time. Our main aim of this work is to increase the database tuple storage capacity and to reduce the number of search states and response time. For future work we are proposing a novel framework which has capability to remove duplicate tuples in early stages and provide results based on user defined probability in less time compared to previous works.

REFERENCES

[1] J. Considine, F. Li, G. Kollios, and J. Byers, "Approximate aggregation techniques for sensor databases," Proc. ICDE, 2004, IEEE, Apr. 2004, pp. 449-460, doi:10.1109/ICDE.2004.1320018.

[2] A. Silberstein, R. Braynard, C. Ellis, K. Munagala, and J. Yang, " A sampling-based approach to optimizing top-k queries in sensor networks," Proc. ICDE, 2006, IEEE, Apr 2006, doi: 10.1109/ICDE.2006.10.

[3] R. Cheng, D. V. Kalashnikov, and S. Prabhakar, "Querying imprecise data in moving object environments," IEEE Trans.on Knowledge and Data Eng., vol. 16, Sept. 2004, pp:1112–1127, doi: 10.1109/TKDE.2004.46.

[4] A. Fuxman, E. Fazli, , and R. J. Miller, "Conquer: Efficient management of inconsistent databases," Proc . SIGMOD, 2005, ACM, pp. 155-166, doi: 10.1145/1066157.1066176.

[5] P. Andritsos, A. Fuxman, and R. J. Miller, "Clean answers over dirty databases: A probabilistic approach," Proc. ICDE, April 2006, IEEE, doi: 10.1109/ICDE.2006.35.

[6] Shichao Zhang and Chengqi Zhang, "A Probabilistic Data Model and Its Semantics," Journal of Research and Practice in Information Technology, vol. 35, Nov 2003, pp. 227-246.

[7] Mohamed A. Soliman, Ihab F. Ilyas and Kevin Chen-Chuan Chang, "Top-k Query Processing in Uncertain Databases," Proc. ICDE, April 2007, IEEE, pp. 896-905, doi: 10.1109/ICDE.2007.367935.

[8] Mauricio A. Hernández and Salvatore J. Stolfo, "Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem," DMKD, SPRINGER, vol. 2, 1998, pp. 9-37, doi: 10.1023/A:1009761603038.

[9] Ming Hua, Jian Pei, Wenjie Zhang and Xuemin Lin, "Ranking queries on uncertain data: a probabilistic threshold approach," Proc. SIGMOD, 2008, ACM, pp. 673-686, doi: 10.1145/1376616.1376685.

◈◈◈

Analyzing The Query Performance Over A Distributed Network of Data Aggregators

P. Prabakar^1 & S. Nageswara Rao^2

Dept. of CSE, Madanapalle Institute of Technology and Science, Madanapalle,
E-mail : prabhakar.mth@gmail.com, nag_sirisala@yahoo.com

Abstract **-** Typically a user desires to obtain the value of some aggregation function over distributed data items, for example AVG of temperatures sensed by a set of sensors. In these queries a client specifies a coherency requirement as part of the query. We present a low-cost, scalable technique to answer continuous aggregation queries using a network of aggregators of dynamic data items. In such a network of data aggregators, each data aggregator serves a set of data items at specific coherencies. Our technique involves decomposing a client query into sub-queries and executing sub-queries on judiciously chosen data aggregators with their individual sub-query incoherency bounds. We provide a technique for getting the optimal set of sub-queries with their incoherency bounds, which satisfies client query's coherency requirement with least number of refresh messages sent from aggregators to the client. For estimating the number of refresh messages, we build a query cost model which can be used to estimate the number of messages required to satisfy the client specified incoherency bound. Performance results using real-world traces show that our cost based query planning leads to queries being executed using less than one third the number of messages required by existing schemes

Keywords: *Content distribution networks, Continuous queries, Distributed query processing, Data dissemination, Coherency, Performance.*

I. INTRODUCTION

Application such as auctions, personal portfolio for financial decisions, sensors based monitoring, route planning based on traffic information, etc., make extensive use of dynamic data. For such applications, data from one or more independent data sources may be aggregated to determine if some action is warranted. Given the increasing number of such applications that make use of highly dynamic data, there is significant interest in systems that can efficiently deliver the relevant updates automatically. Many data intensive applications delivered over the Web suffer from performance and scalability issues. Content distribution networks (CDNs) solved the problem for static content using caches at the edge nodes of the networks. CDNs continue to evolve to serve more and more dynamic applications [1, 2]. A dynamically generated web page is usually assembled using a number of static or dynamically generated fragments. The static fragments are served from the local caches whereas dynamic fragments are created either by using the cached data or by fetching the data items from the origin data sources. One important question for satisfying client requests through a network of nodes is how to select the best node(s) to satisfy the request. For static pages content requested, proximity to the client and load on the nodes are the parameters generally used to select the

appropriate node [3]. In dynamic CDNs, while selecting the nodes node(s) to satisfy the client request, the central site (top-level CDN node) has to ensure that page/data served meets client's coherency requirements also. Techniques to efficiently serve fast changing data items with guaranteed incoherency bounds have been proposed in the literature [4, 5]. Such dynamic data dissemination networks can be used to disseminate data such as stock quotes, temperature data from sensors, traffic information, and network monitoring data. In this paper we propose a method to efficiently answer aggregation queries involving such data items.

In data dissemination schemes proposed in literature [4, 11], a hierarchical network of data aggregators is employed such that each data aggregator serves the data item at some guaranteed incoherency bound. *Incoherency of a data item at a given node is defined as the difference in value of the data item at the data source and the value at that node.* Although CDNs use page-purge [8] based coherency management, we assume that in dynamic data dissemination networks, these messages carry the new data values thereby an invalidation message becomes a refresh message. For maintaining a certain incoherency bound, a data aggregator gets data updates from the data source or some higher level data aggregator so that the data incoherency is not more than the data incoherency bound. In a hierarchical data dissemination network a higher level aggregator guarantees a tighter incoherency bound compared to a lower level aggregator. Thus, data refreshes are pushed from the data sources to the clients through the network of aggregators. **Data incoherency:** data accuracy can be specified in terms of incoherency of a data item, defined as the absolute difference in value of the data item at the data source and the value known to a client of the data. Let $v_i(t)$ denote the value of i^{th} data item at the data source at time t; and let the value the data item known to the client be $u_i(t)$. Then the data incoherency at the client is given by $|u_i(t)\text{-}v_i(t)|$. For a data item which needs to be refreshed at an incoherency bound C a data refresh message is sent to the client as soon as data exceeds C, i.e., $|u_i(t)\text{-}v_i(t)|>C$.

Network of data aggregators: Data aggregators are one kind of secondary server it serves as data sources (data items). The data refreshes can be done using two mechanisms.*(a)Push* based mechanism data source send update messages to client on their own.*(b)Pull* based mechanism data sources send messages to the client only when client makes a request. For scalable handling of push based data dissemination, network of DA's are proposed in the literature [12,15,16]. In such network of DA's, data refreshes occur from data sources to the client through one or more DA's. In this paper we assume that each DA maintains its configured incoherency bounds for various data items. Dissemination networks for various data items can be overlaid over a single network of data aggregators as shown in Figure 1. Thus, From a data dissemination capability point of view, each data aggregator (DA) is characterized by a set of (d_i,c_i) pairs, where d_i is a data item which the DA can disseminate at an incoherency bound C_i.The configured incoherency bound of data item at a DA can be maintained using any of following methods: (a) the data source refreshes the data value of the DA whenever DA's incoherency bound is about to get violated. This method is scalability problems. (b) data aggregators with tighter incoherency bound help the DA to maintain its incoherency bound in scalable manner as explained in [4,7].

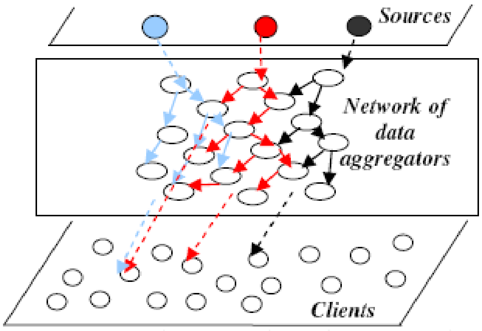


Figure 1: Data dissemination network for multiple data items

1.1.Problem Statement and Contributions

Value of a continuous weighted aggregation query at time *t*, can be calculated as:

$$V_s^q(t) = \sum_{i=1}^{i=n^q} s_i(t) \times w_i^q \tag{1}$$

V_s^q is the value of a client query q involving n^q data items with the weight of the i^{th} data item being w^q_i, $1< i< n^q$. $s_i(t)$ is the value of the i^{th} data item at the data source at time t. Such a query encompasses SQL aggregation operators SUM and AVG besides general weighted aggregation queries such as portfolio queries, involving aggregation of stock prices, weighted with number of shares of stocks in the portfolio. Due to space limitations we are not presenting execution schemes for other aggregation queries such as MIN/MAX. Interested readers are referred to [13] for the extended version of this paper.

Let the value of i^{th} data item, in Equation (1), known to the client/DA be $d_i(t)$. Then the data incoherency is given by $|s_i(t)\text{-}d_i(t)|$. For a data item which needs to be disseminated at an incoherency bound C the data refresh is sent to the client or lower level DA, if the $|s_i(t)\ \text{-}d\ _i(t)|$ is more than C. If user specified incoherency bound for the query q is C^q, then the dissemination network has to ensure that:

$$\left|\sum_{i=1}^{n^q} (s_i(t) - d_i(t)) \times w_i^q\right| \leq C^q \tag{2}$$

Whenever data values at sources change such that query incoherency bound is violated, the updated value(s) is disseminated to the client. If the network of aggregators can ensure that the i^{th} data item has incoherency bound C_i then the following condition ensure that the query incoherency bound C^q is satisfied:

$$\sum_{i=1}^{n^q} C_i \times w_i^q \leq C^q \tag{3}$$

A client specified query incoherency bound needs to be translated into incoherency bounds for individual data items or sub-queries such that Equation (3) is satisfied. It should be noted that Equation (3) is sufficient condition for satisfying the query incoherency bound but not the necessary. This way of translating the query incoherency bound into the sub-query incoherency bounds is required if data is transferred between various nodes using only *push* based mechanism.

We need a method for (a) optimally dividing a client query in to sub-query and (b) assigning incoherency bound to them such that (c) the derived sub-queries can be executed at chosen DA's and (d) total query execution cost, in terms of number of refreshes to the client, is minimized.

Before developing the query cost model we first summarize the model to estimate the number of refreshes required to disseminate a data item at certain incoherency bound. For simulation experiments we use data items from sensor network and stock data domains as explained in our previous work [9]. Stock traces of 45 stocks were obtained by periodically polling *http://finance.yahoo.com*. Sensor network data used were temperature and wind sensor data from Georges Bank Cruises Albatross Shipboard [10]. Due to paucity of space we present results using stock data only but similar results were obtained for sensor data as well [5]. For detailed analysis and simulation results, readers can refer to the extended version of the paper [13]. In this paper we present results using stock data only but similar results were obtained for sensor data as well [8]. Our simulation studies show that for continuous aggregation queries:

→ Our method of dividing query in to sub queries and executing them at individual Das requires less than one third of the number of refreshes required in the existing schemes.

→ For reducing the number of refreshes more dynamic data items should be part of sub-query involving larger number of data items.

Our method of executing queries over a network of DAs is practical since it can be implemented using a mechanism similar to URL-rewriting [14] in content distribution networks (CDNs).

We would like to differentiate the current work with that of designing a network of DAs for a specific set of client queries. Whereas we propose a method to answer a client query using a given network of DAs; if the client queries are fixed, one can use the client query to optimally construct a network of data aggregators as in [5,7,15]. Our aim of minimizing the total number of messages between aggregators and client compliments the works of [5,7,15]. Together, they can be used to minimize the total number of messages between data sources data sources and clients.

II. DATA DISSEMINATION COST MODEL

Cost of disseminating a data item at a certain given incoherency bound *C* can be estimated by combining two models:

2.1: ***Incoherency bound*** *model* is used for estimating dependency of data dissemination cost over the desired incoherency bound. As per this model, we have shown in [13] that the number of data refreshes is inversely proportional to the square of the incoherency bound (1/*C2*). Similar result was earlier reported in [4] where the data dynamics was modeled as a random walk process.

Data dissemination cost ***α*** **1/*C*2** (4)

2.2: ***Data synopsis*** *Model* is used for estimating the effect of data dynamics on number of data refreshes. We define a data dynamics measure called, *sumdiff*, to obtain a synopsis of the data for predicting the dissemination cost. The number of update messages for a data item is likely to be higher if the data item changes more in a given time window. Thus we hypothesize that cost of data dissemination for a data item will be proportional to *sumdiff*, defined as:

$$R_s = \sum_i | s_i - s_{i-1} | \tag{5}$$

where *si* and *si-1* are the sampled values of the data item at *ith* and *(i-1)th* time instances (consecutive ticks). In [13] we corroborate the above hypothesis using simulation over a large number of data items. Pearson product moment correlation coefficient (PPMCC) [11] values, used for quantifying linearity between data *sumdiff* and number of refreshes required to maintain a fixed incoherency bound, were found to be between 0.90 and 0.96 for various values of incoherency bounds. *Sumdiff* value for a data item can be calculated at the data source by taking running average of difference between data values at the consecutive ticks. A data aggregator can also estimate the *sumdiff* value by interpolating the disseminated values. Thus, the estimated dissemination cost for data item *S*, disseminated with an incoherency bound *C*, is proportional to *Rs/C2*. Next we use this result for developing the query cost model.

III. QUERY DISSEMINATION COST

Consider a case where a query consists of two data items *P* and *Q* with weights w_P and w_Q respectively; and we want to estimate its dissemination cost. If data items are disseminated separately, the query *sumdiff* will be:

$$R_{data} = w_p R_p + w_q R_q = w_p \sum | p_i - p_{i-1} | + w_q \sum | q_i - q_{i-1} | \tag{6}$$

Instead, if the aggregator uses the information that client is interested in a query over *P* and *Q* (rather than their individual values), it makes a composite data item w_{pP}+ $w_q q$ and disseminates that data item then the query *sumdiff* will be:

$$R_{query} = \sum | w_p(p_i - p_{i-1}) + w_q(q_i - q_{i-1}) | \quad (7)$$

R_{query} is clearly less than or equal compared to R_{data}. Thus we need to estimate the *sumdiff* of an aggregation query (i.e., *Rquery*) given the *sumdiff* values of individual data items (i.e., R_p and R_q). Only data aggregators are in position to calculate R_{query} as different data items may be from different sources. We develop the query dissemination model in two stages.

3.1 Quantifying correlation between dynamics of Data

From Equations (6) and (7) we can see that if two data items are correlated such that if value of one data item increases, that of the other data item also increases, then R_{query} will be closer to R_{data} whereas if the data items are inversely correlated then R_{query} will be less compared to R_{data}. Thus, intuitively, we can represent the relationship between R_{query} and *sumdiff* values of the individual data items using a correlation measure associated with the pair of data items. Specifically, if ρ is the correlation measure then R_{query} can be written as:

$$R^2_{query} \propto (w_p^2 R_p^2 + w_q^2 R_q^2 + 2\rho w_p R_p w_q R_q) \quad (8)$$

The correlation measure is defined such that $-1 \le \rho \le +1$, so, *Rquery* will always be less than $|w_pR_p+w_qR_q|$ and always be more than $|w_pR_p-w_qR_q|$. The correlation measure ρ can be interpreted as *cosine similarity* [20] between two streams represented by data items *P* and *Q*. Cosine similarity is a widely used measure in information retrieval domain where documents are represented using a vector-space model and document similarity is measured using cosine of angle between two document representations. For data streams *P* and *Q*, ρ can be calculated as:

$$\rho = \frac{\sum(p_i - p_{i-1})(q_i - q_{i-1})}{\sqrt{\sum(p_i - p_{i-1})^2}\sqrt{\sum(q_i - q_{i-1})^2}} \quad (9)$$

3.2 Validating the query cost model

To validate the query cost model we performed simulations by constructing more than 50 weighted aggregation queries using the stock data with each query consisting of 3-7 data items with data weights uniformly distributed between 1 and 10. For each query the number of refreshes was counted for various normalized incoherency bounds between 0.01 and 0.5. *Figure 2* shows that the number of messages is proportional to the normalized query *sumdiff* if their normalized incoherency bounds are same. In this case PPMCC value is found to be 95%. Similarly, *Figure 3* shows the dependence of the number of refreshes on *1/C*^{2} to prove that the relationship that holds between them for single data item also holds for a query with multiple data items. The query cost model can be used in various applications of query assignment, load balancing, optimal order of processing, etc. In the next section, we use this query cost model for our query plan problem to optimally divide a client query into sub-queries and execute it over a network of data aggregators so that the number of refreshes can be minimized

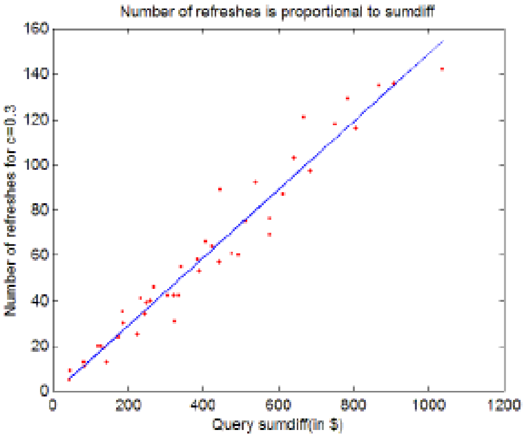


Figure 2: Variation of query cost withquery *sumdiff* (Normalized *C*=0.3)

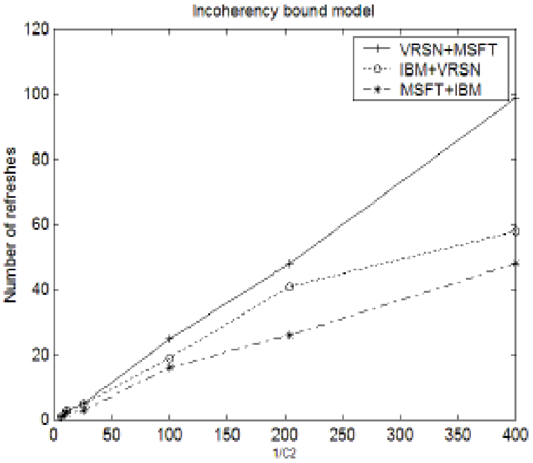


Figure 3: Number of refreshes for varying query incoherency bounds

IV. EXCEUCTING QUERIES USING SUBQUERIES

For executing an incoherency bounded continuous query, a query plan is required which includes the set of sub-queries, their individual incoherency bounds and data aggregators which can execute these sub-queries. We need to find the optimal query execution plan which satisfies client coherency requirement with the least number of refreshes. As explained in Section 1, what we need is a mechanism to:

Task 1: Divide the aggregation query into sub-queries; and

Task 2: Allocate the query incoherency bound among them. while satisfying the following conditions identified in Section 2:

condition 1. Query incoherency bound is satisfied.

condition 2. The chosen DA should be able to provide all the data items appearing in the sub-query assigned to it.

condition 3. Data incoherency bounds at the chosen DA should be such that the sub-query incoherency bound can be satisfied at the chosen DA.

Objective : Number of refreshes should be minimized.

Let the client query be divided into *N* sub-queries {*qk*: *1≤k≤N*};

with R_k being *sumdiff* of *kth* sub-query and C_k being incoherency bound assigned to it. As given is Section 3, the dissemination cost of a sub-query is estimated to be proportional to $R_k\ /C_K^2$. Thus query cost estimate is given by:

$$Z = \sum_{k=1}^{N} (R_k \ / \ C_k^2) \qquad (10)$$

While allocating sub-query incoherency bounds we need to ensure that the query coherency requirement *C* is satisfied (*condition1)*;

$$\sum_{k=1}^{N} C_k \leq C \qquad (11)$$

For satisfying *condition2*, sub-queries should be such that all its data items can be disseminated by the chosen DA. Let X_k be the tightest incoherency bound (defined in Section 2) the chosen DA can satisfy for q_k. For the *condition3*, we have to ensure that $C_k \geq X_k$ for each sub-query q_k and its assigned data aggregator. *Z* needs to be minimized for minimizing the number of refreshes as per the *objective*. Before attempting the *hard* problem of optimizing *Z*, let us first consider a simpler problem where values of C_k are given. In this simpler problem we divide the client query into sub-queries to minimize the estimated execution cost (*Z*) without considering the optimal division of the query incoherency bound into sub-query incoherency bounds.

4.1 Minimum Cost Heuristic

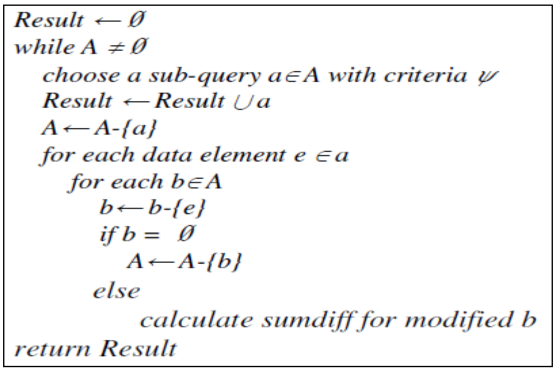
Figure 4 shows the outline of greedy heuristics where different criteria (ψ) can be used to select sub-queries. In this section we describe the case where the estimate of query execution cost is minimized in each step of the algorithm (*min-cost*) whereas in the next section we present the case where gain due to executing a query using sub-queries is maximized (*max-gain*).

4.1.1 Query Plan with Pre-decided Incoherency Bound Allocation

For the given client query (*q*) and mapping between data aggregators and the corresponding {*data-item*, *data incoherency bound*} pairs (*f*: *D*→(*S*, *C*)) maximal sub-queries can be obtained for each data aggregator. Let *A* be the set of such maximal sub queries. In this set, each query *a* € *A* can be disseminated by a designated data aggregator at the assigned incoherency bound. For each sub-query *a*€*A*, its *sumdiff Ra* is calculated. Using the set *A* and sub-query *sumdiffs*, we use the algorithm outlined in Figure 4 to get the set of sub-queries minimizing the query cost. In this Figure each sub-query *a* € *A* is represented by the set of data items covered by it. As we need to minimize the query cost, a sub-query with *minimum cost per data item* is chosen in each iteration of the algorithm i.e., criteria

$$\psi \equiv \text{minimize } (Ra/Ca^2|a|).$$

All data items covered by the selected sub query are removed from all the remaining sub-queries in *A* before performing the next iteration.



```
Result ← Ø
while A ≠ Ø
   choose a sub-query a∈A with criteria ψ
   Result ← Result ∪ a
   A ← A-{a}
   for each data element e ∈ a
      for each b∈A
         b ← b-{e}
         if b =  Ø
            A ← A-{b}
         else
               calculate sumdiff for modified b
return Result
```

Figure 4: Greedy algorithm for query plan selection

4.1.2 Optimizing query execution cost

Now we consider the overall problem to select the optimal set of sub-queries while simultaneously dividing the query incoherency bound among them. In this case we get the set of maximal queries (*A*) without considering the minimum incoherency bounds that the data aggregators can satisfy (i.e., *condition3*). In this algorithm we first get the optimal set of sub-quires without considering the *condition3* and then allocate incoherency bound among them using *condition1* (Equation (10)) and *condition3*. Lagrange multiplier scheme can be used to solve for incoherency bounds (from Equations 10 & 11) so that *Z* is minimized:

$$C_k = R_k^{1/3} C \,/ (\sum_{k=1}^{N} R_k^{1/3}) \qquad (12)$$

i.e., without the constraints of *condition3*, sub-query incoherency bounds should be allocated in proportion to $R_K^{1/3}$. Using Equations (12) and (14) we get:

$$Z^{1/3} = \frac{1}{C^{2/3}} \sum_{k=1}^{N} R_k^{1/3} \qquad (13)$$

From Equation (15), it is clear that for minimizing the query execution cost we should select the set of sub-queries so that $\sum R_k^{1/3}$ is minimized. We can do that by using criteria $\psi \equiv$ *minimize* ($R_a^{1/3}/|a|$) in the algorithm described in Figure 4. Once we get the optimal set of sub-queries we can use the Equation (11) and *condition3* ($C_k \geq X_k$) to optimally allocate the query incoherency bound among them. This allocation problem can be solved by various convex optimization techniques available in the literature such as gradient descent method, barrier method etc. We used gradient descent method (*fmincon* function in MATLAB) to solve this non-linear optimization problem to get the values of individual sub-query incoherency bounds.

V. PERFORMANCE EVALUATION

For performance evaluation we simulated the data dissemination networks of 25 stock data items over 25 aggregator nodes such that each aggregator can disseminate combinations of up to 10 data items with data incoherency bounds chosen uniformly between \$0.005 and 0.02. Then we created 500 portfolio queries such that each query has up to 10 randomly (uniformly) selected data items with weights varying between 2 and 10. These queries were executed with incoherency bounds between 0.3 and 1.0 (i.e., 0.03-0.1% of the query value). In the first set of experiments, we kept the data incoherency bounds at the data aggregators very low so that query satisfiability can be ensured.

5.1 Comparison of algorithms

For comparison with our algorithms, presented in the previous section, we consider various other query plan options. Each query can be executed by disseminating individual data items or by getting sub-query values from DAs. Set of sub-queries can be selected using *sumdiff* based approaches or any other random selection. Sub-query (or data) incoherency bound can either be pre-decided or optimally allocated. Various combinations of these dimensions are covered in the following algorithms:

1. No sub-query, equal data incoherency bound (*naïve*): In this algorithm, the client query is executed with each data item being disseminated independent of other data items in the query. Incoherency bound is divided equally among the data items. This algorithm acts as a baseline algorithm.

2. No sub-query, optimal incoherency bound (*optc*): In this algorithm also data items are disseminated separately but incoherency bound is divided among data items so that total number of refreshes can be minimized. This algorithm is similar to the one presented in [6]. Here, the incoherency bound is allocated dynamically using Equation (14).

3. Random sub-query selection (*random*): In this case, sub queries are generated by randomly selecting one data aggregators and allocating it the maximal sub-query consisting of query data

items which the aggregator can disseminate. Then the process is repeated for the remaining data items until the whole query is covered. This algorithm is designed to see how the sub-query

selection based on query *sumdiff* (Section 3) works in comparison to random selection of sub-queries.

4. Sub-query selection while minimizing *sumdiff* (*min-cost*):

This algorithm is described in Section 4.1.1.

5. Sub-query selection while maximizing gain (*max-gain*): *Figure 5* shows average number of refreshes required for query incoherency bounds of \$0.3, \$0.5 and \$0.8.

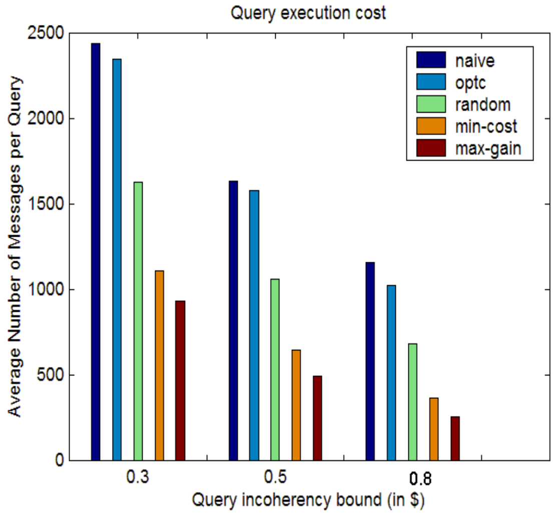


Figure 5: Performance evaluation of algorithms

The naïve algorithm requires more than three times the number of messages compared to *min-cost* and *max-gain* algorithms. For incoherency bound of\$0.8 each query requires 1024 messages if it is executed just by optimizing incoherency bound (*optc*) compared to 255 when we select the query plan using the *max-gain* algorithm. Further, although the optimization problem is similar to the covering a set of data items (query) using its sub-sets (sub-queries) for which the *greedy min-cost* algorithm is considered to be most efficient [7], we see that *max-gain* algorithm requires 20-25% less messages compared to the *min-cost* approach. Reasons for *max-gain* algorithm performing better than other algorithms are explored in the next set of experiments. Although here we presented results for stock traces (man-made data) similar results were obtained for sensor traces (natural data) as well.

VI. RELATED WORK

Various mechanisms for efficiently maintaining incoherency bounded aggregation queries over continuously changing data items are proposed in the literature [6,17]. Our work distinguishes itself by being sub-query based evaluation to minimize number of refreshes. In [6], authors propose using data filters at the sources; instead we assign incoherency bounds to sub-queries which reduce the number of refreshes for query evaluation, as explained in Section 5. Further, we propose that more dynamic data items should be executed as part of larger sub-query. In [15,17], authors present technique of reorganizing a data dissemination network when client requirements change. Instead, we try to answer the client query using the existing network. Reorganizing aggregators is a longer term activity whereas query planning can be done for short as well as long running queries on more dynamic basis.

Pull based data dissemination techniques, where clients or data aggregators pull data items such that query requirements are met, are described in [9,17]. For minimizing the number of pulls, both model the individual data items and predict data values. In comparison, we consider the situation where different sub-queries, involving multiple data items, can be evaluated at different nodes. Further, incoherency bound is applied over the sub-query rather than to individual data items, leading to efficient evaluation of the query. Our work can be extended by using temporal and spatial properties of data items for predicting their correlation measures. A method of assigning clients data queries to aggregators in a content distribution network is given in [17]. We do for client queries consisting of multiple data items what [17] does for client requiring individual data items.

VII. CONCLUSION

This paper presents a cost based approach to minimize the number of refreshes required to execute an incoherency bounded continuous query. For optimal execution we divide the query into sub-queries and evaluate each sub-query at a chosen aggregator. Performance results show that by our method the query can be executed using less than one third the messages required for existing schemes. Further we showed that by executing queries such that more dynamic data items are part of a larger sub-query we can improve performance. Our query cost model can also be used for other purposes such as load balancing various aggregators, optimal query execution plan at an aggregator node, etc. Using the cost model for other applications and developing the cost model for more complex queries is our future work.

REFERENCES

[1] A. Davis, J.Parikh and W. Weihl. “Edge Computing: Extending Enterprise Applications to the Edge of the Internet". WWW 2004

[2] D. Vander Meer, A. Datta, K. Dutta, H. Thomas and K.Ramamritham. Proxy-Based Acceleration of Dynamically Generated Content on the World Wide Web. ACM Transactions on Database Systems (TODS) Vol. 29, June 2004.

[3] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman and B. Weihl. Globally Distributed Content Delivery, IEEE Internet Computing Sept 2002.

[4] S. Shah, K. Ramamritham, and P. Shenoy. Maintaining Coherency of Dynamic Data in Cooperating Repositories. VLDB 2002.

[5] Query cost model validation for sensor data. www.cse.iitb.ac.in/~ravivj/BTP06.pdf.

[6] C. Olston, J. Jiang, and J. Widom. Adaptive Filter for Continuous Queries over Distributed Data Streams. SIGMOD 2003.

[7] D. S. Hochbaum. Approximation algorithms for the set covering and vertex cover problems. SIAM Journal on Computing, vol. 11 (3), 1982.

[8] Zongming Fei. A Novel Approach to Managing Consistency in Content Distribution. WCW 2001

[9] R. Gupta, A. Puri, and K. Ramamritham. Executing Incoherency Bounded Continuous Queries at Web Data Aggregators. WWW 2005.

[10] NEFSC Scientific Computer System http://sole.wh.whoi.edu/~jmanning//cruise/serve1.cgi

[11]Pearson Product moment correlation coefficient. http://www.nyx.net/~tmacfarl/STAT_TUT/correlat.ssi /

[12] S. Agrawal, K. Ramamritham and S. Shah. Construction of a Temporal Coherency Preserving Dynamic Data Dissemination networks”, RTSS 2004.

[13] Optimized Execution of Continuous Queries, APS 2006, www.cse.iitb.ac.in/ ~grajeev/ APS06.PDF

[14] S. Rangarajan, S. Mukerjee and P. Rodriguez. User Specific Request Redirection in a Content Delivery Network, 8th Intl. Workshop on Web Content Caching and Distribution (IWCW), 2003.

[15] R Guptha and K. Ramamritham, "Optimized Query Planning of Continuous Aggregation Queries in Dynamic Data Dissemination Networks", WWW 2007.

[16] R Guptha and K. Ramamritham, "Query Planning for Continuous Aggregation Queries over a network of Data Aggregators", IEEE 2011.

[17] S. Shah, K. Ramamritham, and C. Ravishankar. Client Assignment in Content Dissemination Networks for Dynamic Data. VLDB 2005.

◈◈◈

Privacy Preservation Methods to Confidential Databases With Security

K. Anuradha & S. Nageswara Rao

Dept. of CSE, MITS, Madanapalle
Email:anukanduru@gmail.com, nag_sirisala@yahoo.com

Abstract - Anonymization means to remove personal identifier or converted into non readable form by human to protect private or personal information. Data anonymization can be performed in different ways but in this paper k-anonymization approach is used. Suppose one person A having his own k-anonymous database and needs to determine whether database is still k-anonymous if tuple inserted by another person B. For some applications (for example, Student's record), database needs to be confidential, So access to the database is strictly controlled. The confidentiality of the database managed by the owner is violated once others have access to the contents of the database. Thus, Problem is to check whether the database inserted with the tuple is still k-anonymous without letting the owner A and others (B) to know the content of the tuple and database respectively. In this paper, we propose a protocol solving this problem on suppression based k-anonymous and confidential database.

Keywords*-Anonymization , Privacy ,Confidentiality, Anonymous.*

I. INTRODUCTION

For each application database is important or valuable thing, so their security is important. There are different security control methods are identified and each method have different criteria. For example, FERPA provides privacy protections for such records when held by federally funded educational institutions[1]. FERPA defines an education record as those records, files, documents, and other materials that contain information directly related to a student and are maintained by an educational agency or institution or by a person acting for such agency or institution. Students who are at least 18 years of age, or attending postsecondary institutions or otherwise their parents ,generally have a right to gain access to their education records within 45 days of a written request, seek to amend any information therein considered to be in error, control how information in such records is disclosed to other institutions ,in general, such disclosures must be authorized by the student or parent, with some exceptions and complain to the US Department of Education if these rights appear to have been violated. There is problem of providing security to statistical databases against disclosure of confidential information. There are various security control method classified into four groups: conceptual, query restriction, data perturbation, and output perturbation. Criteria for evaluating the performance of the various security-control methods are identified. A detailed comparative analysis of the most promising methods for protecting dynamic-online statistical databases is also presented. To date no single security-control method prevents both exact and partial disclosures. There is big concern for privacy.

The problem of statistical disclosure control revealing accurate statistics about a population while preserving the privacy of individuals has a vulnerable history. Still, there is a difference between confidentiality and privacy- Confidentiality refers to limiting information access and disclosure to authorized users and preventing access by or disclosure to unauthorized users. Privacy refers to limiting access to individuals' personal information[5].Question is Confidentiality is still required if data have been anonymized-yes because anonymous data have business value for the party owning database or unauthorized disclosure of anonymous data may damage the party owning the data.There have been lots of techniques developed to protect privacy, but here we proposed k-anonymization[4]. K-Anonymity refers to attributes are suppressed or generalized until each row is identical with at least k-1 other rows. At this point the database is said to be k-anonymous. K-Anonymity prevents definite database linkages. The modification of the anonymous database DB can be naively performed as follows: the party who is managing the database or the server simply checks whether the updated database DB is still anonymous. Under this approach, the entire tuple t has to be revealed to the party managing the database server, thus violating the privacy of the patient. Another possibility would be to make available the entire database to the patient so that the individual can verify if the insertion of the data violates their own privacy. This approach however, requires making available the entire database to the patient thus violating data confidentiality. There is a protocol solving this problem on suppression- based k-anonymous and confidential databases. The protocol depends on well-known cryptographic assumptions. The huge numbers of databases recording a large variety of information about individuals makes it possible to find information about specific individuals by simply correlating all theavailable databases. Confidentiality is achieved by enforcing an access policy, or possibly by using some cryptographic tools. Privacy relates to data can be safely disclosed without leaking sensitive information regarding the legitimate owner. The confidentiality is still required once data have been anonymized, if the anonymous data have a business value for the party owning them or the unauthorized disclosure of such anonymous data may damage the party owning the data or other parties.So,problem is that can database owner assure privacy of database without knowing data to be inserted? It is important to assure that database maintains privacy of individual and also who maintain database.So,it needs to check that data entered in database do not violate privacy, and to perform such verification without seeing sensitive information of individual.

II. RELATED WORK

There are some limitations of the protocol, if the database is not anonymous with respect to a tuple that has been inserted, the insertion cannot be performed. Therefore, one of the protocols is extremely inefficient. There are efficient protocols. The first research is based on algorithms for database anonymization. The database is protected by data reduction, data perturbation or generating synthetic data. However, the main concept of k-anonymity to maintain confidentiality of their contents. The problem is to protect privacy of data that has been divided into two groups depending on whether data are continuously released and anonymizaed or data released in different fashion and anonymized. The second research is related to Secure Multiparty Computation protocol which is subfield of cryptography. The third research is related to the private information retrieval, which can be seen as an application of the secure multiparty computation techniques to the area of data management. This allows a user to retrieve an data (or tuple) from database without revealing tuple one is retrieving. Here, main focus is to find efficient techniques to express queries over a database without letting the database know the actual queries [2]. Still, the problem of privately updation of database has not been resolved because these techniques deal with only data retrieval. These approaches that will not address the problem of k-anonymity since their goal is to encrypt the data hence external entities can obtain their data. Thus, the main goal is to protect the confidentiality of the data from the external entities that manages the data. Even though, the data are fully available to the clients that are not the case under our approach. In data anonymization, Insertion cannot be performed if database is not properly anonymized. The problem is private updates to k-anonymous databases The suppression based protocols deals with the problem of updating the databases.

Figure 1 shows Anonymous database system, we assume that information of single student stored in a tuple and database kept confidential at server. The users treated as database of educational record, only institution have right to access to database. Since database is anonymous, the data provider’s privacy is protected from users. Since database have privacy sensitive data, so main aim to protect privacy of student’ data. This can be achieved by anonymization.If database is anonymous, it is not possible to catch student’s identity for database. Suppose new student has to be entered, this means database has to be updated in order to insert a tuple.The modification of anonymous database can be done as follows: the party who is managing the database checks whether the updated database is still anonymous after inserting a tuple. Under this approach, the entire tuple t has to be revealed to the party managing the database server, thus violating the privacy of the student. Another possibility would be to make available the entire database to the student so that the student can verify by himself/herself if the insertion of his/her data violates his/her own privacy. To get solution of these problem, several problem needs to be addressed: The first problem is: without revealing the contents of tuple t to be inserted and database DB, how to preserve data integrity by establishing the anonymity of DB U {t}.The second problem is: once such anonymity is established, how to perform this update?

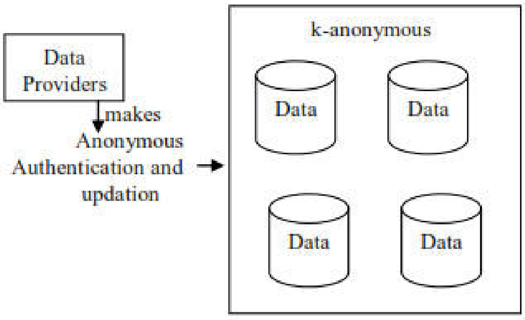


Fig 1: Anonymous Database System

The third problem is: what can be done if database anonymity is not preserved? Finally, The forth problem is: what is the initial content of the database, when no data about users has been inserted yet? In this paper, we propose a protocol solving first Problem, which is the central problem addressed by our paper.

III. PROPOSED TECHNOLOGY

The protocol relies on the fact that anonymity of database does not affected ,if inserting tuple is already in database. then ,the problem of integrity while inserting tuple in database is equivalent to privately checking of inserting tuple with tuple already in database. The protocol is aimed at suppression based anonymous database and it allows the owner of database to properly anonymize the tuple t, without gaining any useful knowledge on its contents and without having to send to its owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. To assure higher level of anonymity to the party inserting a tuple ,we require that the communication between the party and database occurs through anonymous connection, as provided by crowd protocol[3]. Crowd protocol hides each user's communications by routing them randomly within a group of similar users. In order to perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.

A. Prototype Architecture

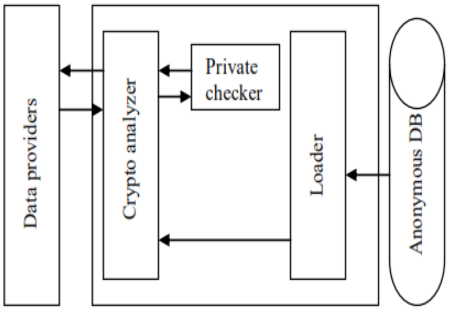


Fig 2 Prototype Architecture

In Fig 2, Data provider enters data is stored in crypto module which perform cryptography operation on all tuples exchanged between user and Private updater, using suppression based method. Loader module read anonymized tuples from k-anonymous database. And checker module checks whether the tuple from the user matches with the tuple in the database. If none of the tuple matches with the user tuple, then loader reads another tuple from k-anonymous database. The functionality provided by the Private Checker. Communication between user and database is carried out by anonymizer and that all the tuples are encrypted.

IV. SUPRESSION BASED PROTOCOL

The suppression based protocol relies on well-known cryptographic techniques. We consider table T={t1,…t2} over the attribute set A. Generally in suppression based method we mask value of some special attributes with *, the value deployed by the user for anonymization. So the main idea behind this protocol is: To form subset of indistinguishable tuples by masking the value of some well chosen attributes.

TABLE 1 Original Dataset

Birth date	Sex	Zip code
21/1/79	male	53715
10/1/79	female	55410
21/2/83	male	02274
19/4/82	male	02237

TABLE 2 Suppressed Data with k=2

Birth date	Sex	Zip code
*/1/79	person	5****
*/1/79	person	5****
//8*	Male	022**
//8*	Male	022**

As shown in table 1 which contains original database (Table T) having three attributes Birth date, Sex, Zipcode.Table-2 shows a suppression based k- anonymization with k=2.As shown in table k=2 means at least k(=2) tuples should be indistinguishable by masking values. Suppression based attributes for every tuple of T is referred as anonymization problem, and finding the anonymization that minimizes the number of masked values.

A Cryptography Primitive

The Diffie Hellmen key exchange algorithm allows two users to establish shared secret key over insecure communication without having any prior knowledge. Here, Diffie Hellmen is used to agree on shared secret key to exchange data between two parties. AES(Advanced Encryption Standard) algorithm is the advanced encryption standard form of algorithm which had been used as a symmetric form of encryption.

There are two encryption schemes, commutative and product homomorphic E to satisfy indistiguishability properly. A commutative,product-homomorphic encryption scheme ensures that the order in which encryptions are preformed is irrelevant(commutativity) and it allows to consistently perform arithmetic operations over encrypted data (homomorphic property). Given a finite set K of keys and a finite domain D, a commutative, product homomorphic encryption scheme E is a polynominal time computable function E: K×D→D satisfying the following properties:

1. *Commutativity*:

In commutative, all key pairs $k1, k2 \in k$ and value $d \in$ D, the following equality holds:

$$\mathrm{E}k1(\mathrm{E}k2\ (d)) = \mathrm{E}k2(\mathrm{E}k1(d))$$

2. *Product-homomorphism:*

In product homomorphic every $k \in k$ and every value pairs $d1, d2 \in D$ the following equality holds:

$$\mathrm{E}(\,d1) \cdot \mathrm{E}k(\,d2) = \mathrm{E}k(d1 \cdot d2)$$

3. *Indistinguishability:*

It is infeasible to distinguish an encryption from a randomly chosen value in the same domain and having the same length. The advantages are high privacy of data even after updation, and an approach that can be used is based on techniques for user anonymous authentication and credential verification.

B Algorithm

Suppose Alice has control over database and Bob is data provider then protocol works as follows:

In step 1, Alice sends Bob encrypted version of tuple containing only non suppressed attributes. At step 2, bob encrypts the information received from Alice and sends it to her, along with encrypted version of each value in his tuple. In final step, Alice examines if the suppressed attributes of tuple is equal to the tuple sent by Bob. If yes then insert tuple in database.

V. CONCLUSION

In this paper, we have proposed secure protocol for privately checking whether a k-anonymous database retains anonymity once a new tuple is being inserted to it. Since the proposed protocol ensures the updated database remains k- anonymous. Thus the database is updated properly using the proposed protocol. The data provider's privacy cannot be violated if user update a table. If updating any record in database violate the k- anonymity then such updating or insertion of record in table is restricted. If insertion of record satisfies the k-anonymity then such record is inserted in table and suppressed the sensitive information attribute by * to maintain the k-anonymity in database. Thus by making such k-anonymity in table that makes unauthorized user to difficult to identify the record.

The important issues for future work are as follows:

- Improve the efficiency of protocols, by the number of messages exchanged and sizes and algorithm used for encryption and decryption.
- The private update to database systems techniques supports notions of anonymity different than k- anonymity.
- In the case of malicious parties by the introduction of an untrusted third party, implementing a real-world anonymous database system.

REFERENCES

[1] U.S. Department of Education. General Family Educational Rights and privacy Act (FERPA).

[2] B.C.M. Fung ,K. Wang, A.W.C. Fu and J. Pei, Anonymity for Continuous Data Publishing Proc. Extending database Technology Conference (EDBT),2008

[3] M.K.Reiter, A. Rubin. Crowds: anonymity with Web transctions.ACM Transactions on Information amd System Security (TISSEC),1(1),1998;66-92

[4] P. Samarati. Protecting respondent's privacy in micro data release, IEEE Transactions on Knowledge and Data Engineering vol. 13,no. 6,pp.1010-1027,Nov/Des.2001

[5] University of Miami Leonard M. Miller School of Medicine, Information Technology.

[6] Dr. Durgesh Kumar Mishra, Neha Koria, Nikhil Kapoor, Ravish Bahety ,A Secure Multi-Party Computation Protocol for Malicious Computation Prevention for preserving privacy during Data Mining, Vol. 3,No. 1,2009

[7] C. Blake and C. Merz, ―UCI Repository of Machine Learning Databases, E. Bertino and R. Sandhu, ―Database Security― Concepts, Approaches and Challenges, IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[8] www.wikipedia.com/wikifiles/

Run-Time Verification in Java Using Race Detection

R. Arjun & C.V. Lakshmi Narayana

Department of Computer Science & Engineering, Annamaacharya Institute of Technology & Science,
Rajampet, Andhra Pradesh, India.
E-mail : arjunmtechit@gmail.com, cvlakshminarayana@gmail.com

Abstract – In the past, researchers have developed specialized programs to aid programmers in detecting concurrent programming errors such as deadlocks, livelocks, starvation, and data races. In this work, we propose a language extension to the aspect-oriented programming language AspectJ, in the form of three new pointcuts, lock(), unlock(), and maybeShared(). These pointcuts allow programmers to monitor program events where locks are granted or handed back, and where values are accessed that may be shared among multiple Java threads. We decide thread locality using a static thread-local-objects analysis developed by others. Using the three new primitive pointcuts, researchers can directly implement efficient monitoring algorithms to detect concurrent-programming errors online. As an example, we describe a new algorithm which we call RACER, an adaption of the well-known ERASER algorithm to the memory model of Java. We implemented the new pointcuts as an extension to the AspectBench Compiler, implemented the RACER algorithm using this language extension, and then applied the algorithm to the NASA K9 Rover Executive and two smaller programs. Our experiments demonstrate that our implementation is effective in finding subtle data races. In the Rover Executive, RACER finds 12 data races, with no false warnings. Only one of these races was previously known.

Keywords - *Race detection, runtime verification, aspect-oriented programming, semantic pointcuts, static analysis.*

I. INTRODUCTION

Programming errors occur frequently in software systems, and therefore, researchers have spent much effort on developing methods to detect and remove such errors as easily and early as possible in the development process. Concurrent programs are even more likely to suffer from programming errors as concurrent programming adds potential sources of failure. In a concurrent program, a programmer has to make sure to avoid deadlocks, to properly protect shared state from data races, and to protect single threads or processes from starvation. Researchers have developed specialized static and dynamic analyses to aid programmers with these tasks. All of these approaches share one common concern. They identify events of interest, such as the acquisition and release of locks or the access to shared state. Static approaches analyze the program source, while dynamic approaches analyze a trace or abstract-state representation generated by executing the program. Up to now, most existing dynamic approaches have used some form of low level bytecode instrumentation library to transform the analyzed program into one that generates those events. However, such libraries, for example, BCEL, are difficult to use and distract efforts from focusing on the more interesting algorithmic aspects of the analyses. Researchers have recognized aspect-oriented programming as a convenient tool to declare instrumentation at a high level of abstraction. Aspect oriented programming allows programmers to use predicates, called point cuts, to intercept certain events of interest at runtime. Unfortunately, in all of the current Java-based aspect-oriented programming languages, programmers can only intercept events such as method calls, field accesses, and exception handling. In particular, none of these languages allows programmers to intercept events that regard the acquisition and release of locks. This precludes programmers from implementing algorithms in Aspect J that are meant to find concurrency related programming errors such as data races. In this work, we hence propose a novel extension to the aspect oriented programming language Aspect J.

The language extension that we propose enhances AspectJ with three new point cuts, to make available to the programmer three additional kinds of events: 1) the acquisition of a lock, 2) the release of a lock, and 3) the event of reading from or writing to a field that may be shared among threads.

For instance, the following pointcut captures the event of locking on object l: lock() && args(l). A programmer can capture the converse event of unlocking l by simply writing unlock() && args(l). Setting a potentially shared field on an object o is captured via the pointcut set(! static *)&&target(o) && may be Shared(). Matching the first two pointcuts against a given program is decidable. The problem of matching the may be Shared() pointcut is, however, generally undecidable. We therefore compute a sound over approximation using a static thread-local-objects analysis. The approximation assures that the pointcut matches every access to a field that is indeed shared. Because of the over approximation, the pointcut may, however, also match accesses to fields that are not actually shared, i.e., fields that only a single thread accesses. Using these three novel pointcuts, programmers can easily implement bug-finding algorithms that detect errors related to concurrency.

The lock() and unlock() pointcuts allow a programmer to uniformly act on any acquisition and release of a lock using synchronized blocks and methods in any Java program. The programmer can use the may be Shared() point cut to gain runtime efficiency by monitoring accesses to only those fields that may be shared among threads. To demonstrate the feasibility of the approach, we implemented the three novel pointcuts as an extension to the Aspect Bench Compiler. To show how programmers can use this language extension, weadapted the ERASER race detection algorithm to Java, and implemented it using the new pointcuts.

The new algorithm is named RACER. Both ERASER and RACER detect program executions which reveal potential for data races in the executed application. We presented a first version of the RACER algorithm at ISSTA 2008. However, we subsequently noted1 that a large number of potential data races that this version of RACER reported were unfortunately false warnings. The initial version of RACER reported these false warnings because it ignored calls to Thread.start(). The improved version of RACER that we present in this paper takes such calls into account and therefore avoids reporting these false positives.

We applied the aspects implementing the RACER algorithm to a plan execution program for the NASA K9 rover and two other multithreaded programs written by computer science researchers. Our results show that the algorithm is effective in finding data races. In the NASA code, RACER found 12 races, 11 of which were previously unknown, although extensive studies had been performed on the K9 rover code before.

The main contributions of this work are:

1. A description of three novel AspectJ pointcuts, lock(), unlock(), and maybeShared().
2. An implementation of these pointcuts in the Aspect-Bench Compiler in the case of the maybeShared() pointcut through a static whole-program analysis,
3. An algorithm for race detection in Java, coined RACER, that improves on ERASER, and an implementation using the three novel AspectJ pointcuts, and
4. An experiment showing that our implementation is effective in finding data races in a plan execution program for the NASA K9 rover.

Module Description

- The K9 Rover and Executive
- The Race on ActionExecution.status
- The Races on syncNum and syncNumOpt
- The Races between the RuntimeExecutive,
- TheExecTimer, and ExecCondChecker

The K9 Rover and Executive:

The K9 Rover is an experimental hardware platform for autonomous wheeled rovers, targeted for the exploration of a planetary surface such as Mars. K9 was specifically used to experiment with new autonomy software. Rovers are traditionally controlled by low-level commands uploaded from Earth. The K9 Executive, a software module, provides a more flexible means of commanding a rover through the use of high-level plans in a domain-specific programming language. High-level plans can, for example, be generated by an on-board AI-based planner. The Executive is essentially an interpreter for the plan language.

The Race on ActionExecution.status:

To indicate a data race on a variable status in class ActionExectution, RACER issues the message. The ActionExecution thread and the Runtime Executive thread cause this race because they both access status without both first acquiring a common lock. This is exactly the error planted in the code during the original verification experiment.

The Races on syncNum and syncNumOpt:

We will not show the error messages from RACER for the remaining data races. The two data races mentioned in this section stem from an experiment performed with the K9 Executive (after the case study from) in order to determine how effectively a static analysis algorithm could reduce the number of locking operations.

TheExecTimer and ExecCondChecker:

We show the situation between the two threads RuntimeExecutive and ExecTimer that cause another seven data races which RACER reported on the K9 Executive. The Main thread starts both these threads.

Overview:

We have proposed a language extension to the aspect-oriented programming language AspectJ. We extend AspectJ with three new pointcuts lock(), unlock(), and maybeShared(). These pointcuts allow researchers to easily implement bug-finding algorithms for errors related to concurrency. As an example, we have implemented RACER, an adaption of the ERASER race-detection algorithm to the Java memory model. We found that, using our AspectJ extension, we were able to implement RACER very easily, in just two aspects with a small set of supporting classes. The RACER algorithm is different from C-based race detection algorithms like ERASER in the way that it treats object initialization. ERASER is very forgiving to programmers in an object's initialization phase. RACER, on the other hand, detects and also reports races that comprise the initialization of an object.

AspectJ Compiler representation:

These approaches are characterized by the roles and obligations of programmer, compiler, and runtime system to determine the correctness of the synchronization.

Terminology

We adopt the terminology and notation from Choi et al. A *program execution* is defined as a sequence *e*0, . . . ,*en* of *events* where *ei* is defined as a tuple h*o*, *f* , *t*,*L*,*k*i:

- *i* is a unique id.
- *o* is the object instance (or class) that is being accessed.
- *f* is the field variable that is accessed inside the object.
- *t* is the thread accessing the object.
- *L* is the set of locks being held during the access.
- *k* is the kind of event (one of {READ,WRITE,LOCK,UNLOCK,START,JOIN}).

  Events shall not only be used to model variable access, but also lock and unlock, as well as thread start and join; for such events, the accessed field variable *f* is void.

Data races

A *critical section* is a statement sequence that should execute without interference of other threads. The concept is useful to guarantee that access from different threads to the same data is ordered, avoiding inconsistency and data corruption. *Races* are used to characterize situations where the guarantee about non-interference of threads accessing shared data is violated. Netzer and Miller distinguish *data races* that refer to unsynchronized access of shared data and *general races* that are sources of nondeterminism in parallel programs in general. Our discussion in this section focuses on data races, while Section discusses synchronization defects related to general races.

Lock-based data races

The difficulty to infer the happened-before relation from a program execution has led to another definition of data races that is not based on the temporal ordering of events but on a *locking policy*. The rationale behind *unique-lock data races* is that accesses to shared mutable variables that are consistently protected by a unique lock cannot be involved in a data race.

Let *E*(*o*, *f*) be the set of events that access field *f* on object *o*. A *unique-lock data race* on the variable *o*. *f* is defined as

uniqueLockRace(*o*, *f*) , 9*ei*,*e j* 2 *E*(*o*, *f*) : *ei*.*t* 6= *e j*.*t* ^ (2.2)

9*ei* 2 *E*(*o*, *f*) : *ei*.*a* = WRITE^ *e*2*E*(*o*, *f*)*e*.*L* = /0.

The data race definitions identify a conservative set of actual data races in a program execution. The definitions are however not operational, i.e., some of the predicates must be approximated from the observations in an execution trace. This can lead to overreporting. Practical mechanisms for data race detection use heuristics to assess the ordering among accesses. Even if this is uncommon and introduce hence unsoundness: Potential of underreporting is accepted in favor of a significant reduction of spurious reports that would be given if a conservative approximation of access ordering was used. Two important sources of inaccuracy of dynamic data race checkers are (1) their approximation of the temporal ordering relation (happened-before based checkers) and (2) the delayed checking until some data is accessed by more than one thread (unique-lock and object race checkers).

II. TRAILS TO CORRECT SYNCHRONIZATION:

The *language-centric approach* is a promising and attractive direction to address the problem of synchronization defects. However, these systems force the software designer to use the synchronization models that can be type-checked, and some popular and efficient, lock-free synchronization patterns are not accommodated.1 In addition, it is unclear if the specific locking discipline that is imposed by the type system and requires, e.g., lock protection for access to all potentially shared mutable data structures, is well amenable to program optimization and high performance in concurrent software systems. Hence it will be a while until language-centric approaches to concurrency control become widely accepted.

(D) In the *synthesized* approach, the view of the programmer is confined to sequential programming and parallelism is automatically synthesized through the compiler. Such autoparallelizing compilers employ dependence analysis, discover computations without data interference that can be executed concurrently, and finally generate efficient parallel programs. Vectorization, e.g., has been successful along this procedure: Vectorizing compilers typically focus on loops and exploit concurrent execution features of synchronous processor architectures (SIMD). The transfer of auto-parallelization to asynchronous multiprocessors (MIMD) has however brought new challenges. Parallel execution is done in separate threads, and current hardware and OS systems cause a critical overhead in managing threads and shared data. The focus on loops is often not sufficient and compilers are compelled to unveil new opportunities and larger scopes for parallel execution. For object-oriented programs, dependence analysis is complicated through the use of dynamic data structures and the effects of aliasing. Hence speculative approaches have been conceived that can be successful to improve the performance of programs where static dependence analysis fails. The *synthesized approach* to parallelism is conceptually appealing because synchronization defects cannot occur by design. The challenges posed through modern processor architectures and applications are however hard and manifold. Hence automated or speculative parallelization are often not as effective as techniques where parallelism is specified explicitly by the programmer. The scope of the dependence analysis and speculation is usually limited to local program constructs, and hence automated parallelization is not successful for programs where independent computational activities can be separated at a high-level in the system design. Kennedy argues that "... it is now widely believed that (auto-) parallelization, by itself, is not enough to solve the parallel programming problem".

Correct synchronization:

We have presented three important classes of synchronization defects (data races, violations of atomicity, and deadlock). Based on their definitions, we would like to conclude with a notion of correctness for parallel programs that reflects the absence of such synchronization defects. For the *synthesized* approach (D), the correctness criterion is to preserve the semantics of the initial sequential program. The presence of data races or atomicity violations is an implementation aspect. In the *language-centric* approach (C), correctness criteria are explicitly stated by the programmer and verified by the type checker. Most type systems focus on specific synchronization defects, and hence the scope of the check is limited, e.g., to the absence of data races and deadlock.

For the *programmer-centric* (A) and the *compiler-and runtime controlled* (B) approach, a precise notion of correctness is difficult to achieve: First, there is no explicit definition of the synchronization discipline in a program, and the structure of synchronization is generally difficult to determine from the program text. Current object-oriented programming languages like Java and C], e.g., offer low-level features that allow for a large variety of inter-thread synchronization mechanism; such mechanisms and their synchronization effect are not easily recognized by a static analysis. Second, the given definitions do not allow to draw a clear line between benign and harmful incidents: The different detection mechanisms of data races have incomparable reporting capabilities that are neither sound nor complete. These aspects make the three classes of synchronization defects a guideline rather than a correctness criterion.

Compiler and runtime controlled concurrency:

This dissertation pursues the compiler and runtime controlled approach to concurrency. The approach is founded on a cooperation of programmer, compiler, and runtime system; the roles and challenges of the individual participants respectively constituents are discussed in the following:

- The *programmer* specifies parallelism and synchronization explicitly and preferably follows certain synchronization patterns that are recognized by the compiler. In cases where the compiler is unable to infer the intention of the programmer, the programmer can provide annotations (e.g., type modifiers like final or volatile) to avoid unnecessary conservatism in the downstream tool chain (compiler and runtime).
- The *compiler* analyzes the program and reports potential incidents for different classes of synchronization defects. The challenges are to avoid underreporting, and at the same time to minimize the number of false positives due to conservatism. Reports presented to the user should be aggregated and concisely specify the class and source of the problem.
- The *runtime* system is responsible to check residual cases of potential faults that are not resolved by the programmer. The challenge is to avoid underreporting while making the checker efficient.

Architecture:

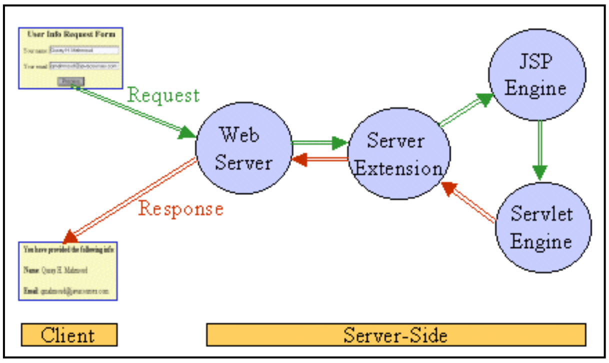


Fig. : Request/Response flow calling a JSP page

The JSP specification presents two approaches for building web applications using JSP pages: JSP Model 1 and Model 2 architectures. These two models differ in the location where the processing takes place. In Model 1 architecture, as shown in below Figure, the JSP page is responsible for processing requests and sending back replies to clients.

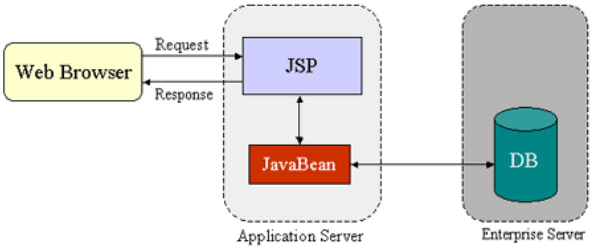


Fig. : JSP Model 1 Architecture

The Model 2 architecture, as shown in below Figure, integrates the use of both servlets and JSP pages. In this mode, JSP pages are used for the presentation layer, and servlets for processing tasks. The servlet acts as a **controller** responsible for processing requests and creating any beans needed by the JSP page. The controller is also responsible for deciding to which JSP page to forward the request. The JSP page retrieves objects created by the servlet and extracts dynamic content for insertion within a template.

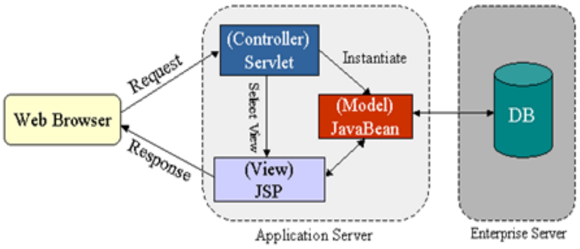
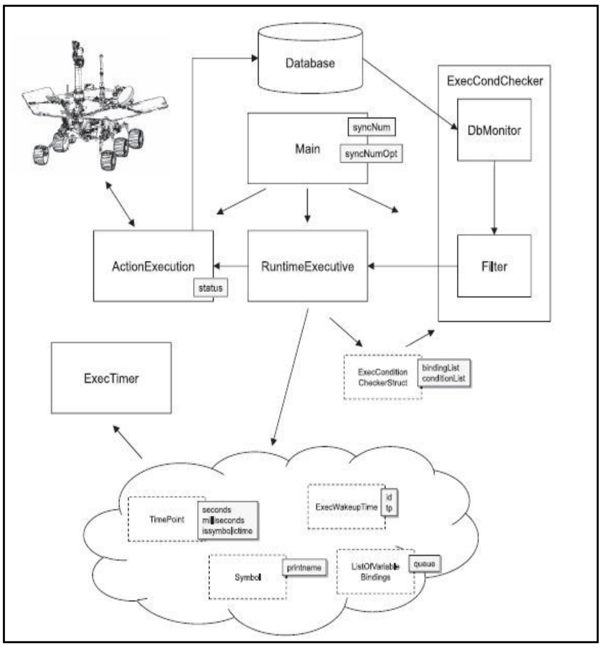


Fig. : JSP Model 2 Architecture

This model promotes the use of the Model View Controller (MVC) architectural style design pattern.. The Apache Struts is a formalized framework for MVC. This framework is best used for complex applications where a single request or form submission can result in substantially different-looking results.



Existing Algorithm Disadvantages:

Concurrent programs are even more likely to suffer from programming errors as concurrent programming adds potential sources of failure.

Proposed System Advantages:

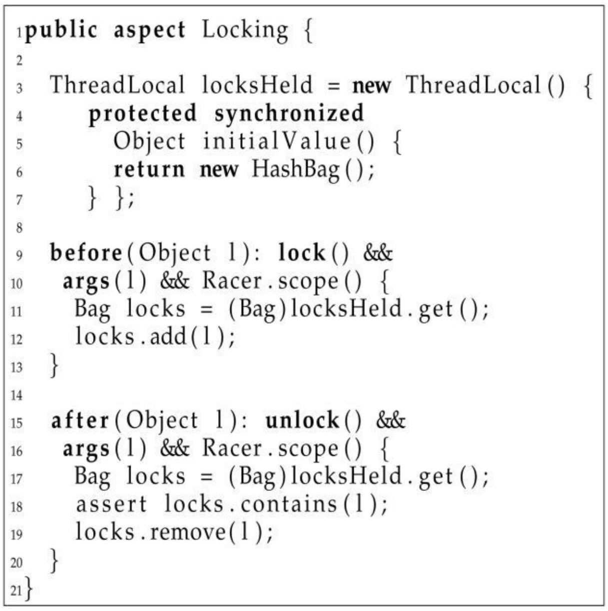
An additional advantage of using AspectJ is that we could easily modify the Locking aspect to take other locking styles into account. For instance, if Reentrant-Locks were used. We could just extend the pointcuts in with an additional disjunct.

Algorithm

Racer Algorithm:

```
✦ when method execution accesses field f
  if (f is uninitialized) {
     if (f is reference field of type T) {
        non-deterministically initialize f to
        – null
        – a new object of class T (with
  uninitialized fields)
        – an object created during prior field
  initialization (alias)
     }
     if (f is numeric/string field)
        initialize f to a new symbolic value
  }
```

Algorithm in case of local lock sets:



```
public aspect Locking {

  ThreadLocal locksHeld = new ThreadLocal() {
     protected synchronized
       Object initialValue() {
       return new HashBag();
     } };

  before(Object l): lock() &&
   args(l) && Racer.scope() {
    Bag locks = (Bag)locksHeld.get();
    locks.add(l);
  }

  after(Object l): unlock() &&
   args(l) && Racer.scope() {
    Bag locks = (Bag)locksHeld.get();
    assert locks.contains(l);
    locks.remove(l);
  }
}
```

III. CONCLUSION AND FUTURE WORK:

In this work, we have proposed a language extension to the aspect-oriented programming language AspectJ. We extend AspectJ with three new pointcuts lock(), unlock(), and maybeShared(). These pointcuts allow researchers to easily implement bug-finding algorithms for errors related to concurrency. As an example, we have implemented RACER, an adaption of the ERASER race-detection algorithm to the Java memory model. We found that, using our AspectJ extension, we were able to implement RACER very easily, in just two aspects with a small set of supporting classes. The RACER algorithm is different from C-based race detection algorithms like ERASER in the way that it treats object initialization. ERASER is very forgiving to programmers in an object's initialization phase. RACER, on the other hand, detects and also reports races that comprise the initialization of an object. This revealed 12 data races in program code of the NASA K9 Rover Executive, 11 of which went previously undetected, although extensive studies of this code had already been performed at a time when nine of these undetected races were already present.

REFERENCES:

[1] J. Harrow, "Runtime Checking of Multithreaded Applications with Visual Threads," SPIN Model Checking and Software Verification, K. Havelund, J. Penix, and W. Visser, eds., pp. 331-342, Springer, 2000.

[2] R. O'Callahan and J.-D. Choi, "Hybrid Dynamic Data Race Detection," Proc. ACM SIGPLAN Symp. Principles and Practice of Parallel Programming, pp. 167-178, 2003.

[3] C. von Praun and T.R. Gross, "Object Race Detection," Proc. Ann. ACM SIGPLAN Conf. Object-Oriented Programming, Systems, Languages, and Applications, pp. 70-82, 2001.

[4] K. Havelund, "Using Runtime Analysis to Guide Model Checking of Java Programs," SPIN Model Checking and Software Verification, pp. 245-264, Springer, 2000.

[5] K. Havelund and G. Ros¸u, "An Overview of the Runtime Verification Tool Java PathExplorer," Formal Methods in System Design, vol. 24, no. 2, pp. 189-215, 2004.

[6] C. Artho, K. Havelund, and A. Biere, "High-Level Data Races," Software Testing, Verification and Reliability, vol. 13, no. 4, pp. 207-227, 2003.

[7] C. Artho, K. Havelund, and A. Biere, "Using Block-Local Atomicity to Detect Stale-Value Concurrency Errors," Automated Technology for Verification and Analysis, F. Wang, ed., pp. 150-164, Springer, 2004.

[8] C. Flanagan and S.N. Freund, "Atomizer: A Dynamic Atomicity Checker for Multithreaded Programs," Proc. 31st ACM SIGPLANSIGACT Symp. Principles of Programming Languages, pp. 256-267, 2004.

[9] L. Wang and S.D. Stoller, "Run-Time Analysis for Atomicity," Electronic Notes in Theoretical Computer Science, vol. 89, no. 2, 2003.

[10] F. Chen, T.F. Serbanuta, and G. Ros¸u, "jPredictor: A Predictive Runtime Analysis Tool for Java," Proc. 30th Int'l Conf. Software Eng., pp. 221-230, 2008.

◈◈◈

Implementation of Shared Memory Synchronization of Multi-core Processor using FPGA

G P Anilkumar[1] & B S Yashodha[2]

[1]VLSI, [2]Dept of ECE,
[1&2]SJBIT, Bangalore-60, India
E-mail : gpanilkumar@gmail.com

Abstract – Multi-core processors are about to defeat embedded systems We present the different methods of using shared memory and its consistency for different types of hardware means. They also support for point-to-point synchronization between the processor cores is realized implementing different hardware barriers. The practical examinations focus on the logical first step from single- to dual-core systems, using an FPGA-development board with two processor cores. Best- and worst-case results, together with intensive bench-marking of all synchronization primitives implemented, show the expected superiority of the hardware solutions. It is also shown that dual-ported memory outperforms single-ported memory if the multiple cores use inherent parallelism by locking shared memory more intelligently using an address-sensitive method.

***Keywords* -** *FPGA, Address sensitive, Shared Memory.*

I. INTRODUCTION

Electronic embedded architectures face a continuous increase in functionality which requires additional memory and computational power.

Changing from a single to multiple processor cores is not without pitfalls and requires prudence. Synchronization is the main topic that must be addressed Data synchronization prevents data from being invalidated by parallel access whereas event synchronization coordinates concurrent execution

Event synchronization forces processes to join at certain point of execution. Barriers can be used to separate distinct Phases of computation and are normally implemented without special hardware using locks and shared memory [4]. An involved process enters the barrier, waits for the other processes and then all processes leave the barrier together

In [6] synchronization primitives are analyzed regarding the amount of energy consumption of busy-waiting vs. blocking methods

In this paper blocking hardware solutions ensuring Synchrozation for a multiple number of processor cores are presented.

II. RELATED WORK

Each processor has a fixed number of channels to send data to the other processors, some bytes can be sent on a channel without blocking the sending processor. Here support of the compiler is needed to coordinate the Execution of the processes on the different processors.

The synchronization primitives' locks, barriers and lock-free data structures are the focus of attention in [9]. The classical implementations of those primitives are compared against hybrid synchronization primitives that use hardware support and the caches to improve efficiency and scalability, yielding promising results that seem to justify hardware acceleration In the specialized multi-core architecture described in [10] a DSP-, RISC- and VLIW-core are connected by a 64-bit AMBA AHB bus. For fast synchronization each pair of the three cores share dual-ported memory on-chip. Caching is not done for the on-chip but for the off-chip memory (SD- RAM). The work in [10] shows some relevance regarding the hardware configurations used and described in this paper In [11] an analysis of how to provide an efficient synchronization by barriers on a shared memory multiprocessor with a shared multi-access bus interconnection is described An innovative, perhaps unorthodox, alternative to ordinary barriers is given in [12]: the waiting of a thread is forced by continuous invalidation of the respective instruction cache.

An example of global event synchronization across parallel processors using a barrier support library is given in [13], a compiler is needed to produce the parallelized binary code.

Besides the performance overhead due to waiting, barriers also have significant power consumption [14] as disadvantage. Different barrier implementations for many-core architectures are analyzed in terms of efficiency and scalability in [15], proving that the scaling behavior of actual hardware implementations can differ to the expected scaling behavior

III. PROCESSOR SYNCHRONIZATION

A hardware-environment based on hard-wired processor Cores and on-chip shared memory is the fundament for the implementation and practical verification of the concepts presented. The main focus is on how to achieve reliable communication between the processor cores using the on-chip shared memory.

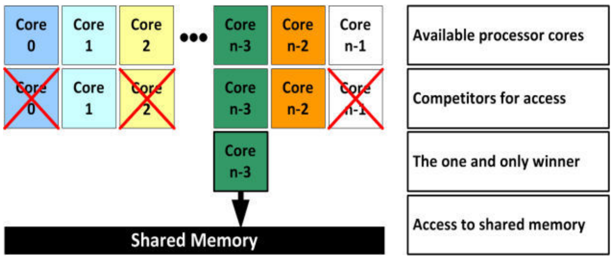


Fig. 1: Scheme of an efficient race for access

A. Problem

Synchronization between the arbitrary number of cores in their access to the shared memory is necessary. An efficient mechanism to resolve arbitrary concurrent requests for our critical resource should fulfill the following demands

Efficiency:

Only cores actually competing for access attend the race and can become its winner

Fairness:

No competitor waits indefinitely to get access. All requirements are met with the synchronization mechanism

B. Concept

In order to fulfill the efficiency and fairness demands a synchronization mechanism has been developed A simple round robin scheme cycling all available processor cores would require minimal resources for implementation but would be very inefficient when only a few cores want to access the shared memory

As shown in Fig. 1 only the cores which really need access are considered for it by our mechanism. Processor cores that have to wait are blocked until it is their turn. The worst case occurs in the situation when all the available cores want access to the shared memory simultaneously resulting in different waiting times. In order to avoid any processor core to be favored or discriminated a dynamic priority scheme is used to choose the access-order.

A global locking scheme making not just arbitrary single but also multiple accesses to the shared memory atomic is present as well. Global locking is offered by a global locking bit that is shared by all processor cores in the system.

Locking the whole shared memory when accessing essentially only a small area of the memory is not very efficient.

Therefore a special form of address-sensitive locking that allows the locking of just blocks instead of the whole shared memory has been developed and implemented. This enables concurrent read- and write-access to regions of shared memory, as is demonstrated for two cores in Fig. 2

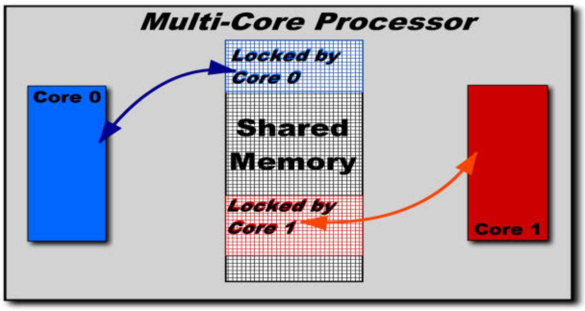
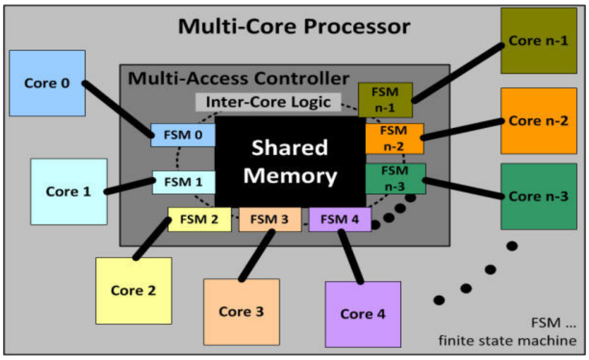


Fig. 2: Parallel access to different memory-regions by address-sensitivity

C. Realization

The developed Multi-Access Controller (MACtrl) consists of core-side and inter-core logic as shown in Figure 3.



A fully generic design of the MACtrl has been developed in the hardware description language VHDL in order to allow easy scaling in terms of the processor cores. The goal to keep the design as compact as possible is achieved by a code optimized algorithm that selects the next core that is allowed to access the shared memory in case of concurrent requests.

The algorithm continuously cycles the highest priority among all available cores. Only cores requesting access are used and the other cores are masked out in the process.

Simple barriers are the easiest method to achieve efficient point-to-point synchronization: each core that wants to meet at a given point of execution writes an arbitrary value to a dedicated barrier register of the MACtrl. Then the respective processor core is blocked until at least one other core writes to its corresponding counterpart-register.

With more than two processor cores in the system, extended simple barriers offer the possibility to define exactly for what other cores to wait for. Each bit in the register corresponds to the fixed number of a processor core in the system. The drawback is that the number of the cores must be known at compile-time.

IV. FINITE STATE MACHINE

At a low level of abstraction, a protocol is often most easily understood as a state machine. Design criteria can also easily be expressed in terms of desirable or undesirable protocol states and state transitions. In a way, the protocol state symbolizes the assumptions that each process in the system makes about the others. It defines what actions a process is allowed to take, which events it expects to happen, and how it will respond to those events. Finite state machine is intuitively the simplest.

In each particular state of the machine there can be zero or more transition rules that are executable. If no transition rule is executable, the machine is said to be in an end state. If precisely one transition rule is executable, the machine makes a deterministic move to a new control state.

If more than one transition rule is executable a nondeterministic choice is made to select a transition rule.

A nondeterministic choice in this context means that the selection criterion is undefined. Without further information either option is to be considered equally likely. A communicating finite state machine can be defined as an abstract demon that accepts input symbols, generates output symbols, and changes its inner state in accordance with some predefined plan.

The formal model of a finite state machine was developed in the early 1950s for the study of problems in computational complexity and, independently, for the study of problems in the design of combinatorial and sequential circuits. There are almost as many variants of the basic model of a finite state machine as there are applications

V. IMPLEMENTATION

All the hardware synchronization methods described in the previous Section were successfully simulated for four, some of them also for eight processor cores. From the beginning on there was no assumption limiting the number of cores.

For the actual hardware implementation the dual-core case could be examined thoroughly using a Xilinx development with a mounted FPGA SPARTAN-3E KIT.

The processor cores are designed using Verilog Code. Including the shared memory is designed using Verilog HDL. The different methods of implementation are done using Verilog HDL and it is tested using modelsim. Then it is implemented on FPGA SPARTAN-3E kit. Then the results are analyzed.

VI. RESULT

The problem of synchronization in multi-core systems with shared memory demands for efficient and reliable solutions, in particular for embedded systems. An approach is to integrate the synchronization mechanisms, which are normally based on locks, into the on-chip hardware.

Ongoing and future research focuses on testing efficient partitions of real embedded software on multi-core systems with hardware synchronization like the one presented here.

REFERENCES

[1] K. Asanovic et al., “A view of the parallel computing landscape,” in Communications of the ACM, Vol. 52, No. 10. ACM Press, October 2009, pp. 56–67.

[2] D. McGrath, “Intel rolls quad-core CPUs for embedded computing.”EE Times, April 2007. [Online]. Available: http://www.eetimes.com

[3] T. Takayanagi et al., “A dual-core 64b UltraSPARC Microprocessor for Dense Server Applications,” Sun Microsystems, Sunnyvale, USA, 2004.

[4] D. E. Culler and J. P. Singh, “Parallel Computer Architecture, a hw/sw approach.” Morgan

Kaufmann Publishers, Inc., Editorial and Sales Office, San Francisco, U. S. A., 1999.

[5] R. Johnson et al., “A New Look at the Roles of Spinning and Blocking,”in Proceedings of the Fifth International Workshop on Data Management on New Hardware, Providence, Rhode Island. ACM Press, June 2009.

[6] C. Ferri, I. Bahar, M. Loghi, and M. Poncino, “Energy-optimal synchronization primitives for single-chip multi-processors,” in GLSVLSI’09, Boston, Massachusetts. ACM Press, May 2009, pp. 141–144.

[7] M. C. August et al., “Cray X-MP: The Birth of a Supercomputer,” Cray Research, 1989.

[8] R. Gupta et al., “The Design of a RISC based Multiprocessor Chip,”University of Pittsburgh, Philips Laboratories New York, 1990.

[9] Nikolopoulos and Papatheodorou, “Fast synchronization on scalable cache-coherent multiprocessors using hybrid primitives,” University of Patras, Greece, 2000.

[10] H.-J. Stolberg et al., “HiBRID-SoC: A multi-core System-on-Chip architecture for multimedia signal processing applications,” Universitaet Hannover, Germany, 2003.

[11] S. Y. Cheung and V. S. Sunderam, “Performance of Barrier Synchronization Methods in a Multi-Access Network,” Emory University, Atlanta, Georgia, 1993.

[12] J. Sampson et al., “Fast synchronization for chip multiprocessors,” in ACM SIGARCH Computer Architecture News, Vol. 33, UCSD, UPC Barcelona, Palo Alto, California, 2005.

[13] A. Marongiu, L. Benini, and M. Kandemir, “Lightweight barrier-based parallelization support for non-cache-coherent MPSoC platforms,” in CASES’07, Salzburg, Austria. ACM Press, Sept. 2007, pp. 145–149.

[14] C. Liu, A. Sivasubramaniam, M. Kandemir, and M. J. Irwin, “Exploiting barriers to optimize power consumption of CMPs,” in Proceedings of IPDPS, 2005.

[15] O. Vila, G. Palermo, and C. Silvano, “Efficiency and scalability of barrier synchronization on NoC based many-core architectures,” in CASES’08, Atlanta, Georgia, USA. ACM Press, October 2007, pp. 81–89.

◈◈◈

Detection of Link Failures and Autonomous Reconfiguration in WMNs

Sumathi P & M. V. Jagannatha Reddy

Dept of CSE, Madanapalle Institute of Technology and Science, Madanapalle.
E-mail : Suma.cse51@gmail.com

Abstract – During their lifetime, multihop wireless mesh networks (WMNs) experience frequent link failures caused by channel interference, dynamic obstacles, and/or applications' bandwidth demands. By reconfiguring these link failures ARS generates an effective reconfiguration plan that requires only local network configuration changes by exploiting channel, radio, and path diversity. ARS effectively identifies reconfiguration plans that satisfy QoS constraints. And ARS's online reconfigurability allows for real-time time failure detection and network reconfiguration. ARS is mainly evaluated in IEEE 802.11a networks. It's design goal is to reconfigure from network link failures accurately. Even then WMNs face some frequent link failures. By overcome these problems we present *L*ocalized s*E*lf-reconfi*G*uration alg*O*rithms (LEGO) to autonomously and effectively recnfigure from wireless link failures. First, LEGO locally detects link failures. Second, it dynamically forms/deforms a local group for cooperative network reconfiguration among local mesh routers in a fully distributed manner. Next, LEGO intelligently generates a local network reconfiguration plan. Finally, by figuring local channel utilization and reconfiguration cost in its planning, LEGO maximizes the network's ability to meet diverse links' QoS demands. LEGO has been implemented on a Linux-based system and experimented on a real life test bed, demonstrating its effectiveness in recovering from link failures and its improvement of channel efficiency by up to 92%.

Keywords **-** *Self-Reconfigurable Networks, Multi-Radio Wireless Networks, IEEE 802.11, WLAN access points (AP).*

I. INTRODUCTION

In WMNs, nodes are comprised of mesh routers and mesh clients. Each node operates not only as a host but also as a router, forwarding packets on behalf of other nodes that may not be within direct wireless transmission range of their destinations. A WMN is dynamically self-organized and self-configured, with the nodes in the network automatically establishing and maintaining mesh connectivity among themselves (creating, in effect, an ad hoc network). Routing in ad-hoc wireless networks has been an active area of research for many years. Much of the original work in the area was motivated by mobile application environments, such as battlefield ad hoc networks. The use of multiple radios is complementary to the use of directional antennas, and we believe that our protocol can be modified for directionality. Specifically, we would have to revisit the assumption that all same-channel links along a path interfere with one another. Another way to improve the capacity of a wireless network is to take advantage of the full spectrum by using rapid channel switching. This approach has been explored by several researchers. However, channel switching can be quite slow with existing 802.11 hardware. With the availability of better hardware, many of the proposed approaches based on rapid channel switching will become feasible. Our approach, however, works with currently available hardware. We also note that even with the ability to switch channels rapidly, a single radio can not transmit and receive simultaneously. Thus, the use of multiple radios can provide a performance improvement even in this case. A novel approach that takes advantage of the inherent multi-radio capability of WMNs. We show that this capability can enable partitioning of the network into sub networks in which simple distributed scheduling algorithms can achieve 100% throughput. The partitioning is based on the recently introduced notion of Local Pooling. Using this notion, we characterize topologies in which 100% throughput can be achieved distributedly. These topologies are used in order to develop a number of channel assignment algorithms that are based on a matroid intersection algorithm[5]. These algorithms partition a network in a manner that not only expands the capacity regions of the sub networks but also allows distributed algorithms to achieve these capacity regions. Finally, we evaluate the performance of the algorithms via simulation and show that they significantly increase the distributedly achievable capacity region. Joint scheduling and routing in a slotted multihop wireless network with a stochastic packet arrival process was considered in the seminal paper by Tassiulas and Ephremides [11]. In that they presented the first centralized policy that is guaranteed to stabilize the network (i.e. provide 100% throughput) whenever the arrival rates are within the stability region. The results of [11] have been extended to various settings of wireless networks and input-queued switches. However, optimal algorithms based on [11] require repeatedly solving a global optimization problem, taking into account the queue backlog information for every link in the network. Obtaining a centralized solution to such a problem in a wireless network does not seem to be feasible, due to the communication overhead associated with continuously collecting the queue backlog information. On the other hand, distributed algorithms usually provide only approximate solutions, resulting in significantly reduced throughput. Using AODV and DSR routing Protocols the packets are reached to destination. Using DSR, the network is completely self-organizing and self-configuring, requiring no existing network infrastructure or administration. Network nodes (computers) cooperate to forward packets for each other to allow communication overmultiple “hops” between nodes not directly within wireless transmission range of one another. As nodes in the network move about or join or leave the network, and as wireless transmission conditions such as sources of interference change, all routing is automatically determined and maintained by the DSR routing protocol. Since the number or sequence of intermediate hops needed to reach any destination may change at any time, the resulting network topology may be quite rich and rapidly changing. The DSR protocol allows nodes to dynamically discover a *source route* across multiple network hops to any destination in the ad hoc network. By including this source route in the header of each data packet, other nodes forwarding or overhearing any of these packets may also easily cache this routing information for future use[8]. The original motivation in the design of DSR came from the operation of theAddressResolution Protocol (ARP) used in the TCP/IP suite of protocols in the Internet. ARP is used on Ethernets and other types of networks to find the link-layer MAC address of a node on the same subnet as the sender. A node sending a packet to a local IP address for which it does not yet have the MAC address cached, broadcasts an ARP REQUEST packet on the local subnet link, giving the IP address of the node it is looking for; that node responds with an ARP REPLY packet, giving its MAC address, and all other nodes ignore the REQUEST. If all nodes in an ad hoc network are within wireless transmission range of each other, this is the only routing protocol needed for the ad hoc network. ABR also adds overhead for periodic beacon packets required to monitor link stability. The Ad Hoc On-Demand Distance Vector routing protocol (AODV) uses mechanisms similar to DSR’s Route Discovery and Route Maintenance, but it uses them to create hop-by-hop routes rather than source routes as is done in DSR[8]; this use of hop-by-hop routes avoids the source routing header overhead of DSR but prevents or makes difficult many of the route caching and other Route Discovery optimizations present in DSR and prevents AODV from supporting uni-directional links between nodes. we establish the capacity of general multi channel networks wherein the number of interfaces, m, may be smaller than the number of channels, c.However, one important exception is a random network with up to O (log n) channels, independent of the number of interfaces available at each node[2]. This implies that it may be possible to build capacity-optimal multi-channel networks with as few as one interface per node. We also extend our model to consider the impact of interface switching delay, and show that capacity losses due to switching delay can be avoided by using multiple interfaces. ARS has been implemented and evaluated extensively via experimentation on our multiradio WMN test-bed as well as via ns2-based simulation. Our evaluation results show that ARS outperforms existing failure-recovery methods, such as static or greedy channel assignments, and local rerouting. ARS’s local reconfiguration improves network throughput and channel efficiency by more than 26% and 92%, respectively, over the local rerouting scheme.

II. PROBLEM DEFINITION

Wireless mesh networks have the potential to deliver Internet broadband access, wireless local area network coverage and network connectivity for stationary or mobile hosts at low costs both for network operators and customers.

We first describe the need for self-reconfigurable mr WMNs. Next, we introduce the network model and assumptions.

A. Why Is Self-Reconfigurability Necessary?

By enabling mr-WMNs to autonomously reconfigure channels and radio1 assignments, as in the following examples.

- *Recovering from link-quality degradation*:

The quality of wireless links in WMNs can degrade (i.e., *link-quality failure*) due to severe interference from other collocated wireless networks.

- *Satisfying dynamic QoS demands*:

Links in some areas may not be able to accommodate increasing QoS demands from end-users depending on spatial or temporal locality.

- Coping with heterogeneous channel availability: Links in some areas may not be able to access wireless channels during a certain time period (*spectrum failures*) due to spectrum etiquette or regulation.

B. Network Model and Assumptions

- Multiradio WMN:

A network is assumed to consist of mesh nodes, IEEE 802.11-based wireless links, and one control gateway. Each mesh node is equipped with radios, and each radio's channel and link assignments are initially made by using (e.g., see Fig. 1) global channel/link assignment algorithms Multiple orthogonal channels are assumed available.

- QoS Support:

During its operation, each mesh node periodically sends its local channel usage and the quality information for all outgoing links via management messages to the control gateway.

- Link Failures:

Channel-related link failures that we focus on are due mainly to narrowband channel failures. These failures are assumed to occur and last in the order of a few minutes to hours, and reconfiguration is triggered in the same order of failure occurrences.

III. SYSTEM ARCHITECTURE

Wireless mesh architecture is a first step towards providing cost effective and dynamic high-bandwidth networks over a specific coverage area. Wireless mesh architectures infrastructure is, in effect, a router network minus the cabling between nodes. It's built of peer radio devices that don't have to be cabled to a wired port like traditional WLAN access points (AP) do. Mesh architecture sustains signal strength by breaking long distances into a series of shorter hops. Intermediate nodes not only boost the signal, but cooperatively make forwarding decisions based on their knowledge of the network, i.e. perform routing. Such an architecture may with careful design provide high bandwidth, spectral efficiency, and economic advantage over the coverage area.

We first present the design rationale and overall algorithm of ARS. Then, we detail ARS's reconfiguration algorithms. Finally, we discuss the complexity of ARS.

A. Overview:

ARS is a communication network that is easily deployable in IEEE 802.11-based mr-WMNs. Running in every mesh node, ARS supports self-reconfigurability via the following distinct features.

- Localized reconfiguration:

Based on multiple channels and radio associations available, ARS generates reconfiguration plans that allow for changes of network configurations only in the vicinity where link failures occurred while retaining configurations in areas remote from failure locations.

- QoS-aware planning:

ARS effectively identifies QoS-satisfiable reconfiguration plans by: 1) estimating the QoSsatisfiability of generated reconfiguration plans; and 2) deriving their expected benefits in channel utilization.

- Autonomous reconfiguration via link-quality monitoring:

ARS accurately monitors the quality4 of links of each node in a distributed manner. Furthermore, based on the measurements and given links' QoS constraints, ARS detects local link failures and autonomously initiates network reconfiguration.

- Cross-layer interaction:

ARS actively interacts across the network and link layers for planning. This interaction enables ARS to include a rerouting for reconfiguration planning in addition to link-layer reconfiguration.

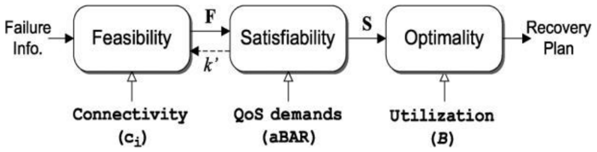


Fig. 1 : Localized reconfiguration planning in ARS. ARS generates a reconfiguration plan by breaking down the planning process into three processes with different constraints.

B. Planning for Localized Network Reconfiguration

The core function of ARS is to *systematically* generate localized reconfiguration plans. A *reconfiguration plan* is defined as a set of links' configuration changes (e.g., channel switch, link association) necessary for a network to recover from a link(s) failure on a channel, and there are usually multiple reconfiguration plans for each link failure. By contrast, ARS systematically generates reconfiguration plans that localize network changes by dividing the reconfiguration planning into three processes—feasibility, QoS satisfiability, and optimality—and applying different levels of constraints. As depicted in Fig. 2, ARS first applies connectivity constraints to generate a *set* of feasible reconfiguration plans that enumerate feasible channel, link, and route changes around the faulty areas, given connectivity and link-failure constraints. Then, within the set, ARS applies strict constraints (i.e., QoS and network utilization) to identify a reconfiguration plan that satisfies the QoS demands and that improves network utilization most.

Feasible Plan Generation*:* Generating feasible plans is essentially to search all legitimate changes in links' configurations and their combinations around the faulty area. However, in generating such plans, ARS has to address the following challenges.

- Avoiding a faulty channel:

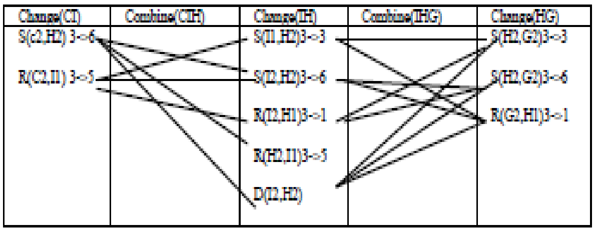
ARS first has to ensure that the faulty link needs to be fixed via reconfiguration.

- Maintaining network connectivity and utilization:

While avoiding the use of the faulty channel, ARS needs to maintain connectivity with the full utilization of radio resources. Because each radio can associate itself with multiple neighboring nodes, a change in one link triggers other neighboring links to change their settings.

- Controlling the scope of reconfiguration changes:

ARS has to limit network changes as *local* as possible, but at the same time it needs to find a locally optimal solution by considering more network changes or scope. To make this tradeoff, ARS uses a –hop reconfiguration parameter. Starting from a faulty link(s), ARS considers link changes within the first hops and generates feasible plans.



Examples of feasible plans generated

P1=[S(C2,I2)3->6,S(I2,H2)3->6,S(H2,G2)3->6],

P2=[S(C2,I2)3->6,D(I2,H2)S(H2,G2)3->3]….P11

Fig.2. Example of network planning.

Let us consider an illustrative example in Fig. 4. Given the failure in link CI, ARS first generates feasible and desirable changes per link (gray columns) using the primitives. Here, the changes must not include the use of a faulty or redundant channel. Next, ARS combines the generated per-link primitives of neighboring links to generate a set of feasible plans. During the combination, ARS has to preserve link and/or radio connectivities. After the two steps, ARS has 11 feasible reconfiguration plans(F) by traversing connected changes of all links considered in the planning.

- *QoS-Satisfiability Evaluation:*

Among a set of feasible plans,ARS now needs to identify QoS-satisfying reconfiguration plans by checking if the QoS constraints are met under each plan.

To filter out such plans, ARS has to solve the following challenges.

- *Per-link bandwidth estimation*:

For each feasible plan, ARS has to check whether each link's configuration change satisfies its bandwidth requirement, so it must estimate link bandwidth. To estimate link bandwidth, ARS accurately measures each link's capacity and its available channel airtime

- *Examining per-link bandwidth satisfiability*:

Given measured bandwidth and bandwidth requirements, ARS has to check if the new link change(s) satisfies QoS requirements. ARS defines and uses the expected busy airtime ratio of each link to check the link's QoS satisfiability.

- *Choosing the Best Plan:*

ARS now has a set of reconfiguration plans that are QoS-satisfiable and needs to choose a plan within the set for a local network to have evenly distributed link capacity.

C. Complexity of ARS

ARS incurs reasonable bandwidth and computation overheads. First, the network monitoring part in the reconfiguration protocols is made highly efficient by exploiting existing data traffic and consumes less than 12 kb/s probing bandwidth (i.e., one packet per second) for each radio. In addition, the group formation requires onlyO(n) message overhead (in forming a spanning tree), where n is the number of nodes in the in the group. Next, the computational overhead in ARS mainly stems from the planning algorithms. Specifically, generating its possible link plans incurs O(n+m) complexity, where n is the number of available channels and m is the number of radios. Next, a gateway node needs to generate and evaluate feasible plans, which incurs search overhead in a constraint graph that consists of O(l(n+m))nodes, where is the number of links that use a faulty channel in the group.

IV. CONCLUSION

An autonomous network reconfiguration system (ARS) that enables a multi-radio WMN to autonomously recover from wireless link failures. ARS generates an effective reconfiguration plan that requires only local network configuration changes by exploiting channel, radio, and path diversity. Furthermore, ARS effectively identifies reconfiguration plans that satisfy applications' QoS constraints, admitting up to two times more flows than static assignment, through QoS aware planning. Next, ARS's online reconfigurability allows for real-time failure detection and network reconfiguration. Based on existing MAC, routing, and transport protocols, network performance is not scalable with either the number of nodes or the number of hops in the network. This problem can be alleviated by increasing the network capacity through using multiple channels/radios per node or developing wireless radios with higher transmission speed. However, these approaches do not truly enhance the scalability of WMNs, because resource utilization is not actually improved. Therefore, in order to achieve scalability, it is essential to develop new MAC, routing, and transport protocols for WMNs.

V. ACKNOWLEDGMENT

Students work is incomplete until they thank the almighty & his teachers. I sincerely believe in this and would like to thank Mr.M.V Jagannatha Reddy, M.Tech(Ph.D), Associate prof., of the department of Computer Science & Engineering, MITS Madanapalle, for his encouragement, motivation and for guiding me to write this paper. Also I am grateful to Dr.Murali Krishna, M.Tech.,Ph.D. Prof., in CSE and Head Of the Department of Computer Science, MITS, Madanapalle.

REFERENCES

[1] Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," Comput. Netw., vol. 47, no. 4, pp. 445–487, Mar. 2005.

[2] P. Kyasanur and N. Vaidya, "Capacity of multi-channel wireless networks: Impact of number of channels and interfaces," in Proc. ACM MobiCom, Cologne, Germany, Aug. 2005, pp. 43–57.

[3] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in Proc. ACM MobiCom, Philadelphia, PA, Sep. 2004, pp. 114–128.

[4] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi channel wireless mesh networks," in Proc. ACM MobiCom, Cologne, Germany, Aug. 2005, pp. 73–87.

[5] A. Brzezinski, G. Zussman, and E. Modiano, "Enabling distributed throughput maximization in wireless mesh networks: A partitioning approach," in Proc. ACM MobiCom, Los Angeles, CA, Sep. 2006, pp. 26–37.

[6] A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in Proc.IEEE INFOCOM, Miami, FL, Mar. 2005, vol. 3, pp. 2223–2234.

[7] S. Chen and K. Nahrstedt, "Distributed quality-of-service routing in ad hoc networks," IEEE J. Sel. Areas Commun., vol. 17, no. 8, pp.1488–1505, Aug. 1999.

[8] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in The Book of Mobile Computing. Norwell, MA: Kluwer, 1996, vol. 353.

[9] A. P. Subramanian, H. Gupta, S. R. Das, and J. Cao, "Minimum interference channel assignment in multiradio wireless mesh networks,"

[10] K.-H. Kim and K. G. Shin, "On accurate and asymmetry-aware management of link quality in wireless mesh networks," IEEE/ACMTrans. Netw., vol. 17, no. 4, pp. 1172–1185, Aug. 2009.

◈◈◈

Business Intelligence Interface for Sales Analysis

Implementation of Business Intelligence for analysis of Sales of an organization – the QlikView way

Ankit Joshi[1] & Sangeet Patil[2]

[1]Department of Advanced Software Technologies, International Institute of Information Technology
Pune, Maharashtra, India
[2]Information Systems (IS), Team Computers, Mumbai, Maharashtra, India
E-mail : ankitj_aug10@ast.isquareit.ac.in[1], sangeet.patil@teamcomputers.com[2]

Abstract – Business Intelligence (BI) forms the backbone of decision making in an organization. It aids the organization in analyzing along with managing the huge amounts of data being generated in a systematic manner. Here, QlikView tool has been put to use, to provide a BI solution for the analysis of sales, which undoubtedly forms the driving force of any organization.

Keywords- *QlikView, Business Inteligence, Sales, Bucket, Projection.*

I. INTRODUCTION

Every business organization's main lifeline is its sales. The success of a business depends majorly on the art it puts to use for selling, and this is entirely reflected in the sales of that organization. Here, we would like to showcase the vital and pivotal role played by QlikView tool in implementation of a Business Intelligence (BI) solution for sales of an organization.

II. QLIKVIEW IN A NUTSHELL

According to the BI Survey 10 – The Customer verdict, QlikView has emerged as the strongest of all the products in its primary peer group and stands tall against other BI Giants. When it comes to buying a software product, if the customers do not go by the price tag rather make their choices on the basis of features being offered, directly complement the product. QlikView has succeeded in this regard, as the companies which buy it, consider its overall capabilities as a strong motivator. In the area of overall agilty, QlikView ranked first. The survey monitored agility over a lot of things. These included the number of full-time administrators needed per user, in what all activities the users handle the product themselves and also, the frequency with which the BI product is used with no help from outside. In comparison to its peer BI products, QlikView offers the fastest implementation time and stands out from others as it helps in going live faster. Next, talking about Performance, it matters a lot in BI products. A customer getting satisfied with the performance of the product is a vital sign showing the worth of the product. Performance affects a lot of things in totality – ranging from outcomes of the project to benefits to the business. The most recent addition as a Key Performance Indicator in this survey was of Suitability. How suitable could a product be to the type of projects it is chosen for, is an important parameter to be considered. QlikView topped this KPI, with it being the best suited to the projects it is chosen for. Overall business achievement has been calculated in this survey by combining business benefit KPI and goal achievement KPI. Not to our astonishment, QlikView ranks first in this category too. It focuses on the combination of bottom-line benefit of a BI project along with measuring the success of the goals which had been set at the beginning of the project. The problems which had been reported by the customers were the least with QlikView, as compared to other BI Giants. Whether we talk about problems related to data, people or technical-related problems – all were the least reported for QlikView. Vendor support, where the quality of support given by QlikView is measured and Implementer support where third-party implementers' support is measured have been taken into consideration too. In both these areas, QlikView scored strongly. QlikView has been reviewed to be the best produvt quality wise, with the fewest reported problems. For three consecutive years, QlikView has been topping the category for being the most loyal product to its customers as compared to its peer group. It shows how comfortable customers are in using this product and use it for their intended purpose. It is said that QlikView has a "viral" appeal, as the existing customers have reported that they expect to purchase more licenses. This is a very positive sign and hence, here too QlikView bags the top position in terms of intent to purchase more licenses.

Aberdeen's research briefs in 2010, presented a detailed research study. The group had surveyed 400 end-user organizations between May and June 2010. They reported that customers using QlikView outperformed the best-in-class with Dashboards against all other customers. It was reported that accuracy of revenue to budget and accuracy of bottom-line budget of QlikView was nearing to best-in-class. The best-in-class referred to the top 20% of the performers in this survey. Also, QlikView users were reported to have per-user costs of dashboards roughly three-quarters of those by the best-in-class, which is another highlight. Apart from this, QlikView has another advantage to boast about, that is the ease with which their dashboards can be configured.

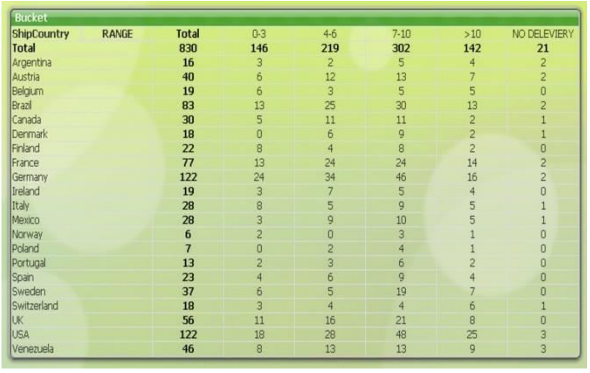
With the changing market, there needs to be a change in the dashboards which are put to use for operational performance management. In this regard, QlikView users out perform others as they are capable of driving their dashboard initiatives according to the needs of Line-of-Business Managers. Also, these dashboards help these managers in their daily decision making quite well. QlikView users most effectively give source data the priority for end user access. They start their Business Intelligence process with source data in comparison to its display, thereby able to deploy dashboards rapidly. Key performance indicators can be defined independently of data sources thus, such users are able to create dashboards quickly, so that IT personnel can work in identifying data sources for particular dashboards without the managers being disturbed.

III. BUSINESS INTELLIGENCE APPLICATION FOR SALES IN AN ORGANIZATION

In this interface, an organization's sales data is being analyzed, highlighting the role which sales hold in the success of an organization. This is a dashboard representation where all the main information pertaining to sales of an organization are well represented. In this, sales can be viewed and analyzed in numerous ways suiting the ease of the user – monthly, quarterly and annually. Sales can be depicted manager wise – i.e. each manager's contribution towards sales can be monitored and their performances thus analyzed. A number of analyses could be made with this application for the sales data available; here the results are based upon the sales information of the years 1996-1998.

A. *Bucket*

It is a wonderful interface of this dashboard which helps in knowing about the countries where the products of the organization get shipped to, the status of the orders placed, number of days it takes for the delivery to take place, which is represented in the form of 0-3, 4-6, 7-10, >10 days taken (Fig. 1). Moreover, if some products remain undelivered, their status is also shown. As the sales would take place all over the world, a feature of currency conversion has also been incorporated in this interface, adding more value to it, being more user friendly.



Bucket

ShipCountry	RANGE	Total	0-3	4-6	7-10	>10	NO DELEVIERY
Total		**830**	**146**	**219**	**302**	**142**	**21**
Argentina		**16**	3	2	5	4	2
Austria		**40**	6	12	13	7	2
Belgium		**19**	6	3	5	5	0
Brazil		**83**	13	25	30	13	2
Canada		**30**	5	11	11	2	1
Denmark		**18**	0	6	9	2	1
Finland		**22**	8	4	8	2	0
France		**77**	13	24	24	14	2
Germany		**122**	24	34	46	16	2
Ireland		**19**	3	7	5	4	0
Italy		**28**	8	5	9	5	1
Mexico		**28**	3	9	10	5	1
Norway		**6**	2	0	3	1	0
Poland		**7**	0	2	4	1	0
Portugal		**13**	2	3	6	2	0
Spain		**23**	4	6	9	4	0
Sweden		**37**	6	5	19	7	0
Switzerland		**18**	3	4	4	6	1
UK		**56**	11	16	21	8	0
USA		**122**	18	28	48	25	3
Venezuela		**46**	8	13	13	9	3

Fig. 1 : Bucket Description

B. *Total Sales*

This pie chart representation in Fig. 2, very vividly describes the part each product plays in adding up to the huge sales figures each year. Percentage wise, it shows the product and the percentage share it holds in the total sales taken place in the organization.

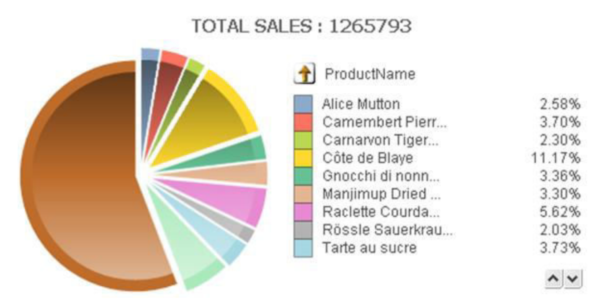


Fig. 2 : Total Sales

C. *Growth Rate*

This helps in knowing the rate at which growth has taken place and can be compared over four quarters in the year. This is being represented in a bar graph format in Fig. 3 and the bars represent – CY: current year chosen, PY: its previous year and GR: growth rate of that particular quarter. This gives an overall view of what the scenario has been across the entire year in terms of sales, split into sub-sections i.e. quarters.

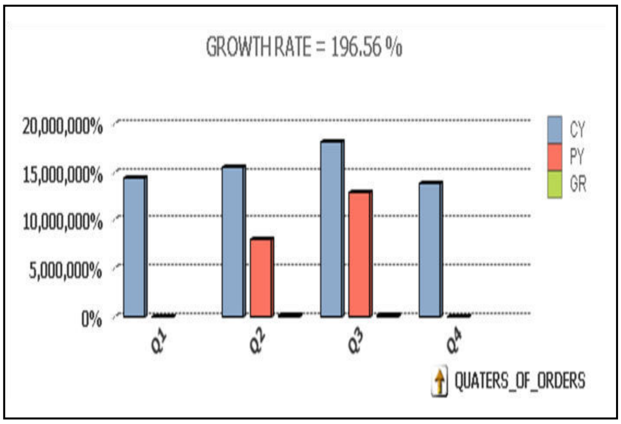


Fig. 3 : Growth Rate

D. *Net Profit*

Fig. 4 depicts the gross i.e. the total profit which has been incurred in a year, and thus can be compared with the other year's profit. Analyzing which year gained the highest profit, would help to know the factors which contributed towards it bagging the highest profit.

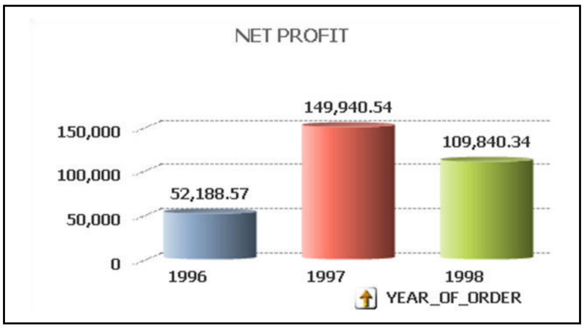
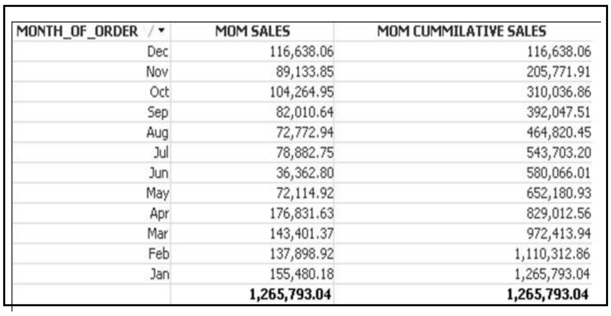


Fig. 4 : Net Profit Gained

E. *Sales Month Wise*

This aids in knowing about the month wise sales which have taken place in a year. Also, every month's cumulative sales are also shown (Fig. 5), which helps in analyzing the sales of the organization after certain months together.



MONTH_OF_ORDER	MOM SALES	MOM CUMMILATIVE SALES
Dec	116,638.06	116,638.06
Nov	89,133.85	205,771.91
Oct	104,264.95	310,036.86
Sep	82,010.64	392,047.51
Aug	72,772.94	464,820.45
Jul	78,882.75	543,703.20
Jun	36,362.80	580,066.01
May	72,114.92	652,180.93
Apr	176,831.63	829,012.56
Mar	143,401.37	972,413.94
Feb	137,898.92	1,110,312.86
Jan	155,480.18	1,265,793.04
	1,265,793.04	**1,265,793.04**

Fig. 5 : Month Wise Sales Presentation

F. *Top Customers*

Over a period of time, who has been the consistent customer, in terms of contributing the maximum to the progress in the sales of the organization, can be known with this. (Fig. 6)

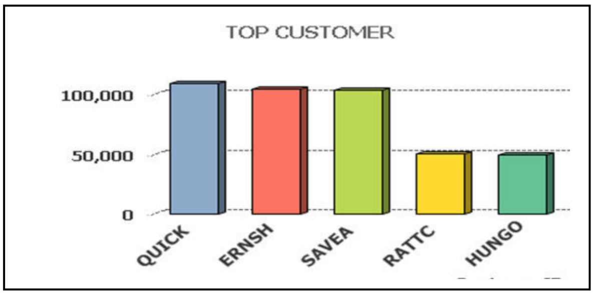


Fig. 6 : Top Customers

G. *Top Employees*

In this, the scenario is according to the performances of the employees (Fig, 7). Based upon their contribution, in enhancing the sales of the organization, top performer employees are recognized.

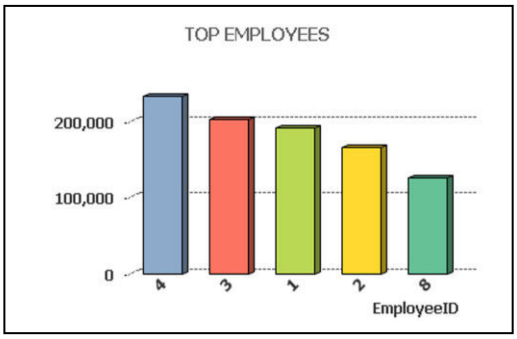


Fig. 7 : Top Employees

H. *Top Products*

These are the ones which have gained maximum popularity over the years amongst the customers (Fig. 8), thus, standing apart from other products and marking their share in promotion of sales of the organization as a whole.

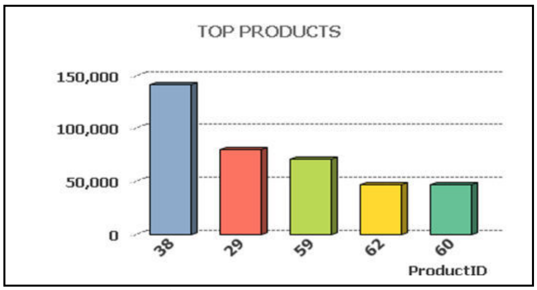


Fig. 8 : Top Products

I. Gross Profit

It highlights the fact from which country is the maximum profit being gained as depicted in Fig. 9. It brings to our knowledge the places where the products get shipped to the most. Hence, getting to know about the sales and moreover, the popularity of the products overseas.

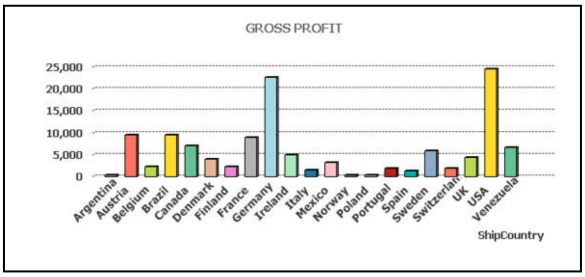


Fig. 9 : Gross Profit Gained

J. Discount

The gross discount which has been given by the organization over a period of time can be known (Fig. 10). It helps in comparing the discounts given in one year with another, thereby helping in analyzing the sales better.

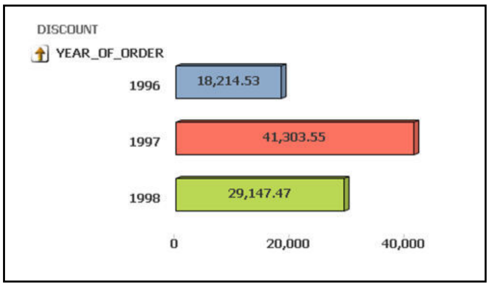


Fig. 10 : Discount Allotted

K. Number of orders

The trend in which the orders are placed on different days on a year as well as across various years can be monitored. Also, the grand total number of orders placed at the end of the analysis is known cumulatively, as shown in Fig. 11.

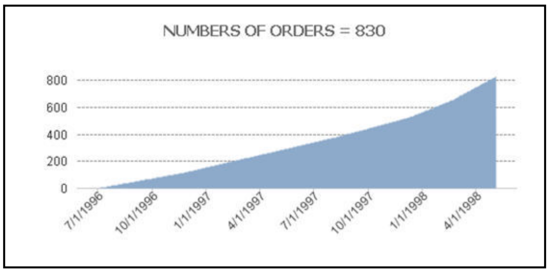


Fig. 11 : Number of Orders Placed

L. Projection

The prospective sales of the organization can be predicted with this, the basis being the previous years' sales figures, presented in Fig. 12. This would help the organization in planning out their activities beforehand according to the predictions made.

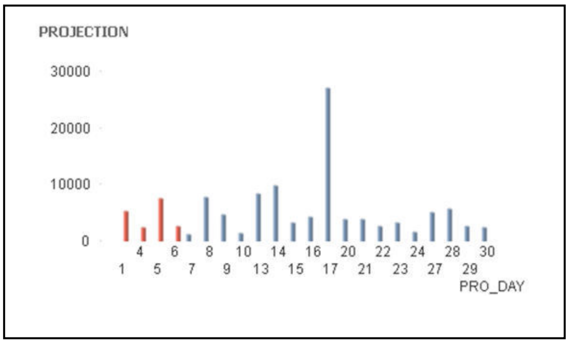


Fig. 12 : Projection

IV. CONCLUSION

It is clearly seen how useful and helpful it becomes, once a Business Intelligence solution, similar to the one explored here, is put into practice by an organization. It helps in easing down a lot of things. Pertaining to analysis of sales performed here, proper and systematic management is ensured. The dashboard representation is the main highlight as it aids in providing with an overall clear picture of what all can be explored. Total sales analysis helps in monitoring the contribution of each of the managers region wise, in addition to the total sales analysis which can be done product wise. One of the most important features of this application is the currency conversion, which eases the analysis especially in overseas dealing with other countries. Calculation and analysis of growth rate, net profit and discount rates are some of the other mainstays. Performance can be measured very easily, as top employees, top products and moreover, a record for top customers is maintained with the help of this interface. Adding another feather in its cap is the information regarding the gross profit being earned. This helps in knowing in which country (overseas sales), the organization's products gain maximum popularity and what parameters can be put to use to increase the sales in other countries.

In order to investigate more about the trend in which orders for various products of the organization are being placed, this platform can be utilized. The most essential thing in sales analysis lies in the power of predicting the trend and figures of the future sales, which are yet to take place. In this BI solution, Projection, provides with this ability, where the organization can foresee and predict an approximate trend of its sales to happen in future, based upon the previous sales figures and trend.

For any organization to stay and moreover, survive in this competitive market scenario, it requires a set plan, a strategy and this decision making as well as strategy making power is endowed upon by Business Intelligence tools. These tools help the organization to analyze and understand its past so well that prediction of its future sets apt.

ACKNOWLEDGMENT

Sincere thanks to Team Computers for guiding me with the analysis through QlikView Tool. Heartfelt thanks to Ms. Aditi Kapoor for helping in the writing and formatting of the manuscript.

REFERENCES

[1] L. Şerbănescu, “Business intelligence applications for human resource outsourcing, “ unpublished.

[2] Aberdeen Group, “QlikView customers outperform the best-in-class with dashboards”, Research Brief, August 2010.

[3] QlikTech International, 2009, QWT Business Intelligence – Professional Layout, http://www.qlikview.com

[4] QlikCommunity, http://community.qlikview.com/index.jspa

◈◈◈

The Campus Security Tracking System Based on RFID and ZigBee Network

M. Sathish Kumar & S. Nagaraj

Sri Venkateswara College of Engineering, R.V.S Nagar, Chittor
E-mail : satish2urs@gmail.com, nagarajsubramanyam@gmail.com

Abstract – This paper contains the campus security tracking system (CST) has been designed and implemented using the RFID and ZigBee network. The CST reads the RFID tags data through FRID & ZigBee node, and then sends it to PC node by a custom wireless protocol on the ZigBee. PC node gives related warning (or) hints by the result of matching master slave RFID tag information. When the warning occurs, its user can logon the web system to get the real-time tracking for valuables with embedded slave RFID where the thief arrives at any one access control node, it would be blocked. User can also manage his own valuables such as lending and recovery operation through the Web manager center. The system can achieve large-scale logistics management and tracking, and has good market prospects and value in use.

I. INTRODUCTION

As the number of valuables owned by college students increases, means of criminal theft becomes more intelligent and hidden. But students' security awareness is relatively weak, which results in frequent campus thefts and has bad effect on teachers, students and the society. To solve the problem, we design an intelligent campus security tracking system based on RFID and ZigBee, which has full range of protection on campus that has small flow of people and wiring is not convenient.

RFID (Radio Frequency Identification) ,which identifies targets within the target area and obtains relevant data automatically through RF signals, has become welcomed by people gradually since it has been proposed. The recognition process can resist harsh environments without human intervention. Based on the technology characteristics and strength of RFID, we combine RFID and campus security tracking system, and use SCM control technology and wireless network technology, to develop an intelligent anti-theft system with networking capability. That's to say, combine RFID and ZigBee to build a wireless network which doesn't need rewiring. And its special high efficiency and convenience will reduce the cost of campus security and make campus safer, which will reduce the fluency of campus theft effectively and benefit students and teachers.

Based on RFID and ZigBee, the intelligent campus security tracking system uses physical methods and electronic technology, automatically detects theft in monitoring areas, generates alarm signals, and tracks targets through the detection point of RFID. Anti-theft alarm system is important facility to prevent robbery, theft, accidents, etc. In the event of emergency, it can show the accurate scene in security control center through sound and light alarm signals to enable emergency measures to be taken. Anti-theft alarm system constitutes a security system together with import and export control system, closed circuit television monitoring system, visitor intercom systems and electronic patrol system.

II. SYSTEM PRINCIPLE AND ERROR ANALYSIS

The intelligent campus security tracking system is based on wireless communication services between nodes provided by RFID sensors and ZigBee, and identifies the RFID tags within the region to prevent thefts and track valuables, so as to protect the property of the teachers and students.

A. *RFID&ZigBee nodes implementation*

As the popularity of RFID sensor technology increases, RFID tags are cheap, and can be reused. Data transmission is convenient and easy, we can deal with tags in accordance with Customized encryption algorithm based on user requests. Then combine RFID sensor with ZigBee(referred to as RFID & ZigBee later) ,and complete the sensor node data collection work. Transceiver inducts RFID tags, sends the Information of the read tags to the micro-controller (Freescale LPC214XX) according to RS232 transmission protocol.

The micro-processor sends the data to wireless module to complete data transmission after simply data processing, as shown in Figure.1.

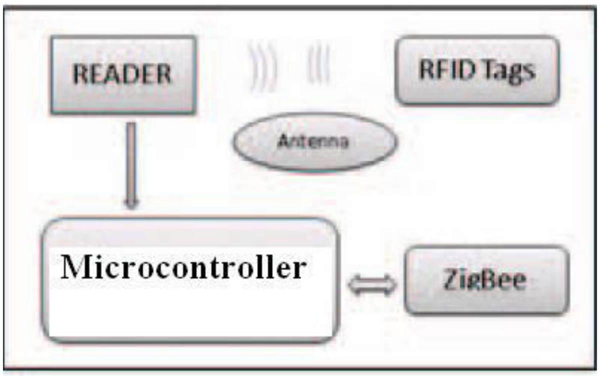


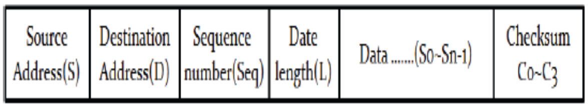
Fig. 1 : Sensor node schematic of RFID&ZigBee

Where ZigBee module provides two-way data transmission services and returns the corresponding control information after passing center node and PC processing. Control information is sent to micro-processor through ZigBee communication module. And the system can give the corresponding early warning tips if the ZigBee module has early warning capability (such as entrance guard).

B. Wireless transmission

The system uses ZigBee to build transmission network, which is used for the transmission of sensor data, and uses customized wireless transmission protocol, which is designed based on simplicity and reliability. In the protocol, considering simplifying microcontroller functions of RFID & ZigBee node and reducing system cost, the wireless transmission protocol mainly achieves the capabilities of error checking, data framing, conflict mechanisms such as retransmission, etc. As for error control, considering that tag data is usually short , the microcontroller of FRID & ZigBee nodes have strong logic computing power, so the original CRC operation is relatively complex and not suitable for the micro-processor. Therefore, this system innovatively uses new type of customized check method: uses and, or, xor operation for error control, along with length check, and insert frame boundary FLAG into the data frame when sending it.

Assuming the data to be sent is S_D~S and source address(one char), destination address(one char), sequence number(one char), date length(one char), checksum(four char) will be sent with data, as shown in Figure.2(ACK is the same as described above, data section is null, L=0).



Source Address(S)	Destination Address(D)	Sequence number(Seq)	Date length(L)	Data.......(So-Sn-1)	Checksum Co~C3

Fig. 2 : Wireless transmission protocol within the system

$C_1 = S \,\&\, D \,\&\, Seq \,\&\, L \,\&\, So \,\&\, ----. \,\&\, Sn\text{-}1$

$C_1 = S \mid D \mid Seq \mid L \mid So \mid ----. \mid Sn\text{-}1$

$C_1 = S \oplus D \oplus Seq \oplus L \oplus So \oplus ----. \oplus Sn\text{-}1$

Where the checksum section provides 4 check methods, besides and, or, xor operation, it allows use of customized method ,where 00 means null. The sequence number is mainly used to prevent the additional overhead of retransmission. In the transmitter, the following steps are taken when sending data.

1. Get the data from the upper layer packed, add source address, destination address, sequence number, date length, checksum to it, then send it to the serial output module of the microcontroller.
2. The serial output module adds FLAG to the data packet, if FLAG has been in the data, add ESC in front of it. After transmission, start the timer. Go to step 3.
3. Return true if receive ACK. If timeout happens, and timeout time is less than the maximum time, go to step 2; else, return false and inform the upper layer that transmission fails.

When the receiver receives data, perform the following steps:

1. Check the data packet according to the check operation said above, if check results are the same as those in the data packet, go to step 2; else, drop the packet.
2. See the destination node number of the data packet, if it's the same with its ID, go to step 3; else, drop the packet.
3. Return the corresponding ACK of the data packet, determine whether it's request for new packets or retransmitted data packets according to the sequence number of data packet, then send it to the upper layer for computing and handling

C. PCnode centre and multi-service platform

The tag information obtained by RFID&ZigBee node is transmitted through ZigBee network, and sent to PC node at last. In this system, PC node gets the data of the ZigBee network through serial ports, parses out the tag information through customized ZigBee protocol. PC node gives the matched results by querying the registered tag information in the database, and returns the corresponding results through the wireless module. When the incoming tag information doesn't match, PC node will give warning message, pass the warning and control message to entrance guard and inform the corresponding owner. In the multi-service platform, users can view the location of their valuables through multi-service platform interface. When valuables are lost, users can find the course of corresponding valuables through web, so as for tracking and getting it back. Multi-service platform only gives prompts warning case in this system, but in real life it can achieve anti-theft and monitoring work through functions reservation and system interface. Specific functions design is shown in Figure.3

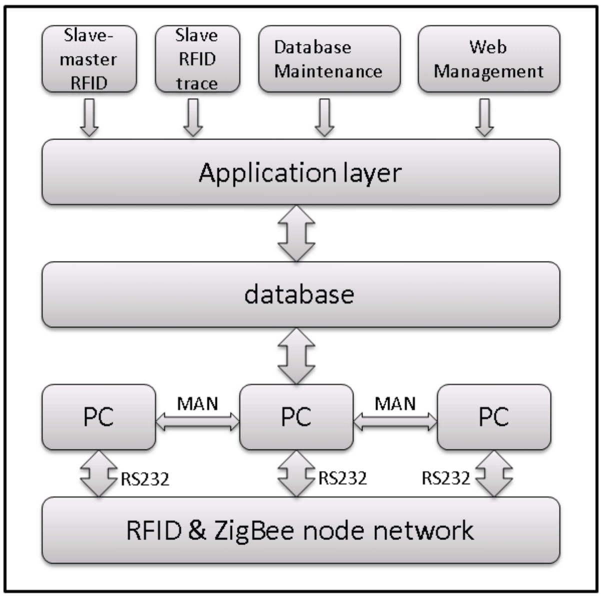


Fig. 3 : Structure of PC node centre service

As can be seen from Figure.3, each PC node is connected through Metropolitan area networks(MAN), thus improving the compatibility of the system greatly. When the RFID sensor network is more widespread, it will greatly improve the tracking range of items.

III. SYSTEM DESIGN AND IMPLEMENTATION

A. Layout of sensor network

Implement RFID & ZigBee sensor network layout within the campus, finish transmission through the ZigBee protocol, set those important nodes(such as the campus house door, the dormitory door)as PC nodes, which are database maintenance, WEB maintenance, RFID information processing servers. And all PC nodes are connected by Metropolitan area networks(MAN). After applying for services, user can get a RFID (Master) tag as a sign of the owner and multiple RFID (slave) tags as signs of valuables. Insert RFID (slave) tag in valuables, and this system can achieve the capability of anti-theft and track lost valuables within the scope of monitoring. Below is the campus sensor network layout in simulation environment. Set the PC-nodes in the dormitory, classroom building (there exists entrance guard at the location of PC nodes, when master RFID and slave RFID don't match, a warning will be sent.).In real life, RFID & ZigBee nodes should be set on the campus as evenly as possible.

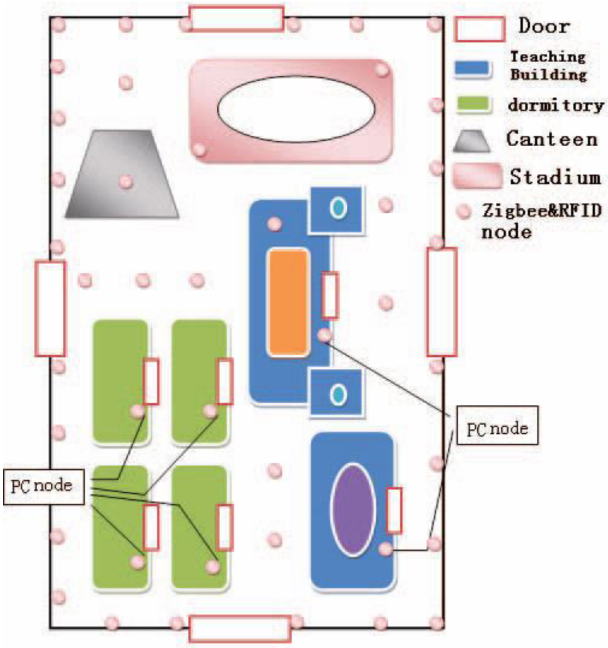


Fig..4 : *Campus sensor network layout*

As shown in Figure .4, there are ZigBee & RFID nodes and PC nodes on the campus, achieving the "Internet of Things" on the campus. PC nodes achieve the collection and processing of server information and database maintenance and query. There is alarm system in front of every door, once it finds that master RFID and slave RFID don't match, it'll give an alarm immediately and inform guards. As for the web form of information display platform, it's also expected to use the campus thumbnails, and words describe the specific location of searched valuables simultaneously.

B. System Architecture and flow chart

As shown in figure.5, the whole system includes two levels design: hardware and software. Hardware includes RFID readers and ZigBee wireless transmission module, and actually microcontroller of FRID & ZigBee nodes are responsible for connecting those two. On the one hand, read the RFID tag data within the RFID reader sensor area, and then send it to PC node in accordance with the wireless transmission protocol.

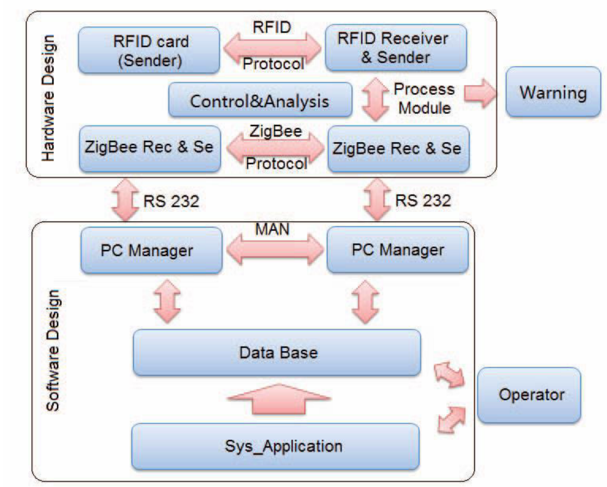


Fig. 5 System Architecture chart

On the other hand, microcontroller in the location of access control should receive message from the computer and give the prompt warning when mismatched RFID tag passes the system Overall, the workflow of the system can be described by the following steps.

1. RFID & ZigBee nodes sense the RFID (master/slave)labels, send information in the labels to the ZigBee network real time, which is then transferred to the PC nodes. PC checks the label information in the database, when master label and slave label match or only master label appears the system recognizes it as legal input. But when only slave label appears or master label and slave label don't match, the system recognize it as illegal input, then go to step 2.

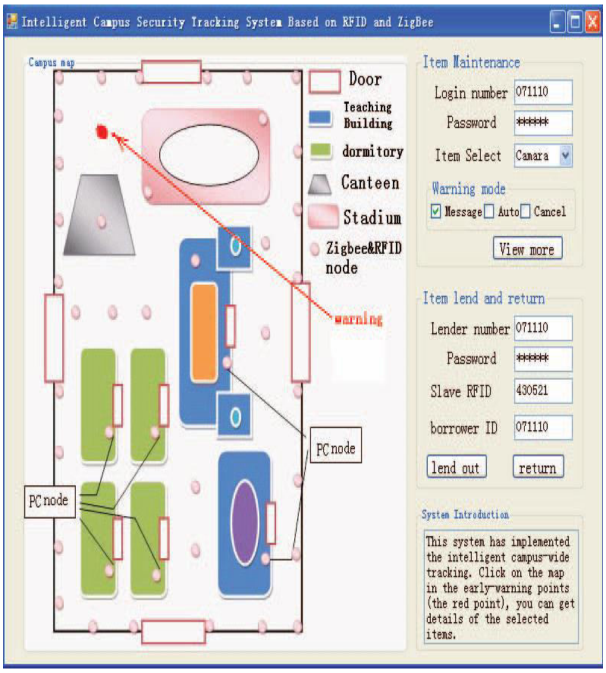
2. Record the position changes of RFID (slave) the tags, track the slave tags, and show warning at PC nodes. Then query the database, look for the owner of items and send confirming information to the owner through the system.

3. The owner logins WEB to search for items. He can see the real-time location of his valuables according to the hints. After confirmation, the valuables will be stopped at the entrance guard. As for false information, the owner can cancel this warning.

IV. TEST RESULT ANALYSIS

In practice, our system's sensor network uses Middle distance RFID reader, SZ05-STD wireless ZigBee, standard RS232, and the mc9s12xs128 of FreeScale as microcontroller of FRID & ZigBee node, .and uses Pentium(4) in the XP operation system to do a small-scale tracking and system verification. In order to save experiment cost, we use simulation method to achieve large-scale test and verification of our sensor network. The simulation in the experiment only replaces the process of reading the RFID tag information, and PC directly simulates the actual tag data acquisition process.

As shown in Figure .6, the system detects an illegal input, and produces early warning information. After receiving a reminder, user logs on the system and sees the following warning tips. From the figure, we can see that it's supported to lend his items to others to enable the daily use of items, so that this warning is cancelled. He can have his items back through this system, too.


Intelligent Campus Security Tracking System Based on RFID and ZigBee

Item Maintenance

Login number 071110

Password ******

Item Select Camara

Warning mode

☑ Message ☐ Auto ☐ Cancel

Item lend and return

Lender number 071110

Password ******

Slave RFID 430621

borrower ID 071110

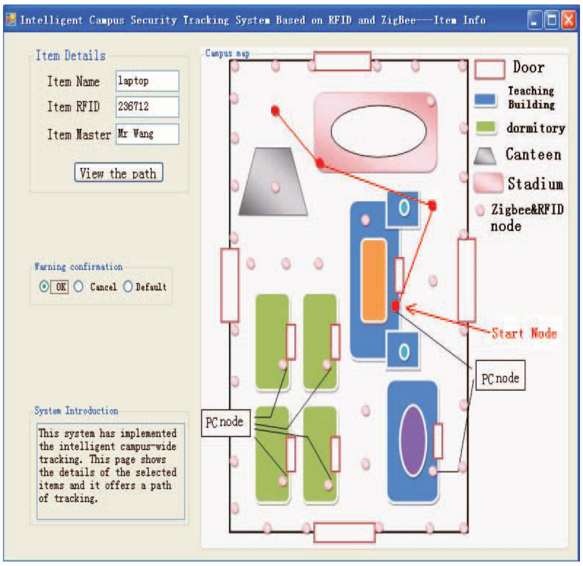
System Introduction

This system has implemented the intelligent campus-wide tracking. Click on the map in the early-warning points (the red point), you can get details of the selected items.

Fig. 6 : Valuables Management

Users click on the red dot Figure .6, enter the chart of tracking items. Users can choose to confirm or cancel this warning in the tracking figure. It's worth noting that, if the user does not confirm this early warning, system will check the identity of the suspect at the entrance guard to ensure campus safety, as is shown in Figure.7.

When the user chooses to confirm or default, valuables will be stopped and checked by the security at the entrance guard so as to achieve security and tracking purposes.



Intelligent Campus Security Tracking System Based on RFID and ZigBee---Item Info

Item Details

Item Name: laptop

Item RFID: 236712

Item Master: Mr Wang

View the path

Warning confirmation

● OK ○ Cancel ○ Default

System Introduction

This system has implemented the intelligent campus-wide tracking. This page shows the details of the selected items and it offers a path of tracking.

Fig. 7 : The course of valuables

V. CONCLUSION

This system studies campus safety and security. By identifying the tags information of master and slave RFID, it not only achieves tracking valuables real time and giving early warning, but also supports users to view the state of their valuables, and lend their valuables to others. As people's awareness of property safety improves, the system can also provide personnel tracking feature to support services related to geographical location. With the further popularization of sensor networks, the system can achieve large-scale logistics management and tracking, and has good market prospects and value in use.

REFERENCES:

[1] Zhu Yuan-jiao, Zhou Ke-qin, Design and Realizing of the Digital Campus Security System, Software Engineering, 2009. WCSE '09. WRI World Congress on IEEE 2009.

[2] Xi Li, Tiyan Shen, Jinjie Zhang, Changmin Shi, A Spatial Technology Approach to Campus Security, Networking, Sensing and Control, 2008.

ICNSC 2008.IEEE International Conference on IEEE 2008.

[3] Elshayeb, S.A., Bin Hasnan, K., Chua Yik Yen, RFID technology and ZigBee networking in improving supply chain traceability Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2009 IEEE International Conference.

[4] Floerkemeier C. , Sarma S. , An Overview of RFID System Interfaces and Reader Protocols, RFID, 2008 IEEE International Conference.

[5] Zhang Ye, Based on RFID technology jewelry and tracking management system Review, E-Business Journal,2010(9) (in Chinese).

[6] Kuan J.H. , Chang J. , Ho J. , A development of information protection system using system engineering and RFID technolog, System Science and Engineering (ICSSE), 2010 International Conference on IEEE 2010.

[7] Chung-Hsin Liu, Jian-Yun Lo, The Study for the ZigBee with RFID Positioning System, Multimedia and Information Technology (MMIT), 2010 Second International Conference on 2010.

[8] Yu Li-e, Deng Xu-dong, The Research about the Application of RFID and 3G Technology in Cargo Transportation Security, Logistics Sci-Tech, 2007(10) (in Chinese).

◈◈◈

Routing in Zigbee Mobile Wireless Networks using Mobility Robust Trees

Hariram R M, Manjunatha S & Jitendranath Mungara

Dept of CSE, CMR Institute of Technology, Bangalore
Email:hariram934@gmail.com, manju02@gmail.com & jmungara@yahoo.com

Abstract - In Wireless Personal Area Networks, the Zigbee protocol as formalized by the IEEE 802.15.4 standard is a specification for low data rate, less cost and low power. The Zigbee network is usually constructed using cluster trees for the purpose of performing data delivery applications among nodes and for power saving. Here the data delivery failures occur due to node movement and topology changes of networks. Route reconstruction in these networks requires high resource utilization. In order to handle topology changes and node movements and to increase data delivery we use mobility-robust tree construction technique. In this paper we utilize the regularity of mobility patterns to reduce the frequency of route reconstructions and achieve higher efficiency in sending data to mobile nodes. The entire setup is developed and simulated by using NS2 network simulator. The result is that we obtain a mobility-robust tree with both uplink and downlink data transfers with a considerable increase in data delivery.

Keywords: *Zigbee wireless networks, Mobility robust trees, Routing.*

I. INTRODUCTION

Zigbee wireless networks are a class of networks which occupy the low power, low cost region in the wireless networks hierarchy. The Zigbee standard initialized by the Zigbee Alliance [3], specifies the network and application layers for sensing data delivery.

There are various applications of Zigbee wireless networks in the real world. Applications include such as wireless light switches, electrical meters with in-home-displays, and other consumer and industrial equipment that require short-range wireless transfer of data at relatively low rates. These applications have increased in the past decades as a result of the widespread growth in wireless communication and sensing [2].

Zigbee is targeted at radio-frequency (RF) range applications that require a low data rate, long battery life, and secure networking. Zigbee has a defined rate of 250 kbps, has a frequency band range of 2.4GHz and supports up to 16 channels, best suited for periodic or intermittent data or a single signal transmission from a sensor or input device. Zigbee based traffic management system have also been implemented. The Zigbee network layer natively supports star, mesh, and tree topology networks. The following figure shows the Zigbee protocol stack.

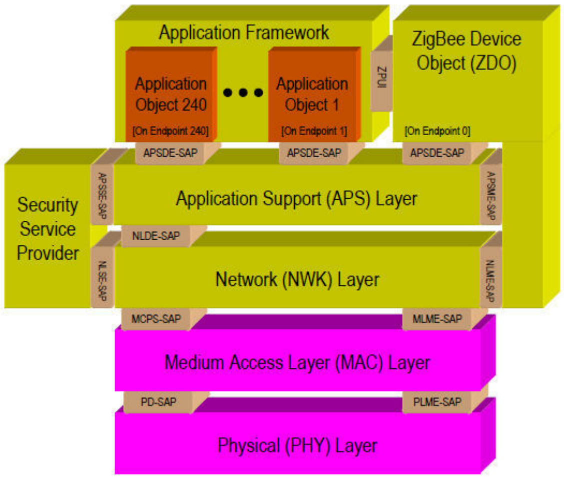


Figure 1: Architecture of Zigbee Protocol Stack

We see in the above figure that Zigbee builds upon both the physical layer and medium access control layers as specified in the IEEE 802.15.4 standard.

There are several types of wireless networks including wireless local area networks (WLANs), mobile ad hoc networks (MANETs), Bluetooth, WiFi etc., and all these different networks have their respective ways for efficient data delivery and handling the mobile nodes in their network. For e.g., cellular networks use the handoff strategy to switch the control of cellular or mobile devices from one base station to another. Similarly WLANs use mobile IP to configure mobile nodes in multiple LAN sub-networks. MANETs use a special multicast protocol which is adaptive to dynamic network topologies and resources.

These different strategies and protocols are specific to their type of wireless network. This causes a problem for Zigbee based networks. Because the Zigbee networks are unique in their own way, by having low power, low cost etc, there has been several research experiments conducted regarding the mobility patterns and issues present in Zigbee networks. Our researchers have finally come up with a unique characteristic method called "mobility-robustness" [1].

In this paper, we discuss the mobility-robust tree construction in Zigbee networks with relevant enhancements, so as to further increase the effectiveness of this technique. According to this technique, the focus is on improving the downlink data delivery ratio, by collecting information about the mobility or movement patterns of the mobile nodes across the network, and utilizing this information for constructing mobility-robust trees. The aim of this approach is to construct a tree topology such that mobile nodes move into their data forwarding path with a high probability. These mobile nodes, while they are moving across the network can send and receive data from routers as long as they follow their highly probable data path. The initially available Zigbee tree network is refined using graph optimization so as to form a more useful and compact network for faster and reliable robust data delivery. We use NS2 simulator for setting up the environment and simulating the "mobility-robust" tree structure.

II. SYSTEM MODEL

In a Zigbee network there are three types of devices: the Zigbee Coordinator(ZC), Zigbee Router (ZR) and Zigbee End Device (ZED). All these devices follow a certain hierarchy model in the network. Basically a Zigbee network is formed by one Zigbee coordinator and multiple Zigbee routers and Zigbee end devices. The following is a brief description [3] about their roles and functions in the network. They are:

i. Zigbee Coordinator (ZC)

A Zigbee Coordinator performs the task of initializing, maintaining and controlling the other nodes in the network. There is only one ZC node for every Zigbee network. It also acts as a router once a network is formed. It is not necessarily a dedicated device, can also perform application tasks.

ii. Zigbee Router (ZR)

It is used for storing and forwarding (routing) data between the ZED (end device/mobile node) and ZC (coordinator).It manages the local address allocation/de-allocation. It participates in multi-hop routing of messages.

iii. Zigbee End Device (ZED)

They are the mobile nodes which discover and associate with ZC or ZR. They can be optimized for very low power operation.

Though we can form Zigbee based star or mesh networks, we still opt for cluster topology because it provides for power saving and supports a very light-weighted protocol without the need to maintain a routing table.

The following figure shows a Zigbee Network.

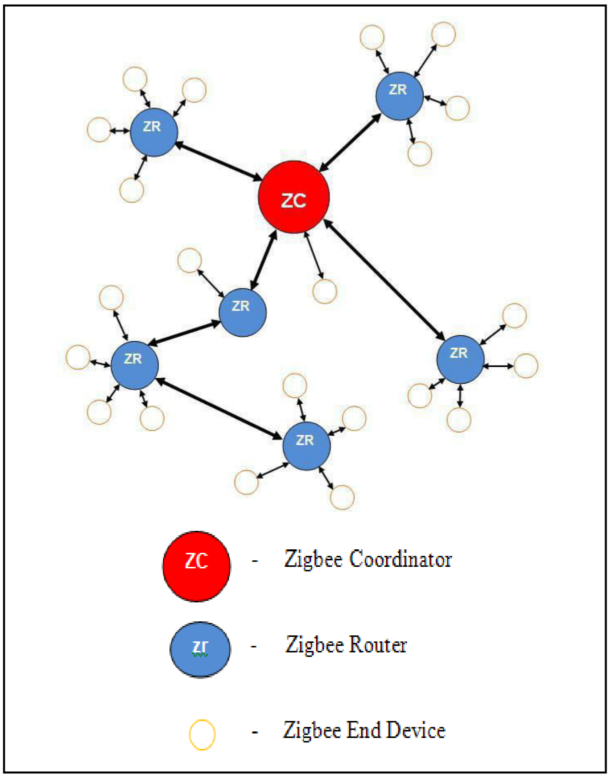


Figure 2: Zigbee Network Structure

Address assignment is different from that done in a conventional Zigbee network. Here every mobile end device in our network is assigned a random address, so as to uniquely identify every mobile node in our network.

III. RELATED WORK

There is a constraint put on the maximum number of child routers of a router/the coordinator (R_m) and the depth of the network (L_m).Whenever a mobile end device has a packet to send, it just sends it. On receiving a packet, the router forwards it to its parent in the network

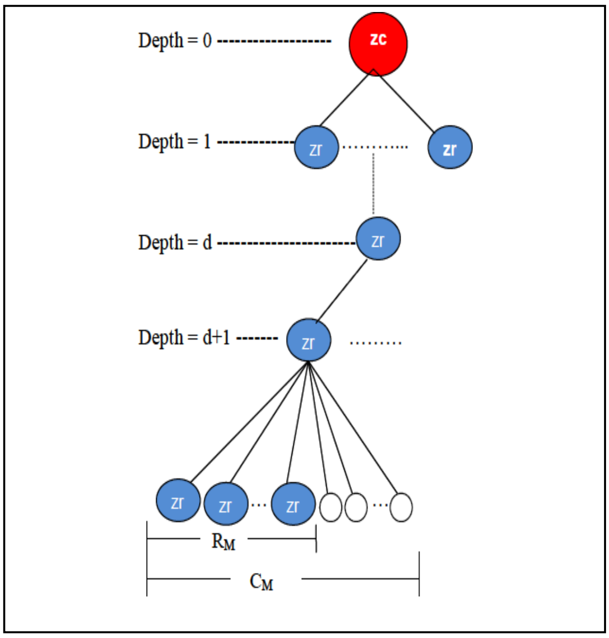


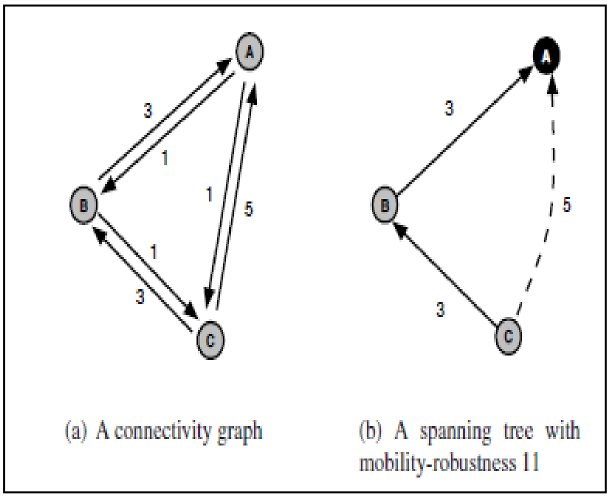
Figure 3: Tree structure of Zigbee

When a downlink packet is destined to a mobile end device, the coordinator delivers the packet by using the previously recorded location (i.e., the last router receiving the packet from the mobile end device) as the destination. Upon receiving the downlink packet, the router then simply sends it out, and expects to receive an acknowledgement from the mobile end device. If the destined mobile node has already moved out of its previous location, the data delivery fails, and the system searches for the mobile node by sending a broadcast message to inquire its current location.

Mobility-robust tree construction can be formulated as a graph problem, in which a vertex represents an immobile node, i.e., the coordinator or a router, and a directed edge represents a possible transmission link from one immobile node to another. That is, a Zigbee network is represented as $G = (V, E)$, where V is a set of immobile nodes and E is a set of transmission links in the network. With the movement historical data collected among immobile nodes, each edge $e = (u,v) \in E$ is associated with a weight, $W(e)$, which represents the number of transitions of all mobile nodes moving from the transmission range of immobile node u to that of v in the collected data. For any directed edge $e = (u,v)$, there exists a directed edge $e = (v,u) \in E$ in the reverse direction.

The weights of these edges are non-negative. Our method is to construct a Zigbee cluster tree T *in* the bi-directed weighted graph G. In the edge $e = (v,u)$, the node v is the parent of node u. Movements from u to v follows the upstream of down-link data forwarding from v to u. Our objective is to maximize the total counts of movements toward the upstream of data forwarding paths, so that we can minimize missed data deliveries caused by mobile device mobility.

To achieve this objective, we define the mobility-robustness of the constructed tree T as the sum of the weights in the chosen directed edges, $W(e)$, $\forall \bar{e} \in T$ and the weights of those unselected edges, $W(e)$, $e /\in T$, which connect all descendants-to parent pairs of vertices in the same branch in T.



(a) A connectivity graph (b) A spanning tree with mobility-robustness 11

Figure 4: Example of mobility-robustness of a tree

IV. PROPOSED APPROACH

The proposed approach [1] involves constructing the maximum-mobility robust tree in the following manner:

Instance: An instance includes a bi-directed graph

$G = (V,E)$ with edge weight $W(e) >= 0$, $\forall e \in E$, and two positive constraint integers R_m and L_m.

Objective: The objective is to find a rooted spanning tree T in G such that the mobility-robustness of T is maximized among all possible trees in G. Also, the out-degree of every vertex in T does not exceed R_m, and the depth of T does not exceed L_m.

A. Assumptions:

1) We assume that the coordinator maintains the location of the mobile end device when it sends an uplink data packet to the coordinator.

2) In order to adapt to quick topology changes there is no association mechanism between mobile end devices and routers.

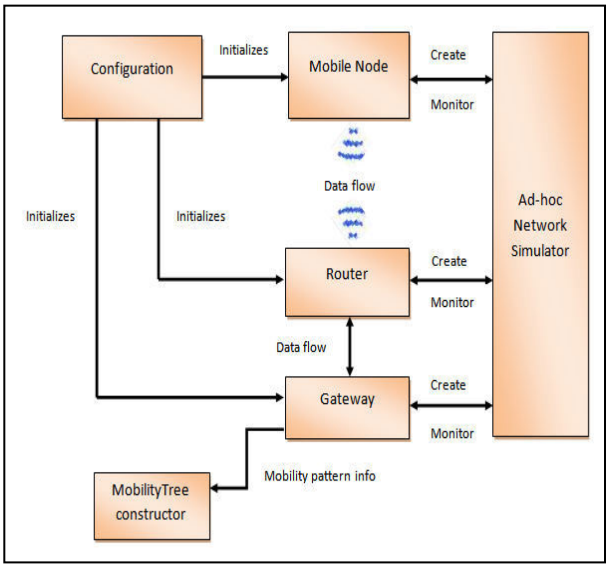


Figure 4: System Architecture

There are six modules in the system architecture diagram. They are:

- The Configuration
- Gateway (Zigbee Coordinator)
- Router(Zigbee Router),
- Mobile Node(Zigbee End Device),
- Mobility Tree Constructor and
- Ad-hoc Network Simulator.

The modules work as follows:

User can configure the number of nodes in the simulation and their mobility movement. Mobile Nodes are created by Configuration module. Mobile Node uses the Ad-hoc Network Simulator to communicate with other Mobile Nodes. Mobile Node move from one router to another. Gateway Node keeps track of mobility pattern and constructs the mobility tree. It uses the mobility tree to deliver the messages to the nodes i.e., messages are delivered along the mobility tree.

B. Algorithms Used:

There are two algorithms used in the construction of mobility-robust Zigbee trees [1]. They are: the ZTG phase algorithm and the FIX phase algorithm.

1) **ZTG phase**: The ZTG phase searches and connects the vertices that add the most mobility-robustness to the tree. The output of this phase is a forest of trees which is sent to the next phase/algorithm.

2) **FIX phase**: The FIX phase merges the trees constructed in the ZTG phase. These trees are directed trees, which mean that we can only connect one tree to another by their roots.

The result after applying these two algorithms is a Maximum-mobility robust tree with reliable data flow.

V. SIMULATION

NS2 simulator is used to develop the environmental setup with a set of parameters as shown below:

Specification	Setting
Network Standard	IEEE 802.15.4
Deployment Environment	100m × 60m Rectangle with Walls
Antenna Type	F-shaped Antenna
Frequency	2.4GHz ISM Band
Data Rate	250Kbps
Media Access Control	CSMA/CA
Propagation Model	TwoRayGround
Transmission Power	0dBm
Receiver Sensitivity	-94dBm
Average Communication Range	48m
Average Carrier Sense Range	120m
Traffic Pattern	Poisson Destination
Packet Size	70 bytes
Packet Inter-arrival Rate	0.1 packets/second per node

Table 1: Parameter settings in NS2

The input for the graphics is stated above. The simulation tool used is NS 2.34. It provides substantial support for simulation of TCP, routing and multicast protocols. The simulator is written in C++ and script language is OTcl. The user writes the script in OTcl to define a network (nodes and links), the traffic in the network (source, destination and type of traffic) and which protocols it will use. The results of the simulations are an output trace file that can be used to data processing and to visualize simulations with a program called Network Animator (NAM) [9, 10].

VI. CONCLUSION

We have used the concept of mobility-robustness of a tree topology. We have come out with a method to increase the data delivery ratio of the constructed Zigbee cluster tree by using the high probability data forwarding path information of the mobile nodes so that both uplink and downlink data can be sent and received by the mobile nodes, in accordance with the working of coordinator and router nodes.

REFERENCES

[1] Wei-Ho Chung and Pi-Cheng Hsiu, Yuan-Ya Shih, Ai-Chun Pang, Yu-Kai Huang, and Kuan-Chang Hung, “Mobility-Robust Tree Construction in Zigbee Wireless Networks”, IEEE conference paper 2011.

[2] Wikipedia (http://www.wikipedia.com)

[3] ZigBee Alliance (http://www.zigbee.org) 2006.

[4] I.F. Akyildiz, J. Xie, and S. Mohanty. “A survey of mobility management in next-generation all-ip-based wireless systems”. In IEEE Wireless Communications, volume 11, pages 16–28, 2004.

[5] Y.K. Huang et al. “An integrated deployment tool for Zigbee-based wireless sensor networks”. In IEEE/IFIP EUC, volume 1, pages 309–315, 2008.

[6] I.F. Akyildiz, W. Su, Y. Sankara subramaniam, and E. Cayirci. “A survey on sensor networks.” In *IEEE Communications Magazine*, volume 40, pages 102-114, 2002.

[7] Yu-Kai Huang, Ai-Chun Pang, Pi-Cheng Hsiu, WeihuaZhuang, and Pangfeng Liu’s ”Distributed Throughput Optimization for Zigbee Cluster-Tree Networks” IEEE transactions on parallel and distributed systems,2011

[8] B. An and S. Papavassiliou.“A mobility-based clustering approach to support mobility management and multicast routing in mobile ad-hoc wireless networks”. In *International Journal of Network Management*, volume 11, 2001.

[9] Teerawat Issariyakul and *Ekram* Hossain “Introduction to Network Simulator NS2” 2009. [11] Teerawat Issariyakul and *Ekram* Hossain “Introduction to Network Simulator NS2” 2009.

[10] De los Andes, Merida, venezuela and ESSI “NS simulator for beginners” Lecture notes: 2003-2004.

◈◈◈

FPGA Implementation of Hummingbird Cryptographic Algorithm

Sridevi Galagali & Renuka Korti

Dept of Electronics and communication Engineering, SDM College of Engineering and Technology
Dharwad-580 002, Karnataka, India
E-mail : Shridevi12@gmail.com & rhh_korti@yahoo.com

***Abstract* -** Hummingbird is a new ultra-lightweight cryptographic algorithm targeted for resource-constrained devices like RFID tags, smart cards, and wireless sensor nodes. In this paper, we describe efficient hardware implementations of a stand-alone Hummingbird component in field-programmable gate array devices. We implement an encryption only core and an encryption/decryption core on the low-cost Xilinx, Virtex5 and compare our results with other reported lightweight block cipher implementations on the same series. Our experimental results highlight that in the context of low-cost FPGA implementation Hummingbird has favorable efficiency and low area requirements.

Keywords*—Lightweight cryptographic primitive, resource constrained***.**

I. INTRODUCTION

Hummingbird is a recently proposed ultra-lightweight cryptographic algorithm targeted for low-cost smart devices like RFID tags, smart cards, and wireless sensor nodes [3].

Hummingbird enciphers a plain-text block which is the \width" of the rotors, e.g. Hummingbird enciphers sixteen-bit-wide words, or two ASCII characters at a time and produces a cipher block sixteen bits wide. Thus each rotor is a substitution table for all possible sixteen-bit words, and so is effectively 65,536 entries long. It has a hybrid structure of block cipher and stream cipher and was developed with both lightweight software and lightweight hardware implementations for constrained devices in mind. Moreover, Hummingbird has been shown to be resistant to the most common attacks to block ciphers and stream ciphers including birthday attack, differential and linear cryptanalysis, structure attacks, algebraic attacks, cube attacks, etc. [3]. In practice, Hummingbird has been implemented across a wide range of different target platforms [3], [5]. Those implementations demonstrate that Hummingbird provides efficient and flexible software solutions for various embedded applications. However, the hardware performance of Hummingbird has not yet been investigated in detail. As a result, our main contribution in this paper is to close this gap and provide the first efficient hardware implementation of Hummingbird encryption/decryption cores on low cost FPGAs.

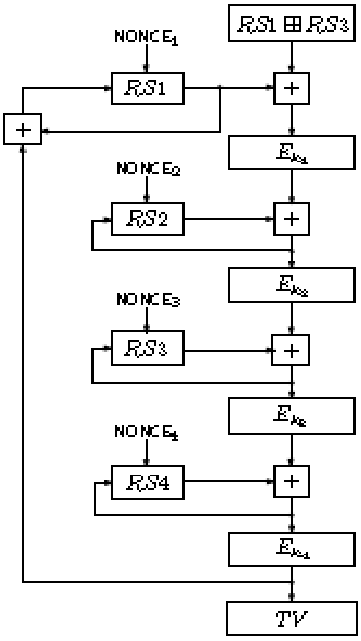
II. THE HUMMINGBIRD CRYPTOGRAPHIC ALGORITHM

Hummingbird is neither a block cipher nor a stream cipher, but a rotor machine equipped with novel rotor-stepping rules. The design of Hummingbird is based on an elegant combination of a block cipher and stream cipher with 16-bit block size, 256-bit key size, and 80-bit internal state. Hummingbird is neither a block cipher nor a stream cipher; it is a rotor machine, both in structure and in function. That is, it operates on individual plain-text words to produce individual cipher-text words. A “word” in this case is a 16-bit chunk, i.e. two ASCII characters. Here is not the place to discuss the myriad modes of using stream or block ciphers suffice it to say that none of them can fully duplicate the operation of a rotor machine, and that, for certain applications, the differences are important.

Hummingbird's encryption procedure resembles that of classic rotor machines, each word is acted upon by each of four rotors, in succession each action depends on both the particular rotor and its rotor-setting at that time. The final action produces a word of cipher text. But the resemblance to physical rotor machines goes only so far, Hummingbird has some significant departures from typical practice. Most conspicuously central to the design is a particular pattern of using successive internal states to drive the “motion” of the rotors. The rotors, as usual, execute permutations, like what are often named “S-boxes”,but here these are extraordinarily large ones since they operate on 16-bit words. In the usual cryptographic implementations of S-boxes, such a size would be close to impossible to use. In Hummingbird it is made practical and efficient, both in generation and in implementation, by substituting for each rotor a special-purpose 16-bit block cipher.

A. Initialization

A nonce is used to initialize the rotors. The initialization process is illustrated in Figure 1. When using Hummingbird in practice four 16-bit random nonce NONCE*i* are first chosen to initialize the four internal state registers *RSi* ($i = 1; 2; 3; 4$), respectively, followed by four consecutive encryptions on the message $RS1 \boxplus RS3$ by Hummingbird running in the initialization mode (Figure 1). The final 16-bit ciphertext *TV* is used to initialize the LFSR. Moreover, the 13th bit of the LFSR is always set to prevent a zero register. The LFSR is also stepped once before it is used to update the internal state register RS_3. The exact initialization procedure as well as the internal state updating of Hummingbird are illustrated in below



B. Encryption Process

The basic encryption of a 16-bit word is accomplished by first adding the word modulo 2^16 to the rotor setting and then encrypting that sum with the first rotor. This process is repeated in a similar manner by passing the result of each rotor through the remaining rotors in turn. Then each rotor is stepped in various ways based on the states of different parts of the machine. The encryption process is illustrated in Figure 2.

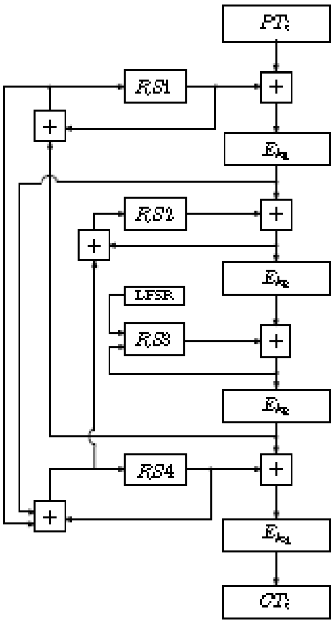


Figure2. Encryption Process

$PT_{i:}$ the *i*-th plaintext block, $i = 1; 2; : : : ; n$

$CT_{i:}$ the *i*-th ciphertext block, $i = 1; 2; : : : ; n$

K: the 256-bit secret key

Rs1: the *i*-th 16-bit internal state register, $i = 1; 2; 3; 4$

$E_k(.)$: the encryption function of Hummingbird with 256-bit secret key K.

$D_k(.)$:the decryption function of Hummingbird with 256-bit secret key KLFSR: a 16-stage Linear Feedback Shift Register with the characteristic polynomial

$$f(x) = x^{16} + x^{15} + x^{12} + x^{10} + x^{7} + x^{3} + 1$$

C. Decryption Process

Decryption is similar to encryption and is shown in the Figure 2. The additional term used in this figure is F^{-1}which is the inverse permutation, i.e., FF^{-1} = identity, for each of the particular instances of F

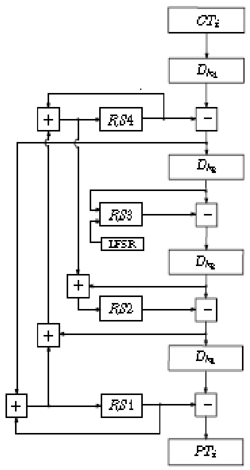


Figure3.Decryption Process

C. 16 bit block cipher

Four identical 16-bit block ciphers are employed in a consecutive manner in the Hummingbird encryption scheme. The 16-bit block cipher is a typical substitution permutation (SP) network with 16-bit block size and 64-bit key as shown in Figure 4. It consists of four regular rounds and a final round that only includes the key mixing and the S-box substitution steps. The 64-bit subkey *ki* is split into four 16-bit round keys $K_1(i)$, $K_2(i)$, $K_3(i)$ and $K_4(i)$ which are used in the four regular rounds, respectively. Moreover, the final round utilizes two keys $K_5(i)$ and $K_6(i)$ directly derived from the four round keys. Like any other SP network, one regular round comprises of three stages: a key mixing step, a substitution layer, and a permutation layer. For the key mixing, a simple exclusive operation is used in this 16-bit block cipher for efficient implementation in both software and hardware. The substitution layer is composed of 4 Serpent-type S-boxes [1] with 4-bit inputs and 4-bit outputs. The permutation layer in this 16-bit block cipher is given by the linear transform

$$L : \{0;\ 1\}16 \rightarrow \{0;\ 1\}16$$

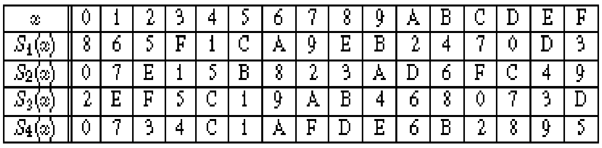
defined as follows:

$$L(m) = m \oplus (m \ll 6) \oplus (m \ll 10);$$

where $m = (m0;m1; \cdots ;m15)$ is a 16-bit data block.

Remark 1. To further reduce the consumption of the memory, are and power of Hummingbird in both software and hardware implementations, four S-boxes used in Hummingbird can be replaced by a single S-box, which is repeated four times in the 16-bit block cipher. The compact version of Hummingbird can achieve the same security level as the original Hummingbird and will be implemented on wireless sensor nodes in this paper.

Table1
Four S-boxes in Hexadecimal Notation



x	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$S_1(x)$	8	6	5	F	1	C	A	9	E	B	2	4	7	0	D	3
$S_2(x)$	0	7	E	1	5	B	8	2	3	A	D	6	F	C	4	9
$S_3(x)$	2	E	F	5	C	1	9	A	B	4	6	8	0	7	3	D
$S_4(x)$	0	7	3	4	C	1	A	F	D	E	6	B	2	8	9	5

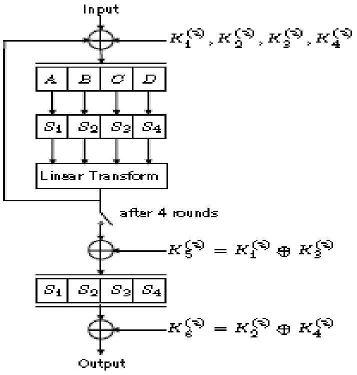

Input
$K_1^{(i)}, K_2^{(i)}, K_3^{(i)}, K_4^{(i)}$
A
B
C
D
S_1
S_2
S_3
S_4
Linear Transform
after 4 rounds
$K_5^{(i)} = K_1^{(i)} \oplus K_3^{(i)}$
S_1
S_2
S_3
S_4
$K_6^{(i)} = K_2^{(i)} \oplus K_4^{(i)}$
Output

Figure 4: 16-bit block cipher

III. RESULTS

A. Encryption

Device utilization summary:

Virtex5,xc5vlx110t-1ff1136

Table2

Logic Utilization	**Used**	**Available**	**Utilization**
Number of slices registers	79	69120	0%
Number of slice LUTs	776	69120	1%
Number of fully used LUT-FF pairs	63	792	7%
Number of bonded IOBs	11	640	1%
Number of BUG/BUFGCTRLs	1	32	3%

B. Decryption

Device utilization summary

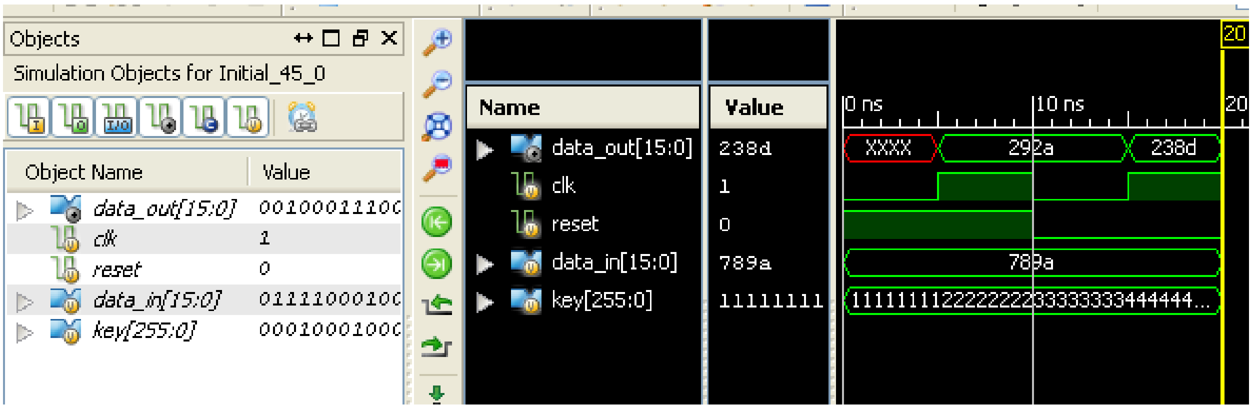
Table3

Logic Utilization	Used	Available	Utilization
Number of slices	79	69120	0%
Number of slices LUTs	700	69120	1%
Number of fully used LUT-FF pairs	56	723	7%
Number of bonded IOBs	11	640	1%
Number of BUG/BUFGCTRLs	1	32	3%

.

Simulation results:

A. Encryption



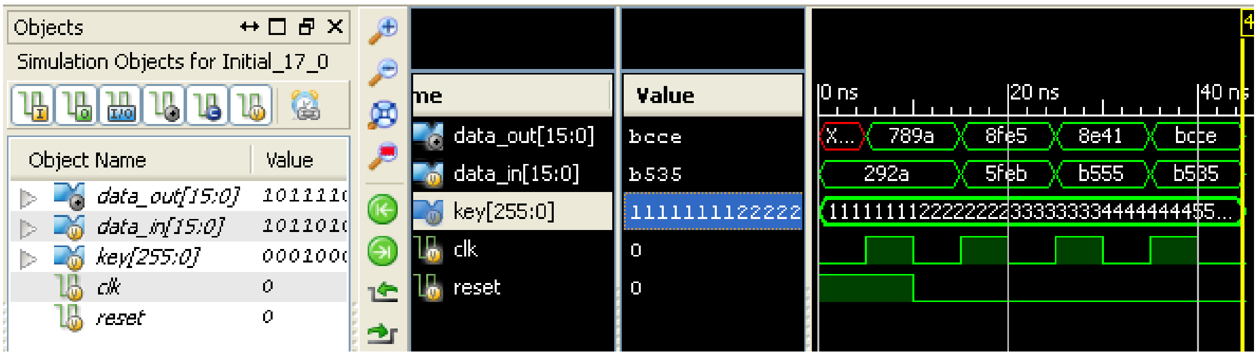
Objects

Simulation Objects for Initial_45_0

Object Name	Value
data_out[15:0]	00100011100
clk	1
reset	0
data_in[15:0]	01111000100
key[255:0]	0001000100

Name	Value
data_out[15:0]	238d
clk	1
reset	0
data_in[15:0]	789a
key[255:0]	11111111

B. Decryption



Objects

Simulation Objects for Initial_17_0

Object Name	Value
data_out[15:0]	101111(
data_in[15:0]	101101(
key[255:0]	000100(
clk	0
reset	0

me	Value
data_out[15:0]	bcce
data_in[15:0]	b535
key[255:0]	111111112222
clk	0
reset	0

.

IV. CONCLUSION

In this paper we present a novel ultra-lightweight cryptographic algorithm, Hummingbird, which is a combination of block cipher and stream cipher. This paper presented the first efficient FPGA implementations of the ultra-lightweight cryptographic algorithm Hummingbird. The proposed speed optimized Hummingbird encryption/ decryption cores can encrypt or decrypt a 16-bit message block with 4 clock cycles, after an initialization process of 20 clock cycles. Compared to other lightweight FPGA implementations of block ciphers XTEA, ICEBERG, SEA and AES, Hummingbird can achieve larger throughput with smaller area requirement. Consequently, Hummingbird can be considered as an ideal cryptographic primitive for resource-constrained environments.

ACKNOWLEDGMENT

The authors wish to thank the authorities of SDMCET Dharwad, for facility extended to carry out the present work.

REFERENCES

1. P. Bulens, F.-X. Standaert, J.-J. Quisquater, and P. Pellegrin, “Implementation of the AES-128 on Virtex-5 FPGAs”, *Progress in Cryptology - AFRICACRYPT 2008*, LNCS 5023, pp. 16-26, 2008.

2. P. Chodowiec and K. Gaj, “Very Compact FPGA Implementation of the AES Algorithm”, *The 5th International Workshop on Cryptographic Hardware and Embedded Systems - CHES 2003*, LNCS 2779, pp. 319- 333, 2003.

3. D. Engels, X. Fan, G. Gong, H. Hu, and E. M. Smith, "Hummingbird: Ultra-Lightweight Cryptography for Resource- Constrained Devices", to appear in the proceedings of *The 14th International Conference on Financial Cryptography and Data Security - FC 2010*, 2010.

4. N. N. Espresso. Available at http://embedded.eecs.berkeley.edu/pubs/ downloads/espresso/index.htm, November 1994.

5. X. Fan, H. Hu, G. Gong, E. M. Smith and D. Engels, "Lightweight Implementation of Hummingbird Cryptographic Algorithm on 4-Bit Microcontrollers", *The 1st International Workshop on RFID Security and Cryptography 2009 (RISC'09)*, pp. 838-844, 2009.

6. T. Good and M. Benaissa, "AES on FPGA from the Fastest to the Smallest", *The 7th International Workshop on Cryptographic Hardware and Embedded Systems - CHES 2005*, LNCS 3659, pp. 427-440, 2005

◈◈◈

The Map Reduce Approach to Detect Malwares Retrospectively in Cloud Network

Rashmi S & Mahendra Pratap Singh

Dept. of Computer Engineering,National Institute of Technology Karnataka, Surathkal 575025, India
E-mail : rashmi.sambrama@gmail.com, mahoo15@gmail.com

Abstract - Nowadays malwares are widely spreading in the internet and there is drastic increase in its negative impact on computers. From the security point of view, it is required to detect the malwares and affected ones, as early as possible. Enterprises today collect and generate more data than ever before. Hence it is required to safeguard the data from these malware attacks. Although it will be difficult to process the huge amount of data, with the evolution of cloud computing technologies, this aspect can be achieved.

In this paper, we propose a new mapreduce technique , to detect malwares in retrospective fashion, using Hadoop platform. It is used to track the attack based on process log information from the computers in the network. We are using the large number of small files to process the data in hadoop platform. We are using the 7 real world malwares which affects the removable drives. As a result, we are able to detect and report the affected computers by obtaining the process-count information effectively, using mapreduce technique.

Keywords— *malware, cloud computing, Hadoop, mapreduce , retrospective detection, removable drives.*

I. INTRODUCTION

Malware is a malicious software that usually installs itself on user's computer unknowingly. Malwares are created for various purposes, these include intrusion of privacy for various unethical reasons, destroying the registry, vandalism, crimes, spying or just for prank .Their purpose is to monitor and deliver information regarding user's computer habits to the remote monitoring agents. The process results in endless pop up ads displaying on user's computer screen which finally slows down the computer performance.

Retrospective Detection is a mechanism which provides the ability to track the attack based on historical logs by specifying files or registry keys. Since the cloud computing technique provides both time and storage intensive mechanism for large datasets, the malware detection can be achieved on a large scale.

In spite of having anti-malware softwares and remedial measures, due to day to day increasing activities of evolving malwares, it has become difficult to combat the malware attacks. There are many prevalent vectors through which malwares can transmit to the information systems. We believe that malwares can enter the systems only while installing or updating the softwares. Hence, we will install a system process monitor tool in all the client systems, run to collect the log information and send it to the server to keep track of the malware affected processes in the network.

Hadoop [6] is an open-source implementation of Google's MapReduce for reliable, scalable, distributed computing and data storage. It can be used for running applications on large clusters of commodity hardware. Hadoop implements a computational paradigm named map/reduce, where the application is divided into many small fragments of work, each of which may be executed or re-executed on any node in the cluster. The approach on Hadoop and mapreduce are used to prove effectiveness and efficiency in retrospective detection of malware attacks.

II. PROBLEM STATEMENT

A. *System Model*

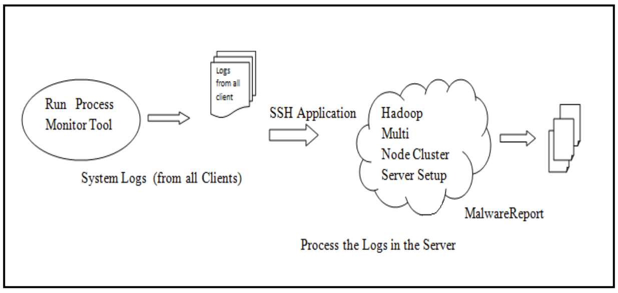

Run Process Monitor Tool
Logs from all client
System Logs (from all Clients)
SSH Application
Hadoop Multi Node Cluster Server Setup
MalwareReport
Process the Logs in the Server

Fig 1 : Block Diagram of the System Model

In this system, we develop a Hadoop Cluster Setup to implement the retrospective approach to detect malware attacks. Here the focus is on two aspects, one is the malware detection using retrospective approach and the other is implementation of the approach in the cloud computing using mapreduce technique in distributed systems. Hence the system is designed to collect the logs from each client systems in the network by invoking the *Process Monitor* [10] tool and send the logs to the Hadoop server setup. The logs are sent using *SSH Secure File Transfer Client* application. The *Hadoop Server Setup* are meant to process the input data logs to index files. These index files are used for malware detection and helps in finding out which host systems are affected by the malware processes.

III.EXPERIMENTAL SETUP

A. *System Configuration*

To evaluate the effectiveness of our approach, we have collected 7 real-world malwares, as shown in Table 1. It consists of worms and Trojans/virus. Client computer will run Process Monitor tool on a Windows XP Professional computer with a 2.80 GHz Pentium CPU and 1GB RAM . The Server is setup in master slave model, one machine is used as Hadoop master and rest are used as Hadoop slaves, ran on Ubuntu 10.04, Hadoop 0.20.2 with 2.80 GHz Pentium CPU and 1GB RAM.

B. *Work Flow*

Considering the system configuration which is explained in the section 3.1, we have built the experiment setup ,as follows.,

- The 3 client systems are considered. Each client computer has installed a monitor agent viz *Process Monitor* to collect the Portable Executable format file created logs and send them to Hadoop server.
- The log files generated are uploaded to the server by invoking SSH Secure Shell Client application.
- Hadoop Multinode Cluster Setup is created using 7 linux machines where hadoop 0.20.2 is running in all the machines.
- The Process index and ProcessCount index mapreduce techniques are used to read the logs to build specific indexes for retrospective detection.
- Malware Search and Process Count mapreduce techniques are used to search and report the specified malware to the user and reveals that which all hosts have been affected by the malware processes in the network

Table 1: Malwares Used for Test

Type	Malware
Trojan/Backdoor virus	astry.exe
Trojan virus	Trojan.Falupan(Winlogon.exe)
Trojan virus	GameThief.Win32.Tibia virus
Trojan virus	fotoku.exe
Worm	Recycler.exe

IV. BASIC IDEA BEHIND PROPOSED ALGORITHMS FOR MALWARE DETECTION IN CLOUD NETWORK

A. *Existing Methods*

In Splunk [7] is an Information Technology, search engine for log management, operations management, security. It traces the malware by the registry key and file name. In “Retrospective Detection of Malware Attacks by Cloud Computing" [6], a retrospective detection approach based on Portable Executable (PE) format file relationships has been proposed. They implemented the system in a Hadoop platform and Lucene[5] which is a high-performance, scalable Information Retrieval (IR) library is used as core data structure and used 18 real-world malware to do the tests. Their approach has a higher detection rate as well as a lower false positive rate than the Splunk tool. The Mapreduce approach uses Lucene[5] to index and for querying the indexes. This index approach is designed to find the file attributes and suspicious file chains based on the given parent, child or target file attributes as well. These given attributes are file name, size, and hash of the suspicious files.

B. *Proposed Methods*

1) ***Process Index Mapreduce Technique:***

Algorithm of Process Index MapReduce Function :

1: Input:

Xml log files to be processed, collected from all clients.

2: Output:

User, ProcessName, Hash value of process_path

3: Process:

- Map function
- KeyValue :

<User, ProcessName, Hash value of process_path>

- Reduce function :- Null

– XmlInputFormat:

Map function intakes the xml input format to divide the task among the slave datanodes. Here each <event> tag of a xml file is parsed by mapper.

2) *Malware-Search Mapreduce Technique:*

Algorithm of MalwareSearch MapReduce Function:

1: Input:

Indexed files.

2: Output:

Specified malware and its host computer information is obtained.

3: Process:

– MalwareSearch from FileIndex output files

– Map function:

Search for a malware based on <ProcessName>.

– Reduce function:

Output to a file with specified malware records,

in following format.,

<User - ProcessName>

3) *Processcount Index Technique*

Algorithm of Process count Index MapReduce Function:

1: Input:

Xml log files to be processed, collected from all clients.

2: Output:

Indexed file stores each record with

<User-ProcessName >

3: Process:

– Map task:
– KeyValue :
<User-ProcessName>
– Reduce Task: Null
– XmlInpuFormat:

Map function intakes the xml input format to divide the task among the slave datanodes. Parse the contents based on each <event> tag in xml file by mapper function.

4) *ProcessCount Mapreduce Technique*

Algorithm of Process count MapReduce Function:

1: Input:

Indexed files.

2: Output:

Count of each processes are displayed ,

<User - ProcessName, total count of the process>

3: Process:
– Map task:
– KeyValue - <User- Process Name>
– Reduce Task:

Count the occurrence of each process and displays with host information.

V. RESULTS

C. *Performance improvement Obtained*

In the system, we have considered and tested different client system log information on single and multi node cluster using 3 and 7 machines. From the experimental results we found execution time is drastically reduced when huge amount of data is processed in multi node cluster with increasing number of slave nodes. The efficiency of the proposed Process index, ProcessCount index, Malware Search and Process Count mapreduce techniques are illustrated in Fig 2, Fig 3, Fig 4 and Fig 5 respectively.

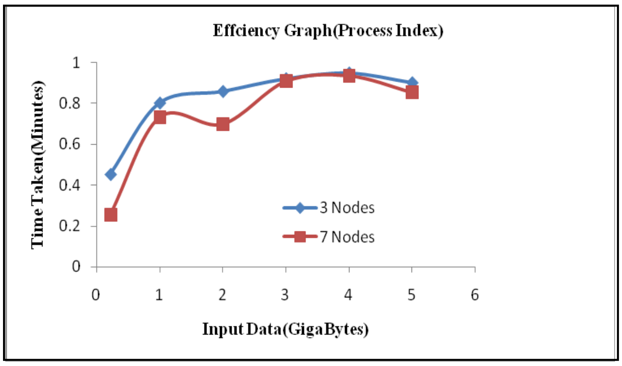


Fig 2: Efficiency Graph- Process Index

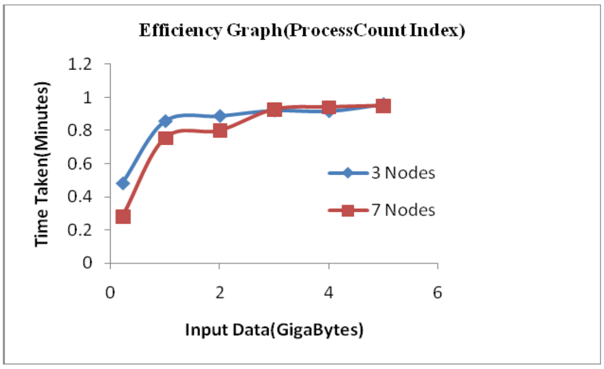


Fig 3:Efficiency Graph- ProcessCount Index

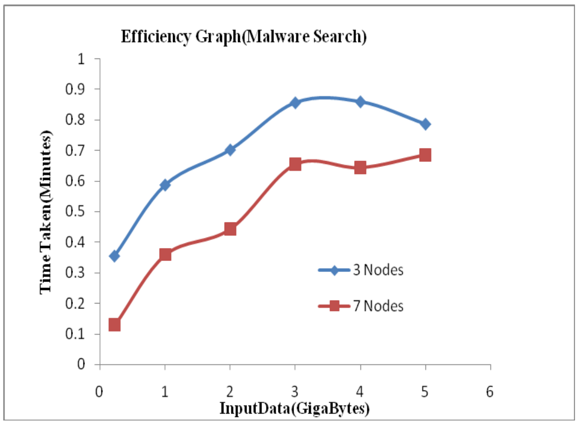


Fig 4: Efficiency Graph-Malware Search

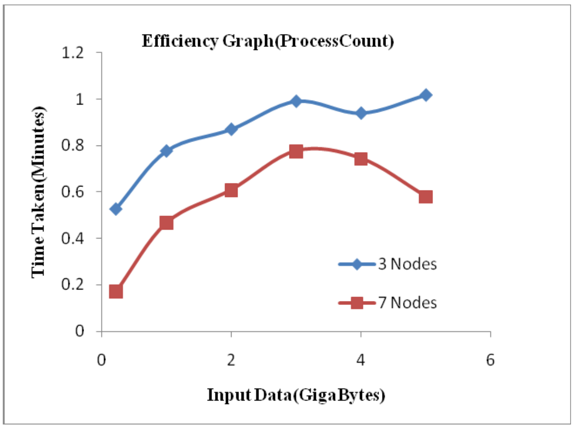


Fig 5:Efficiency Graph-ProcessCount

Table 2: Malware Detection Results

MALWARE	CLIENTS CONSIDERED	RESULTS OBTAINED	REMARKS
Recycler.exe (e5188982.exe)	3	3	Detected
Recycler.exe (tikno.exe)	3	Nil	Not Detected
astry.exe	3	2	Detected
Trojan.Falupan (winlogon.exe)	3	3	Detected
fotoku.exe	3	Nil	Not Detected
Trojan GameThief.Win32.Tibia virus(svchost.exe)	3	3	Detected
rundll32.exe	3	3	Detected

From the above results, we can infer that 5 out of 7 ie., 72% of malwares, which are considered for testing are detected by using the proposed mapreduce techniques.

D. *Meeting security issues*

We are sending the logs from the client to the cloud servers using the SSH application. The communication between the machines in the server setup is carried out through the SSH . It transmits the data in the encrypted format, hence the metadata stored in the datanodes will be safe and secure. The Hadoop platform is fault tolerant and data is distributed and stored in all the datanodes. Hence there are less or no traces of loss of metadata in the system network.

V. CONCLUSION

The concept of proposed method is conceived from the process of retrospective detection mechanism. By tracking the portable executable format file processes, malware are detected. This method can be effectively used in any organization, which efficiently detect the affected victim computers and results are reported. The setup is tested and evaluated for a specific type of malwares which affects the removable drives. These techniques can be extended to detect different types of malwares to be indexed and searched. The Hadoop system can be scaled to large cluster environment to improve and increase the processing speed and reduce the processing time. Advance search mechanism can be provided by using search libraries which is compatible with hadoop and implemented.

REFERENCES

[1] "Apache Lucene," 2010 , [Online]. Available: http://lucene.apache.org.[Accessed Sep 15,2011].

[2] Cluck Lam, *Hadoop in Action, by Manning Publications Co*, Chapter1, Chapter 2, Chapter 3 and Chapter 4, Sep 2011.

[3] J. Dean and S. Ghemawat : "*MapReduce: Simplified data processing on large clusters*", in Communications of the ACM, vol. 51, pp. 107-113, 2008.

[4] Jimmy Lin and Chris Dyer , *Data-Intensive Text Processing with MapReduce* , Chapter 2 and Chapter 3, April 11, 2010.

[5] J. Oberheide ,"*Cloudav: N-version antivirus in the network cloud*" , in the Proceedings of 17th USENIX Security Symposium, 2008.

[6] "Hadoop" 2010. [Online]. Available: http://hadoop.apache.org.[Accessed Sep 15, 2011]

[7] Shun-Te Liu and Yi-Ming Chen ,” *Retrospective Detection of Malware Attacks by Cloud Computing”*, in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, cyberc, pp.510-517, 2010.

[8] "Splunk," [Online]. Available: http://www.splunk.com.[Accessed Jun, 5, 2010].

[9] T. White, *Hadoop: The Definitive Guide*: O'Reilly Media, 2009.

[10] Vinod P.and V.Laxmi,M.S.Gaur,” *Survey on Malware Detection Methods*”.

[11] Windows Sysinternals, Process Monitor Tool [Online].Available: http://technet.microsoft.com/enus/sysinternals/bb545021.aspx.

◈◈◈

An Approach for Hybrid Signature Scheme for Multicast Communication

Navya L & Manjunath S.S

M.Tech Student Department of CSE, Dayananda Sagar College of Engineering, Bangalore, India
E-mail : navya.lakshman@gmail.com, mnj_ss200@yahoo.co.in

Abstract - Most of the Multicast Authentication schemes suffer from drawbacks such as vulnerability to packet loss and authentication latency at the sender and/or receivers. In internet and wireless networks the packets are vulnerable to loss due to correlation among them. They are vulnerable to packet injection by malicious attackers leading to DOS. In this paper we propose an efficient cryptographic primitive called batch signature to verify the signatures of any number of packets at the same time. Here BLS and DSA signature schemes are combined to generate signatures. It is efficient in terms of latency, authentic and provides DOS defense.

Keywords*-DOS, Authentication, Signature, Multicast, Pairing, Generator.*

I. INTRODUCTION

In multicasting, a message or information is delivered to a selected group of receiver nodes in single transmission from source node. But in traditional multicast authentication schemes the relationship between packets tends to packet loss which is very common in internet and wireless communication. Authentication is critical to ensure the origin of a multicast stream. The loss of a certain number of packets can result in the failure of authentication of other received packets. In an extreme case, the loss of the signature of one block makes the whole block of packets unable to be authenticated. Though existing schemes allow increasing the resilience to packet loss by attaching more authentication information to each packet, this results in more computational and communication overhead, which is undesirable in resource constrained scenarios such as mobile and wireless communications.

Multicast Authentication may provide the following security services [1]:

1. Data Integrity: The assurance that data received are exactly as sent, with no duplications, insertion, modification, reordering or replays by an authorized entity.

2. Data Origin Authentication: The assurance that the communicating entity is the one that it claims to be. Assuring that a communication is authentic.

3. Non-repudiation: It prevents either sender or receiver from denying a transmitted message. It provides protection against denial by one of the entities involved in a communication of having participated in all or part of the communication.

All the three services can be supported by an asymmetric key technique called signature. Figure 1 shows asymmetric signature scheme [2]. The sender generates a signature for each packet with its private key, which is called signing, and each receiver checks the validity of the signature with the sender's public key, which is called verifying. If the verification succeeds, the receiver knows the packet is authentic.

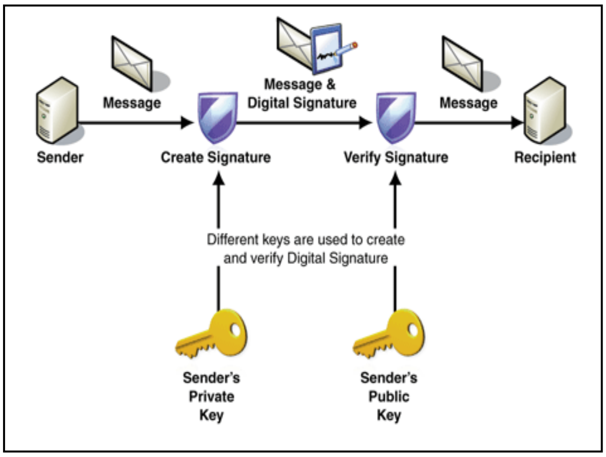


Figure 1: Asymmetric Signature Scheme

Some of the issues in design of Multicast authentication are latency, communication and computation overhead. Packet loss is inevitable. Existing digital signature algorithms are expensive. They are vulnerable to packet injection by malicious attackers leading to DOS [3]. In order to achieve a perfect resilience to packet loss and alleviate the DOS impact the batch signature primitive [4] is proposed for providing authentication in multicast communication. Each packet is attached with a signature, and a receiver authenticates any number of packets simultaneously by verifying their signatures through only one verification operation. It is efficient in terms of latency and computation cost. Both the BLS and DSA signature scheme is combined for generating a powerful signatures.

This paper is organized in the following way: firstly, the related work is described; secondly, the overview of signature schemes is described; thirdly, the design and implementation of the system are described; finally, the results followed by conclusions.

II. RELATED WORK

In recent years, many schemes have been proposed to provide multicast communication, which overcomes the issues like packet loss, DOS. IBS [11] uses batch verification method to verify multiple signatures simultaneously such that the time for the verifications is significantly reduced. But it works on only random oracle models. TESLA [12] is based on symmetric-key cryptography and loose time synchronization between the sender and the receivers, followed by delayed release of keys by the sender. The basic TESLA scheme provides delayed authentication. But it cannot support non repudiation of the data source to the third parties. An improved batch verification signature scheme [13] based on HCC. It is composed of interactive protocol and non-interactive protocol for different application. It achieves the same security with less storing space, smaller communication bandwidth and less overhead of the system.

III. OVERVIEW OF SIGNATURE SCHEMES

The details of signature schemes are as follows:

A. BLS Signature Scheme

BLS stands for Boneh-Lynn-Shacham

BLS [5] signature scheme allows a user to verify that a signer is authentic. The scheme uses a pairing function for verification and signatures are group elements in some elliptic curve. Signatures are often referred to as short signatures, BLS short signatures, or simply BLS signatures.

The pairing function [8] is a process to uniquely encode two natural numbers into a single natural number. It can be defined as a map over two cyclic groups G_1 and G_2, where map e: $G_1 \times G_1 \rightarrow G_2$ satisfying following properties:

1. Bilinear maps [7]: Consider cyclic groups G_1, G_2 and G_t having same order. A bilinear map from $G_1 \times G_2$ to G_t is a function defined as

$$e: G_1 \times G_2 \rightarrow G_t$$

such that for all $u \in G_{1,}$ $v \in G_2$, $a \in Z$, $b \in Z$ we have

$$e\,(u^a, v^b) = e\,(u, v)^{ab}$$

2. Non-degenerate[9]: For the generator g_1 of G_1 that is

$$g_1{}^p = 1 \in G_1$$

where p is order of G_1, we have

$$e\,(g_1, g_1) \neq 1 \in G_2$$

The BLS signature scheme consists of three phases as shown in Figure 2:

1. Key Generation: The key generation algorithm selects a random integer x in the interval [0, *Z*]. x is private key. The holder of the private key publishes the public key, g^x where g is a generator of G.

2. Signing: Given the private key x and some message m, compute the signature by hashing the message m, as h=H (m), then signature

$$\sigma = h^x.$$

3. Verification: Given a signature σ and a public key g^x, we verify that

$$e\,(\sigma , g) = e\,(H(m) , g^x)$$

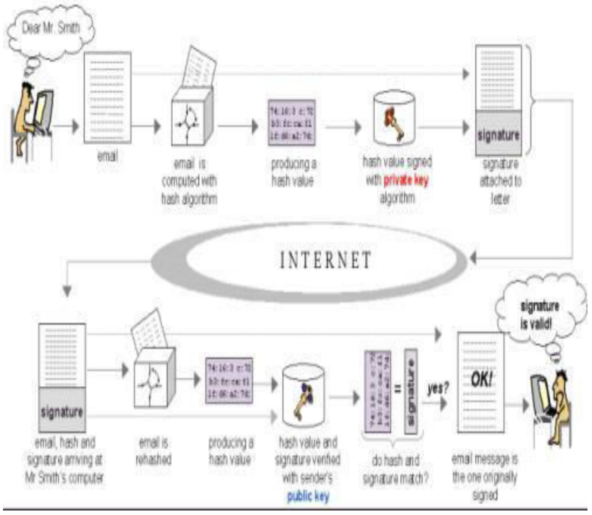

Dear Mr. Smith
email
email is computed with hash algorithm
74:16:3 e:70 b3:fd:ca:11 1d:06:a2:7d
producing a hash value
hash value signed with private key algorithm
signature
signature attached to letter
INTERNET
signature
email, hash and signature arriving at Mr Smith's computer
email is rehashed
producing a hash value
hash value and signature verified with sender's public key
=
signature
do hash and signature match?
yes?
OK!
email message is the one originally signed
signature is valid!

Figure 2: BLS Signature Scheme Phases

B. DSA Signature Scheme

DSA stands for Digital Signature Algorithm

DSA [6] is a standard and one of the popular signature algorithms.

The steps in design of DSA [10] are as follows:

1. Choosing of global public key components like:

p- random l bit prime, $512 \leq l \leq 1024$

q- random 160 bit prime dividing p-1

$r = h^{(p-1)/q} \bmod p$, where h is a random primitive

element of Z_p, such that $r > 1$

2. Choosing of User's private key and computing public key:

x- random integer, $0 < x < q$

$y = r^x \bmod p$

3. Key is K = (p, q, r, x, y)

The steps in DSA signature generation are as follows:

1. Choose a random k, $0 < k < q$ such that

GCD (k, q) = 1

2. Compute $a = (r^k \bmod p) \bmod q$

3. Compute $b = k^{-1}(m + xa) \bmod q$ where m is a message to sign and $kk^{-1} \equiv 1 \pmod q$

4. Signature sig(m, k) = (a, b)

The steps in DSA signature verification are as follows:

1. Compute $z = b^{-1} \bmod q$

2. Compute $u_1 = wz \bmod q$, $u_2 = az \bmod q$

3. Verification: $VER_K(m, a, b) = true \Leftrightarrow (r^{u1}\ y^{u2} \bmod p) \bmod q = a$

IV. DESIGN AND IMPLEMENTATION OF THE SYSTEM

The design process for software system has often two levels. At the first level the focus is on deciding which modules are needed for the system, the specifications of these modules, and how these modules should be interconnected. This is called the system design or top-level design. In the second level, the internal design of the modules, or how the specifications of the module can be satisfied, is decided. This design level often called detailed design or logic design.

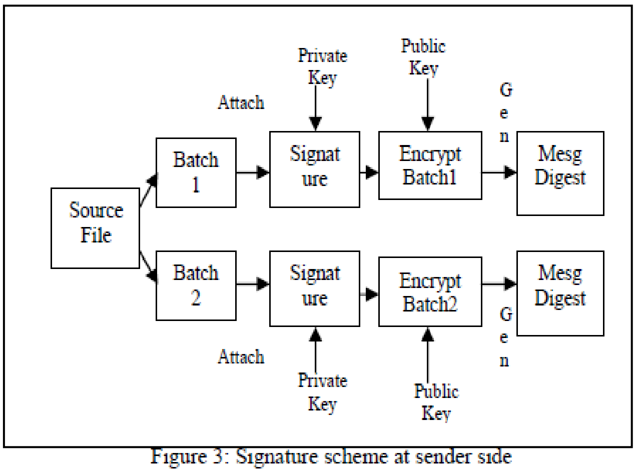


Figure 3: Signature scheme at sender side

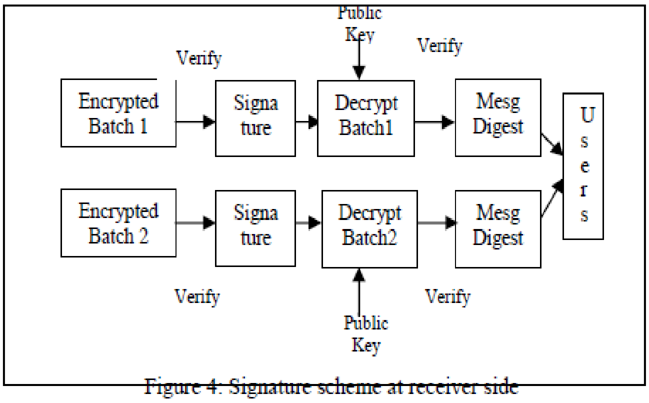
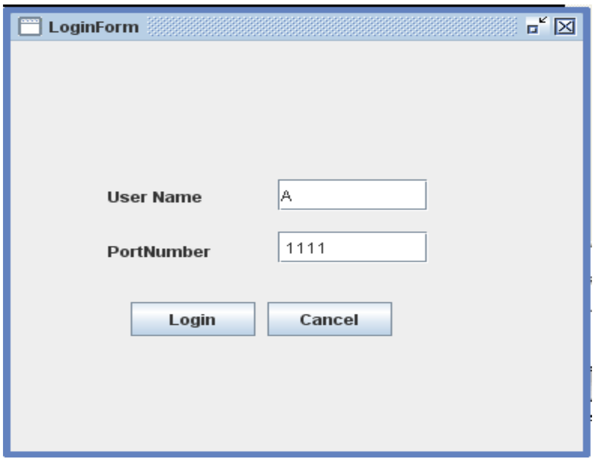


Figure 4: Signature scheme at receiver side

V. RESULTS

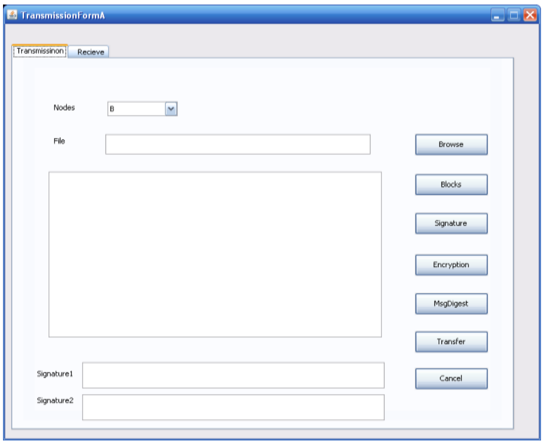
The obtained results of this implementation are as follows.


LoginForm

User Name A

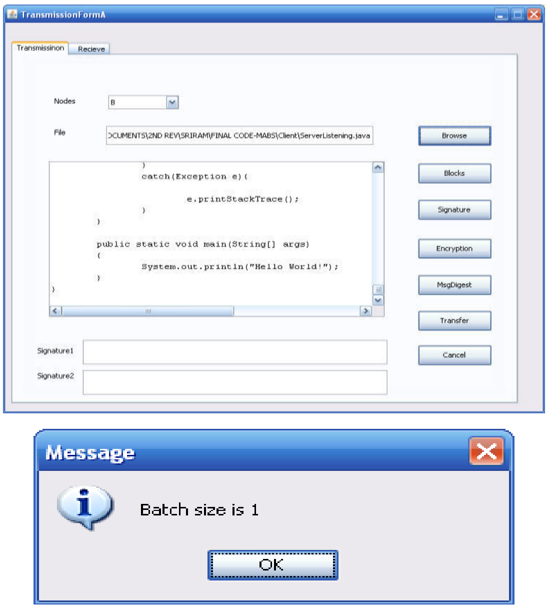
PortNumber 1111

Figure 5: Login window



Nodes: B

Figure 6: Sender window



Nodes: B

File: ƆCUMENTS\2ND REV\SRIRAM\FINAL CODE-MABS\Client\ServerListening.java

```
                }
                catch(Exception e){

                        e.printStackTrace();
                }
        }

        public static void main(String[] args)
        {
                System.out.println("Hello World!");
        }
}
```

Batch size is 1

Figure 7: Browsing the file and blocks is created, showing block size

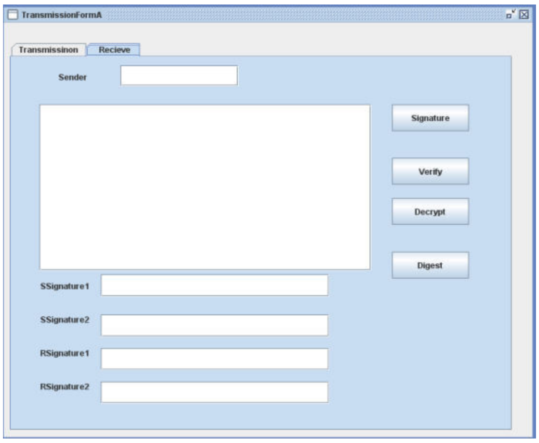
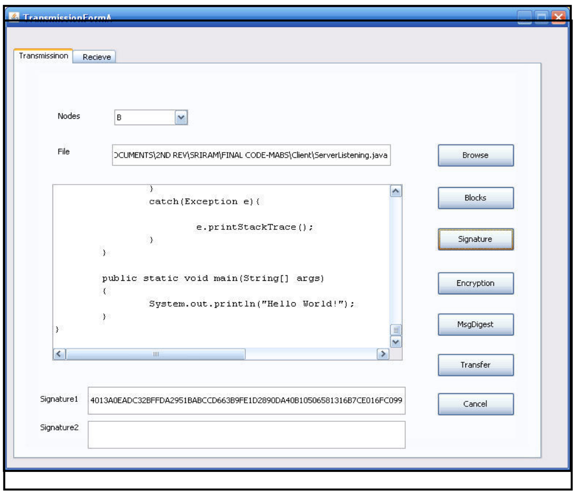


Figure 8: Reciever window



Nodes: B

File: ƆCUMENTS\2ND REV\SRIRAM\FINAL CODE-MABS\Client\ServerListening.java

```
                }
                catch(Exception e){

                        e.printStackTrace();
                }
        }

        public static void main(String[] args)
        {
                System.out.println("Hello World!");
        }
}
```

Signature1: 4013A0EADC32BFFDA2951BABCCD663B9FE1D2890DA40B10506581316B7CE016FC099

Figure 9: Signature generation

VI. CONCLUSIONS

The problems like latency, packet loss and packet forgery are solved. The use of batch signature provides efficiency in terms of latency and computation cost. The two new batch signature schemes are more efficient then previous schemes.

ACKNOWLEDGMENT

This paper would not have existed without my guide Associate Prof. Manjunath.S.S. I also would like to thank our head of the department Dr. Ramesh Babu D.R

REFERENCES

[1] Yun Zhou, Xiaoyan Zhu, and Yuguang Fang, “MABS: Multicast Authentication Based on Batch Signature,” IEEE Transactions on Mobile Computing, vol. 9, no. 7, July 2010.

[2] A. Menezes, P. van Oorschot, and S.Vanstone. Handbook of Applied Cryptography. CRC Press, 1996.

[3] P. Rohatgi, “A Compact and Fast Hybrid Signature Scheme for Multicast Packet,” Proc. Sixth ACM Conf. Computer and Comm. Security (CCS ’99), Nov. 1999.

[4] C. Boyd and C. Pavlovski, “Attacking and Repairing Batch Verification Schemes” Proc. Sixth Int’l Conf. Theory and Application of Cryptology and Information Security Advances in Cryptology (ASIANCRYPT’00),pp.58-71,Dec.2000.

[5] D. Boneh, B. Lynn, and H. Shacham, “Short Signatures from the Weil Pairing,” Proc. Seventh Int’l Conf. Theory and Application of Cryptology

and Information Security Advances in Cryptology (ASIACRYPT ’01), pp. 514-532 Dec. 2001.

[6] FIPS PUB 186, Digital Signature Standard (DSS) May 1994.

[7] D. Boneh, C. Gentry, B. Lynn, and H.Shacham,“Aggregate and Verifiably Encrypted Signatures from Bilinear Maps,” International Association for Cryptologic Research 2003.

[8] Arnold L. Rosenberg, “Efficient Pairing Functions- and why you should Care,” International journal of foundations of Computer, IEEE Transactions 2002

[9] Gr. Tsagas and P. Nerantzi, “Symmetric Invariant Non-Degenerate bilinear forms on nilpotent lie algebras,” Proceedings of the Workshop on Global Analysis, Differential Geometry and Lie Algebras, 1996, 124-129.

[10] T. ElGamal, “A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms,” IEEE Trans. Information Theory, vol. IT-31, no. 4, pp. 469-472, July 1985.

[11] S. Cui, P. Duan, X.Cheng and C.W. Chan, “An Efficient Identity-Based Signature Scheme and Its Applications,” International Journal of Network Security, Vol.5, No.1, PP.89–98, July 2007.

[12] A. Perrig, R. Canetti, D. Song, and J. Tygar, “Efficient and Secure Source Authentication for Multicast,” Proc. Network and Distributed System Security Symp. (NDSS ’01), 2001.

[13] Xuanwu Zhou, “Cryptanalysis schemes against batch verification signature,” Chinese Control and Decision Conference (CCDC 2009).

◈◈◈

User and Query-Dependent Ranking for Web Databases

V. Chandra Sekhar

CSE DEPARTMENT, KSRMCE, KADAPA, A.P.
E-mail : chandra.ksrm@gmail.com

***Abstract* -** With the emergence of the deep Web, searching Web databases in domains such as vehicles, real estate, etc. has become a routine task. One of the problems in this context is ranking the results of a user query. Earlier approaches for addressing this problem have used frequencies of database values, query logs, and user profiles. A common thread in most of these approaches is that ranking is done in a user- and/or query-independent manner.

This paper proposes a novel query- and user-dependent approach for ranking query results in Web databases. We present a ranking model, based on two complementary notions of user and query similarity, to derive a ranking function for a given user query. This function is acquired from a sparse workload comprising of several such ranking functions derived for various user-query pairs. The model is based on the intuition that similar users display comparable ranking preferences over the results of similar queries. We define these similarities formally in alternative ways and discuss their effectiveness analytically and experimentally over two distinct Web databases.

I. INTRODUCTION

The emergence of the deep Web has led to the proliferation of a large number of Web databases for a variety of applications (e.g., airline reservations, vehicle search, real estate scouting). These databases are typically searched by formulating query conditions on their schema attributes. When the number of results returned is large, it is time-consuming to browse and choose the most useful answer(s) for further investigation. Currently, Web databases simplify this task by displaying query results sorted on the values of a single attribute (e.g., Price, Mileage, etc.). However, most Web users would prefer an ordering derived using multiple attribute values, which would be closer to their expectation.

II. EXISTING SYSTEM

Where a large set of queries given by varied classes of Users is involved, the corresponding results should be ranked In a user and query dependent manner.The current sorting Based mechanisms used by web databases do not perform such ranking.While some extensions to sql allow manual specification of Attribute weights, this approach is cumbersome for most Web Users.Automated ranking of database results has been Studied in the context of relational databases, and although a number of techniques perform query dependent ranking,they do not differentiate between users and hence,provide a single ranking order for a given query across all users.In contrast, techniques for building extensive user profiles as well as requiring users to order data tuples.

III. PROPOSED SYSTEM

We propose a user and query dependent approach for Ranking query results of web databases. we develop a Ranking model, based on two complementary measures of query similarity and user similarity, to derive functions from a workload containing ranking functions for several user-query pairs .We present experimental results over two web databases supported by google base to validate our approach in terms of efficiency as well as quality for real-world use. We present a discussion on the approaches for generating a workload ,and propose a learning method for the same with experimental results.

IV. RELATED WORK

Although there was no notion of ranking in traditional database, It has existed in the context of information retrieval for quite some time.

Ranking in Databases

Although ranking query results for relational and Web databases has received significant attention over the past years, simultaneous support for automated user and query-dependent ranking has not been addressed in this context. For instance, address the problem of query-dependent ranking by adapting the vector model from infor-mation retrieval, whereas do the same by adapting the probabilistic model. However, for a given query, these techniques provide the same ordering of tuples across all users.

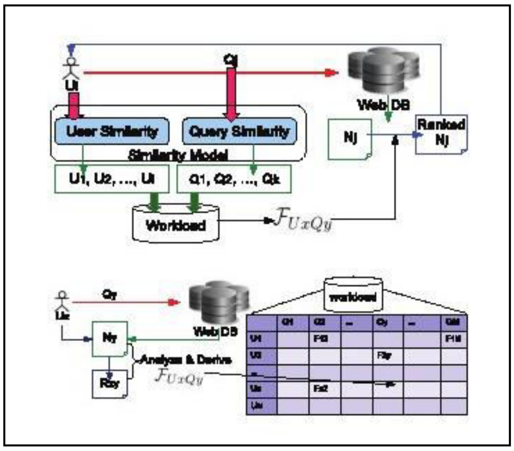
Ranking in Information Retrieval

Ranking has been ex-tensively investigated in the domain of information retrieval.The cosine-similarity metric is very successful in practice,and we employ its variant for establishing similarities between attribute-value pairs as well as query results in our framework. The problem of integrating information retrieval system and database systems have been attempted with a view to apply the ranking models (devised for the former)to the latter; however, the intrinsic difference between their underlying models is a major problem.

Relevance Feedback

Inferring a ranking function by analyzing the user's interaction with the query results originates from the concepts of relevance feedback in the domain of document and image retrieval systems.

RANKING ARCHITECTURE



The Similarity model (shown in above Figure) forms the core component of our ranking framework. When the user Ui poses the query Qj , the query-similarity model determines the set of queries ({Qj , Q1 , Q2 , ..., Qp }) most similar to Qj .

Likewise, the user-similarity model determines the set ofusers ({Ui , U1 , U2 , ...Ur }) most similar to Ui . Using thesetwo ordered sets of similar queries and users, it searchesthe workload to identify the function FUx Qy such that the combination of Ux and Qy is most similar to Ui and Qj .FUx Qy is then used to rank Qj 's results for Ui . The workload used in our framework comprises of ranking functions for several user-query pairs. Figure shows the highlevel view of deriving an individual ranking function for auser-query pair (Ux , Qy). By analyzing Ux 's preferences (interms of a selected set of tuples (Rxy)) over the results (Ny),an approximate ranking function (FUx Qy) can be derived.

The composite Similarity Model

In order to derive a user's (Ui) ranking function for a query (Qj), we have proposed two independent approaches based on user and query similarity. However, given the scale of Web users and queries, and the sparseness of the workload, applying only one model may not be the best choice at all times. Considering Example-1 and Workload-B (Table 2), we want to identify a ranking function to rank Q1 's results for U1 .

Algorithm 1 Deriving Ranking Functions from Workload INPUT: Ui , Qj , Workload W (M queries, N users) OUTPUT: Ranking Function Fxy to be used for Ui , Qj

```
STEP ONE:
for p = 1 to M do
  %% Using Equation 2 %%
    Calculate Query Condition Similarity (Qj , Qp )
end for
%% Based on descending order of similarity with Qj %%
Sort(Q1 , Q2 , .... QM )
Select QKset i.e., top-K queries from the above sorted set
STEP TWO:
 for r = 1 to N do
%% Using Equation 7 %%
   Calculate User Similarity (Ui , Ur ) over QKset
end for
%% Based on descending order of similarity with Ui %%
Sort(U1 , U2 , .... UN ) to yield Uset
STEP THREE:
for Each Qs ∈ QKset do
   for Each Ut ∈ Uset do
      Rank(Ut ,Qs ) =
      Rank(Ut ∈ Uset ) + Rank(Qs ∈ QKset )
   end for
end for
Fxy = Get-RankingFunction()
```

Using only the query-similarity model, F13 will be selected since Q3 is most similar to Q1 . In contrast, applying only user-similarity model will yield F21 as U2 is most similar to U1 . It would be meaningful to rank these functions (F13 and F21) to choose the most appropriate one. Furthermore, in a more practical setting, the workload is likely to have a ranking function for a similar query (to Q1) derived for a similar user (to U1). For instance, the likelihood of F22 existing in the workload would be higher than the occurrance of either F13 or F21 . Hence, it would be meaningful to combine the two measures into a single Similarity Model.

The goal of this composite model is to determine a ranking function (Fxy) derived for the most similar query (Qy) to Qj given by the most similar user (Ux) to Ui to rank Qj 's results.

The process for finding such an appropriate ranking function is given by the Algorithm 1.

The input to the algorithm is a user (Ui) and a query (Qj) along with the workload matrix (W) containing ranking functions. The algorithm begins by determining the query-condition similarity (STEP ONE) between Qj and every query in the workload. It then sorts all these queries (in descending order) based on their similarity with Qj and selects the set (QKset) of the top-K most similar queries to Qj that satisfy the conditions for the top-K user similarity model.

Based on these selected queries, the algorithm determines the user-similarity (STEP TWO) between Ui and every user in the workload. All the users are then sorted (again, in descending order) based on their similarity to Ui . We then generate a list of all the user-query pairs (by combining the elements from the two sorted sets), and linearise these pairs by assigning a rank (which is the sum of query and use similarity ranks) to each pair (STEP THREE). For instance, if Ux and Qy occur as the xth and y th elements in the respective ordering with the input pair, the pair (Ux , Qy) are assigned an aggregate rank.

In this case, a rank of “x + y” will be assigned. The “Get- RankingFunction” method then selects the pair (Ux , Qy) that (Fxy) in the workload. Then, in order to rank the results (Nj) the corresponding attribute weights and value weights obtaine for Fxy will be individually applied to each tuple in Nj (using Equation 1), from which a general ordering of all tuples will be achieved.

REFERENCES

1. S. Agrawal, S. Chaudhuri, G. Das, and A. Gionis. Automated ranking of database query results. In CIDR, 2003.

2. S. Amer-Yahia, A. Galland, J. Stoyanovich, and C. Yu. From del.icio.us to x.qui.site: recommendations in social tagging sites. In SIGMOD Conference, pages 1323–1326, 2008.

3 R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999.

4 M. Balabanovic and Y. Shoham. Content-based collaborative recom-mendation. ACM Communications, 40(3):66–72, 1997.

5 J. Basilico and T. Hofmann. A joint framework for collaborative and content filtering. In SIGIR, pages 550–551, 2004.

6. S. Chaudhuri, G. Das, V. Hristidis, and G. Weikum. Probabilistic ranking of database query results. In VLDB, pages 888–899, 2004.

7 T. Kanungo and D. Mount. An efficient k-means clustering algorithmAnalysis and implementation. IEEE Transactions of Pattern Analysis in Machine Intelligence, 24(7):881–892, 2002.

8. A. Telang, C. Li, and S. Chakravarthy. One size does not fit all: Towards user- and query-dependent ranking for web databases

9. K. Werner. Foundations of preferences in database systems. In VLDB, pages 311–322. VLDB Endowment, 2002.

10. S.-W. Hwang. Supporting Ranking For Data Retrieval. PhD thesis, University of Illinois, Urbana Champaign, 2005.

11. databases on the web: Observations and implications. SIGMOD Record, 33(3):61–70, 2004.

12. C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In International conference on World Wide Web

◈◈◈

Low-Power and Area-Efficient Carry select Adder

Shaik. Shafi & M.Chennakesavulu

JNTUA, Dept. of ECE, R G M College of Engineering & Technology, Nandyal, India
E-mail : shafirgm@gmail.com, onlinechenna@yahoo.com:

Abstract – In digital adders, the speed of addition is limited by the time required to propagate a carry through the adder. Carry select adders are used to reduce the Carry propagation delay by independently generating multiple carries and then select a carry to generate the Sum. A simple and efficient gate-level modification is done to significantly reduce the area and power of the Carry Select Adder. In this proposed method Binary to Excess-1 Converter is used instead of Ripple Carry Adder with carry input 1.The performance of the proposed design is evaluated in terms of delay, area, power, power-delay product. The reduced number of transistors in this work offers great advantage in reduction of area and total Power. The Schematics of both regular Square root Carry Select Adder and Proposed Square root Carry select Adder are Simulated at transistor level using Micro Wind 3.1 Tool in DSM Technology. It is proposed to show that the proposed CSLA structure will be better than the regular SQRT CSLA.

Keywords - *Carry Propagation Delay, Ripple Carry Adder, Power-Delay Product, low Power, area-efficient,SQRT CSLA.*

I. INTRODUCTION

Addition is one of the fundamental arithmetic operations. It is used in many VLSI systems such as application-specific DSP architectures and microprocessors. Its main task is adding two binary numbers; it is also the nucleus of many other useful operations such as subtraction, multiplication, division, address calculation, etc. In most of the systems adder is critical path which determines overall performance of the system. Design of power and area efficient data path logic systems are one of most important areas of research inVLSI system design. Carry Select Adder (CSLA) is used to reduce the problem of Carry propagation Delay. CSLA generates multiple carries and then select a carry to generate the sum.CSLA is not area efficient, Because of using pair of Ripple Carry Adders (RCA) to generate partial sum and carry. Partial sum and carry are generated by considering carry input Cin=0 and Cin =1, then the final sum and carry are selected by the multiplexers (mux). The Proposed method is to use Binary to Excess- 1 converter (BEC) instead of Ripple Carry Adder (RCA) With Cin = 1 in the regular CSLA to achieve lower area and power consumption.

II. CARRY SELECT ADDER (CSLA):

Carry Select Adder (CSLA) is implemented with dual Ripple-carry adder (RCA) with the carry-in of Cin=0 and Cin=1, respectively. Depending on the configuration of block length, CSLA is further classified as either Linear or Square root. In linear CSLA uses Linear -sized blocks, where as square root CSLA uses variable-sized blocks and ripple-carry addition in each block. The basic idea of CSLA is Anticipatory Parallel Computation. The two Ripple Carry Adders with Cin=0 and Cin=1 produces two partial sums and carries, the correct sum and carry-out signals will be simply selected by a set of multiplexers. Regular CSLA is shown in Fig 4.

III. BEC LOGIC :

The main idea of this proposed method is to use BEC Logic Instead of the RCA with carry in ,Cin = 1 in order to reduce the area and power consumption of the regular CSLA. To replace the n-bit RCA, an n+1 bit BEC is required. 4-bit BEC is shown in Fig 3 and Table 1. Fig 3 illustrates how the basic function of the CSLA is obtained by using the 4-bit BEC with the mux .One input of the 8:4 mux gets as it input (B3, B2, B1, and B0) and another input of the mux is the BEC output. This produces the two possible partial results in parallel and the mux is used to select either the BEC output or the direct inputs according to the control signal Cin. The importance of the BEC logic is the large silicon area reduction when the CSLA with large number of bits are designed. The Boolean expressions of the 4-bit BEC are shown below (note the functional symbols NOT, & AND, XOR)

X0 = ~B0

X1 = B0 ^ B1

X2 = B2 ^(B0 & B1)

X3 = B3 ^ (B0 & B1 & B2)

IV. DELAY AND AREA EVALUATION METHODOLOGY OF BASIC BLOCKS OF CARRY SELECT ADDER

The AND,OR, and Inverter (AOI) implementation is shown in Fig 1.Parallel operations are performed by the gates in between the dotted lines, the numeric value represents delay contributed by that gate. All gates are to be made up of AND,OR and Inverter, each having delay equal to 1 unit and area equal to 1 unit. The number of gates in longest path of block which contributes maximum delay is then added. Area is evaluated by counting the total number of AOI (AND-OR-INVERTER)gates for each logic block. This is shown in Table 2.

V. DELAY AND AREA EVALUATION METHODOLOGY OF REGULAR 16-BIT SQUARE ROOT CSLA:

The delay evaluation of each group depends upon the arrival time of multiplexer selection input and the arrival time of data outputs from the RCA's. For Group 2 the arrival time of multiplexer selection input C1[time(t) = 7] of 6:3 mux is earlier than s3[(t) = 8] and later than s2[(t) = 6].

{c1, sum [3]} = s3 [(t) =8] + mux

[time (t) =3] =Sum3 [11]

{c1, sum [2]} = C1 [(t) =7] + mux

[time(t) =3] =Sum2 [10]

Except for group2, the arrival time of mux selection input is always greater than the arrival time of data outputs from the RCA's. Thus, the delay of group3 to group5 is determined, respectively as follows:

{c6, sum [6:4]} = c3 [t = 10] + + mux [time (t) =3]

{c10, sum [10:7]} = c6 [t = 13] + mux [time (t) =3]

{Cout, sum [15:11]} = c10 [t = 16] + mux [time (t) =3]

The one set of 2-bit RCA in group2 has 2 FA for Cin = 1 and the other set has 1 FA and 1 HA for Cin = 0.Based on the area count of Table 2, the total number of gate counts in group2 is shown below. Similarly the maximum delay and area of the other groups in the regular

SQRT CSLA is shown in below Table 3.

Gate count = 57 (FA + HA + Mux)

FA = 39(3 X13)

HA = 6(1 X6)

Mux = 12 (3 X 4)

VI. DELAY AND AREA EVALUATION METHODOLOGY OF MODIFIED 16-BIT SQUARE ROOT CSLA:

The group2 has one 2bit RCA which has 1 FA and 1 HA for RCA with Cin = 0. Instead of another 2bit RCA with Cin = 1 a 3-bit BEC is used which adds one to the output from 2bit RCA. Based on the delay values of Table 2, the arrival time of selection input C1[time(t) = 7] of 6:3 mux is earlier than s3[t = 9] and c3= [t = 10] and later than the s3[t = 4]. Thus, the sum3 and final c3 (output from mux) are depending on s3 and mux and partial c3 (input to mux) and mux, respectively. The sum2 depends on c1 and multiplexer

For the remaining group's the arrival time of mux selection input is always greater than the arrival time of data inputs from the BEC's. Thus, the delay of the remaining groups depends on the arrival time of mux selection input and the mux delay.

The area count of group2 is determined as follows:

Gate count = 43 (FA + HA + Mux + BEC)

FA = 13(1X13)

HA = 6(1X6)

AND = 1

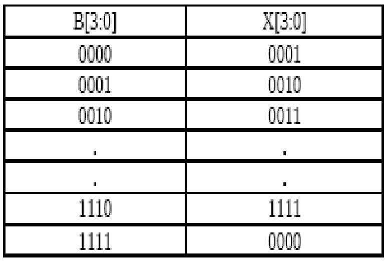
NOT = 1

XOR = 10(2X5)

Mux =12(3X4)

VII. FIGURES AND TABLES :

Table1
FUNCTION TABLE OF THE 4-b BEC



B[3:0]	X[3:0]
0000	0001
0001	0010
0010	0011
.	.
.	.
1110	1111
1111	0000

Table 2

DELAY AND AREA COUNT OF THE BASIC BLOCKS OF CSLA

Adder blocks	Delay	Area
XOR	3	5
2:1 Mux	3	4
Half Adder	3	6
Full adder	6	13

Table 3

DELAY AND AREA COUNT OF THE Regular SQRT CSLA

Groups	Delay	Area
Group2	11	57
Group3	13	87
Group4	16	117
Group5	19	147

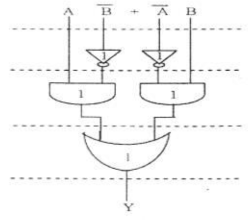


Fig. 1. Delay and Area evaluation of an XOR gate.

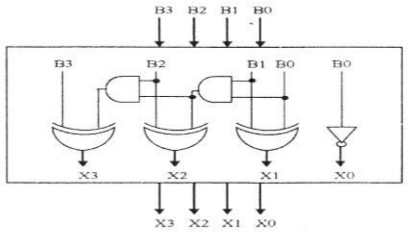


Fig. 2. 4-b BEC.

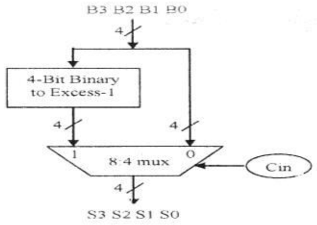


Fig. 3. 4-b BEC with 8:4 mux.

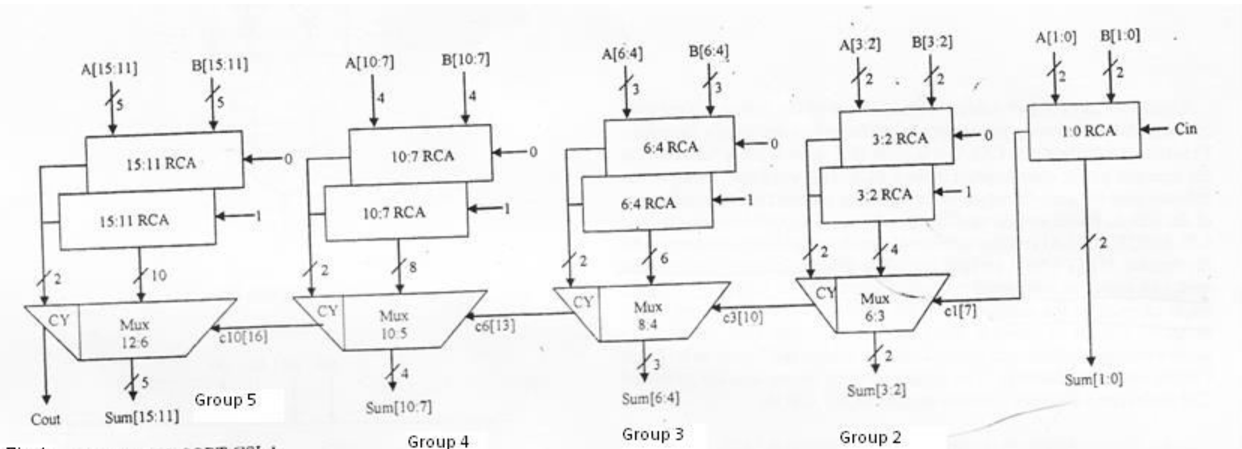


Fig 4: Regular 16-b SQRT CSLA.

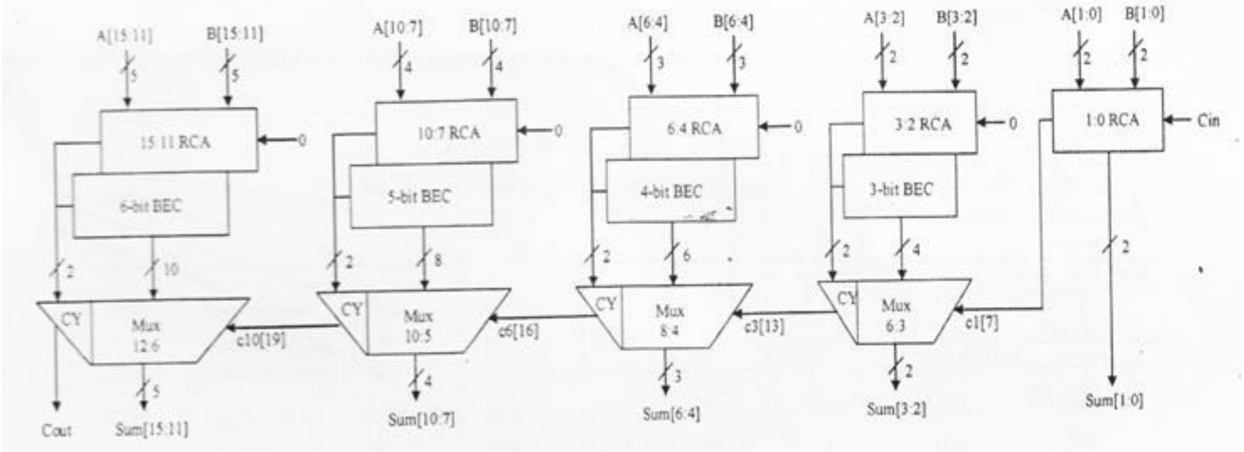


Fig 5 : Modified 16-b SQRT CSLA. The parallel RCA with Cin = 1 is replaced with BEC.

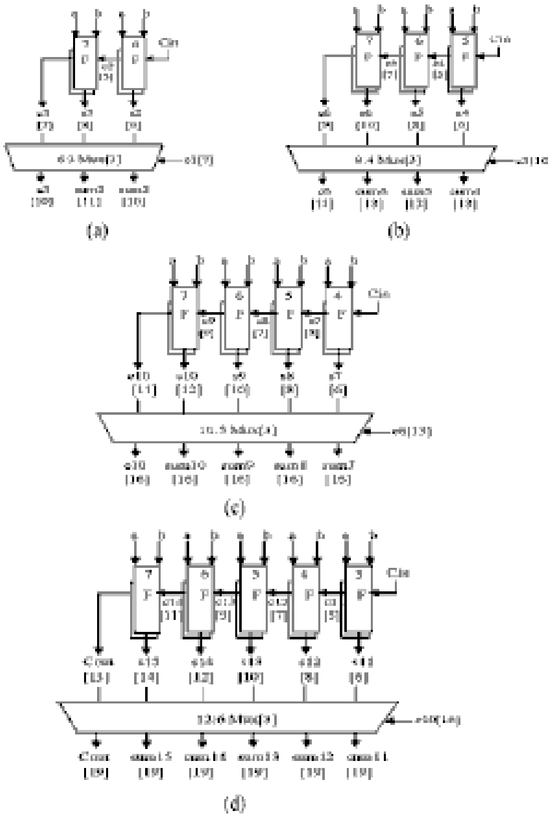


Fig 5: Delay and area evaluation of regular SQRT CSLA: (a) group2, (b) group3, (c) group4, and (d) group5. F is a Full Adder.

THE SIMULATION RESULTS

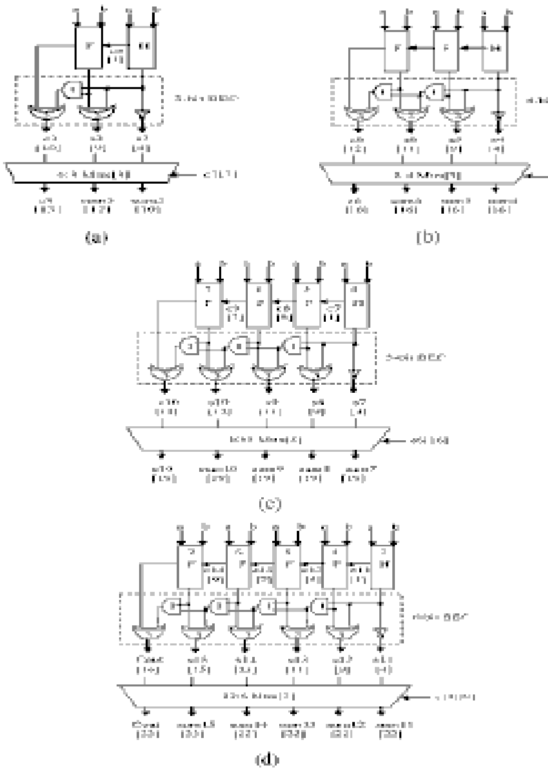


Fig 6: '. Delay and area evaluation of modified SQRT CSLA: (a) group2, (b) group3, (c) group4, and (d) group5. H is a Half Adder.

VIII.CONCLUSION:

A simple approach is proposed in this paper to reduce the area and power of SQRT CSLA architecure.The compared results show that the modified SQRT CSLA has a slightly larger delay, but the area and power are significantly reduced.The modified CSLA architecture is therefore, low area ,low power, simple and efficient for VLSI hardware.

Table 4: Comparison between Regular and Modified Carry Select Adder (CSLA)

Adder(16-bit word size)	Delay(ns)	Area (um^2)	Power (mW)	Power -Delay Product(10^{-12})
Regular CSLA	0.256	9397.51	0.205	0.052
Modified CSLA	0.269	8870.71	0.174	0.046

REFERENCES:

[1] O. J. Bedrij, "Carry-select adder," IRE Trans. Electron. Comput.,pp. 340–344, 1962. Volume: EC – 11, issue: 3,ISSN: 0367- 9950

[2] B. Ramkumar, H.M. Kittur, and P. M. Kannan, "ASIC implementation of modified faster carry save adder," Eur. J. Sci.Res., vol. 42, no. 1, pp. 53-58, 2010.

[3] Akhilesh Tyagi, "**A** Reduced-Area Scheme for Carry-Select Adders" Electron. Lett., vol. 34, no. 22, pp. 2101–2103, Oct. 1993.

[4] Y. Kim and L.S. Kim, "64-bit carry-select adder with reduced area," Electron. Lett., vol. 37, no. 10, pp. 614–615, May 2001. This paper appears in: Electronics Letters ,Issue Date : 10 May2001, Volume : 37 , Issue:10, On page(s): 614 - 615 , ISSN : 0013-5194

[5] Carry-select adder using single ripple-carry adder Chang, T.-Y. ;Hsiao, M.-J. ;Dept. of Electr. Eng., Nat. Tsing Hua Univ., Hsinchu This paper appears in: Electronics Letters Issue Date : 29 Oct1998 Volume : 34 , Issue:22 On page(s): 2101 – 2103 ISSN: 0013-5194

6] Y. He, C. H. Chang, and J. Gu, "An area efficient 64-bit square root carry-select adder for low power applications," in Proc. IEEE Int. Symp. Circuits Syst., 2005, vol. 4, pp. 4082-4085.

[7] Rawat, Tarek Darwish and Magdy Bayoumi "A Low Power and Reduced Area Carry Select Adder" This paper appears in: Circuits and Systems, 2002. The Issue Date : 4-7 Aug. 2002 Volume : 1 On page(s): I - 467-70 vol.1.

[8] G.A.Ruiz, M. Granda "An area efficient static C MOS CSLA based on a compact carrylook ahead unit".

[9] Behnam, Ameliford Farzan massoud pedram "Closing the gap between carry select adder and ripple carry adder: A new class of low power high performance adder"2004

[10] Padma devi,ashima girdher,Balwindher singh"Improved carry select adder with reduced Area and low power consumption. Volume 3 – No.4, June 2010

◈◈◈

Minimization of Data Replica in Mobile Adhoc Network Routing

M. K. Nageswari & M. V. Jagannatha Reddy

Dept of CSE, Madanapalle Institute of Technology and Science, Madanapalle.
E-mail : nageswarimk816@gmail.com

Abstract - Adhoc networks have characteristics such as flexibility ,easy deployment,robustness which makes them an intersresting technology for various applications.Adhocnetworks are considered as the most promising terminal networks in future mobile communications.A novel Position based Opportunistic Routing protocol, in which several forwarding candidates cache the packet that has been received using MAC interception. If the best forwarder does not forward the packet in certain time slots, suboptimal candidates will take turn to forward the packet according to a locally formed order. In this way, all candidates receives the packet, the data transmission will not be interrupted. Potential multi-paths are exploited on the-fly on a per-packet basis. We propose minimization of data replica at forwarding candidates in Mobile Adhoc network routing.The forwarding candidates will be ranking ,based on the location variance of candidate within the time factor has given to it.

General Terms- This paper exposes the minimization of data replica in MANETS routing.

Keywords - *Opportunistic routing, geographic routing,greedy forwarding,moile adhoc network.*

I. INTRODUCTION

Adhoc network is a dynamoc multihop wireless network that is established bya set of mobile nodes on a shared wireless channel.Each mobile host performs local broadcasts in order to identify its existence to the surrounding hosts.Sorrounding hostsare nodes that are in close proximity to the transmitting host.In that way each mobile hosts becomes potentially a router and it is possible to dynamically esatablish routes between itself and nodes to which a route exists .Adhoc networks wre initially proposed for military applications such as batlefield communications and Emergency application, i.e. emergency rescue operations, police, earthquakes.

The primary challenge in building a MANET is equipping each device to continuously maintain the information required to properly route traffic.MANETs are a kind of wireless adhoc networks that usually has a routable networking environment on top of a link layer ad hoc network. The growth of laptop and 802.11/Wi-Fi wireless networking have made MANETs a popular research topic since the mid 1990s. Many academic papers evaluate protocols and their abilities, assuming varying degrees of mobility within a bounded space, usually with all nodes within a few hops of each other. Different protocols are then evaluated based on measure such as the packet drop rate, the overhead introduced by the routing protocol, end-to-end packet delays, network throughput etc.

Traditional topology based MANET routing protocols such as AODV,DSR,DSDV etc are the susceptible to node mobility.one of the main reason is due to the predetermination of an end-to-end route before data transmission .Fast changing network topology ,it is very difficult to maintain determinstic route.once the path breaks data packets will get lost or be delayed for a long time untill the reconstructoon of the route,causing transmission interruption.

Geographic routing (also called georouting or position-based routing) is a routing principle that relies on geographic position information. It is mainly proposed for wireless networks and based on the idea that the source sends a message to the geographic location of the destination instead of using the network address. The idea of using position information for routing was first proposed in the 1980s in the area of packet radio networks and interconnection networks. Geographic routing requires that each node can determine its own location and that the source is aware of the location of the destination. With this information a message can be routed to the destination without knowledge of the network topology or a prior route discovery.

There are various approaches, such as single-path, multi-path and flooding-based strategies.Most single-path strategies rely on two techniques: greedy forwarding and face routing. Greedy forwarding tries to bring the message closer to the destination in each step using only local information. Thus, each node forwards the message to the neighbor that is most suitable from a local point of view.

The opportunistic forwarding, which was proposed to increase the network throughput, also shows its great power in enhancing the reliability of data delivery. a novel Position based Opportunistic Routing protocol in which several forwarding candidates cache the packet that has been received using MAC interception. If the best forwarder does not forward the packet in certain time slots, suboptimal candidates will take turn to forward the packet according to a locally formed order. In this way, as long as one of the candidates succeeds in receiving and forwarding the packet, the data transmission will not be interrupted. Potential multi-paths are exploited on the-fly on a per-packet basis, leading to POR's excellent. A position based opportunistic routing mechanism which can be deployed without complex modification to MAC protocol and achieve multiple reception without losing the benefit of collision avoidance provided by 802.11. The concept of in-the-air backup significantly enhances the robustness of the routing protocol and reduces the latency and duplicate forwarding caused by local route repair. POR which takes advantage of the stateless property of geographic routing and the broadcast nature of wireless medium.

This work focuses on minimization of data replica at forwarding candidates based on the dispalcement of nodes.So data will be delivered to only few candidates by using the time factor ,cumulative variance,ratings.

II. RELATED WORK

2.1 Destination-Sequenced Distance-Vector Routing Protocol:

The Destination-Sequenced Distance-Vector (DSDV) Routing Algorithm is based on the idea of the classical Bellman-Ford Routing Algorithm with certain improvements. Every mobile station maintains a routing table that lists all available destinations, the number of hops to reach the destination and the sequence number assigned by the destination node. The sequence number is used to distinguish stale routes from new ones and thus avoid the formation of loops.

2.2 Ad hoc On-demand Distance Vector Routing:

Ad hoc On-demand Distance Vector Routing (AODV) is an improvement on the DSDV algorithm. AODV minimizes the number of broadcasts by creating routes on-demand as opposed to DSDV that maintains the list of all the routes. To find a path to the destination, the source broadcasts a route request packet. The neighbors in turn broadcast the packet to their neighbors till it reaches an intermediate node that has a recent route information about the destination or till it reaches the destination. A node discards a route request packet that it has already seen. The route request packet uses sequence numbers to ensure that the routes are loop free and to make sure that if the intermediate nodes reply to route requests, they reply with the latest information only. When a node forwards a route request packet to its neighbors, it also records in its tables the node from which the first copy of the request came. This information is used to construct the reverse path for the route reply packet. AODV uses only symmetric links because the route reply packet follows the reverse path of route request packet. As the route reply packet traverses back to the source, the nodes along the path enter the forward route into their tables. If the source moves then it can reinitiate route discovery to the destination.

2.3 Dynamic Source Routing Protocol:

The Dynamic Source Routing Protocol is a source-routed on-demand routing protocol. A node maintains route caches containing the source routes that it is aware of. The node updates entries in the route cache as and when it learns about new routes. The two major phases of the protocol are: route discovery and route maintenance. When the source node wants to send a packet to a destination, it looks up its route cache to determine if it already contains a route to the destination. If it finds that an unexpired route to the destination exists, then it uses this route to send the packet. But if the node does not have such a route, then it initiates the route discovery process by broadcasting a route request packet. The route request packet contains the address of the source and the destination, and a unique identification number. Each intermediate node checks whether it knows of a route to the destination. If it does not, it appends its address to the route record of the packet and forwards the packet to its neighbors. To limit the number of route requests propagated, a node processes the route request packet only if it has not already seen the packet and it's address is not present in the route record of the packet. A route reply is generated when either the destination or an intermediate node with current information about the destination receives the route request packet. A route request packet reaching such a node already contains, in its route record, the sequence of hops taken from the source to this node.

2.4 Position based opportunistic routing protocol:

A novel Position based OpportunisticRouting protocol (POR) is proposed, in which several forwarding candidates cache the packet that has been received using MAC interception. If the best forwarder does not forward the packet in certain time slots, suboptimal candidates will take turn to forward the packet according to a locally formed order. In this way, as long as one of the candidates succeeds in receiving and forwarding the packet, the data transmission will not be interrupted. Potential multi-paths are exploited on the- fly on a per-packet basis, leading to POR's excellent robustness.

Opportunistic Forwarding in Multi-Rate and Multi-Hop Wireless Networks

MULTI-HOP wireless networks have attracted a lot of research interest in recent years since they can be easilydeployed at low cost without relying on the existing infrastructure. Routing in such networks is very challenging mainly due to variable and unreliable wireless channel conditions .Traditional routing schemes for multi-hop wireless networks have followed the concept of routing in wired networks by abstracting the wireless links as wired links, and finding the shortest path between a source and destination. However, the traditional shortest path approach is not ideal for wireless environment, because fluctuations in the quality of any link along the predetermined path can cause excessive retransmissions at the link layer or reroutings at the network layer, thus consume precious network resources, such as bandwidth and energy. Recently, a new routing paradigm, known as opportunistic routing , was proposed to mitigate the impact of link quality variations by exploiting the broadcast nature of the wireless medium and the spatial diversity of network topology.

The general idea behind these schemes is that, for each destination, a set of next-hop forwarding candidates are selected at the network layer and one of them is chosen as the actual relay at the MAC layer on a per-packet basis according to its availability and reachability after the transmission. As more forwarding candidates are involved in helping relay the packet, the probability of at least one forwarding candidate having correctly received the packet increases, which results in higher forwarding reliability and lower retransmission cost. Some variants of opportunistic routing schemes use nodes' location information to define the forwarding candidate set and prioritize candidates.

In this paper, we mainly focus on this kind of opportunistic routing by assuming that nodes' location information are available. Two important issues in opportunistic routing are candidate selection and relay priority assignment. The existing works on opportunistic routing typically address these issues in the network with a single channel rate.

Position-Based Routing inMobileAdHocNetworks

These *ad hoc networks*, as they are commonly called, consist of autonomous nodes that collaborate in order to transport information. Usually, these nodes act as end systems and routers at the same time.Ad hoc networks can be subdivided into two classes: *static* and *mobile*. In static ad hoc networks the position of a node may not change once it has become part of the network. Typical examples are rooftop networks . For the remainder of this work we will solely focus on mobile ad hoc networks.

In mobile ad hoc networks, systems may move arbitrarily.Examples where mobile ad hoc networks may be employed are the establishment of connectivity among handheld devices or between vehicles. Since mobile ad hoc networks change their topology frequently and without prior notice, routing in such networks is a challenging task. We distinguish two different approaches: *topology-based* and *position-based* routing. Topology-based routing protocols use the information about the links that exist in the network to perform packet forwarding.

They can be further divided into *proactive*, *reactive*, and *hybrid* approaches.Proactive algorithms employ classical routing strategies such as distance-vector routing (e.g., DSDV)or link-state routing. They maintain routing information about the available paths in the network even if these paths are not currently used.

The main drawback of these approaches is that the maintenance of unused paths may occupy a significant part of the available bandwidth if the topology of the network changes frequently . In response to this observation, reactive routing protocols were developed. Reactive routing protocols maintain only the routes that are currently in use, thereby reducing the burden on the network when only a small subset of all available routes is in use at any time.

III. PROPOSED WORK

3.1 Minimization of data replica in MANET Routing:

In existing routing protocol,source forwards data to all forwarding candidates with in its forwarding area based on the conditions.It unnecessarly sends data to forwarding candidates those are not stable. we are proposing pruining strategy to minimize data replica at forwarding candidates.Time limit will be given for forwarding candidates ,during this time period location variance will be observed for each forwarding candidate. The forwarding candidates will be assigned with rating**.**

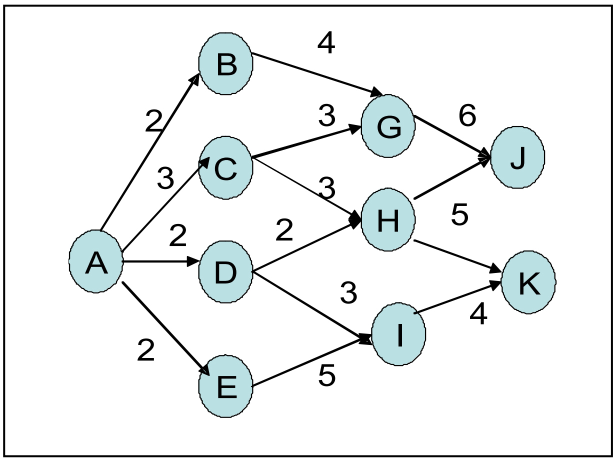


Figure 1. selection of few forwarding candidates

Pruning Strategy based on displacement of nodes

TABLE 1

Time (in units)	Forwarding candidates							
	B		C		D		E	
	d	v	d	v	d	v	d	v
1	2		3		2		2	
2	2	0	4	1	4	2	2	0
3	4	2	5	1	2	2	2	0
Cumulative variance	2		2		4		0	
Rating	3		2		4		1	

Candiddate selects the next forwarding candidate based on the rating to forward the packet.

One of the key problems in POR is the selection and prioritization of forwarding candidates. Only the nodes located in the forwarding area [14] would get the chance to be backup nodes. The forwarding area is determined by the sender and the next hop node. A node located in the forwarding area, satisfies the following two conditions: i) it makes positive progress towards the destination; ii) its distance to the next hop node should not exceed half of the transmission range of a wireless node (i.e. R/2) so that ideally all the forwarding candidates can hear from one another. In Fig. 1, the area enclosed by the bold curve is defined as the forwarding area. The nodes in this area, besides node A (i.e. node B, C), are potential candidates. According to the required number of backup nodes, some (maybe all) of them will be selected as forwarding candidates. The priority of a forwarding candidate is decided by its distance to thedestination. The nearer it is to the destination, the higher priority it will get. When a node sends or forwards a packet, it selects the next hop forwarder as well as the forwarding candidates among its neighbors. The next hop and the candidate list comprise the forwarder list.

Algorithm 1 shows the procedure to select and prioritize the forwarder list. The candidate list will be attached to the packet header and updated hop by hop. Only the nodes specified in the candidate list will act as forwarding candidates. The lower the index of the node in the candidate list, the higher priority it has. Every node maintains a forwarding table for the packets of each flow (identified as source-destination pair) that it has sent or forwarded. Before calculatinga new forwarder list, it looks up the forwarding table, an example is illustrated in Table 1, to check if a valid item for that destination is still available. The forwarding table is constructed during data packet transmissions and its maintenance is much easier than a routing table. It can be seen as a tradeoff between efficiency and scalability. As the establishment of the forwarding table only depends on local information, it takes much less time to be constructed. Therefore we can set an expire time on the items maintained to keep the table relatively small. In other words, the table records only the current active flows, while in conventional protocols, a decrease in the route expire time would require far more resources to rebuild.

Forwarding candidate table

Source , destination	next hop	Candidate list
N1,N11	N4	N5,N6
N2,N1 2	N7	N8,N9

. . . .

IV. SIMULATION SETUP

To evaluate the performance, we simulate the Agorithm in a variety of mobile network topologies in NS- 2 [19] and compare it with AOMDV [20] (a famous multipath routing protocol) and GPSR [5] (a representative geographic routing protocol).

The improved random way point (RWP) [21] without pausing is used to model nodes' mobility. The minimum node speed is set to 1 m/s and we vary the maximum speed to change the mobility degree of the network. The following metrics are used for performance comparison:

• *Packet delivery ratio*. The ratio of the number of data packets received at the destination(s) to the number of data packets sent by the source(s).

• *End-to-end delay*. The average and the median endto-end delay are evaluated, together with the cumulative distribution function (CDF) of the delay.

• *Path length*. The average end-to-end path length (number of hops) for successful packet delivery.

• *Packet forwarding times per hop (FTH)*. The average number of times a packet is being forwarded from the perspective of routing layer to deliver a data packet over each hop.

• *Packet forwarding times per packet (FTP)*. The average number of times a packet is being forwarded from the perspective of routing layer to deliver a data packet from the source to the destination.

V. RESULTS AND ANALYSIS

5.1 Memory Consumption & Duplicate Relaying

One main concern of POR is its overhead due to opportunistic forwarding, as several copies of a packet

Need to be cached in forwarding candidates. However, it will be presented need to be cached in the forwarding candidates, leading to more memory consumption, and duplicate relaying would possibly happen if the suppression scheme fails due to node mobility. In our analysis, we only consider the effect of node mobility as stated at the very beginning of this section. Then we look into the overhead due to duplicate relaying.

Duplicate relaying will be minimzed by giving the time factor and observing the location variance of forwarding candidates..Based on the ranking only limited candidates will receive the packets.

VI. CONCLUSION

In this paper, we address the problem of duplicate relaying in dynamic mobile ad hoc networks.

Constantly changing network topology makes conventional ad hoc routing protocols incapable of providing satisfactory performance. In the face of frequent link break due to node mobility, substantial data packets would either get lost, or experience long latency before restoration of connectivity. Inspired by opportunistic routing,. we propose minimization of data replica at forwarding candidates in MANET routing . The efficacy of the involvement of forwarding candidates against node mobility, as well as the overhead due toopportunistic forwarding is analyzed. Through simulation, we further confirm the effectiveness of POR: high packet delivery ratio is achieved while the delay and duplication are the lowest.

ACKNOWLEDGMENT

Students work is incomplete until they thank the almighty & his teachers. I sincerely believe in this and would like to thank Mr.M.V.Jagannatha Reddy, M.Tech(Ph.D), Associate prof., of the department of Computer Science & Engineering, MITS Madanapalle, for his encouragement, motivation and for guiding me to write this paper. Also I am grateful to Dr.A.Nagaraja Rao, M.Tech.,Ph.D. Prof., in CSE and Head Of the Department of Computer Science, MITS, Madanapalle .

REFERENCES

1. Z. Ye, S. Krishnamurthy, and S. Tripathi, "A framework for reliable routing in mobile ad hoc networks," in *INFOCOM 2003*, vol. 1, March-3 April 2003, pp. 270–280 vol.1.
2. A. Valera, W. Seah, and S. Rao, "Improving protocol robustness in ad hoc networks through cooperative packet caching and shortest multipath routing," *Mobile Computing, IEEE Transactions on*, vol. 4, no. 5, pp. 443–457, Sept.-Oct. 2005.
3. M. Mauve, A. Widmer, and H. Hartenstein, "A survey onposition-ba sed routing in mobile ad hoc networks," *Network, IEEE*, vol. 15, no. 6, pp. 30–39, Nov/Dec 2001.
4. "Location-aided opportunistic forwarding in multirate and multihop wireless networks," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 6, pp. 3032–3040, July 2009.
5. D. Chen, J. Deng, and P. Varshney, "Selection of a forwarding area for contention-based geographic forwarding in wireless multi-hop networks," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 5,pp. 3111–3122, Sept. 2007.
6. F. Wu, T. Chen, S. Zhong, L. E. Li, and Y. R. Yang, "Incentivecompatible opportunistic routing for wireless networks," in *MobiCom '08*, 2008, pp. 303–314.
7. R. Groenevelt, "Stochastic models for mobile ad hoc networks,"Ph.D. dissertation, Universite de Nice, Sophia Antipolis, France,2005.
8. S. Mueller, R. Tsang, and D. Ghosal, "Multipath routing in mobile ad hoc networks: Issues and challenges," *Lecture Notes in ComputerScience*, vol. 2965, pp. 209–234, 2004

◈◈◈

A Dynamic Watermarking Model for Medical Image Authentication

Priya. H. K & Anitha. S

Dept. of ECE, EWIT, Bangalore, India

Abstract – This paper proposes a dynamic watermarking model for the purpose of medical authentication. While transferring the data through a public network there is jittering or tampering of data. This is a matter of concern as any jitter or tampered data is not desirable in the medical field. It is noted that there is loss of life due to corrupted data received leading to wrong diagnosis. The proposed dynamic model proves that the medical image watermarked with the proposed system provides near lossless original image. Since the watermark is generated dynamically it is unique to the images considered therefore enhances the security of the images. The Proposed scheme is in the TIFF (Tagged Image File Format) using RGB colour space. The given watermark is embedded inside the image by expanding intraplane difference between any two colour planes of images.

Keywords - *Dynamic watermark, Integer Transform, Fundus Camera, Intra plane Expansion*

I. INTRODUCTION

Image processing is a fast developing field which can be used for the purpose of medical authentication. This research work presents a technique for the purpose of medical authentication. Authentication involves a challenge to determine whether the image and embedded data is received without any modification to the original image and embedded data

The embedded information watermarking techniques that embed information into a host image in a block-wise independent fashion is vulnerable to a vector quantization (VQ) counterfeiting attack [2-3]. Specifically, given a watermarked image, one can forge the watermark it contains into another image without knowing the secret key.

The digital fundus images are one particular class of medical images which has been chosen for simulation and analysis of the proposed scheme. These images are given in Tagged Image File (TIF) format in RGB colour. Correlation values are compared from different portions of the image, the technique enables us to distinguish malicious changes, such as replacing or adding features from no malicious changes resulting from common image processing operations space [1]

The proposed scheme dynamically generates the watermark using dynamic models. And, it is embedded inside the image by expanding intra plane difference between any two colour planes of images. It is known as intraplane difference expanding.

II. FRAMEWORK FOR DYNAMIC WATERMARKING SCHEME

The proposed scheme in this paper works in four stages.

- ➢ The first stage selects the reference color plane for generating watermark.
- ➢ The second stage uses the proposed dynamic model and generates the watermark using the reference color plane.
- ➢ The third stage involves embedding. This process is carried out using Integer transform.
- ➢ The fourth stage performs the extraction and verification process.

A. *Selection of Reference Colour Plane*

The green color is the selected reference plane to generate a watermark. A fundus camera is used which uses the special green filter for photograph of the fundus area. The image in the green channel contains all details along with other color plane.

B. *Watermark Using Dynamic Model*

The dynamic system is defined by the following equation,

$$x_{n+1=}f(x_n) \tag{1}$$

In this system the dynamic image changes with time. Though the behaviour is random it is deterministic.

These changes are very sensitive to the initial conditions. The sensitivity of the image increase exponentially to the growth of perturbations in the initial conditions. Therefore, the watermark is generated through dynamic system using the reference color plane as initial condition [1]. Thereby, the watermark is generated dynamically.

In the Proposed system, a hybrid optical bistable dynamic system is used which is defined by

$$f(x_n) = \sin^{(2)} (x_n - 2.5) \qquad (2)$$

C. Embedding by Intra Plane Difference Expanding

The embedding process is carried out using integer transform by using Intra Plane Difference Expanding. In the Embedding stage, the original imaging (I,J,K) is divided into colour planes. Here I denote number of rows, J denotes number of columns and K denotes number of planes. Since, the input image is in RGB (Red, Green, Blue) mode, k=3 in the proposed scheme. The green color plane will be used as seed to generate the watermark in messy system. Since, the watermark is generated dynamically; it will be unique to the images. Then, pixel pair is formed from the red and blue color planes of the images. By checking overflow and underflow condition for pixel pair, the watermark is embedded in the difference of the pixel pair by expanding the difference. This is known as intra-plane difference expanding.

The watermark is generated through dynamic system by using prominent pixel values of reference color plane of the image as seed. The initial values to the messy system is designed by

EffectiveKey (:,k1) = KeyIn((1+(k1-1)*M) : (((k1-1)*M)+M)); (3)

Where,(k1) refers the pixel values of reference color plane of the image. I refers embedding depth. The position information (pos) and secret key (key) is also used in the initial condition. The dynamic sequence is generated by substituting EffectiveKey (:,k1) value for Xn in Eqn.2.For the kth pixel the sequence is referred as EffectiveKey (:,k1), i=l, 2, 3 ...1. The reasonable number of iteration (I) is performed for the pixel to attain the dynamic status. This sequence contains floating numbers that is converted in to binary sequence in the proposed scheme. Hence, the thresholding T is introduced here to convert the sequence c_seq (k, i) from floating to binary sequence w (k, i). The w (k, i) is obtained by

$$W(k,i) = \begin{cases} 1 & c\_seq(k,i) > T \\ 0 & elsewhere \end{cases} \qquad (4)$$

Where, T is set to 8/3 by the number of test to bring equal number of zeros and ones. The length of sequence G is combined to one bit w (.) by applying XOR operation. Thus, the watermark is generated for the kth pixel. By repeating the same procedure for remaining pixels of the reference color plane of the image,the watermark is generated for the whole image for the Fig 1.

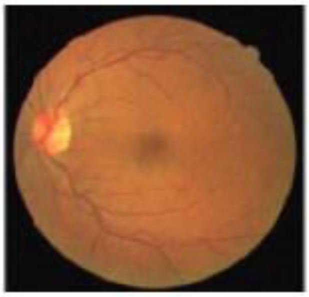


Fig.1 : Original Retina Image

Difference expansion transform is a remarkable breakthrough in reversible data-hiding schemes. The difference expansion method achieves high embedding capacity and keeps distortion low. The difference expansion method with the simplified location map and new expandability can achieve more embedding capacity while keeping the distortion at the same level as the original expansion method. This improvement can be possible by exploiting the quasi-Laplace distribution of the difference values.

Integer Transform: For a 8 bit gray scale pixel pair (x, y), 0::; x, y ::; 255, the integer transform is given by the pair (m, d).Where m refers integer average and d refers difference

$$M = \left\lfloor \frac{x+y}{2} \right\rfloor \qquad (5)$$

$$d = x-y \qquad (6)$$

The inverse transform is given by

$$x = m + \left\lfloor \frac{d+1}{2} \right\rfloor \qquad (7)$$

$$y = m - \left\lfloor \frac{d}{2} \right\rfloor \qquad (8)$$

Where L, J refers floor operation which rounds the value to nearest integer, in the integer transform, the difference (d) is modified based on the watermark bit (bit) to hide the bit into the pixel pair. The modification of difference (d ') is given by

$$d' = 2 * d + \text{bit} \tag{9}$$

The modification process checks two conditions. They are overflow and underflow. It is done to ensure that the difference is expandable or not. The expandable difference should satisfy the following condition.

$$|d| \le 2 * (255 - m) \;\; \text{if } 128 \le m \le 255$$

$$|d| \le 2 * m + 1 \;\; \text{if } 0 \le m \le 127 \tag{10}$$

Only expandable difference can be used for embedding. If all the expandable differences are used, the capacity will reach its limit. Let N and Ne denote the number of differences and the number of expandable differences, respectively. The hiding capacity of an image is defined as:

$$c = \frac{Ne}{N} \tag{11}$$

D. *Extraction and Verification*

In the extraction process, the watermarked image is processed in the same way as original image processed for embedding. The extraction process is complete blind. Both original image and original watermarks are not used for the extraction process. Extraction process produces the reference sequence using messy system and green color plane as seed. The embedded watermark is extracted by applying inverse integer transform using Eqn. 6 and Eqn.7. Where, the LSB (Least Significant Bit) of the difference value gives the embedded watermark bit.

The reference sequence and the extracted watermark sequence are compared to check that whether the given volume of the image is tampered or not. The difference between reference sequence and the extracted watermark sequence will show the tampered volumes in the image as shown in fig.3. Thus, the extraction process works in complete blind way and enhances the security. The extraction process in this paper is reversible. It means that the original image should be retrieved without any loss after removing the watermark at the extraction stage. The medical images were exchanged from one place to another for diagnosis purposes. Hence, the loss in the quality of images is not accepted here.

Results and Snapshots

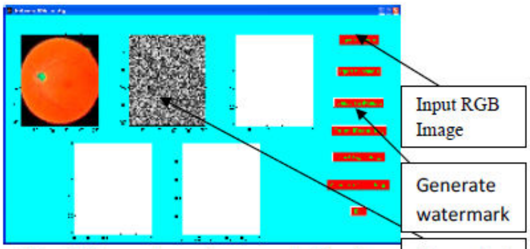


Fig. 2 The watermark generated for the whole image

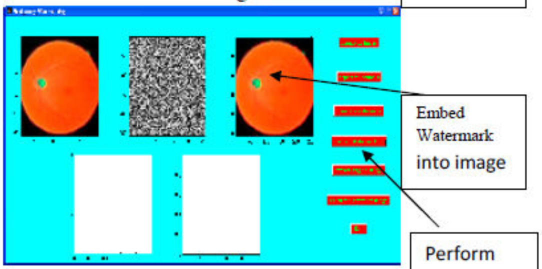


Fig. 3 The watermarked image (Wimg) generated for the given original image (Img).

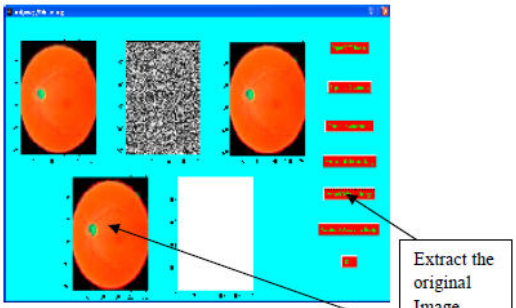


Fig.4 Tampered Image

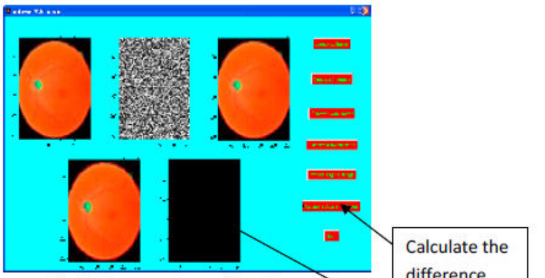


Fig 5 Error watermarking image

Conclusions

III. CONCLUSIONS

- Test results performed on retina image i.e. JPEG format, RGB color mode confirms that we can authenticate whether the received image is with or without any modification.
- Enhances the security of medical image.
- By the test results we can conclude that "A Dynamic watermarking Model for Medical Image Authentication" can be used to detect authenticity of received image precisely.

REFERENCES

[1] S.Poonkuntran, R.S.Rajesh," A Messy Watermarking for Medical Image Authentication", IEEE 2011

[2] M.Wu, B. Liu,"Watermarking for image authentication", in:Proceedings of the IEEE International Conference on Image Processing, Chicago, IIIinoise,US, (1998), pp. 437-441.

[3] M. Holliman, N. Memon, "Counterfeiting attacks on oblivious blockwise independent invisible watermarking schemes", IEEE Trans. Image Process. 9 (2000) 432-441.

[4] M.U. Celik, G. Sharma, E. Saber, AM. Tekalp, "Hierarchical watermarking for secure image authentication with localization", IEEE Trans. Image Process. 11 (2002) 585-595.

◈◈◈

A Novel Method for Watermarking using Opaque & Translucent Methods in Videos

Kuna Dhilli & Y. Rajyalakshmi

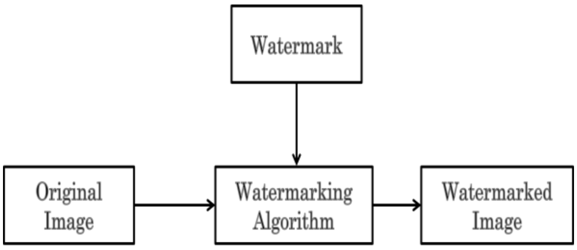
Dept. of E.C.E., SSCE, Chilakaplem Jn, Srikakulam, India
E-mail : dhillikuna@gmail.com, gnraoaitam@gmail.com

Abstract – Watermarking is a popular technique for discouraging illegal copyright and distribution of copyrighted digital image information. One of the important features of the watermarking technique is the lossless visible watermarking, which will preserve the quality of the watermark and watermarked image. A novel method for generic visible watermarking with a capability of lossless video image recovery is proposed. The method is based on the use of deterministic one-to-one compound mappings of image pixel values for overlaying a variety of visible watermarks of arbitrary sizes on video images. The compound mappings are proved to be reversible, which allows for lossless recovery of original images from watermarked images. The mappings may be adjusted to yield pixel values close to those of desired visible watermarks. Various types of visible watermarks, including opaque Monochrome and translucent full color ones, are embedded as applications of the proposed generic approach. A two-fold monotonically increasing compound mapping is created and proved to yield more distinctive visible watermarks in the watermarked image. Security protection measures by parameter and mapping randomizations have also been proposed to deter attackers from illicit image recoveries. Videos are playing key role in broadcasting now. Experimental results demonstrating the effectiveness of the proposed approach are also included in videos.

Keywords - *Lossless reversible visible watermarking, mapping randomization, one-to-one compound mapping, parameter randomization, translucent watermark.*

I. INTRODUCTION

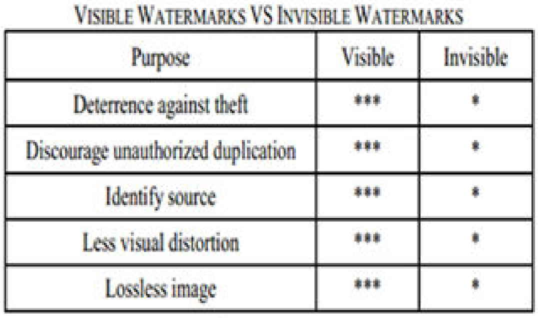
Copy right Protection of intellectual properties has become an important topic with the advance of computer technologies and the proliferation of the Internet which have made reproduction and distribution of digital information easier than ever before. Digital Watermarking is one of the ways for copyright protection. Embedding of certain specific information about the copyright holder (company logos, owner-ship descriptions, etc.) into the media to be protected is called Digital Watermarking.



Watermark embedding process

Digital watermarking methods for images are usually categorized into two types: *invisible* and *visible*. The invisible type of digital watermarking aims to embed copyright information imperceptibly into host media such that in cases of copyright infringements, the hidden information can be retrieved to identify the ownership of the protected host.

A watermarked image must be resistant to common image operations to ensure that the hidden information is still retrievable after such alterations. Methods of the visible type, on the other hand, yield visible watermarks which are generally clearly visible after common image operations are applied. In addition, visible watermarks convey ownership information directly on the media and can determined attempts of copyright violations. For the security reasons we are using visible watermarking in broadcasting.



VISIBLE WATERMARKS VS INVISIBLE WATERMARKS

Purpose	Visible	Invisible
Deterrence against theft	***	*
Discourage unauthorized duplication	***	*
Identify source	***	*
Less visual distortion	***	*
Lossless image	***	*

Number of "*" means the degree of importance

Watermarks' embedding, either visible or invisible, results in the degradation of the quality of the host media in general. A group of techniques, named *reversible* watermarking, allow legitimate users to remove the embedded watermark and restore the original content as needed. However, not all reversible watermarking techniques guarantee *lossless image recovery*, which means that the recovered image is identical to the original, pixel by pixel. Lossless recovery is important in many applications where serious concerns about image quality arise. Some examples include forensics, medical image analysis, historical art imaging, or military applications.

There are relatively few mentions of lossless visible watermarking in the literature compared with their invisible counterparts. In the past several lossless invisible watermarking techniques have been proposed. The most common approach is to compress a portion of the original host and then embed the com-pressed data together with the intended payload into the host. The other approach is to superimpose the spread-spectrum signal of the payload on the host so that the signal is detectable and removable. The third approach to embed a bit of information by manipulating a group of pixels as a unit. Although one may use lossless invisible techniques to embed re-movable visible watermarks, the low embedding capacities of these techniques obstruct the possibility of implanting large-sized visible watermarks into host media.

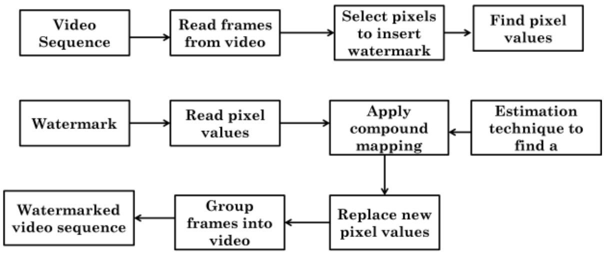
The most common approach for lossless visible watermarking is to embed a monochrome watermark using reversible and deterministic mappings of coefficients or pixel values in the watermark region. The other method for this is to embed a visible watermark by rotating consecutive watermark pixels. One advantage of these approaches is that watermarks of arbitrary sizes can be embedded into any host image. However, using these approaches, only *binary* visible watermarks can be embedded which is too restrictive since most company logos are colorful.

In this paper, a novel method for lossless visible watermarking is proposed by using appropriate *compound mappings* that allow mapped values to be controllable. For lossless recovery of the original image the mappings are proved to be *reversible*. There is a possibility of embedding different types of visible watermarks into cover images, an approach called *generic*. Two applications of the proposed method are demonstrated, where opaque monochrome watermarks and non-uniformly translucent full-color ones are respectively embedded into colorimages. The compound mappings which are more specific are also created and proved to be able to yield visually more distinctive visible watermarks in the water marked image.

In the remainder of this paper, the proposed method for deriving one-to-one compound mappings is described in section II. Related lemmas and theorems are also provided and security protection measures described. Applications of the proposed method for embedding opaque monochrome and translucent color watermarks into color images are described. The specific compound mapping for yielding more distinctive visible watermarks is described. Experimental results are presented to demonstrate the effectiveness of the proposed method finally; a conclusion with some suggestions for future work is also included.

II. PROPOSED NEW APPROACH TO LOSSLESS VISIBLE WATERMARKING

This section deals with the proposed approach to lossless reversible visible watermarking, based on which appropriate one-to-one compound mappings can be designed for embedding different types of visible watermarks into images. From the resulting watermarked image, the original image can be recovered without any loss by using the corresponding reverse mappings.



Reversible One-to-One Compound Mapping

First, we propose a generic *one-to-one compound mapping f* for converting a set of numerical values, P={p1,p2,…….,pm} to another set Q={q1,q2……qm} such that the respective mapping from p_i to q_i for all i=1,2….., M is *reversible*.

Here, for the copyright protection applications investigated in this study, all the values p_i and q_i are image pixel values (grayscale or color values). The compound mapping f is governed by a one-to-one function F_X with one parameter x=a or b in the following way:

$$q = f(p) = F_b^{-1}(F_a(p)) \tag{1}$$

Where F_x^{-1} is the inverse of which, by the one-to-one property, leads to the fact that if $F_a(p) = p'$, then $F_a^{-1}(p') = p$ for all values of a and p. On the other hand, $F_a(p)$ and $F_b(p)$ generally are set to be *unequal* if $a \neq b$. The compound mapping described by (1) is indeed *reversible*, that is p, can be derived exactly from q using the following formula:

$$p = f^{-1}(q) = F_a^{-1}(F_b(q)) \qquad (2)$$

as proved below.

Lemma 1 *(Reversibility of Compound Mapping):*

If $q = F_b^{-1}(F_a(p))$ for any one-to-one function F_x with a parameter x, then $p = F_a^{-1}(F_b(q))$ for any values of a, b, p and q.

***Proof*:**

Substituting (1) into $F_a^{-1}(F_b(q))$, we get $F_a^{-1}(F_b(q)) = F_a^{-1}\left(F_b\left(F_b^{-1}(F_a(p))\right)\right)$.

By regarding $F_a(p)$ as a value c, the right-hand side becomes $F_a^{-1}(F_b(F_b^{-1}(c)))$, which, after F_b and F_b^{-1} are cancelled out, becomes $F_a^{-1}(c)$. But $F_a^{-1}(c) = F_a^{-1}(F_a(p))$, which just p is after F_a and F_a^{-1} are cancelled out. That is, we have proved $p = F_a^{-1}(F_b(q))$.

As an example, if $F_x(p) = xp + d$, then $F_x^{-1}(p') = (p' - d)/x$.

Thus

$$q = F_b^{-1}(F_a(p)) = F_b^{-1}(ap + d)$$
$$= (ap + d - d)/b = ap/b$$

and so, we have

$$F_a^{-1}(F_b(q)) = F_a^{-1}(b(ap/b) + d) = F_a^{-1}(ap + d)$$
$$= [((ap + d) - d)/a] = (ap/a) = p$$

as expected by Lemma 1.

Lossless Visible Watermarking Scheme based on Lemma 1, we will now derive the proposed generic lossless visible watermarking scheme in the form of a class of one-to-one compound mappings, which can be used to embed a variety of *visible watermarks* into images. The watermark can be removed to recover the original image losslessly. This makes the embedding reversible. For this aim, a preliminary lemma is first described as follows.

Lemma 2 *(Preference of Compound-Mapped Value):*

It is possible to use the compound mapping $q = F_b^{-1}(F_a(p))$ to convert a numerical value to another value close to a *preferred* value l.

Proof:

Let $F_x(p) = p - x$ where x is the parameter for F. Then $F_x^{-1}(p') = p' + x$. Also, let a=p- ε and b=l, where ε is a small value. Then, the compound mapping $F_b^{-1}(F_a(p))$ of p yields q as

$$q = F_b^{-1}(F_a(p)) = F_b^{-1}(p - a) = F_b^{-1}(\varepsilon)$$
$$= \varepsilon + b = \varepsilon + l$$

which means that the value q is close to the preferred value l. The above lemma relies on two assumptions. The first is that is a is close to p, or equivalently, that a=p- ε. The reason why we derive the above lemma for instead of for a=p is that in the reverse mapping we want to recover p from q *without knowing p*, which is a requirement in the applications of reversible visible watermarking investigated in this study.

Although the value of p cannot be known in advance for such applications, it can usually be estimated, and we will describe some techniques for such estimations in the subsequent sections.

The second assumption is that $F_x(p)$ yields a small value if a and p are close. Though the basic difference function $F_x(p) = p - x$ used in the above proof satisfies this requirement for most cases, there is a possible problem where the mapped value may exceed the range of valid pixel values for some values of a, b and p. For example, when a=255, b=255 and p=253, we have q=255-253+255=257>255. It is possible to use the standard modulo technique (i.e., taking q=257 mod 256=1) to solve this issue; however, such a technique will make q far from the desired target value of b, which is 255. Nevertheless, we will show in that using such a standard modulo function, $F_x(p) = (p - x)_{\text{mod } 256}$, can still yield reasonable experimental results. Furthermore, we show in a more sophisticated one-to-one function that is free from such a wraparound problem.

Security Considerations

The Advance of computer technologies and the proliferation of the internet have made reproduction and distribution of digital information easier than ever before. Copyright protection of intellectual properties has, therefore become an important concern. As mentioned previously, although we want legitimate users to be able to recover the original image from a watermarked one, we do not want an attacker to be able to do the same. Herein, we propose some security protection measures against illicit recoveries of original images.

First, we make the parameters and in the above algorithms to be dependent on certain secret keys that are known only by the creator of the watermarked image and the intended receivers. One simple technique to achieve this is to use a secret key to generate a pseudo-random sequence of numerical values and add them to either or both of a and b for the pixels in the watermarking area. This technique is hereinafter referred to as *parameter randomization*.

Another way of security protection is to make the choices of the *positions* for the pixels to be dependent on a secret key. Specifically, we propose to process *two* randomly chosen pixels (based on the security key) in P simultaneously as follows. Let the two pixels be denoted as X1 and X2 with values p1 and p2, respectively.

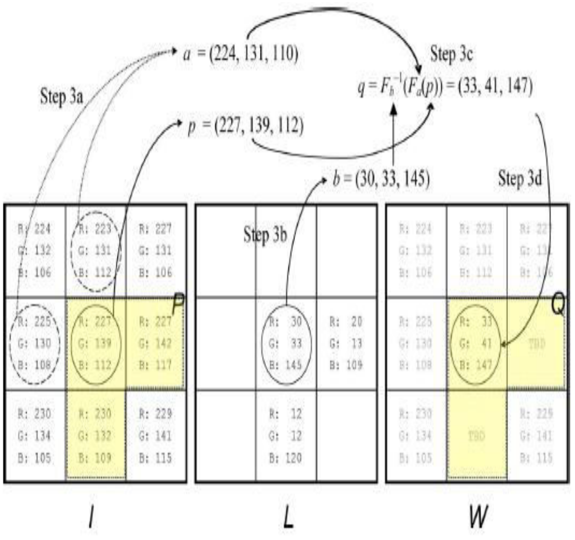


Fig. 1. Illustration of mapping the center pixel of a 3 x 3 image. Only the mapping of the center pixel is shown for clarity; the east and south pixels are depicted as TBD (to be determined) in W.

The color estimates a1and a2 corresponding to X1 and X2, respectively, are individually derived as before using their respective neighbors. The parameters b1 and b2 are set to be the values l1 and l2 of the respective watermark pixels y1 and y2.

Then, instead of setting the values of the watermarked pixels Z1 and Z2 to be $q_1 = F_{b_1}^{-1}(F_{a_1}(p_1))$ and $q_2 = F_{b_2}^{-1}(F_{a_2}(p_2))$ as before, we *swap* the parameters and set $q_1 = F_{b_1}^{-1}(F_{a_2}(p_2))$ and $q_2 = F_{b_2}^{-1}(F_{a_1}(p_1))$.

This parameter exchange does not affect the effectiveness of lossless recoverability, because we can now recover the original pixel values by the following compound mappings:

$$p_1 = F_{a_1}^{-1}(F_{b_2}(q_2)) \quad \text{and} \quad p_2 = F_{a_2}^{-1}(F_{b_1}(q_1)).$$

We will refer to this technique in the sequel as *mapping randomization.* We may also combine this technique with the above mentioned parameter randomization technique to enhance the security further. Last, the position in the image where a watermark is embedded affects the resilience of the watermarked image against illicit image recovery attempts. In more detail, if the watermark is embedded in a smooth region of the image, an attacker can simply fill the region with the background color to remove the watermark irrespective of the watermarking technique used. To counter this problem, an appropriate position should be chosen, using, for example, the adaptive positioning technique when embedding a watermark. However, for ease of discussions and comparisons, we always embed a watermark in the lower right-hand corner of an image in this study.

III. LOSSLESS VISIBLE WATERMARKING OF OPAQUE MONOCHROME WATERMARK

We describe now how we embed a losslessly-removable opaque monochrome watermark L into a color image I such that the watermark is *visually distinctive* in the watermarked image W, as an application of the proposed generic approach to lossless visible watermarking.

First, we denote the sets of those pixels in I corresponding spatially to the *black* and *white* pixels in L by P and P' , respectively. An illustration of such areas of p and P' is shown in Fig. 2. We define Q and Q' in a similar way for the watermarked image W, which correspond to P and P' , respectively. Then, we adopt the simple one-to-one function $F_a(p) = p - a$, and use the same pair of parameters a and b for *all* mappings of pixels in P. Also, we apply the "modulo-256" operation to the results of all computations so that they are within the valid range of color values. Our experiments show that this method still yields reasonable results.

For the values of parameters a and b, we set a to be the *average* of the color component values of the pixels in P' . This average value presumably is close to the value of pixel in P, satisfying the condition $a = p - \varepsilon$ mentioned previously. To ensure that the watermark is distinctive in W, we do not simply embed black values for pixels in watermarking area P (that is, we do not embed for l=0 for P), but set l to be a value which is distinctive with respect to the pixel colors in the surrounding region P'. To achieve this, we set $b = l = a + 128$, which is a value *distinctive* with respect to a. As a result, the value of a pixel in Q, according to Lemma2, becomes $q = F_b^{-1}(F_a(p)) = b + \varepsilon = a + 128 + \varepsilon$, meaning that the pixel values of Q are also distinctive with respect to those of the surrounding pixels in Q' as desired.

Since both a and b are derived from P' during watermark embedding, the exact same values of a and b can be derived during watermark removal because Q' is identical to P'. The original image can, therefore, be recovered without any loss using Algorithm 2.

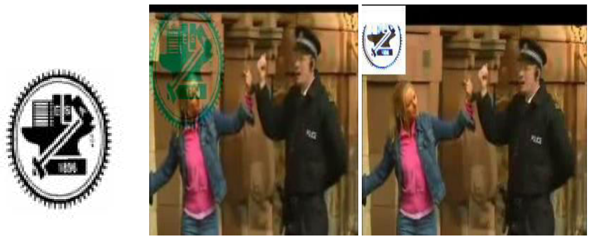


Fig. 2 : Illustration of images in watermark. (a) Logo (b) Opaque (c) Translucent

We embedded the watermark of Fig. 2(a) into the video frames, respectively through which we have demonstrated the effectiveness of the proposed method, in one of our experiments. We applied both the mapping randomization and the parameter randomization techniques described in Section 2, for security protection. Specifically, for the latter technique we added random integer values in the range of -12 to +12 to the parameter. The images recovered by using correct keys for the parameter b. That the embedded opaque watermarks are distinctive with respect to their surroundings and can be removed completely when the input key is correct. On the contrary, when the key was incorrect, the inserted watermark cannot be removed cleanly, with noise remaining in the watermarking area. The videos are converted to frames, frames are treated as image and same methods are applied.

IV. LOSSLESS VISIBLE WATERMARKING OF TRANSLUCENT COLOR WATERMARKS

As another application of the proposed approach, we describe now how we embed more complicated *translucent color watermarks*. A translucent color watermark used in this study is an arbitrary RGB image with each pixel being associated with an *alpha component value* defining its *opacity*. The extreme alpha values of 0 and 255 mean that the watermark pixel is *completely transparent* and *totally opaque*, respectively. A translucent full-color watermark is visually more attractive and distinctive in a watermarked image than a traditional transparent monochrome watermark, as mentioned previously. Such a kind of watermark can better represent trademarks, emblems, logos, etc., and thus is more suitable for the purpose of advertising or copyright declaration.

V. CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

In this paper, a new method is proposed for reversible visible watermarking with lossless image recovery capability. The method uses one-to-one compound mappings that can map image pixel values to those of the desired visible watermarks. Relevant lemmas and theorems are described and proved to demonstrate the reversibility of the compound mappings for lossless reversible visible watermarking. The compound mappings allow different types of visible watermarks to be embedded, and two applications have been described for embedding opaque monochrome watermarks as well as translucent full-color ones. A translucent watermark is clearly visible and visually appealing, thus more appropriate than traditional transparent binary watermarks in terms of advertising effect and copyright declaration. The two-fold monotonically increasing property of compound mappings was defined and an implementation proposed that can provably allow mapped values to always be close to the desired watermark if color estimates are accurate. We have also described parameter randomization and mapping randomization techniques, which can prevent illicit recoveries of original images without correct input keys. Experimental results have demonstrated the feasibility of the proposed method and the effectiveness of the proposed security protection measures. Future research may be guided to more applications of the proposed method and extensions of the method to other data types other than bitmap images, like in JPEG images and MPEG videos.

REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, “Information hiding—A survey,” Proc. IEEE, vol. 87, no. 7, pp. 1062–1078, Jul. 1999.

[2] N. F. Johnson, Z. Duric, and S. Jajodia, Information Hiding. Steganography and Watermarking—Attacks and Countermeasures,

sBoston, MA: Kluwer, 2001.

[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” IEEE Trans. Image Process. , vol. 6, no. 12, pp. 1673–1687, Jun. 1997.

[4] M. S. Kankanhalli, Rajmohan,and K.R. Ramakrishnan, “Adaptive visible watermarking of images,” in Proc. IEEE Int. Conf. Multimedia Computing and Systems , 1999, vol. 1, pp. 568–573.

[5] Y.HuandS.Kwong,“Wavelet domain adaptive visible watermarking,” Electron. Lett. , vol. 37, no. 20, pp. 1219–1220, Sep. 2001.

[6] S. P. Mohanty, K. R. Ramakrishnan, and M. S. Kankanhalli, “A DCT domain visible watermarking technique for images,” in Proc. IEEE Int. Conf. Multimedia and Expo , Jul. 2000, vol. 2, pp. 1029–1032.

[7] G. Braudaway, K. A. Magerlein, and F. Mintzer, “Protecting publicly available images with a visible image watermark,” in Proc. SPIE Int. Conf. Electronic Imaging , Feb. 1996, vol. 2659, pp. 126–133.

[8] Y. J. Cheng and W. H. Tsai, “A new method for copyright and integrity protection for bitmap images by removable visible watermarks and irremovable invisible watermarks,” presented at the Int. Computer Symp.—Workshop on Cryptology and Information Security, Hualien, Taiwan, R.O.C., Dec. 2002

[9] P. M. Huang and W. H. Tsai, “Copyright protection and authentication of grayscale images by removable visible watermarking and invisible signal embedding techniques: A new approach,” presented at the Conf. Computer Vision, Graphics and Image Processing, Kinmen, Taiwan, R.O.C., Aug. 2003.

[10] Y. Hu, S. Kwong, and J. Huang, “An algorithm for removable visible watermarking,” IEEE Trans. Circuits Syst. Video Te chnol. , vol. 16, no. 1, pp. 129–133, Jan. 2006.

[11] Y. Hu and B. Jeon, “Reversible visible watermarking and lossless recovery of original images,” IEEE Trans. Circuits Syst. Video Te chnol. , vol. 16, no. 11,pp. 1423–1429, Nov. 2006.

[12] B. Macq, “Lossless multiresolution transform for image authenticating watermarking,” presented at the European Signal Processing Conf., Tampere, Finland, Sep. 2000.

◈◈◈

Real Time Earthquake Prediction Algorithm Through Spatial-Temporal Outlier Detection

Dipika Kalyani & Setu Kumar Chaturvedi

Computer Technology and Application, Technocrats Institute of Technology, Bhopal (M.P), India
E-mail : dipika.mtech@gmail.com, setukchaturvedi@gmail.com

***Abstract* -** This paper investigates various precursors of the earthquake. The purpose of this work is to discuss and analyze these precursors and generate the Earthquake Model for predicting earthquake. In this paper, we will develop a promising algorithm for predicting earthquake and implement outlier detection techniques. The model proposed for earthquake prediction is based on detecting spatio-temporal outliers.

***Keywords* -** *earthquake precursors; spatiotemporal outliers; outlier detection techniques*

I. INTRODUCTION

Among all disasters, earthquake is considered as a natural disaster that includes more dangers therefore it is more noteworthy. Earthquake is always counted as a threat to species life. All distinct natural events such as big forest fires, earthquakes and volcanic activities, traffic accidents, hurricanes and floods can be regarded as spatio-temporal outliers. **Spatio-temporal outlier** is an instance whose non-spatial attribute is significantly different from its spatio-temporal neighborhood. To have a better understanding or for better modeling of the spatial phenomena, STOs should be detected [1]. STO detection can be examined under distance, wavelet analysis, clustering, visualization and graph based approaches. Clustering approach detects STOs as instances which are not lying in any cluster.

II. FEATURE EXTRACTIONS

In efforts to forecast earthquakes, seismologists have investigated the parameters which show anomalous variation in their values such as seismicity patterns, crustal movements, ground water level in wells, radon or hydrogen gas emissions from the Earth, changes of seismic wave velocities, electromagnetic fields, largescale changes in soil temperature, and changes in ion concentration in the ionosphere. The International Association of Seismology and Physics of the Earth's Interior appointed a subcommission which defined **precursor** as a quantitatively measurable change in an environmental parameter that occurs before main shock and is thought to be linked to the preparation of the main shock [3]. Various precursors which we will use in this work are explained below.

TABLE1. CLASSIFICATION OF PRECURSORS

S. No	Types of Precursors	Precursors
1	Seismic	Preshocks
2	Borehole	Water level changes
3	Hydrochemical	Radon concentration in water changes
4	Geodetic	Strain rate changes
5	Geophysical	Ground Tilting

These are the features used for earthquake prediction:

A. Seismicity Patterns - Preshocks

Preshocks and other patterns of seismicity are an important class of physical phenomena that can be readily observed by seismic networks and have been extensively investigated as precursors. A preshock is just an earthquake in a sequence that happens to be proximate to, but precedes, a larger event. Preshocks are medium-sized earthquakes that precede major quakes. An increase in seismic activity in an area has been used as a method of predicting earthquakes.

B. Water Level Changes

The borehole precursor is fluctuations in underground water level. Evidence from multiple studies indicates that earthquakes alter water levels. Changes in groundwater occurs pre-, coand postseismically and many researchers attempt to determine if these alterations can be used to predict major events. The changes in underground water level are due to the deformation of the ground water system caused by crustal movement resulting in the tilting, expansion, or contraction of the ground. After the earthquake, the underground water level would return to normal.

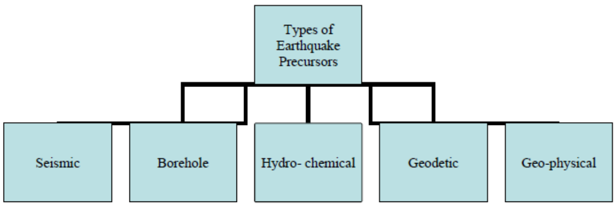


Fig. 1: Types of earthquake precursors

C. Changes in Radon Concentration

Radon is a rare gas with atomic number 86, and has 28 isotopes. Variation of radon concentration in ground water, as a tracer, has been known as a means of earthquake prediction. Recently, several typical methods have been proposed for the measurement of radon concentration. Units for measuring radon concentration are Becquerel per liter and picocuries per liter.

Two different devices were used for the measurements:

- an electrostatic radon detector i.e. RAD-7
- a pulse ionization chamber named AlphaGUARD,

Monitoring of radon concentration has been found to have a great potential as a reliable precursor.

D. Strain Rate Changes

Change in the strain rate is another very important precursor for earthquake prediction. A **strainmeter** is an instrument used by geophysicists to measure the deformation of the earth. Linear strainmeters measure the changes in the distance between two points, using either a solid piece of material over a short distance or a laser interferometer. Strainmeter records show signals from the earth tides, and seismic waves from earthquakes. The most extensive network of strainmeters is installed in Japan; it includes mostly quartz-bar instruments in tunnels and borehole strainmeters, with a few laser instruments. Borehole strainmeters with resolution exceeding a part per billion — would need to be within a few kilometres of the impending earthquake's epicentre to detect this aseismic strain.

E. Ground Tilting

The geophysical precursor is ground uplift and tilt. Tiltmeters can be used to measure the amount of change in slope on the surface. A tiltmeter is an instrument that measures its own rotation and, therefore, the rotation of the structural element or portion of ground to which it is connected. **Tiltmeters** are instruments used to measure the change in elevation or slope on the surface. Originally designed as part of the guidance and control system for military missiles, a variety of electronic tiltmeters are now available [6]. RST tiltmeters are available in either portable or inplace versions utilizing force balanced servo-accelerometer to measure in either one or two axial planes to the surface of the base plate. A water-tube tiltmeter can also be used to measure ground uplift and tilt.

III. EARTHQUAKE DATA AND DETECTION TASK

The data are collected by a network of sensors that use a number of physical principles to detect earthquake. These quantities are normally observed and aggregated over a time period. The detection task is to observe the data stream and check when the readings indicate an earthquake [5]. The time when the sensors detect the anomaly depends on the deviation of data recorded from threshold. In this section, we look at the available data and define the earthquake detection task.

The values that do not follow the characteristic distribution of the data are referred as outliers. Outliers are not necessarily error values as they can indicate unusual behavior within the underlying process and highlight anomalies. Identification of outliers is one of the main tasks in data mining. *Outlier analysis* is a mathematical concept, whose main useful role is to extract the most similar or dissimilar separated sets of objects according to a given similarity or dissimilarity measure [7]. Nowadays outlier detection and other feature extraction algorithms are recognized as important tools for revealing coherent features in the earth sciences and in data mining. Depending on the data structures and goals of classification, different outlier detection methods must be applied because:

- Each method has different measures to detect outliers.
- Expected outlier percentages change differently according to the sample size or distribution type of the data.

There are two kinds of outlier detection methods: formal and informal methods. Most **formal methods** are usually based on assuming some well-behaving distribution, and test if the target extreme value is an outlier of the distribution, i.e., weather or not it deviates from the assumed distribution. Even though formal tests are quite powerful under well-behaving statistical assumptions such as a distribution assumption, most distributions of real-world data may be unknown or may not follow specific distributions such as the normal, gamma, or exponential. Another limitation is that they are susceptible to masking or swamping problems [8].

On the other hand, **informal methods** generate an interval or criterion for outlier detection instead of hypothesis testing, and any observations beyond the interval or criterion is considered as an outlier. In our study we are using informal methods because informal methods find the extreme values away from the majority of the data regardless of the distribution. Most informal methods present the interval using the location and scale parameters of the data. In our work we will use two different informal outlier detection methods:

1. Sigma Approach
2. Median Absolute Deviation method

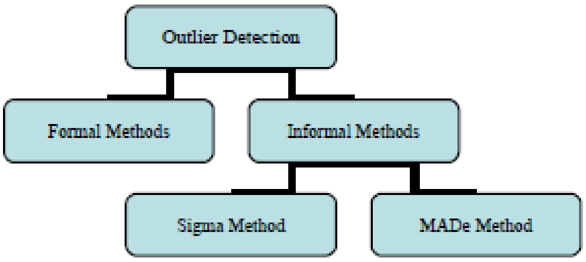


Fig. 2 : Classification of Outlier Detection methods

IV. PROPOSED APPROACH

The proposed approach includes spatial-temporal data mining and outlier detection technique. To detect earthquake, the system will generate spatial-temporal earthquake models based on various precursors and then identify the outliersbased on comparing real time data with the historical models. The earthquake model is a model with incremental learning ability. The outlier will illustrate the possibility of earthquake [2]. At first, all the historical precursory data are collected from the archive. Then each time for any location the precursor values are received from the sensors, a comparison is made between the vector of current data from all sensors and the vector of average data from the model. If the current vector varies greater than a predetermined threshold range of earthquake model, it reports possibility of earthquake.

Considering the correlation of the sensors, Sigma Approach is used to measure the deviation in the earthquake data. This is a simple classical approach to screen outliers. With the sigma approach, a value is identified as outlier if it lies outside the mean by + or – “x” times sigma. Where x is an integer and sigma is standard deviation for the variable. Sigma approach helps to identify the probability of occurrence of a data point, which will be used to determine the outlier.

Second method we used in our work for earthquake prediction is Median Absolute Deviation. The MADe method calculates median and Median Absolute Deviation (MAD). It is one of the basic robust methods which are largely unaffected by the presence of extreme values of the data set. This approach is similar to the sigma method. However, the median and MADe are employed in this method instead of the mean and standard deviation. The MADe method is defined as follows;

2 MADe Method: Median ± 2 MADe

3 MADe Method: Median ± 3 MADe,

where MADe=1.483×MAD for large normal data.

MAD is an estimator of the spread in a data, similar to the standard deviation, but has an approximately 50% breakdown point like the median.

MAD= median (|xi – median(x)| i=1,2,…,n)

When the MAD value is scaled by a factor of 1.483, it is similar to the standard deviation in a normal distribution. This scaled MAD value is the MADe.

The earthquake detection process can be divided into four steps as listed below.

Step 1: Pre-processing

In this task, the raw data retrieved from sensors is cleaned and organized for the mining stage. As the data received from detectors contains noise and missing values caused by malfunction of the detector or sensors, data cleaning must be performed to identify and remove such data to assure the data quality.

Step 2: Earthquake Model Generation

In the model generation task, earthquake data are analyzed to construct earthquake model. The earthquake model is then generated as an average of the historical data. Thus this step calculates the average value for earthquake data and threshold range. All this calculation is done through sigma and MADe methods.

Step 3: Detection

This task is executed in real-time to discover potential earthquake possibility based on the earthquake model. The earthquake data are collected from the detectors and are cleaned in runtime. Detection is performed by checking the deviation of real-time data from the upper and lower thresholds in the earthquake model. If the data is out of threshold range, a possibility of earthquake is identified.

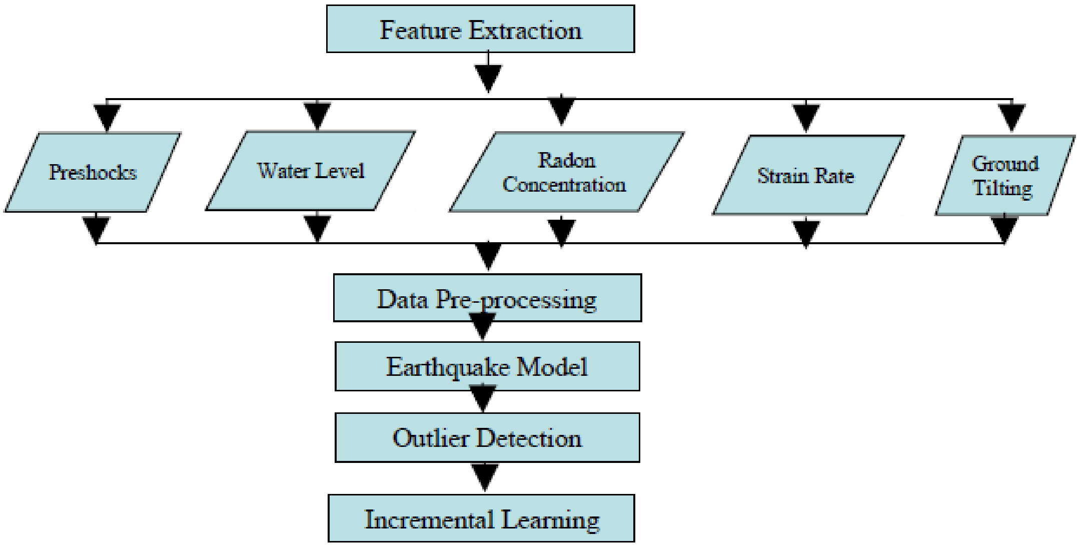


Fig. 3 : Flow of Algorithm

.

Step 4: Incremental Learning Model

Studies in India and elsewhere suggest that earthquakes are chaotic in nature. The technique for earthquake prediction can therefore be based on a outlier approach. The system requires the ability to dynamically learn from incoming earthquake data to be adaptable. The continuously coming earthquake data without true incidents will be used to refine the model. If the true earthquake is verified, the precursory data collected during the earthquake period will not be used in the model. By merging the new precursory data of false earthquake prediction into the original model, the earthquake model can be refined and the accuracy of detection can be improved.

V. METHODOLOGY ANALYSIS

All the data used for the analysis is stored in Microsoft SQL Server 2005; its management software is developed with Microsoft Visual Studio 2008. The database consists of data obtained from sensors i.e. data describing precursor values. Microsoft Visual Studio 2008 is used as the environment for development of application in VB.Net. Connection to the SQL Server 2005 is established to fetch or update the data. Sigma approach and MADe method is implemented with stored procedures and functions written in Transact-SQL for Microsoft SQL Server 2005.

We have presented a simple system that performs earthquake prediction and our results suggest that the method works effectively. In the experiment presented here we used two spatiotemporal outlier detection techniques by concentrating on few earthquake precursors. In the proposed earthquake prediction approach, the value of the precursors greatly impacts the effectiveness of prediction. We do not use any standard software package. Our goal is to construct a simple system for data-mining, which allows one to match the most appropriate outlier detection schemes on the structure of actual precursory data. Our data-mining techniques, include not only various outlier detection algorithms but also feature extraction. This present approach is different from the earlier work done in this field.

VI. TEST RESULT

The performance of an earthquake prediction system is determined on two levels: data collection and data processing algorithms. Data collection refers to the detection technologies that are used to obtain earthquake data. Data processing refers to the algorithms used for detecting and classifying earthquake by analyzing the precursor values obtained from detectors for predicting the occurrence of an earthquake. Combined, the two levels provide a technical platform on which a variety of algorithms can be designed and applied. The “mixing and matching” of data collection technologies and data processing methodologies results in a variety of solutions for earthquake detection.

The experimental results show that the method works effectively to detect changes in values of precursors based on training data Moreover, most algorithms are established based on the measurements from detectors, partially because detector systems have been the most widely used nationwide and are of relatively low cost compared to other detection technologies. The STO detection approach we proposed i.e sigma and MADe methods both used sensors and detectors for data collection phase. This outlier analysis helps greatly in detecting subtle earthquakes, which escape the classical techniques.

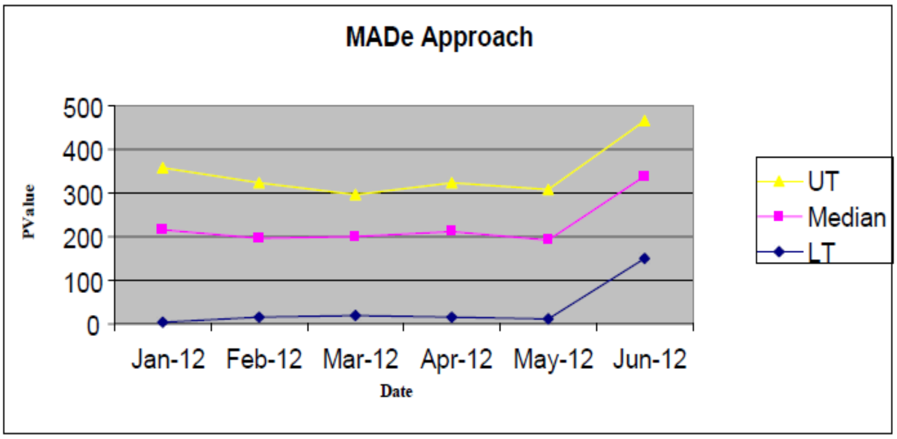


Fig. 4 : Upper Threshold, Mean And Lower Threshold Using Made Approach.

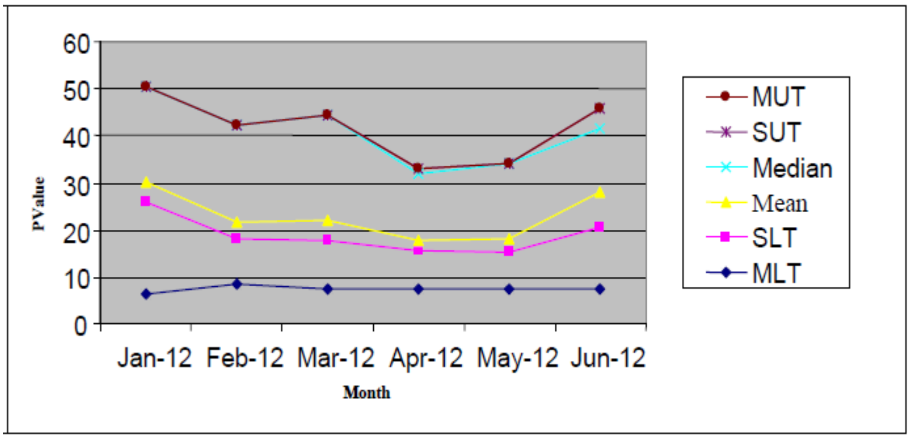


Fig. 5 : Comparative Graph For Sigma And Made Methods

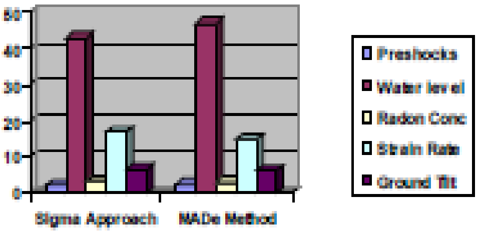


Fig. 6 : Values of Precursors Using Sigma And Made Methods

As the model is relatively simple and based on incremental learning, the computational costs are significantly lower than many other methods. Different density regions may bias the standard deviation and mean of the data. Biased mean and standard deviation will consequently result in misleading STOs. So MADe method is more accurate. Although both these methods are very effective, a brief comparison between them is given below.

TABLE 2. COMPARISION OF OUTLIER DETECTION METHODS

Sigma Method	MADe Method
Based on Mean and Standard Deviation	Based on Median and Median Absolute Deviation
Percentage of outliers get affected largely by skewness of data	High breakdown point so not unduly affected by skewness of data
Biased in different density regions	More accurate
Simple classical approach	Robust approach

VII. CONCLUSION AND FUTURE WORK

The IASPEI Sub- Commission chair concluded, "It is not clear that any of these proposed precursors are understood to the point that they can now be used for prediction; they are simply a collection of phenomena which have a better than average chance of becoming useful in earthquake prediction someday". Research on precursory behavior has contributed substantially to the understanding of earthquake processes, and it should be part of a fundamental research program on earthquake predictability.

In this paper, we have demonstrated a promising approach for predicting earthquake. It is based on spatial-temporal data view and analyzes various earthquake precursors. We applied two of the outlier detection techniques in data processing algorithm. We applied Sigma approach and MADe method to consider the correlation of earthquake data from time to time. We suggested this algorithm to check the possibility of earthquake by finding spatiotemporal outliers. To evaluate the performance of the earthquake detection model, we have used two detection methods which check and compare earthquake probability. This comparisonnprovides real-time warnings about earthquake.

Future efforts will be needed to refine the precursors used in this approach. Opportunities exist on both levels – data collection technologies and data processing algorithms -- to improve the reliability and effectiveness of earthquake prediction system. Efforts will be needed to develop and improve detectors and algorithms as most algorithms are based on the measurements from detector systems. This approach can also be applied to other applications which consider both temporal and spatial features, such as disease control, weather monitoring or any other natural disaster.

REFERENCES

[1] Kunal Tiwari, Krishna Mehta, Nitin Jain, Ramandeep Tiwari, Gaurav Kanda, "Selecting the Appropriate Outlier Treatment for Common Industry Applications", Statistics and Data Analysis NESUG 2007.

[2] Ying Jin, Jing Dai, Chang-Tien Lu, "Spatial-Temporal Data Mining in Traffic Incident Detection", Department of Computer Science, Virginia Polytechnic Institute and State University.

[3] Songwon Seo, "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets", University of Pittsburgh.

[4] Yong-Kul Ki, "Accident Detection System using Image Processing and MDR", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.3, March 2007

[5] Warner Marzocchi, Laura Sandri, and Enzo Boschi, "On the Validation of Earthquake-Forecasting Models: The Case of Pattern Recognition Algorithms", Bulletin of the Seismological Society of America, Vol. 93, No. 5, pp. 1994–2004, October 2003.

[6] Witold Dzwinel, David A.Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka, Takeo Ito, "Cluster Analysis, Data-Mining, Multidimensional Visualization of Earthquakes over Space, Time and Feature Space", Submitted to Earth and Planetary Sci. Letters, August, 2003.

[7] Elizabeth Wu, Wei Liu, Sanjay Chawla, "Spatio-Temporal Outlier Detection in Precipitation Data", 200X ACM.

[8] Derya Birant, Alp Kut, "Spatio-Temporal Outlier Detection in Large Databases", Journal of Computing and Information Technology - CIT 14, 2006, 4, 291–297 doi:10.2498/cit.2006.04.04.

◈◈◈

A User Oriented Image Retreival System Based On Interactive Genetic Algorithm

Vasundara C & B.Krishna Sagar

Department of Computer Science & Engineering
Madanapalle Institute of Technology and Science, Madanapalle,Andhrapradesh,India.
E-mail : vasuraju19@gmail.com

Abstract – Nowadays, content-based image retrieval (CBIR) is the mainstay of image retrieval systems. To be more profitable, relevance feedback techniques were incorporated into CBIR such that more precise results can be obtained by taking user's feedbacks into account. However, existing relevance feedback-based CBIR methods usually request a number of iterative feedbacks to produce refined search results, especially in a large-scale image database. This is impractical and in efficient in real applications. Inthis paper, we propose a novel method, Navigation-Pattern-based Relevance Feedback (NPRF), to achieve the high efficiency and effectiveness of CBIR in coping with the large-scale image data. In terms of efficiency, the iterations of feedback are reduced substantially by using the navigation patterns discovered from the user query log. In terms of effectiveness, our proposed search algorithm NPRFSearch makes use of the discovered navigation patterns and three kinds of query refinement strategies, Query Point Movement (QPM), Query Reweighting (QR), and Query Expansion (QEX), to converge the search space toward the user's intention effectively. By using NPRF method, high quality of image retrieval on RF can be achieved in a small number of feedbacks. The experimental results reveal that NPRF outperforms other existing methods significantly in terms of precision, coverage, and number of feedbacks.

Index Terms**—**Content-based image retrieval, relevance feedback, query point movement, query expansion, navigation pattern mining.

I. INTRODUCTION

MULTIMEDIA contents are growing explosively and the need for multimedia retrieval is occurring more and more frequently in our daily life. Due to the complexity of multimedia contents, image understanding is a difficult but interesting issue in this field. Extracting valuable knowledge from a large-scale multimedia repository, so-called multimedia mining.

Unfortunately, this kind of textual-based image retrieval always suffers from two problems: high-priced manual annotation and inappropriate automated annotation. On one hand, high-priced manual annotation cost is prohibitive in coping with a large-scale data set. On the other hand, in appropriate automated annotation yields the distorted results for semantic image retrieval.

As a result, a number of powerful image retrieval algorithms have been proposed to deal with such problems over the past few years. Content-Based Image Retrieval (CBIR) is the mainstay of current image retrieval systems. In general, the purpose of CBIR is to present an image conceptually, with a set of low-level visual features such as color, texture, and shape.

The hidden problem is that the extracted visual features are too diverse to capture the concept of the user's query. To solve such problems, in the QBE system, the users can pick up some preferred images to refine the image explorations iteratively. The feedback procedure, called Relevance Feedback (RF), repeats until the user is satisfied with the retrieval results.

Although a number of RF studies [1], [11], [12], [16] have been made on interactive CBIR, they still incur some common problems, namely redundant browsing and exploration convergence. First, in terms of redundant browsing, most existing RF methods focus on how to earn the user's satisfaction in one query process.

To resolve the aforementioned problems, we propose a novel method named Navigation-Pattern-based Relevance Feedback (NPRF) to achieve the high retrieval quality of CBIR with RF by using the discovered navigation patterns.

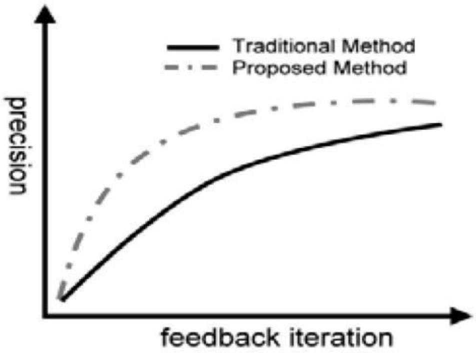


Fig: The expected scenario for effectiveness versus efficiency.

II. RELATED WORK

Relevance feedback [5], [17], [25], in principle, refers to a set of approaches learning from an assortment of users' browsing behaviors on image retrieval [10].

1. Query Reweighting:

The notion behind QR is that, if the ith feature fi exists in positive examples frequently, the system assigns the higher degree to fi. QRlike approaches were first proposed by Rui et al. [14], which convert image feature vectors to weighted-term vectors in early version of Multimedia Analysis and Retrieval System (MARS).

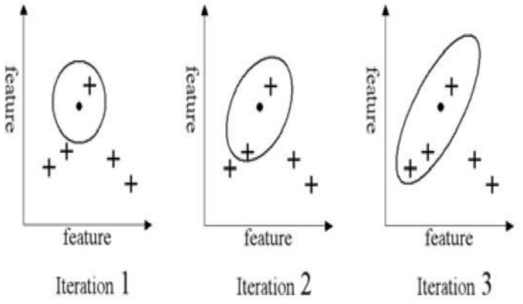


Fig:Relevance feedback with generalized QR technique

2. Query Point Movement

Another solution for enhancing the accuracy of image retrieval is moving the query point toward the contour of the user's preference in feature space. QPM regards multiple positive examples as a new query point at each feedback

3. Query Expansion Because QR and QPM cannot elevate the quality of RF, QEX has been another hot technique in the solution space of RF recently.

4. Hybrid RF

One of the hybrid RF strategies is IRRL. IRRL, proposed by Yin et al. [18], addresses the important empirical question of how to precisely capture the user's interest at each feedback.

III PROPOSED APPROACH

Our proposed approach NPRF integrates the discovered navigation patterns and three RF techniques to achieve efficient and effective exploration of images.

3.1 Overview of Navigation-Pattern-Based Relevance Feedback

The major difference between our proposed approach and other contemporary approaches is that we approximate an optimal solution to resolve the problems existing in current RF, such as redundant browsing and exploration convergence.

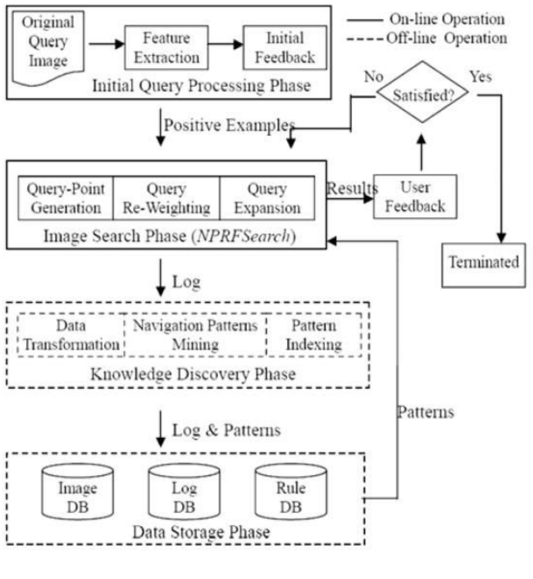


Fig: Workflow of NPRF.

For online operation, once a query image is submitted to this system, the systemfirst finds the most similar images without considering any search strategy, and then returns a set of the most similarimages. The first query process is called initial feedback.

Next, the good examples picked up by the user deliver thevaluable information to the image search phase, includingnew feature weights, new query point, and the user'sintention. Then, by using the navigation patterns, three search strategies, with respect to QPM, QR, and QEX, arehybridized to find the desired images.

3.1.1 Online Image Retrieval:

Initial Query Processing Phase: Without considering the feature weight, this phase extracts the visual features from the original query image to find the similar images.

Image Search Phase: Behind the search phase, ourintent is to extend the one search point to multiplesearch points by integrating the navigation patternsand the proposed search algorithm NPRFSearch.

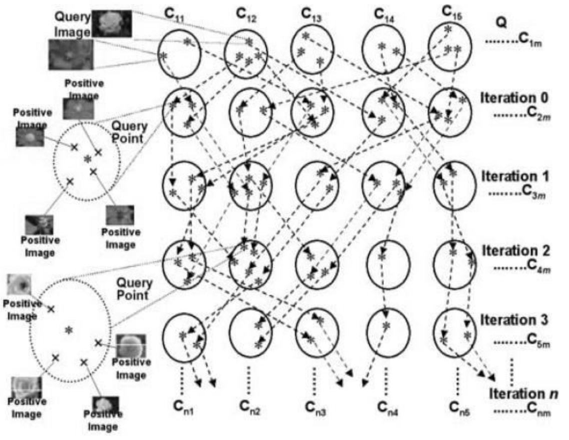
3.1.2 Offline Knowledge Discovery:

Knowledge Discovery Phase: Learning from users'behaviors in image retrieval can be viewed as one typeof knowledge discovery.

Data Storage Phase: The databases in this phase can be regarded as the knowledge marts of a knowledge warehouse, which store integrated, time-variant, and nonvolatile collection of useful data including images, navigation patterns, log files, and image features.

3.2 Offline Knowledge Discovery

In fact, usage mining has been made on how to generateusers' browsing patterns to facilitate the web pages retrieval. Similarly, for web image retrieval, the user hasto submit a query term to the search engine, so-called textual-based image search.



3.2.1 Data Transformation

To date, very few significant studies have succeeded insemantic image retrieval or image recognition because of the complicated visual contents

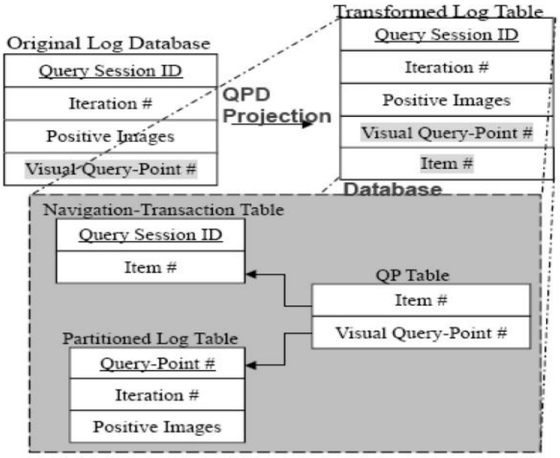


Fig:The entity-relationship data model for partitioning the log data.

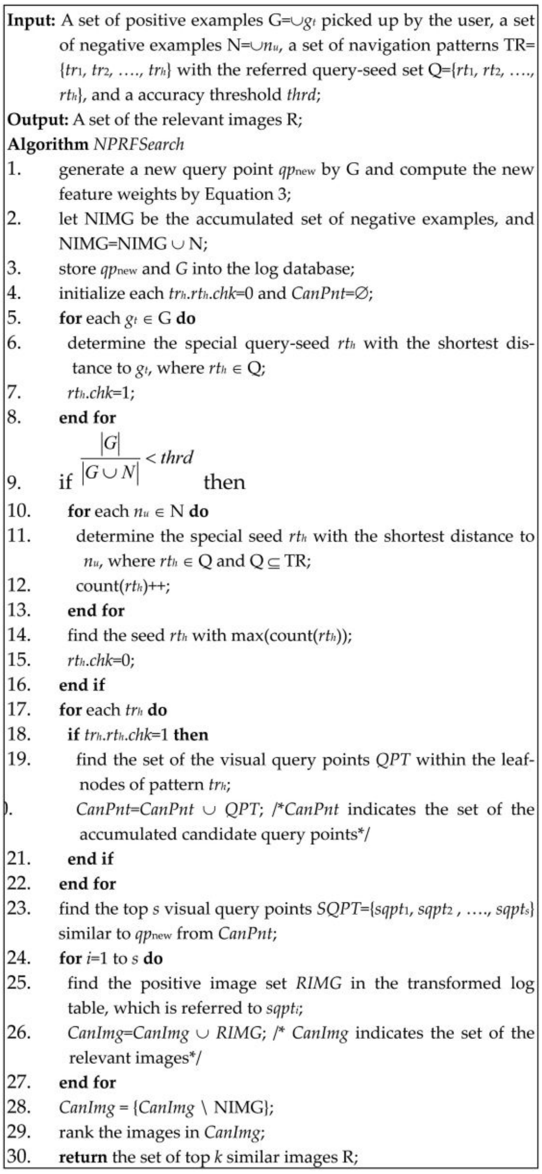
3.3 Online Image Search:

3.3.1:Algorithm for NPRF Search:

As already described above, NPRFSearch is proposed to reach the high precision of image retrieval in a shorter queryprocess by using the valuable navigation patterns. In this section, we explain the details of NPRFSearch

The iterative search procedure can be decomposed into several steps as follows:

1. Generate a new query point by averaging the visualfeatures of positive examples.
2. Find the matching navigation pattern trees by determining the nearest query seeds (root).
3. Find the nearest leaf nodes (terminations of a path) from the matching navigation pattern trees.
4. Find the top s relevant visual query points from the set of the nearest leaf nodes.
5. Finally, the top k relevant images are returned to the user



```
Input: A set of positive examples $G=\cup g_t$ picked up by the user, a set
    of negative examples $N=\cup n_u$, a set of navigation patterns $TR=\{tr_1, tr_2, ...., tr_h\}$
    with the referred query-seed set $Q=\{rt_1, rt_2, ...., rt_h\}$, and a accuracy threshold $thrd$;
Output: A set of the relevant images R;
Algorithm NPRFSearch
    generate a new query point $qp_{new}$ by G and compute the new feature weights by Equation 3;
    let NIMG be the accumulated set of negative examples, and $NIMG=NIMG \cup N$;
    store $qp_{new}$ and $G$ into the log database;
    initialize each $tr_h.rt_h.chk=0$ and $CanPnt=\varnothing$;
    for each $g_t \in G$ do
      determine the special query-seed $rt_h$ with the shortest distance to $g_t$, where $rt_h \in Q$;
      $rt_h.chk=1$;
    end for
    if $\frac{|G|}{|G \cup N|} < thrd$ then
      for each $n_u \in N$ do
        determine the special seed $rt_h$ with the shortest distance to $n_u$, where $rt_h \in Q$ and $Q \subseteq TR$;
        count($rt_h$)++;
      end for
      find the seed $rt_h$ with max(count($rt_h$));
      $rt_h.chk=0$;
    end if
    for each $tr_h$ do
      if $tr_h.rt_h.chk=1$ then
        find the set of the visual query points $QPT$ within the leaf-nodes of pattern $tr_h$;
        $CanPnt=CanPnt \cup QPT$; /*CanPnt indicates the set of the accumulated candidate query points*/
      end if
    end for
    find the top $s$ visual query points $SQPT=\{sqpt_1, sqpt_2 , ...., sqpt_s\}$ similar to $qp_{new}$ from $CanPnt$;
    for $i$=1 to $s$ do
      find the positive image set $RIMG$ in the transformed log table, which is referred to $sqpt_i$;
      $CanImg=CanImg \cup RIMG$; /* CanImg indicates the set of the relevant images*/
    end for
    $CanImg = \{CanImg \setminus NIMG\}$;
    rank the images in $CanImg$;
    return the set of top $k$ similar images R;
```

Fig: Algorithm for NPRFSearch.

IV CONCLUSION

To deal with the long iteration problem of CBIR with RF, we have presented a new approach named NPRF byintegrating the navigation pattern mining and a navigation- pattern-based search approach named NPSearch.

In summary, the main feature of NPRF is to efficiently optimize the retrieval quality of interactive CBIR. On one hand, the navigation patterns derived from the users' longterm browsing behaviors are used as a good support for minimizing the number of user feedbacks.

Onthe other hand, the proposed algorithm NPRFSearch performs the navigation- pattern-based search to match the user's intention by merging three query refinement strategies. As a result, traditional problems such as visual diversity and exploration convergence are solved.

For navigation-pattern-based search, the hierarchical BFS-based KNN is employed to narrow the gap between visual features and human concepts effectively. In addition, the involved methods for special data partition and pattern pruning also speed up the image exploration.

The experimental results reveal that the proposed approach NPRF is very effective in terms of precision and coverage. Within a very short term of relevance feedback, the navigation patterns can assist the users in obtaining the global optimal results. Moreover, the new search algorithm NPRFSearch can bring out more accurate results than otherwell-known approaches.

In the future, there are some remaining issues toinvestigate. First, in view of very large data sets, we will scale our proposed method by utilizing parallel and distributed computing techniques. Second, we will integrate user's profile into NPRF to further increase the retrieval quality. Third, we will apply the NPRF approach to more kinds of applications on multimedia retrieval or multimedia recommendations.

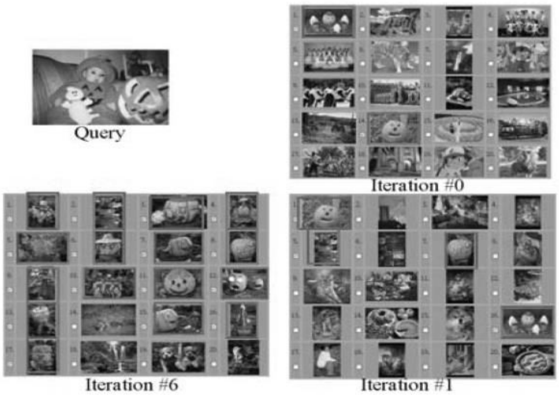


Fig:The resulting example for QPM.

V FUTURE WORK

The main goals we have achieved in this work are 1) mine all frequent association rules without imposing any a- priori restriction on the structure and the content of the rules;

2) store mined information in XML format

3) use extracted knowledge to gain information about the original datasets. We have not discussed the updatability of both the document storing TARs and their index. As an ongoing work, we are studying how to incrementally update mined TARs when the original XML datasets change and how to further optimize our mining algorithm; moreover, for the moment we deal with a (substantial) fragment of XQuery; we would like to find the exact fragment of XQuery which lends itself to translation into intentional queries.

ACKNOWLEDGMENT

I sincerely believe in this and would like to thank B.Krishna Sagar, M.Tech.,Asst Professor., and the Department of Computer Science & Engineering, MITS, Madanapalle for his encouragement, motivation and for guiding me to write this paper. Also I am grateful to Mr.M.V Jagannatha Reddy, M.Tech(Ph.D), Associate prof., the department of Computer Science & Engineering, MITS, Madanapalle.

[1] M.D. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Steele, and P. Yanker, “Query by Image and Video Content: The QBIC System,” Computer, vol. 28, no. 9, pp. 23-32, Sept. 1995.

[2] R. Fagin, “Combining Fuzzy Information from Multiple Systems,” Proc. Symp. Principles of Database Systems (PODS), pp. 216-226, June 1996.

[3] R. Fagin, “Fuzzy Queries in Multimedia Database Systems,” Proc. Symp. Principles of Database Systems (PODS), pp. 1-10, June 1998.

[4] J. French and X-Y. Jin, “An Empirical Investigation of the Scalability of a Multiple Viewpoint CBIR System,” Proc. Int’l Conf.

[5] D. Harman, “Relevance Feedback Revisited,” Proc. 15th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 1-10, 1992.

[6] Y. Ishikawa, R. Subramanya, and C. Faloutsos, “MindReader: Querying Databases through Multiple Examples,” Proc. 24th Int’l Conf. Very Large Data Bases (VLDB), pp. 218-227, 1998.

[7] X. Jin and J.C. French, “Improving Image Retrieval Effectiveness via Multiple Queries,” Multimedia Tools and Applications, vol. 26, pp. 221-245, June 2005.

[8] D.H. Kim and C.W. Chung, “Qcluster: Relevance Feedback Using Adaptive Clustering for Content-Based Image Retrieval,” Proc. ACM SIGMOD, pp. 599-610, 2003.

[9] K. Porkaew, K. Chakrabarti, and S. Mehrotra, “Query Refinement for Multimedia Similarity Retrieval in MARS,” Proc. ACM Int’l Multimedia Conf. (ACMMM), pp. 235-238, 1999.

[10] J. Liu, Z. Li, M. Li, H. Lu, and S. Ma, “Human Behaviour Consistent Relevance Feedback Model for Image Retrieval,” Proc. 15th Int’l Conf. Multimedia, pp. 269-272, Sept. 2007.

[11] A. Pentalnd, R.W. Picard, and S. Sclaroff, “Photobook: Content- Based Manipulation of Image Databases,” Int’l J. Computer Vision (IJCV), vol. 18, no. 3, pp. 233-254, June 1996.

[12] T. Qin, X.D. Zhang, T.Y. Liu, D.S. Wang, W.Y. Ma, and H.J. Zhang,“An Active Feedback Framework for Image Retrieval,” PatternRecognition Letters, vol. 29, pp. 637-646, Apr. 2008.

[13] J.J. Rocchio, “Relevance Feedback in Information Retrieval,” TheSMART Retrieval System—Experiments in Automatic Document Processing, pp. 313-323, Prentice Hall, 1971.

[14] Y. Rui, T. Huang, and S. Mehrotra, “Content-Based Image Retrieval with Relevance Feedback in MARS,” Proc. IEEE Int’l Conf. Image Processing, pp. 815-818, Oct. 1997.

[15] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, “Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval,” IEEE Trans. Circuits and Systems for Video Technology, vol. 8, no. 5, pp. 644-655, Sept. 1998.

[16] J.R. Smith and S.F. Chang, “VisualSEEK: A Fully Automated Content-Based Image Query System,” Proc. ACM Multimedia Conf., Nov. 1996.

[17] G. Salton and C. Buckley, “Improving Retrieval Performance by Relevance Feedback,” J. Am. Soc. Information Science, vol. 41, no. 4, pp. 288-297, 1990.

[18] H.T. Shen, S. Jiang, K.L. Tan, Z. Huang, and X. Zhou, “Speed up Interactive Image Retrieval,” VLDB J., vol. 18, no. 1, pp. 329-343, Jan. 2009.

[19] V.S. Tseng, J.H. Su, J.H. Huang, and C.J. Chen, “Integrated Mining of Visual Features, Speech Features and Frequent Patterns for Semantic Video Annotation,” IEEE Trans. Multimedia, vol. 10, no. 2, pp. 260-267, Feb. 2008.

[20] V.S. Tseng, J.H. Su, B.W. Wang, and Y.M. Lin, “Web Image Annotation by Fusing Visual Features and Textual Information,” Proc. 22nd ACM Symp. Applied Computing, Mar. 2007.

[21] K. Vu, K.A. Hua, and N. Jiang, “Improving Image Retrieval Effectiveness in Query-by-Example Environment,” Proc. 2003 ACM Symp. Applied Computing, pp. 774-781, 2003.

[22] L. Wu, C. Faloutsos, K. Sycara, and T.R. Payne, “FALCON: Feedback Adaptive Loop for Content-Based Retrieval,” Proc. 26th Int’l Conf. Very Large Data Bases (VLDB), pp. 297-306, 2000.

[23] P.Y. Yin, B. Bhanu, K.C. Chang, and A. Dong, “Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1536-1551, Oct. 2005.

[24] H. You, E. Chang, and B. Li, “NNEW: Nearest Neighbor Expansion by Weighting in Image Database Retrieval,” Proc. IEEE Int’l Conf. Multimedia and Expo, pp. 245-248, Aug. 2001.

[25] X.S. Zhou and T.S. Huang, “Relevance Feedback for Image Retrieval: A Comprehensive Review,” Multimedia Systems, vol. 8, no. 6, pp. 536-544, Apr. 2003.

◈◈◈

Design, Implementation and Analysis of Flash ADC Architecture with Differential Amplifier as Comparator Using Custom Design Approach

Channakka Lakkannavar, Shrikanth K. Shirakol & Kalmeshwar N. Hosur

Department of E&CE., S.D.M. College of Engineering & Technology, Dharwad-02
E-mail : channi.sdm@gmail.com , shrikanthks@yahoo.com, kalmeshwar10@rediffmail.com

Abstract – Analog-to-Digital Converters (ADCs) are useful building blocks in many applications such as a data storage read channel and an optical receiver because they represent the interface between the real world analog signal and the digital signal processors. Many implementations have been reported in the literature in order to obtain high-speed analog-to-digital converters (ADCs).

In this paper an effort is made to design 4bit Flash Analog to Digital Converter [ADC] on 180nm technology. For high-speed applications, a flash ADC is often used. Resolution, speed, and power consumption are the three key parameters for an analog-to-digital converter (ADC). The integrated flash ADC is operated at 4-bit precision with analog input voltage of 0 to 1.8V. The ADC has been designed, implemented & analysed in standard gpdk180nm technology library using Cadence tool.

Keywords - *Flash ADC, Resolution, Cadence, power consumption, gpdk180.*

I. INTRODUCTION

With the rapid growth of modern communications and signal processing systems, handheld wireless computers and consumer electronics are becoming increasingly popular. Mixed-signal integrated circuits have a tendency in the design of system-on-chip (SOC) in recent years. SOC designs have made possible substantial cost and form factor reductions, in part since they integrate crucial analog interface circuits, such as ADCs with digital computing and signal processing circuits on the same die. The interfaces only occupy a small fraction of the chip die and for SOC designs, the technology selection and system design choices are mainly driven by digital circuit requirements [1].

For high-speed applications, a flash ADC is often used. Resolution, speed, and power consumption are the three key parameters for an analog-to-digital converter (ADC). These parameters cannot be changed once an ADC is designed. While one can use 6-bit precision from an 8-bit ADC, it is non-optimal resulting in slower speed and extra power consumption due to full 8-bit internal operation. In this paper, a new flash ADC design is proposed that is a true variable-power and variable-resolution ADC.

It can operate at higher speed and will consume less power when operating at a lower resolution. Such features are highly desirable in many wireless and mobile applications. For example, the strength of a radio frequency (RF) signal varies greatly depending on geographic location. Optimally, the ADC resolution can be reduced upon the reception of strong signal and can be increased upon the reception of weak signal. Substantial reduction in power consumption at lower resolution will prolong the battery life [1]. Low power ADC architectures are implemented with pipelined, successive approximation, and sigma-delta modulators. These are all useful for the medium speed conversion and high resolution applications. On the other hand, the flash architecture is suitable for high speed conversion and low resolution applications due to its parallel architecture.

II. BACKGROUND

The paper on “The CMOS Inverter as a Comparator in ADC Designs”, spinger Analog Integrated Circuits and Signal Processing, Vol.39, pp.147-155, 2004 by Tangel A. Choi K discuss about the advancement of technology, digital signal processing has progressed dramatically in recent years.

Signal processing in digital domain provides high level of accuracy, low power consumption and small silicon area besides providing flexibility in design and programmability. The design process is also quite faster and cost effective. Furthermore, their implementation makes them suitable for integration with complex digital signal processing blocks in a compatible low-cost technology, particularly CMOS [1].

This evolution of technology provides much faster transistors with smaller sizes, making it possible to have very high clock rate in digital circuits. In the end, it leads us to design a very high speed as well as systems with small die area called System on a chip (SoC), with a smaller number of chips using increased integration level.

However, "CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters", 2nd Edition, 2005 by Rudy J. van de Plassche et al, deals with the evolution of technology has not provided same level of benefit for the analog circuit design. So to extract the advantages of digital signal processing, there is a trend of shifting signal processing from analog to more efficient digital domain and dealing with the analog signals only in the input-output stages. This has resulted in the requirement of smart converters between analog and digital signals to cope up with the evolution of technology [3].

Section III discuss about the ADC architecture, in the mean while Section IV about its Implementation. Section V, VI describes the experimental results and Conclusion respectively.

III. FLASH ADC ARCHITECTURE

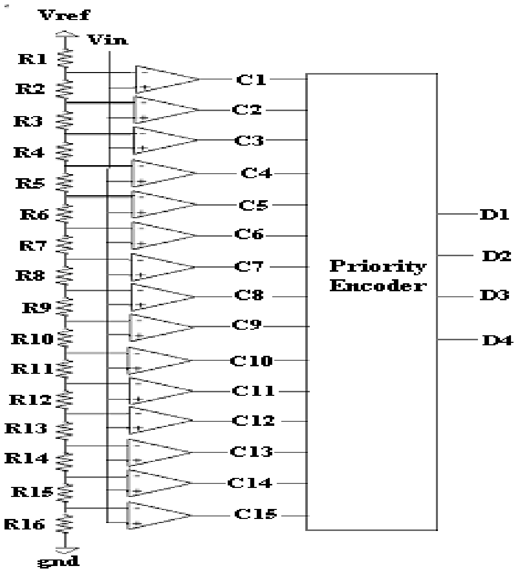


Fig.1: Flash ADC architecture

The above Fig.1., shows a typical flash ADC block diagram. For an "4" bit converter, the circuit employs 2^4-1= 15 comparators. A resistive divider with 2^4 = 16 resistors provides the reference voltage The reference voltage for each comparator is one least significant bit (LSB) greater than the reference voltage for the comparator immediately below it. Each comparator produces a "1" when its analog input voltage is higher than the reference voltage applied to it. Otherwise, the comparator output is "0".

The Flash ADC consists of the following components which are given below:

*1. **Resister string*** : In an 'n' bit flash ADC, 2^n resistances are necessary. The two extreme resistors are calculated to delimit the voltage input range. Each resistor divides the reference voltage to feed a comparator. The higher the resistance value is, the weaker the current is consumed in the device. That is why a high resistance will minimize power dissipation. Nevertheless, we have to put a reasonable value for this resistor string: it should stand lower than the input resistance of the comparators.

*2. **Comparator*** : Here 2^n-1 differential amplifiers are used as comparators in"n" bits flash-ADC architecture. We first tried to implement a complex type of differential amplifier. But this element was not easy enough to understand and use for beginners. We consequently decided to prefer a basic design to realize our ADC. This decision made us loose several advantages as an improved gain, or power savings that would have benefit to our ADC precision and efficiency. When the input signal voltage is less than the reference voltage, the comparator output is at logic '0'.when the input signal voltage is higher than the reference voltage, the comparator output is at logic '1'. The comparators give the 2^n-1 levels of outputs in terms of reference voltage.

*3. **Priority encoder*** : The output of the comparators is in the encoded form. Therefore a priority encoder has to be designed in order to convert the encoded signal into n bits data (digital) which is unipolar binary code[4][6].

IV. DESIGN AND IMPLEMENTATION

This section deals with implementation of three components as discussed in section III.

1. The resistor string

The 4 bit flash ADC, needs 2^4 resistances [fig.,1]. The two extreme resistors are calculated to delimit the voltage input range. Each resistor divides the reference voltage to feed a comparator. The higher the resistance value is, the weaker the current is consumed in the device. That is why a high resistance will minimize power dissipation. Nevertheless, we have to put a reasonable value for this resistor string: it should stand lower than the input resistance of the comparators. We expected to convert any voltage between 0 and 1.8 V.

In general, the voltage division takes place as follows:

$$V_a = (M*V_{ref})/2^n \qquad (1)$$

Where,

M = No., of resistors at which voltage division occurs.
n = No., of bits.

2^n = Total No., of resistors used.

The design of resister string for proposed Flash ADC is done using schematic approach in cadence as shown in fig., 2.

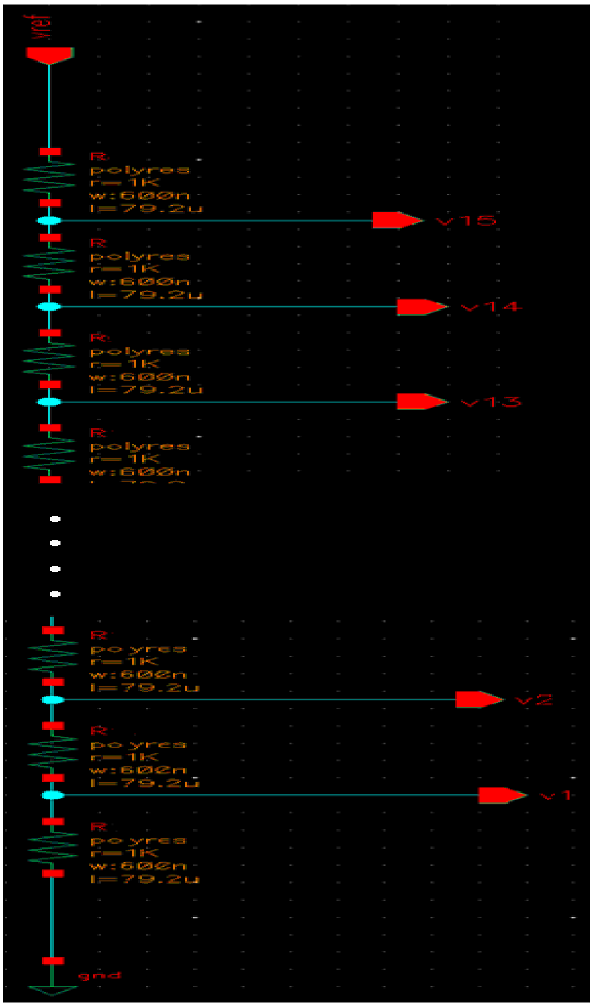


Fig. 2: Resistor string for proposed flash ADC.

The Table I show the voltage division for resistor string with reference voltage is taken to be 1.8 V.

Table I: Voltage division occurs as follows:

M *	V_a*= $(M*V_{ref})/2^n$
Tap 1	0.1125V
Tap 2	0.225V
Tap 3	0.3375V
Tap 4	0.45V
Tap 5	0.5625V
Tap 6	0.675V
Tap 7	0.7875V
Tap 8	0.9V
Tap 9	1.0125V
Tap 10	1.125V
Tap 11	1.2375V
Tap 12	1.35V
Tap 13	1.4625V
Tap 14	1.575V
Tap 15	1.6875V
Tap 16	1.8V

*M= Resistor tap Number & V_a=Voltages of each resistor tap

2. *The Comparator*

The proposed flash ADC consists of comparator as one of the important component, this comparator is designed in such a way that which is less immunity for noise and with high common mode rejection ratio. Hence, the Differential Amplifier is used to achieve the same.

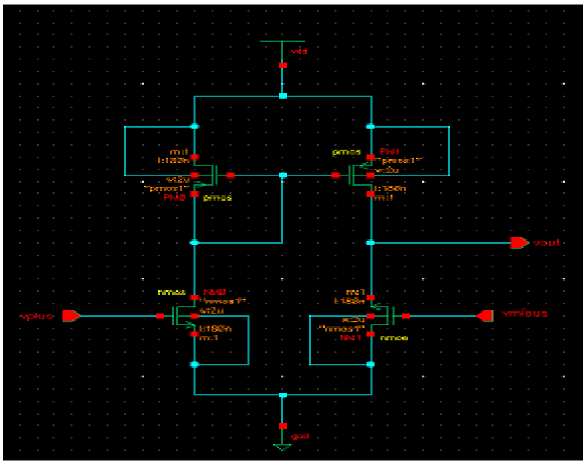


Fig. 3. Proposed comparator design

Hence forth the design of comparator is shown in fig. 3., uses 2^4-1=15 differential amplifiers are used as comparators in 4bit flash-ADC architecture.

Working of Comparator:

When the input signal voltage is less than the reference voltage, the comparator output is at logic ‘0’.when the input signal voltage is higher than the reference voltage, the comparator output is at logic ‘1’. The comparators give the 15 levels of outputs in terms of reference voltage. In transient response, during 0 to 5ns time V1 is ‘0’,V2 is ‘1’,so output will be ‘0’ because V1<V2 and during 5 to 10 ns time V1 is ‘1’,V2 is ‘0’, so output will be ‘1’ because V1>V2.

3. *Priority encoder stage:*

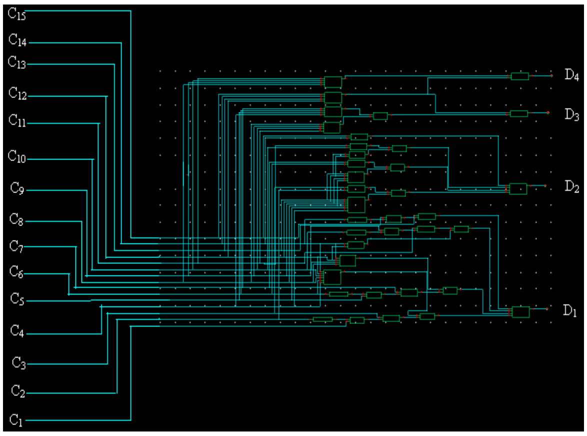


Fig. 4: Priority encoder stage

The output of the comparators is in the encoded form. Therefore a priority encoder has to be designed in order to convert the encoded signal into 4 bits data (digital) which is unipolar binary code.

The logic employed in designing the priority encoder is explained as follows,

$$D1=C_1C'_2C'_4C'_6C'_8C'_{10}C'_{12}C'_{14}+C_3C'_4C'_6C'_8C'_{10}C'_{12}C'_{14}+C_5C'_6C'_8C'_{10}C'_{12}C'_{14}+C_7C'_8C'_{10}C'_{12}C'_{14}+C_9C'_{10}C'_{12}C'_{14}+C_{11}C'_{12}C'_{14}+C_{13}C'_{14}+C_{15}.$$

$$D2=C_2C'_4C'_5C'_8C'_9C'_{12}C'_{13}+C_3C'_4C'_5C'_8C'_9C'_{12}C'_{13}+C_6C'_8C'_9C'_{12}C'_{13}+C_7C'_8C'_9C'_{12}C'_{13}+C_{10}C'_{12}C'_{13}+C_{11}C'_{12}C'_{13}+C_{14}+C_{15}.$$

$$D3=C_4C'_8C'_9C'_{10}C'_{11}+C_5C'_8C'_9C'_{10}C'_{11}+C_6C'_8C'_9C'_{10}C'_{11}+C_7C'_8C'_9C'_{10}C'_{11}+C_{12}+C_{13}+C_{14}+C_{15}.$$

$$D4 = C_8+C_9+C_{10}+C_{11}+C_{12}+C_{13}+C_{14}+C_{15}.$$

V. EXPERIMENTAL RESULTS

This Section clearly disscuss about the simulation results of above said three important components of flash ADC, the work is carried out on cadence virtuoso the simulation is done using spectre and layout using assura.

Fig.5., discuss about transient response for resistor string of voltage at resister taps. These tap voltages become inputs to comparator stage.

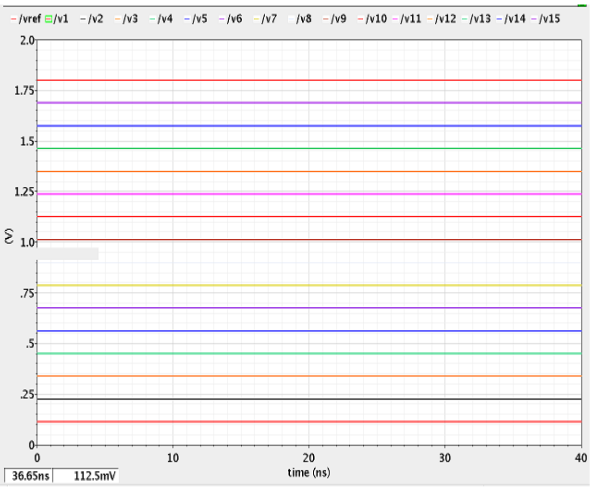


Fig. 5: Transient response for resistor string.

When the input signal voltage is less than the reference voltage, the comparator output is at logic ‘0’,when the input signal voltage is higher than the reference voltage, the comparator output is at logic ‘1’. The comparators give the 15 levels of outputs in terms of reference voltage. In transient response, during 0 to 5ns time V_1 is ‘1’,V_2 is ‘0’, hence output will be ‘1’ because $V_1>V_2$ and during 5 to 10 ns time V_1 is ‘0’,V_2 is ‘1’, hence output will be ‘0’ because $V_1<V_2$ shown in Fig.6.,

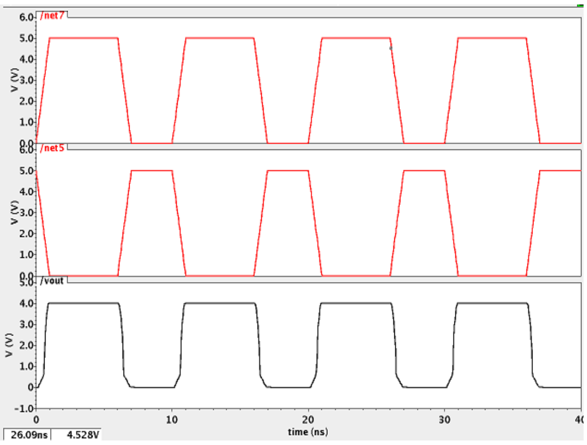


Fig.6: Transient response for comparator

The schematic simulation of 4 bit Flash ADC using Cadence tool is shown in Fig.7-11.

Fig.7., shows transient response for Vin = 0V, so here Vin will be in range between 0≤Vin≥0.112, so comparator output will be “000000000000000” and priority encoder output will be “0000”.

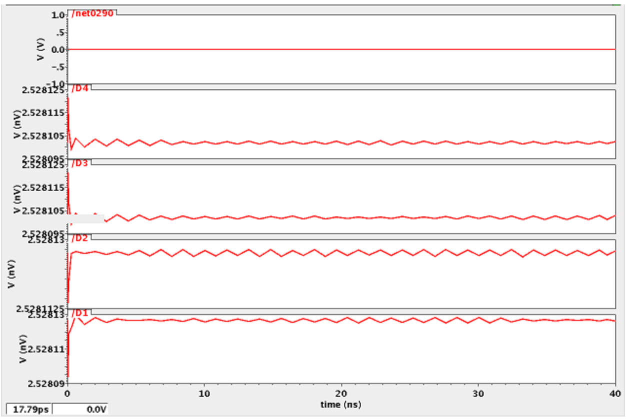


Fig.7: Transient response for V_{in}=0V, output will be 0000

Fig. 8 shows transient response for Vin = 0.8V, so here Vin will be in range between 0.7875≤Vin≥0.9, so comparator output will be "000000011111111" and priority encoder output will be "0111".

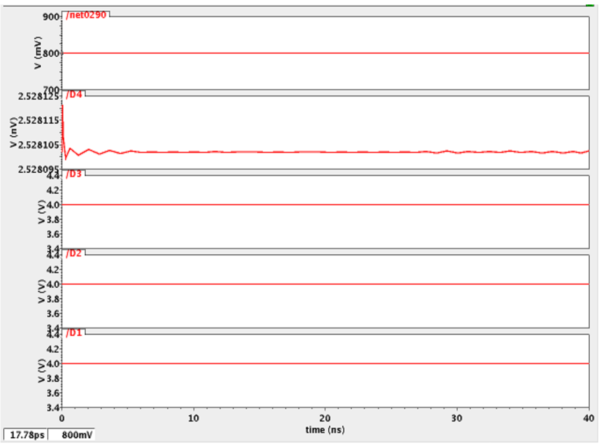


Fig 8: Transcient response for V_{in}=0.8V, output will be 0111

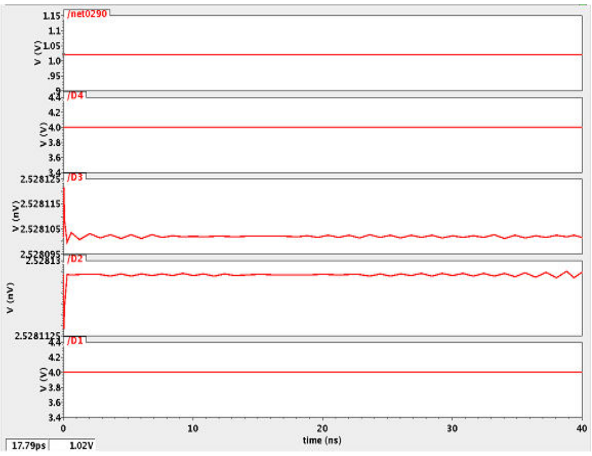


Fig.9: Transient response for V_{in} = 1.02V, output will be 1001

Fig. 9 shows transient response for Vin = 1.02V, so here Vin will be in range between 1.0125≤V_{in}≥1.125, so comparator output will be "000000111111111" and priority encoder output will be "1001".

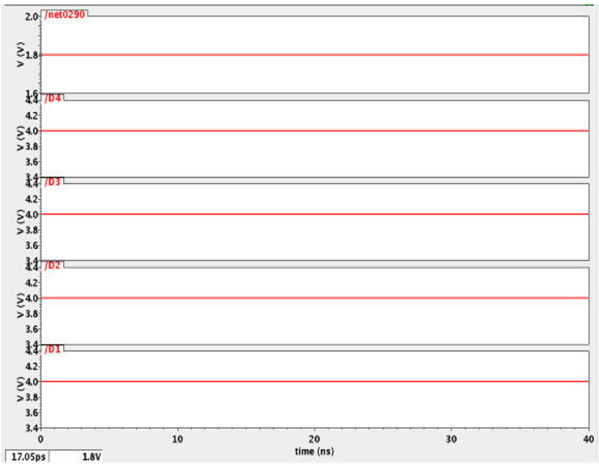


Fig.10: Transient response for V_{in} = 1.8V, output will be 1111

Fig.10. shows transient response for Vin = 1.8V, so here Vin will be in range between 1.6875≤V_{in}, so comparator output will be "111111111111111" and priority encoder output will be "1111".

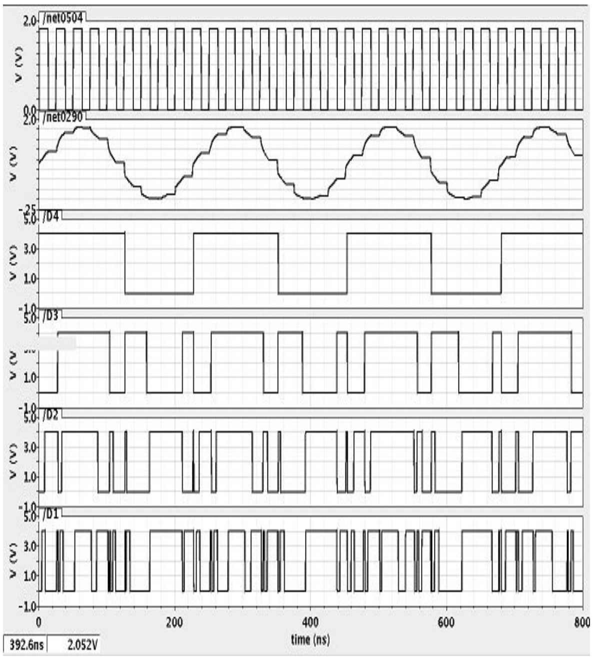


Fig. 11: Flash ADC outputs for analog input of 0 to 1.8V and 4.4M Hz Input frequency

Fig.11., shows outputs for analog input of 0 to 1.8V and 4.4M Hz input frequency using sample and hold circuit.

VI. LAYOUT DESIGN FOR PROPOSED ADC DESIGN

Layout design of resister string made up of 16 polyresisters metals, which is connected to 15 outputs with Metal1-poly, this design is made inside PR (Place and Route) boundary, each Resister is constructed from polycrystalline silicon and poly is having width of 600ηm, segment length of 79.2μm, sheet resistivity 7.5Ω, body resistance 990Ω, contact resistance 10Ω and end resistance 0Ω. Layout design of comparator, totally it contains two pmos & nmos of gpdk180 libraries, where pmos is connected with nwell & nmos with psubstrate,each comparator is constructed from 32 pcapacitor & 13 presister pmos and nmos of width 2μm and length 180ηm. Calculated layout area is 821634.1415 (ηm)^2.

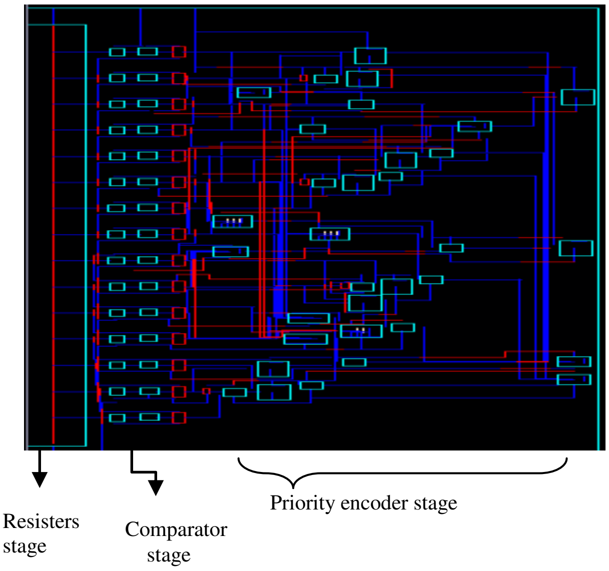


Fig. 12: Layout design of Flash ADC

Table II. Specifications Summary of ADC

Technology	180ηm
Analog voltage, V_{in}	0 to 1.8V
Reference voltage, V_{ref}	1.8V
V_{dd}	4V
Resolution	4-bits
Speed	3.8 GS/sec
Power Dissipation	49.94mW
SNR	25.84 dB
Standard Deviation	12ηs
Mean	19ηs
Calculated layout area	821634.145 (ηm)^2

VII. CONCLUSION

The schematic and layout of register stage, comparator stage, priority encoder stage are designed and integrated. From table II Gives brief design summary of flash ADC i.e., the integrated flash ADC is operated at 4-bit precision with analog input voltage of 0 to 1.8V, supply voltage 4V, Resolution 4bits, SNR 25.84dB, consumes 49.94mW power, speed is 3.8GS/s and layout Area is 0.821634μm^2. The ADC is designed and implemented in standard gpdk180nm technology of version – IC 6.1 using Cadence virtuoso tool.

ACKNOWLEDGMENT

We thank the Management, the Principal/Director and Staff of Sri Dharmasthala Manjunatheshwara College of Engineering and Technology, Dhavalgiri, Dharwad, Karnataka, India for encouraging us for this research work.

First and second Authors express there heartily thanks to Prof. Kotresh E. Marali, for his kind support in carrying out this research work.

REFERENCES

[1] A. Tangel, "VLSI implementation of the threshold inverter quantization (TIQ) technique for CMOS flash A/D converter applications." Ph.D. Thesis, the Pennsylvania State University, Aug. 1999.

[2] J. Yoo, "A TIQ Based CMOS Flash A/D Converter for System-on-Chip Applications", PhD Thesis, the Pennsylvania State University, May 2003.

[3] Tangel, A.; Choi, K, “'The CMOS Inverter as a Comparator in ADC Designs”, spinger Analog Integrated Circuits and Signal Processing, Vol.39, pp.147-155, 2004.

[4] E. Säll, "Implementation of Flash Analog-to-Digital Converters in Silicon- on-Insulator Technology," Linköping Studies in Science and Technology, Thesis No. 1213, ISBN 91-85457-79-5, Linköping, Sweden, Dec. 21,2005.

[5] J. M. Rabaey, A. Chandrakasan, and B. Nikolic′, “Digital Integrated Circuits”, 2nd Edition, 2003.

[6] Maxim Integrated Products, INL/DNL Measurements for High-Speed Analog to-Digital Converters (ADCs).

[7] Maxim Integrated Products. Defining and Testing Dynamic Parameters in High-Speed ADCs, 2001

[8] Rudy J. van de Plassche, “CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters”, 2nd Edition, 2005.

[9] K.L.Lin, T.van den Boom, Stevanovic. N, “Basic design guide for CMOS olding and interpolating A/D converters Overview and case study”, IEEE International Conference on Circuits and Systems, vol.1, pp. 529 – 532.

[10] Cadence NCLaunch User Guide, Product Version 8.1, May 2008.

[11] IEEE, "Std 1241-2000 IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters," 2009.

[12] Sreehari Veeramachaneni, A.Mahesh Kumar , Venkat Tummala, M.B.Srinivas, “Design of a Low Power Variable-Resolution Flash ADC”, In the Proceedings of the 22nd IEEE/ACM International Conference on VLSI Design and Embedded Systems (VLSI DESIGN -2009), New Delhi , India.

◈◈◈