International journal of scientific and engineering research

Contents

Contents

Contents

Establishing bridges between UML, HAD and GRAFCET Metamodels for the Modelling of Dynamic Systems

M. Nkenlifack, E. Tanyi and F. Fokou

Abstract— This article shows the scope and limits of UML as a tool for modelling Automatic Control Systems. An alternative metamodel, Hybrid Activity Diagram (HAD), is proposed and applied to a concrete example, in order to illustrate its efficiency in comparison to the limits of UML diagrams. The article also presents bridges which interlink UML, HAD and GRAFCET and establish compatibilities between the three models. Specifications for the development of a HAD simulator and the results obtained from the implementation of the simulator, are also presented in the article.

Index Terms—HAD, UML, Grafcet, Hybrid Dynamic Systems, Modelling and Simulation.

———————————————— ◆ ————————————————

1 INTRODUCTION

The Hybrid Activity Diagram (HAD) is a synthesis and fusion of concepts from two domains – Automatic Control Engineering and Software Engineering. The design and implementation of the formalism has been the subject of intense research [1], [2], [3], [4], [5]. The formalism is designed to model both the discrete and continuous parts of a system unlike single-paradigm tools like Grafcet [6] which model only one type of system.

Based on concepts of Object Orientation, HAD is designed to convert UML into a tool which is capable of modelling hybrid control systems, in order to facilitate the development of a simulator of such systems. In this regard, the focus of the article is to outline the design of HAD, as indicated in [1], [3], and to present its simulator which is known under the acronym SIMHAD. The functionality of SIMHAD is illustrated through an application which models and simulates a liquid-level control system. The model describes the causal relationships governing the operation of the system, the types of signals and the operational conditions of the system.

The primary advantage of the HAD metamodel is convertibility to both UML (Unified Modelling Language) and GRAFCET models through bridges which are described in the article.

The article is structured in six parts. Section 2 presents the fundamental concepts associated with hybrid systems and UML. The modelling of a liquid-level control system using UML is then presented in section 3 and this serves as a springboard for understanding and defining the properties of HAD. The bridges or mechanisms for converting the HAD model into Grafcet and UML Activity Diagram are presented in section 4. The specifications of the hybrid simulator SIMHAD and the results obtained from the simulation of the liquid-level control system are presented in section 5. The last part of the article, section 6, presents the conclusions and perspectives for further work.

2 HYBRID DYNAMIC SYSTEMS AND UML FUNDAMENTALS

A hybrid dynamic system is one which incorporates both continuous and discrete subsystems. The continuous subsystems are characterized by continuous-time variables while the discrete subsystems are event-driven and usually sequential in operation. More ample information on such systems can be obtained in [1], [3], [7], [8], [9]). A hybrid dynamic system can, thus, be represented as shown in figure 1.
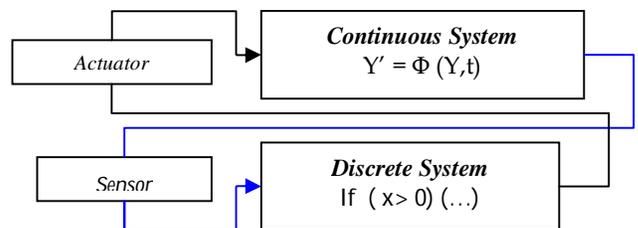
Fig. 1. Structure of Hybrid Dynamic Systems [9]

When the logic condition defined on the variable X is true, the actuator immediately triggers the operation (**Y′ = Φ(Y, t)**), resulting in a "sequential - continuous" interaction between the discrete and continuous subsystems [2]. When the continuous subsystem executes the specified action, this is detected by the sensor which provides information on the state or status of the actuator and the sequential subsystem is switched to the next state of the sequence. The result is a "continuous-sequential" interaction [2].

A hybrid dynamic system is, thus, characterized by interactions between the continuous and discrete subsystems. The evolution of the discrete subsystem from one state to another is predicated on the state of the logic va-

riables while the operation of the continuous component is based on physical laws which are usually described by differential or algebraic equations expressing cause-and-effect relationships. However, most of the tools and paradigms developed for the modeling and simulation of dynamic systems are either continuous or discrete ([1], [3], [8], [10]).

The primary focus of this research project is, therefore, to model and simulate hybrid dynamic systems using the Unified Modeling Language (UML) which is a universal, implementation-independent modeling language providing such properties as modularity, structured programming, re-usability and extensibility [11]. Based on Object Orientation, UML provides a multiplicity of diagrams which can be used to define the inheritance of attributes between related classes of objects as well as relationships between the components of a system [3], [11] [10].

3 ANALYSIS OF A LIQUID-LEVEL CONTROL SYSTEM TO HIGHLIGHT THE LIMITS OF UML

3.1 System Functionality

The system in figure 2 is taken from [12]. Each of the two tanks is fitted with two sensors, one of which detects when the tank is empty and the other detects when the tank is full. Sensor b1, in tank 1, detects when tank1 is empty while sensor h1 detects when the tank is full. Similarly, b2 and h2 have the same functions in tank 2. The sensors are, therefore, level-detectors and are modeled as binary rather than continuous variables.
The sequence of operations is as follows:

- Initially, both tanks are empty

- When the operator presses the switch m, the inlet valves V1 and V2 open and the tanks start filling up

- Once a tank is full, the corresponding inlet valve (V1 or V2) closes, to stop filling the tank, and the corresponding outlet valve (W1 or W2) opens, to start evacuating the contents of the tank.

- When a tank is empty, the corresponding outlet valve is closed

When both tanks are empty, the sequence repeats, to start refilling the tanks, when the operator presse the button *m*.

3.2 Modeling of the system using UML

3.2.1. Analysis of system functionality

The system functionality described in section 3.1 requires the following Use-cases: starting the system, opening of the valves (V1, V1, W1, W2), activation of the sensors (h1, h2, b1, b2), filling of the tanks and evacuation of the contents of the tanks. The actors include the operator, motor and tank, for each filling operation.

m : push-button to start the system
M : motor
h : maximum level detection sensor
b : maximum level detection sensor
W and V : valves
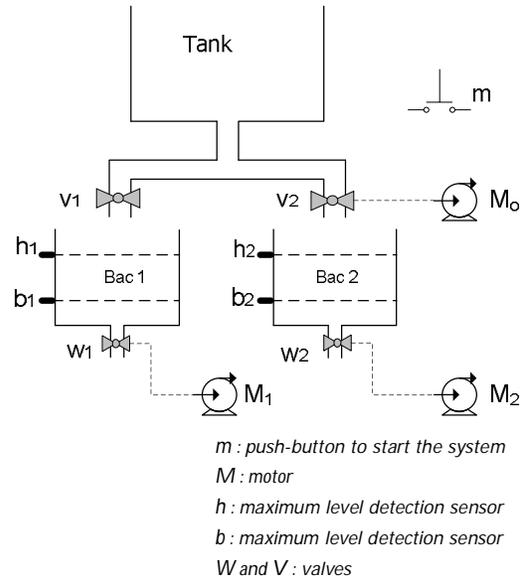
Fig. 2. Liquid-level control system [12].

The Use-case diagram (a fundamental UML construct for the description of the functions of a system), which associates the actors to their actions, is shown in figure 3.

This Use-case diagram shows the different sequences and highlights the role of each actor.

3.2.2. Other UML Diagrams required for the Modeling of the System

The Diagram of Classes in figure 4 shows both the static links between the actors as well as the action associated with each link.

The Diagram of Classes in figure 4 describes the attributes of the system components and the communication between them, but it does not describe the sequence of execution of the various tasks, neither does it define the conditions associated with the different tasks as is commonly the case with control systems modeling tools such as Grafcet and Petri nets. For this reason, Sequence Diagrams are required to describe the chronology in task execution. However, in order to avoid any ambiguity, more detailed description of the sequences are required.

We consider two scenarios.

Scénario 1 : *Sequence of actions in the filling of tank 1*

1. When the system operator presses the button m, motor Mo is turned on.
2. When motor Mo reaches its operational speed, it opens the inlet valve V1, to start filling tank 1.
3. When tank 1 is full (h=1), motor Mo reverses its direction of rotation. Mo is, thus, a bi-directional motor.
4. Motor Mo, rotating in the reverse direction, closes the inlet valve V1.

The Sequence Diagram for this sequence of operations is shown in figure 5.
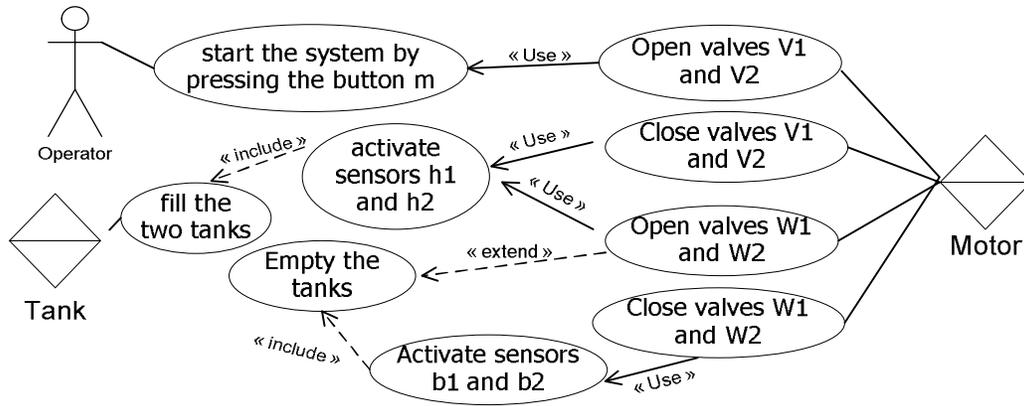
Fig. 3. Use-case diagram of the liquid-level control system
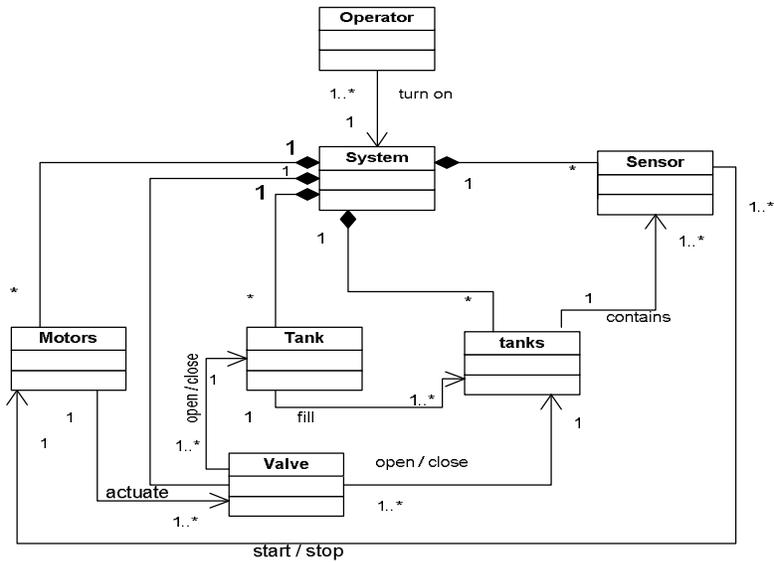
Fig. 4. Diagram of Classes of the liquid-level System
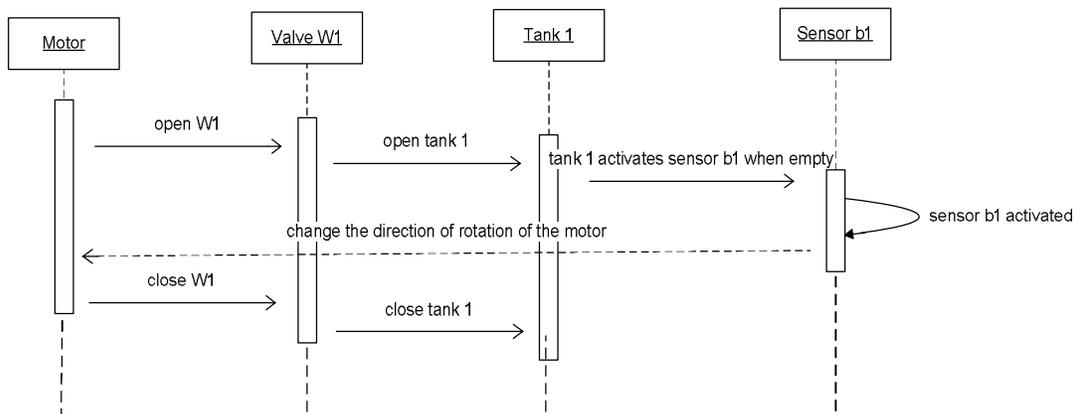
Fig. 5. Sequence Diagram for the filling of tank 1

Scénario 2 : *Sequence of actions in the evacuation of the contents of tank 1*

1. The activation of sensor h1 (last step in scenario 1) turns on motor M1
2. Motor M1 opens the outlet valve W1 for evacuation of the liquid
3. When tank 1 is empty (b1=1), motor M1 reverses its direction of motion
4. The rotation of M1, in the reverse direction, closes valve W1.

The corresponding Sequence Diagram is shown in figure 6.

The two Sequence Diagrams (figures 5 and 6) provide detailed information on the order in which the different actions are executed during the filling and evacuation of the contents of tank 1. These Sequence-diagrams are also valid for tank 2, by simply substituting the variables of tank 1 with those of tank 2. However, they do not describe sequence selection and simultaneous (parallel) sequences. The appropriate structure for such sequences is the Activity Diagram [3],[11]. Figure 7 presents the Activity Diagram for the overall system functionality involving the filling and evacuation of the contents of the two tanks.
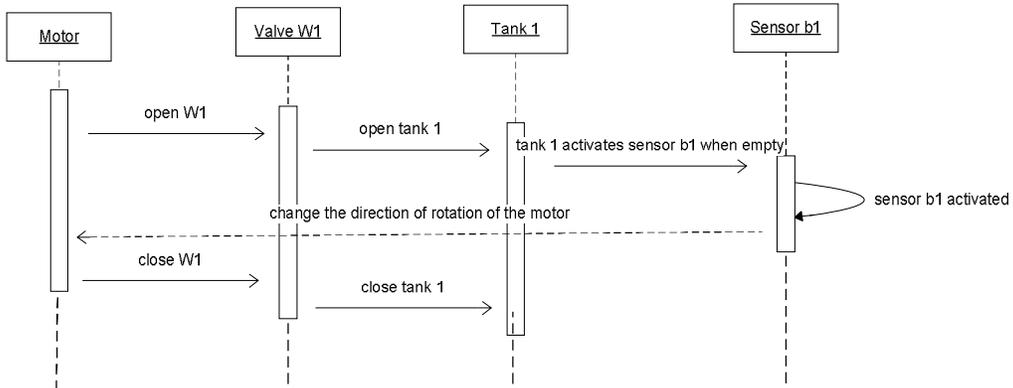
Fig. 6. Sequence Diagram for the evacuation of the contents of tank 1.
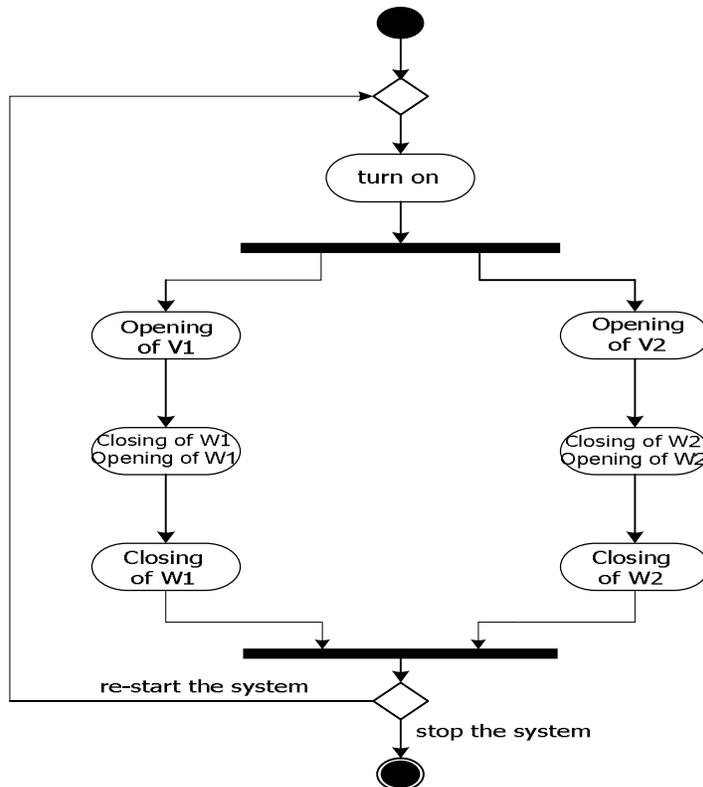
Fig. 7. Activity Diagram for the overall system functionality involving the filling and evacuation of the contents of the two tanks

The diagram in figure 7, in addition to combining the information in figures 5 and 6, shows the conditions for

simultaneous and conditional activation of sequences as well as the order of execution of the actions in a sequence. However, the conditions (state of the sensors) driving the sequence from one state to another, are not shown.

The various limits of UML highlighted in its application to the modeling of the liquid-level system confirm the assertion in [1], [2] that UML does not model the cause-and-effect phenomena in dynamic systems. This creates the need for the HAD metamodel which allows sequences to be modeled as Object-oriented entities incorporating cause-and-effect relationships and the capacity to interchange messages between objects.

4. EXTENSION OF UML TO THE MODELING OF HYBRID SYSTEMS: APPLICATION TO THE LIQUID-LEVEL SYSTEM

Inspired by the Activity Diagram [4], [11], [13], HAD is designed to model a system as a collection of localized, autonomous objects having both static and dynamic properties and incorporating a representation of the physical phenomena associated with an object. For the most part, the phenomena are described by differential equations, but algebraic and logic equations are sometimes used.

4.1. Analysis of the filling of the two tanks

The opening of a valve requires the actuation of the valve stem by the motor. The motion of the motor, coupled to the valve stem, is described by:

$$\frac{J_t}{f_r}\frac{d\Omega}{dt} + \Omega = \frac{T_{em} - T_r}{f_r} \qquad (1)$$

And the electromagnetic torque which drives the motor shaft and valve stem is given by :

$$T_{em}(n) = K.V_1^2 . \frac{-(R_2 \cdot n_s)n + R_2 n_s^2}{X_2^2 \cdot n^2 - 2(n_s \cdot X_2^2)n + n_s^2(R_2^2 + X_2^2)} \qquad (2)$$

Where,

f_r *frictional torque constant ;*

J_t *is the combined moment of inertia of the motor shaft and load ;*

R_2 *and* X_2 *are the resistance and reactance of one phase of the motor ;*

n_s *is the synchronous speed*

Details of these equations are available in [17].

The block diagram in figure 8 regroups and organizes the causes and effects of the system into a single entity. The sensors and push-button "m" are the causes while the valves which are the controlled components are the effects.
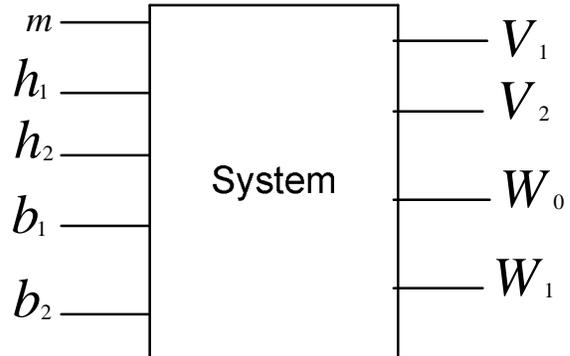
Fig. 8. Block Diagram for the filling of the two tanks

The corresponding HAD diagram is shown in figure 9.

The HAD diagram is an abstract representation of the filling process and provides a more realistic description of the system since there is greater visibility of the signals (discrete and continuous) interchanged and a better expression of the alternation between the sequential and continuous subsystems. Consequently, it describes the physical phenomenon (through equations) which govern the behavior of the system. It also facilitates the analysis of the system at three levels : visual, syntaxic and semantic. This, in turn, facilitates the development of the simulation model.

4.2. Bridges between HAD and other Metamodels

The bridges are pre-defined rules which are embedded in the HAD diagram for the automatic generation of equivalent UML or Grafcet models. Since this article is not entirely devoted to these bridges, only a subset of these rules and their implementation will be presented here.

4.2.1. The HAD – UML Bridge

Table 1 presents an excerpt of the rules for converting a HAD model to UML.

These rules of conversion have been succesfully tested on several systems ([14], [15], [5]). The application of the rules to the HAD diagram in figure 9 generates the Activity Diagram in figure 10, which conforms to the UML formalism [13].

OU: ``OK''>S1>

S1

IN:< BE < T/S <
S:< BE< OR <T/S <
OU: ``OK'' > **0** >

0

IN:< 0 <
S:< 0 <
OP: V1=0 ; V2=0 ; W1=0 ; W2=0 ; ()
OU: ``OK'' >T1>

IN:< 0 <E1<
S:< 0 <AND<E1<
OP: opning the valves V1 = V2=1 ()

OU:`` turn on''m >T1>

E1   [m=1]   **T1**

$$\begin{cases} \dfrac{J_t}{f_r}\cdot\dfrac{d\Omega}{dt} + \Omega = \dfrac{T_{em} - T_r}{f_r} \\[2ex] T_{em} = K.V_1^2\cdot\dfrac{-(R_2\cdot n_s)n + R_2 n_s^2}{X_2^2\cdot n^2 - 2(n_s\cdot X_2^2)n + n_s^2(R_2^2 + X_2^2)} \end{cases}$$

OU: ``OK''>1>; ``OK''>2>;

OU: ``action'', swich h1>1> [h1=1]
   ``actioin'', swich h2>2> [h2=1]

E2   **1**

IN:<TI<E2<
S:<T1<AND<E2<
OP: Closing V1=0; Opening W1=1 ()

2

IN:<T1<E2
S:<T1<AND<E2
OP: Closing v2=0; Opening W2=1 ()

$$\frac{J_t}{f_r}\cdot\frac{d\Omega}{dt} + \Omega = \frac{T_{em} - T_r}{f_r}$$

$$T_{em} = K.V_1^2\cdot\frac{-(R_2\cdot n_S)n + R_2 n_S^2}{X_2^2\cdot n^2 - 2(n_S\cdot X_2^2)n + n_S^2(R_2^2 + X_2^2)}$$

OU:``OK''>3>

$$\frac{J_t}{f_r}\cdot\frac{d\Omega}{dt} + \Omega = \frac{T_{em} - T_r}{f_r}$$

$$T_{em} = K.V_1^2\cdot\frac{-(R_2\cdot n_S)n + R_2 n_S^2}{X_2^2\cdot n^2 - 2(n_S\cdot X_2^2)n + n_S^2(R_2^2 + X_2^2)}$$

OU:``OK''>4>

OU: ``action'', swich b1>3>   [b1=1]
   ``actioin'', swich b2>4>   [b2=1]

E3   **3**

IN:<1<
S:<1< AND <E3<
OP: Closing W1=0 ()

4

IN:<2<
S:<2< AND < E3<
OP: Closing W2=0 ()

$$\frac{J_t}{f_r}\cdot\frac{d\Omega}{dt} + \Omega = \frac{T_{em} - T_r}{f_r}$$

$$T_{em} = K.V_1^2\cdot\frac{-(R_2\cdot n_S)n + R_2 n_S^2}{X_2^2\cdot n^2 - 2(n_S\cdot X_2^2)n + n_S^2(R_2^2 + X_2^2)}$$

OU: ``OK'' >T/S>

$$\frac{J_t}{f_r}\cdot\frac{d\Omega}{dt} + \Omega = \frac{T_{em} - T_r}{f_r}$$

$$T_{em} = K.V_1^2\cdot\frac{-(R_2\cdot n_S)n + R_2 n_S^2}{X_2^2\cdot n^2 - 2(n_S\cdot X_2^2)n + n_S^2(R_2^2 + X_2^2)}$$

OU: ``OK'' >T/S>

T/S

IN:<3<4<
S:<3<AND<4<
OU:``OK''>S1> [re-start the system]
   ``OK''>END> [stop the system]

IN : <T/S<

Fig. 9. HAD Model for the filling of the two tanks

TABLE 1
EXCERPT OF THE RULES USED IN CONVERTING HAD MODELS TO UML

Beginning of simultaneous sequences: "AND" - divergence	**T** IN : S: OU:	Object without an operation	
	T IN: S: OP: op1 () , op 2 () OU:	Object with operation (op 1 and op 2 become activity 1 and activity 2 respectively)	Activité 1 Activité 2

End of simultaneous sequences : "AND" - convergence	T — IN: S: OU:	Object without an operation	(diagram)
	T — IN: S: OP: op1(), op2() OU:	Object with operation (op 1 and op 2 become activity 1 and activity 2 respectively)	Activity 1, Activity 2 (diagram)
"AND" - divergence and convergence	T — IN: S: OP: op1(); op2() OU:	If there is no operation, the two activities after the bar are deleted	Activity 1, Activity 2 (diagram)
"OR" - divergence and convergence	S — IN : S: OP: op1(), op2() OU: > 'ok '> [garde 1] >'ok'> [garde 2]	If there is no operation, the activities which are present are deleted	[condition 1], [condition 2], Activity 1, Activity 2 (diagram)
"OR" - convergence "AND" – divergence	S/T — IN: S: OP: opération ( ) OU:	If there is no operation, the activity which figures on the object is deleted	activity (diagram)
"AND" - convergence "OR" – divergence	T/S — IN: S: OP: opération () OU: >'ok' > [garde 1] > 'ok' > [garde 2]	If there is no operation, the activity which figures on the object is deleted	activity, [condition1], [condition2] (diagram)
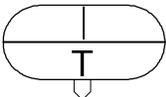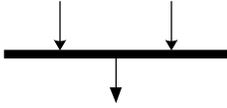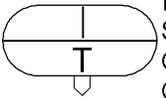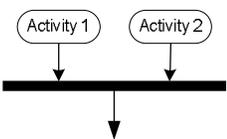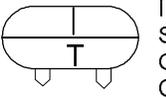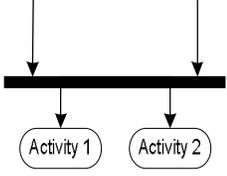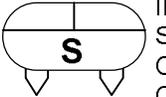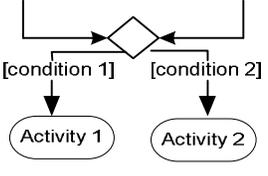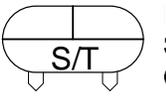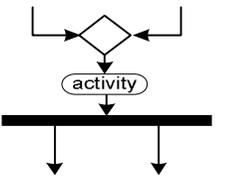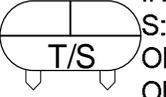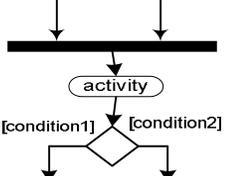
4.2.2. The HAD – Grafcet Bridge

The HAD-Grafcet bridge is designed to perform a two-stage transformation. The HAD source-code is first transformed into an intermediate model, which is then transformed into Grafcet

Intermediate Model for the transformation of HAD into grafcet

Only an excerpt of rules is presented here. More details of the relationships between the two metamodels are presented in [14], [15].

- Eliminate the object which marks the beginning and end of the HAD metagraph

- Eliminate all mathematical equations assigned to the attribute "*operation*" which describes the dynamics of an object

External influences (ActivityCauses, ActivityEffects) and internal influences (ActivityClasses) to which the system is subjected are equivalent to conditions on Grafcet transitions.

Transformation of Intermediate Model into Grafcet

An excerpt of the rules in [14], [15] is presented here. Analyzing the Grafcet from top to bottom :

- Delete all parallel or conditional sequences (sequence selection) which do not have steps between them ;

- Merge any two consecutive transitions which are not separated by a step

The application of this bridge to the HAD model for the filling of the two tanks generates the Grafcet in figure 11, which conforms to the Grafcet principle of the alternation of steps and transitions.

These models illustrate the fact that HAD models provide greater visibility and abstraction of actions, operations and interchange of signals of a hybrid dynamic sys-

tem, compared to the other modeling paradigms used in Automatic Control Engineering. HAD extends UML and adapts it to the requirements of Automatic Control Engineering. This property makes it an interesting tool. In the next section, we show that its syntax and structure facilitate the development of software for the simulation of hybrid dynamic systems.
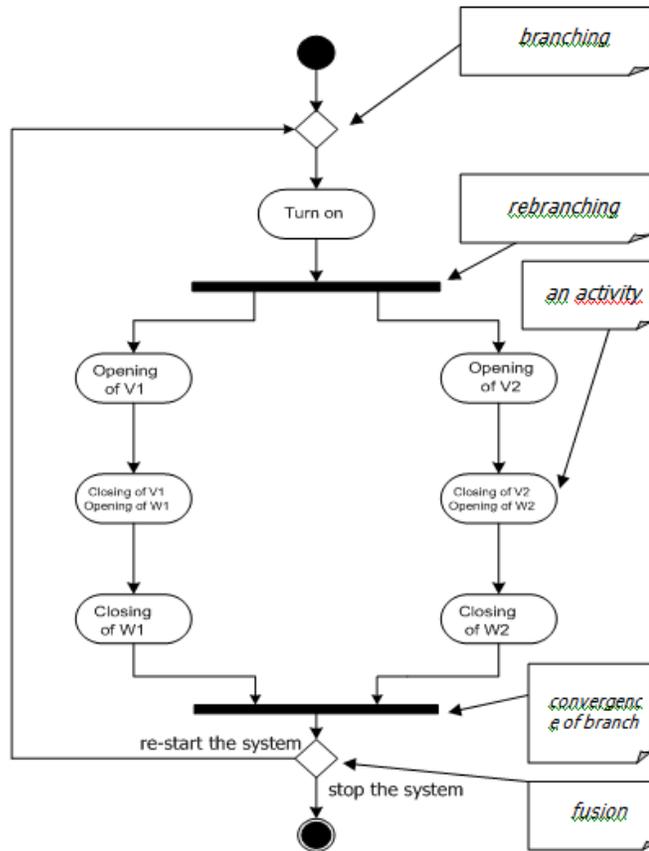
Fig. 10. Activity Diagram generated from the HAD model for the filling of a single tank.
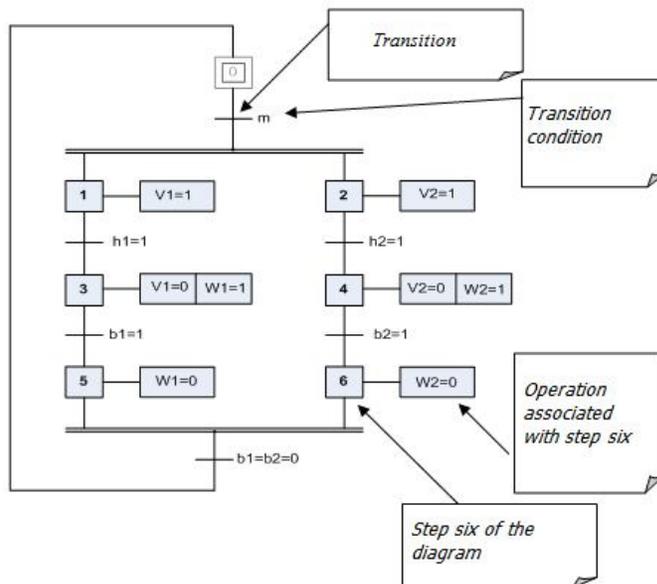
Fig. 11. Grafcet generated from the HAD model for the filling of two tanks,

5. TECHNICAL SPECIFICATIONS AND ABSTRACT

HYBRID SIMULATION OF SYSTEMS

5.1. Technical Specifications

The HAD metamodel has the advantage of being an extension of UML which is widely used in software engineering [13]. This is the motivation and justification for the development of a simulator of HAD models.

This aspect requires a specification of dozens of constraints. Some of these constraints are highlighted in this section. The full details are available in [14], [15].

5.1.1. Constraints on the Structure of HAD Models

- CH 1: An object is either active or inactive at any given time.

- CH 2: An object is associated with one or more operations.

- CH 3: Two objects communicate through influences (sending and reception of messages) which are either internal or external.

- CH4: An object can be subjected to internal and external influences at the same time. This is a corollary of CH3.

- CH 5: The alternation of transmission and reception of influences must be observed at all times.

- CH 6: The initial and terminal objects must be at the appropriate places.

5.1.2 Constraints on the interconnection of objects:

-CH 7: The ActivityCause – ActivityClass or ActivityNoEffect must be represented as horizontal dotted lines.

-CH 8: The ActivityEffect – ActivityClass or ActivityNoEffect must equally be represented as horizontal dotted lines.

-CH 9: The links between the objects : ActivityClass, ActivityNoeffect, ActivitySelect, ActivityThread, ActivitySelect/thread, ActivityThread/Select, must be represented as vertical solid lines.

5.1.2. Constraints on the description of entities

- CH 10: All objects are inactive during their description.

- CH 11: The order of a differential equation must be greater than or equal to 1.

- CH 12: Algebraic equations are valid if and only if their coefficients are correct.

- CH 13: The fields reserved for the parameters of an object must always contain data.

- CH 14: For every ''ActivityConnect'' and *ActivityModul (except ActivityCause), provide a connection point to ActivityCause* except for input/output connection points.

- CH 15: All inputs and outputs of an object must be connected.

- CH 16: The terminal object has only one input and no output while the initial object has only one output and no input.

- CH 17: The system must allow the reusability of values entered into the simulation or calculated by it and stored in memory.

- CH 18: Objects should have access to parameters entered into the simulation or calculated by it and stored in memory.

- CH 19: The interface for model construction must provide graphical objects in a menu, allowing for their reusability, modification and extensibility.

- CH 20: A logical consequence of CH19 is to provide a minimum configuration of generic objects which can be used to construct all other objects.

5.1.3. Constraints on the simulation

- CH 21: The simulation of a HAD model resulting from the reception of a message from one object to another can only evolve when the message is validated and its content is true.

- CH 22: The reception of a message causes the simultaneous activation of immediately following objects and the deactivation of all immediately preceding objects.

- CH 23: Every object has a unique number

-CH 24 : To represent or describe HAD, the following messages are required:

"*Displace*", "*Duplicate*", "*Delete*", "*Modify*", "*Mark*", "*Search*", "*Select*", «*Read parameter* ». Consequently, these methods will be declared as "*public*" in the source code of the simulator.

It is easy to implement the simulator by applying all of these constraints.

5.2. The Simulator of HAD Models (SIMHAD)

The simulator has been modelled using UML 2 [13], to ensure the extensibility and reuseability of its source code. It is implemented in multithread Java [16]. The design details of the simulator will not be presented in this article. The simulator incorporates six (6) different interfaces, three of which include the model construction interface, the model description interface and the simulation interface.

When the simulation is in progress, a dialog box displays any errors in the model. Such errors may result from the absence of parameter values, inappropriate connectivity of objects or some other violation of a constraint. The dialog box also shows the time that has elapsed since the simulation was launched. A red dot moves from the top of the diagram towards the bottom, to show the order in which the objects are activated. This is the sequential phase of the system. By double-clicking on an object, the

curves representing the continuous-time dynamics of the object are displayed.

Note that the curves represent the operations or methods associated with the object. Figure 12 shows the simulation interface during the simulation of the liquid-level control system. Figure 13 illustrates the simulation of the HAD model of a system.

The curves are plotted by first resolving equation (1), presented in section 4, to give the result,

$$\Rightarrow \Omega(t) = (1 - e^{-\frac{f_{r_t}}{J_t}t})\frac{T_d - c}{f_r} \qquad (3)$$

This expresses the speed as a function of time.

The simulator then plots graphs of equations 2 (section 4), 3 and 4.

$$T_r(n) = a.n^2 + b.n + c \qquad (4)$$

Equation 4 represents the torque of the motor shaft and load.

The different curves obtainable from a SIMHAD simulation are presented in [14], [15]. For example, by double-clicking on object number 2 of the diagram in figure 12, the results in figure 13 are obtained.

The speed curves obtained from the simulator are consistent with those found in the literature. The simulator also highlights one of the advantages of HAD – encapsulation of the actions and interactions of the hybrid system.

Details of the functionality and implementation of SIMHAD will be the subject of the next article to be published by the authors.
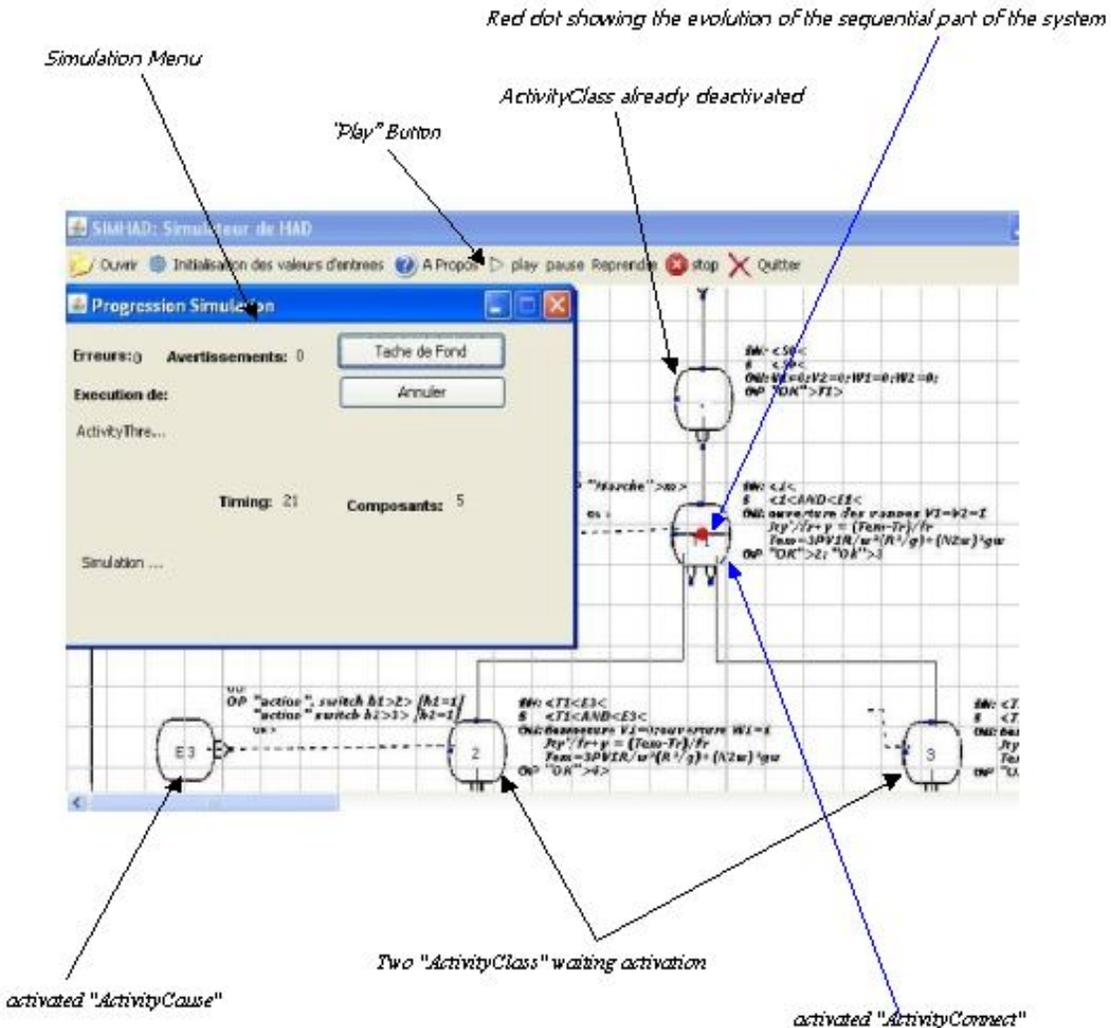
Fig. 12. Simulation Interface showing the liquid-level control system
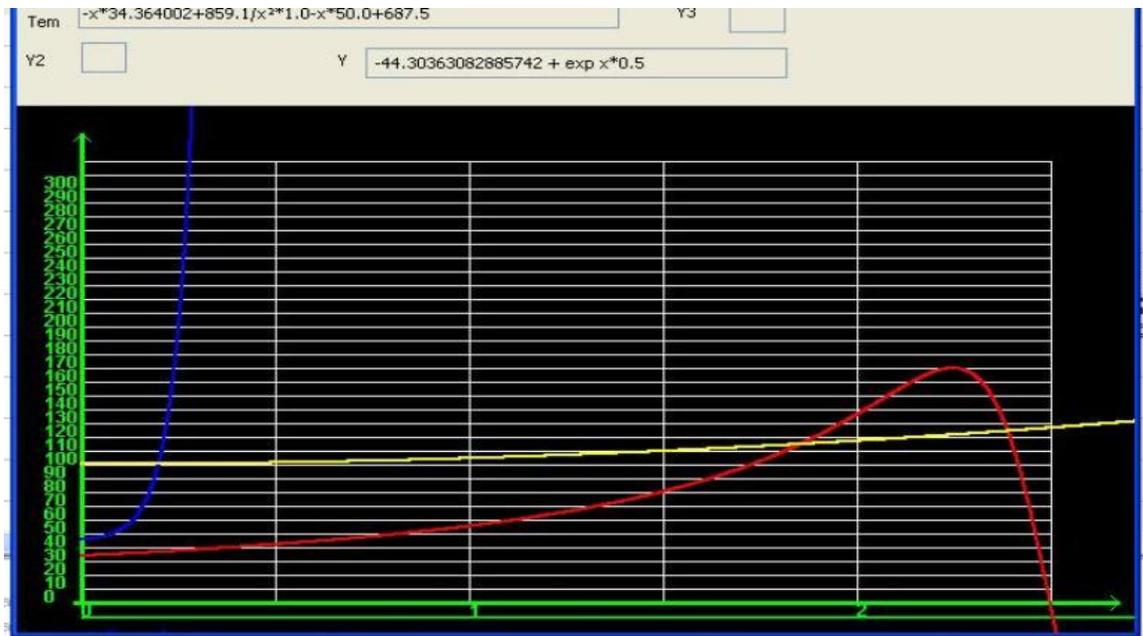
Fig. 13. Curves obtained from SIMHAD

6. CONCSLUSIONS AND PERSPECTIVES

The current phase of our research project, described in this paper, has a triple objective :

- Illustrate the functionality of the HAD metamodel as a tool which is specifically adapted to the modeling of hybrid dynamic systems such as the liquid-level control system

- Present the bridges which serve as mechanisms for converting HAD models to UML and Grafcet

- Present SIMHAD – a simulator of HAD models

All of these three objectives have been fulfilled.

The next phase of the research project will focus on the extension of SIMHAD. Two aspects merit particular attention:

- Integration of the bridges within the simulator, to facilitate the automatic generation of Grafcet and Activity Diagrams from a HAD model

- Solution of nonlinear differential equations, to extend the range of continuous-time models which can be processed by the system.

REFERENCES

[1] Tanyi E., Nkenlifack M., "An extended UML for the modeling of hybrid control systems", in Burnham K.J., Haas O.C.L.(Editors), *Proc. of the sixteenth International Conference on Systems Engineering (ICSE2003)*, Coventry, UK, 9-11 September 2003, Vol.2, pp.681-686, ISBN 0-905949-91.

[2] Tanyi E., Nkenlifack M., "Hybrid Activity Diagrams: Extending UML for the Modeling of Hybrid Systems", Poster in *ECOOP'03, 17th European Conference on Object –Oriented Programming*, July 21-25, 2003, Darmstadt University of Technology, Germany, http://www.st.informatik.tu-darmstadt.de:8080/ecoop/posters/index.phtml - Site Web poster.

[3] [3] M. Nkenlifack, "Modélisation objet et développement d'un atelier de simulation des automatismes industriels hybrides", PhD Thesis, National Advanced School of Engineering (Ecole Nationale Supérieure Polytechnique), University of Yaounde I, Cameroon, 2004.

[4] E. Tanyi, M. Nkenlifack, "An object oriented simulation platform for hybrid control systems", *Analysis and Design of hybrid systems (ADHS) 2003, Proc. of the IFAC International Conference*, St Malo, France, June 16-18 2003, Edited by S.Engell, H Gueguen&J.Zaytoon, ISBN 0-08-04044094-0.

[5] T. Noulamo, Modèle Métier et Architectures Génériques pour la Commande et la Surveillance des Systèmes Dynamiques, PhD Thesis, Faculty of Sciences, University of Yaounde 1, Cameroon, nov. 2010.

[6] CEI-IEC (International Electrotechnic Commission), Grafcet specification language for sequential function charts, International Standard, IEC 60848, 2002.

[7] T. Noulamo, E. Tanyi, M. Nkenlifack, J. Lienou, « Domain Specific Model and Generic Architectures for Control and Monitoring of Dynamic Systems », Advances in Computer Science and Engineering Journal, Volume 4, Issue 1, Pages 55 - 75 (February 2010), Pushpa Publishing House, India.

[8] J. Zaytoon, *Systèmes Dynamiques Hybrides, Traité Systèmes automatisés, Information Commande et Communication*, Hermes, Paris, 2001.

[9] O. Bouissou, M. Martel, "Analyse Statique par Interprétation Abstraite de Systèmes Hybrides Discrets-Continus", *CEA LIST* Laboratoire de Sûreté des Logiciels 22 septembre 2005.

[10] H. Brenier, *Métamodèle UML Les spécifications fonctionnelles des automatismes industriels et temps réel*, 303–362, Dunod, Paris, 2001.

[11] M. Flower, *UML Distilled*, Second edition, Addison-Wesley Longman, Inc, 2000.

[12] R. David, Alla H., *Du Grafcet au Réseaux de Petri, série automatique*, 2em édition revue et augmentée, édition Hermes 1992, 1997 500 pages (271).

[13] OMG, UML 2.0 Reference Manual, //www.omg.org, 2008.

[14] L. Domche, "Metamodele HAD-Grafcet, Analyse et Mise en Œuvre dans le cadre de la Commande des Réseaux Electriques : cas du Réseau Sud AES SONEL Cameroun", Bachelor Thesis, Electrical Engineering, IUT FV of Bandjoun, University of Dschang, Cameroon, 2009.

[15] F. Fokou, "Metamodele UML–HAD, Analyse et Mise en Œuvre dans le cadre de la Commande des Réseaux Electriques : cas du Réseau Sud AES SONEL Cameroun", Bachelor Thesis, Electrical Engineering, IUT FV of Bandjoun, University of Dschang, Cameroon, 2009.

[16] C. Delannoy, Programmer en Java, Eyrolles, France, 2001.

[17] T. Wildi and G. Sybille, Electrotechnique, 3em edition, DeBoeck Université.

Study of Microstructure and Mechanical Properties of Human Cortical Bone

S. Biswas, P. C. Pramanik, P. Dasgupta, A. Chanda

Abstract— The microstructure along with the micro-mechanical properties of bone have been extensively observed in this study. It has also been attempted to correlate the mechanical properties with the microstructural aspects. The load deformation pattern of human cortical bone (both male and female) under uniaxial compressive and tensile loading was studied. The physical properties like density and porosity of the whole femur as well as different parts of it was studied to get a mapping of the femur. Microstructure of bone, including the osteon structure, its distribution and its deformation under different stress fields have been thoroughly studied with the help of Scanning electron microscopy and reported here.

Index Terms—bone porosity, fractography, hardness anisotropy, load-deformation curve, Osteon deformation.

———————————— ◆ ————————————

1 INTRODUCTION

Bone is a highly hierarchical structure and its mechanical properties are dependent on its structure and composition. Previous studies have focused on the macroscopic properties and their changes with respect to age and diseased condition. One such study showed with increasing tissue age, mineralization increases at certain portions and these hypermineralised portions become stiffer to some extent [1]. In another study about the post yield behavior it was found that middle-aged bone specimens demonstrated higher yield strain and yield stress than those from the elderly ones [2]. Recent studies are focusing on the microstructural effect [3], [4]. Due to the presence of haversian canals as an integral part of its structure, bone is quite porous in nature. Studied have been done to find out the effect of porosity on mechanical properties [5]. A significant negative correlation was found between the elastic anisotropy and porosity of cortical bone [6]. Density also varies from person to person and also with age. Osteoporosis has become a serious issue all over the world and efforts have been made to predict the risk of fracture through different techniques [7], [8]. Nanoindentation studies have also become popular for determining the hardness and elastic modulus of bone [9]. Variation of properties at specific sites like lamella under different physiological conditions was also studied by few people [10]. Fracture toughness was another area of interest for many researchers who tried to correlate it with age and microdamage accumulation [11], [12]. In one of the very recent studies with the help of finite element models and wavelet transforms, complete mapping of the energy descipation in bone after nanoindentation was shown. [13]. Laser engineered net shaping (LENS) is a rapid prototyping technique in which the machine gets instructions from CT Scan data and can produce or mimic a part of bone with powdered materials and without using any other tools. Very recent studies indicate that the in vivo life of implants depends on the porosity and mechanical properties of the material which should match with that of bone [14]. In this study we

have tried to map the apparent density and the porosities of different parts of femur both in males and in females. The load-deformation curve was also studied in details for both male and female under compression and tension to get the idea of its mechanical properties and the unde deformation of osteons under stress field was observed. Hardness was measured in both males and females throughout the femur shaft and its anisotropy was observed.

2 METHODOLOGY

Femurs from both male and female human cadavers were collected, pretreated and cleaned. Density and porosity of the whole femur (shown in Fig-1a) were calculated by applying Archimedean principle. Then after drying they were sectioned into different parts as shown in Fig-1b.
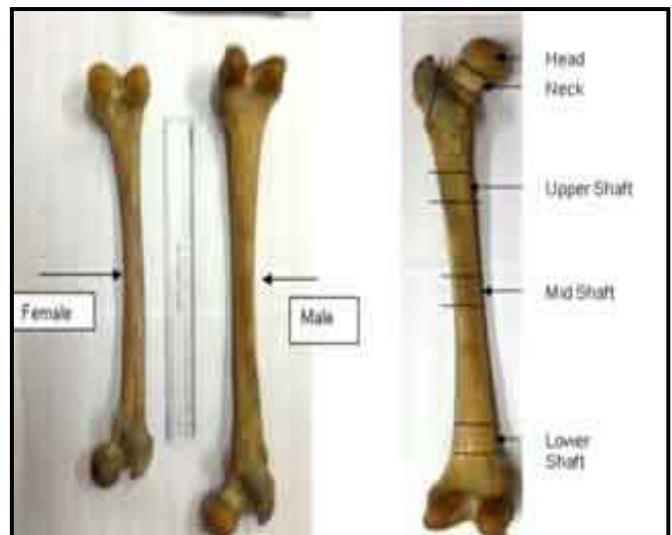
Fig. 1. Human Femur and the different parts in which it was sectioned for measurements.

Then density and porosity of individual parts were also measured in the same way. A variation between the parts was very distinct. For observing the mechanical properties, samples of different sizes were cut and prepared following the ASTM standard (ASTM F 451 RE).The load-deformation curve of the cortical parts was obtained for both tension and compression from Universal Testing Machine (UTM: INSTRON 4204, U.K.). Young's modulus and work of fracture was calculated from the load-deformation curve itself. The loading was done till fracture for few samples, while in others it was stopped at various intermediate points, e.g. in the linear zone, in the plastic flow zone and then at the point of fracture to observe the deformation of the osteons at different stages of loading. This was mainly performed under compression as preparation of samples for tensile testing was little bit difficult due to some machine constraints. Experiments were also performed at different crosshead speeds of UTM i.e., with varying strain rate to determine its effect on stiffness of the bone. The cross section of the samples was polished in the polishing machine (LECO Spectrum System 1000, USA) with different grits until the value for the average centerline reached around 0.2534 micro meters.

Hardness was measured in both longitudinal and transverse direction to check the anisotropy using Vicker's Hardness testing machine (LV 700AT, LECO, USA). The structures of osteons were observed from the cross sectional images of the cortical bone (of both male and female) by a Reflecting inverted optical microscope (Olympus GX51, Japan). Using image analysis software (Analysis five, Olympus, Japan) the aspect ratio, lamellar thickness, diameter of haversian canals were calculated.

3 RESULTS AND DISCUSSIONS

The structure of bone was thoroughly observed under the optical microscope under different magnifications. Osteons were clearly visible in the polished cross sections of the cortical part. Osteonal density was slightly different in different regions of the cross section. The average number of osteons per mm^2 was nearly 19 (1 .9 x 10^{-5} per sq. micron) in the inner part (towards the bone cavity) to about 14 (1.3 x 10^{-5} per sq. micron) in the outer part (Fig-2). With increasing magnification, the lamella, which are the concentric layers of tissue around the Haversian canal, were clearly visible. The thickness of the lamella ranged from 6.25 micron to 10.55 micron from inward to outward with a scatter of ±10%. Almost all the osteons had a combination of thin and thick lamella. The diameter of the Havesian canal had an average value of 78.18 micron towards the periphery as compared to 54.66 micron in the inner part with a large scatter in the values all over the region. Towards the inner side, the osteons were found to be more closely spaced which may be indicative of comparatively higher rate of bone formation in that region. Pictures taken from the scanning electron microscope revealed even minute structures called the lacune which are

the sites for the bone cells (osteocytes).
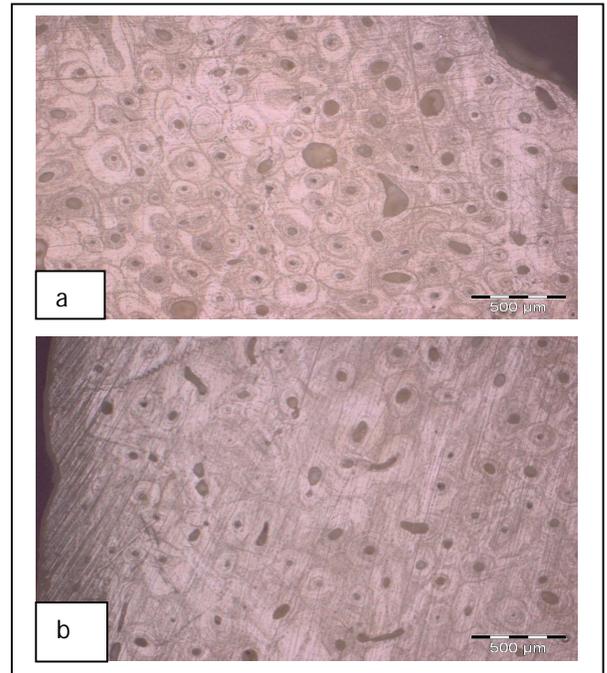
Fig. 2. Cross section of cortical bone (a) inner part (b) outer part man

The lacunae were visible along the layers of lamella. Interconnections between them are established by the canaliculi. Thus the structure reveals that such discontinuities are integral part of the bone structure. It will be shown later how they contribute to the fracture of bone by promoting the coalescence of cracks under different types of loading. Therefore concentration of porosity is to be taken into consideration when properties of bone are concerned. We measured the porosity of the whole femur and found it different when compared to the porosity of different parts of the same femur considered separately. Fig-3 shows the porosity distribution of in two male and two female samples. The cortical part is quite less porous than the head or the trabecular part in general. The density in females was found to be slightly lower in males as shown in Fig-4.
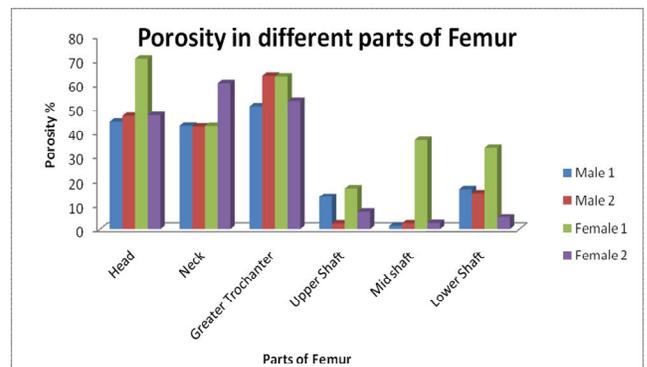
Fig. 3. Porosity distribution in Femur
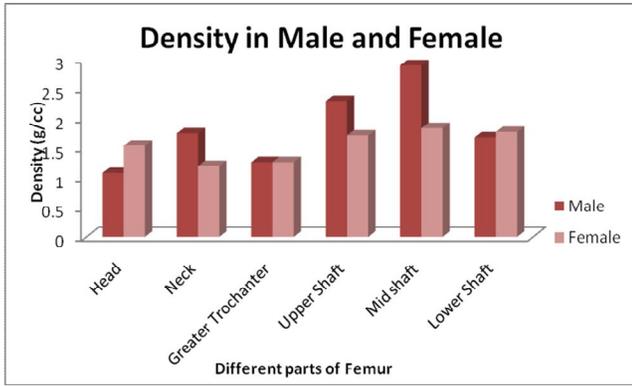
Fig. 4. Density distribution in male and female femur

Fig. 5. Representative Load-Deformation curves a) For Male samples b) For Female samples of Femur

But the results of a students t-test performed on both the sets of data showed that the difference in males and females were not significant at 95% confidence level. In our samples, porosity of one female femur was extremely high than others. This may be due to increased suscepti-bility to osteoporosis at higher age.

The load-deformation curves were obtained from compressive and tensile loading to study the mechanical properties. The initial part showed a stunning similarity for all the samples both in male and female. The load-deformation curve could be divided into two zones i.e., the elastic zone which was linear in all cases, and a nonli-near zone. Under compression, after reaching the peak load the curve was more flat in females than in males. The nonlinear zone revealed number of kinks representing initiation of fracture, progressive fracture and finally complete breakage occurred with a comparatively sharp drop of load. However during experiment, at these in-termediate points no major splitting was observed, there were some primary signs of cracking. The parts of load-deformation curve corresponding to fracture zone were not truly identical in all the cases but the general pattern was comparable. Figure-5 shows a representative curve for the load deformation curves in both male and female with error bars showing 5-10% scatter at each data point.

The equations fitting the individual curves of these two zones are given in the Table1. In the elastic zone, they are all fitted into linear equations with $R^2 > 0.95$ indicating a good amount of fitting, although slopes may have been different. For the non-linear part, most of the curves fitted with a 2nd degree polynomial and few with 3rd degree, all having R^2 values greater than 0.9. The ultimate strength was in the order of 158.66 MPa with standard deviation of 5.9 in females and 150.76 MPa with standard deviation of 11.9 in males. The elastic modulus was also calculated. The stiffness value determined in our study ranged from 10 to 15 GPa which was also verified by the nanoindenta-tion technique. Work of fracture, which is the energy ab-sorbed by the bone before complete fracture was obtained separately from the area under each curve and was slightly more in females in most of the cases than in males.

TABLE 1
EQUATIONS DESCRIBING THE LINEAR AND NON-LINEAR PART OF THE LOAD-DEFORMATION CURVES

Sample	Linear part	R² Value	Non-Linear Part	R² Value
Male 1	y=3.320x-32.39	0.97	y=0.183x³-10.4x²+192x-1144	0.957
Male 2	y=3.135x-31.5	0.94	y=0.113x³-6.622x²+126.2x-774.9	0.908
Male 3	y=2.611x-27.09	0.947	y=0.021x²-1.595x+35.62	0.952
Male 4	y=6.2635x-6.7718	0.976	y=1.5574x³-11.456x²+25.687x-14.891	0.962
Female 1	y=3.196x-34.25	0.967	y=-0.038x²+1.286x+5.264	0.96
Female 2	y=3.627x-38.50	0.962	y=-0.094x²+3.682x-20.67	0.965
Female 3	y=5.919x-6.243	0.962	y=11.85x³+65x²-121.8x+78.02	0.98
Female 4	y=6.8574x+7.6309	0.946	y=-2.6055x²+8.0409x-2.667	0.955

The difference in the average values of work of fracture has been around 10% which is well within possible experimental scatter. Rather we should say the values were quite comparable. This strongly contradicts the prevailing hypothesis of higher vulnerability of female bone due to increased porosity particularly at higher age. The samples taken were rectangular in shape and in case of both tensile (Fig-6a) and compressive (Fig-6b) fracture it was observed that the fractured surface made an inclination with the vertical axis. The crack propagated making almost a 30° angle with the vertical axis of the bone.
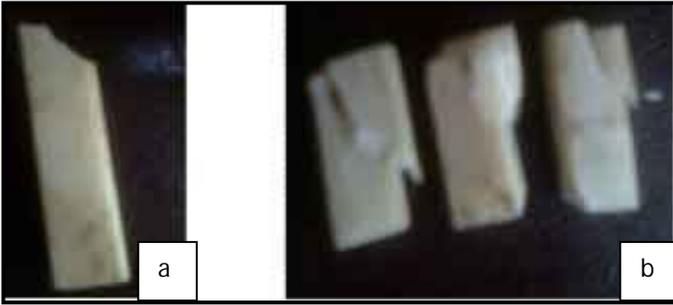
Fig. 6. Fracture plane making an angle with the vertical axis (a) under tension (b) under compression

In few cases of compressive loading for female bone samples, the samples experienced a crushing pattern where instead of forming asharp edge it got smashed from the top up to the bottom surface. The slope of the load-deformation curve changed as the crosshead speed was varied from 0.5mm/min to 4 mm/min. It was thus in accordance with the previous findings that stiffness of the bone increases with increasing loading rate or strain rate in case of tensile loading [15].Fig-7 shows the variation of stiffness with varying crosshead speed, or in other words varying strain rate. Thus from this figure it can be observed that not only under tension,
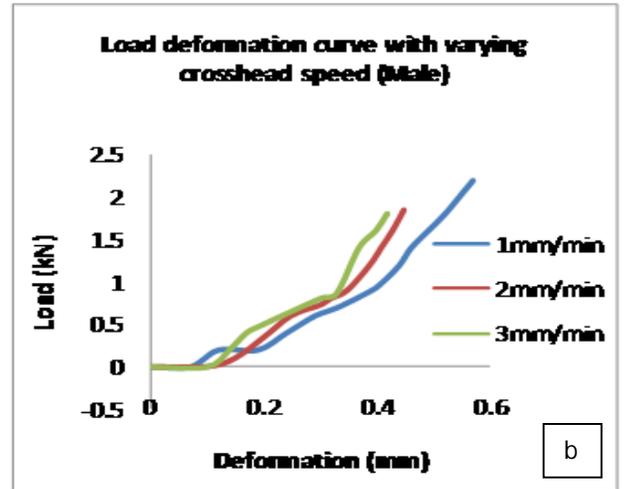
Fig. 7. Load-Deformation curve with varying cross head speed
(a) for Female (b) for Male

effect of varying strain rate on the slope of the load-deformation curve was prominent in case of compressive loading also.

SEM images (Fig-8) of the fractured surface both under compression and tension showed the deformation of osteons very clearly. In case of compressive fracture, the osteons deformed to such an extent that the aspect ratio of the osteons reached about 1:2. But in case of tensile fracture, osteons were seen to be protruding from the surface like tubular structures without much change in the aspect ratio. Again, fractography showed the fracture in the cementing zone appeared to be comparatively smooth or cleavage in nature in contrast with fibrous fracture in lamellar zones. As observed in case of porosities, for other mechanical properties also, there is a high chance that their values will differ when the whole bone is considered than when different parts are considered separately.

Hardness in the cortical part along the femur length did not vary significantly. But values of the indentations over the lamella were little higher (around 0.38 GPa) than those in the cementing zone (0.33GPa). The hardness of the cortical part of female bones was found to be lower as compared to males. Not much difference could be observed in longitudinal and transverse values. More or less it was observed that the microstructural features like lamella and interstitial zone had some effect on the hardness and also on crack propagation. Also the typicality in fracture and deformation of osteons under various types of loading had been identified quite clearly but the quantitative estimate of their (collagenous lamella or mineralised interstitial zone) individual contribution in controlling load-deformation as well as minute fracture events is yet to be understood in a comprehensive manner. It needs controlled set of experimentation, that are being done at present.
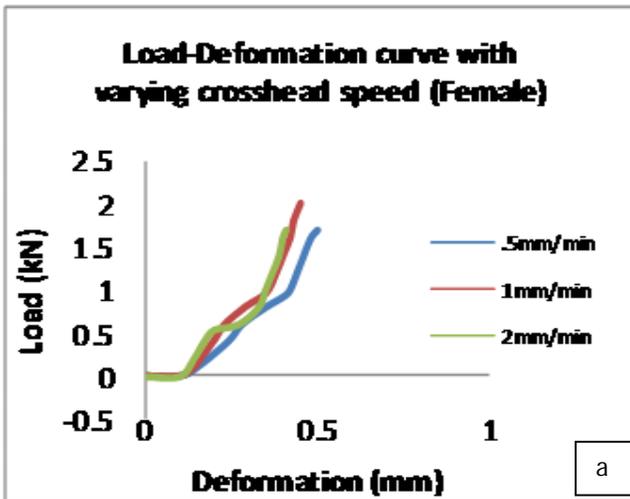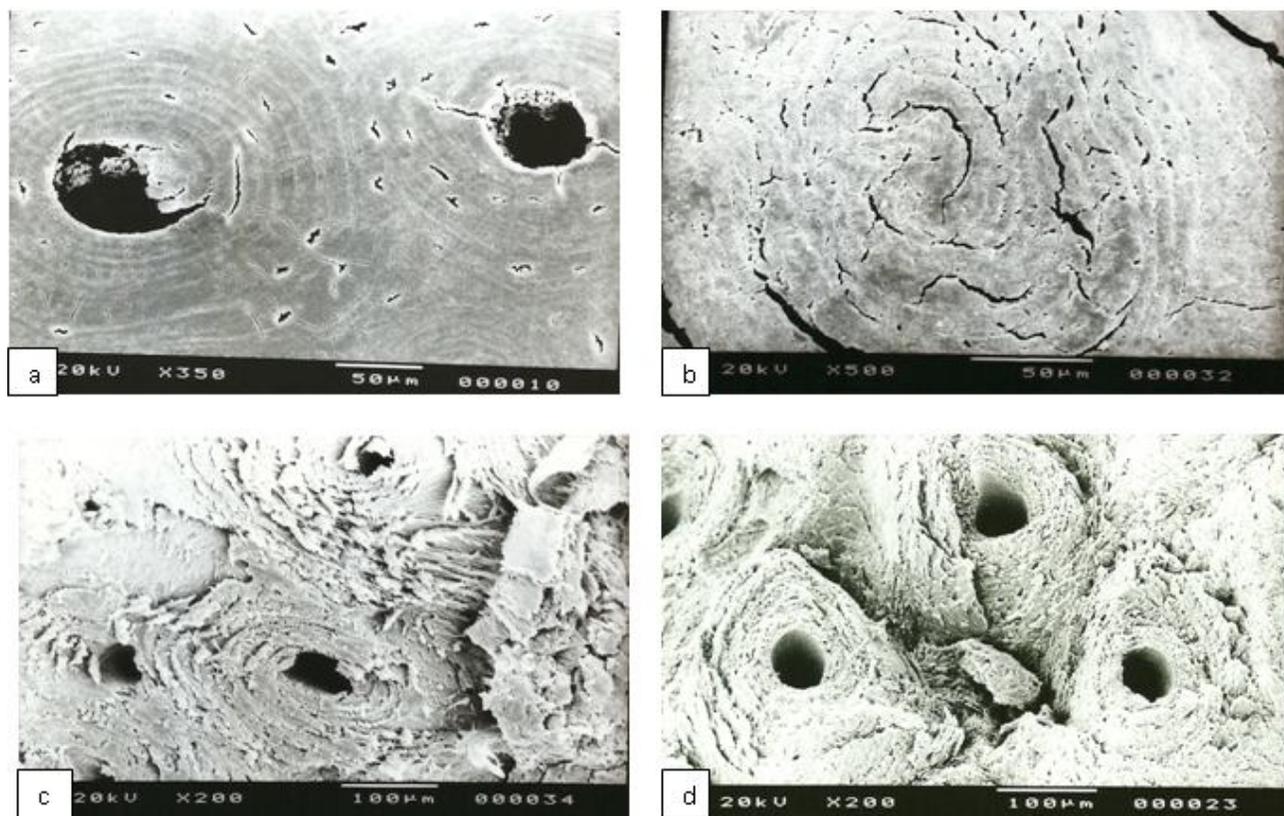
Fig. 6. SEM images of (a) undeformed osteons (b) Cracks developed along the lamellae well as across it.(c) Surface under compressive fracture (d) Surface under tensile fracture

4 CONCLUSION

From the study done so far, it has been observed specifically that the osteons suffer different types of damage and fracture in tensile and compressive loading. More deformation of microstructure was found in compression while in tension, there were more signs of shearing and tearing in-between the osteons. From the load-deformation curves it was evident that both male and female bones behaved similarly before yield point but post yield characteristics were different. It may be attributed to more energy absorption capacity of female bones. Although the lamellar structure is known to provide ductility to bone, cracks propagating across the lamella under high stress may be indicative of some inherent brittleness of bone. Finally it may be noted that mechanical behavior of only the mid-shaft cortical part is considered here for detailed study, the other portions are also to be studied before we make any generalization about the micro-structure-property relationship of bone.

ACKNOWLEDGMENT

REFERENCES

[1] J.Y. Rhoa, P. Ziouposb, J.D. Curreyc, G.M. Pharr. Microstructural elasticity and regional heterogeneity in human femoral bone of various ages examined by nano-indentation. Journal of Biomechanics 35 (2002) 189–198.

[2] Huijie Leng, X.Neil Dong, XiaoduWang. Progressive post-yield behavior of human cortical bone in compression for middle-aged and elderly groups. Journal of Biomechanics 42 (2009) 491–497

[3] Jeffry S. Nyman, Michael Reye, Xiaodu Wang. Effect of ultrastructural changes on the toughness of bone. Micron 36 (2005) 566–582

[4] R.K. Nalla a, J.J. Kruzic b, J.H. Kinney c, M. Balooch c, J.W. Ager III a, R.O. Ritchie,a. Role of microstructure in the aging-related deterioration of the toughness of human cortical bone. Materials Science and Engineering C 26 (2006) 1251 – 1260

[5] Igor Sevostianov, Mark Kachanov. Impact of the porous microstructure on the overall elastic properties of the osteonal cortical bone. Journal of Biomechanics 33 (2000) 881-888

[6] X. Neil Dong1, X. Edward Guo. The dependence of transversely isotropic elasticity of human femoral cortical bone on porosity. Journal of Biomechanics 37 (2004) 1281–1287

[7] M. Muller, D. Mitton, P. Moilanen, V. Bousson, M. Talmant, P. Laugier. Prediction of bone mechanical properties using QUS and pQCT: Study of the human distal radius.

[8] G. Dougherty. Quantitative CT in the measurement of bone quantity and bone quality for assessing osteoporosis. Medical Engg Physics Vol.18 no. 7 pg 557-568. 1996.

[9] Philippe K. Zysset, X. Edward Guo, C. Edward Hoffler, Kristin E. Moore, Steven A. Goldstein. Elastic modulus and hardness of cortical and trabecular bone lamellae measured by nanoindentation in the human femur. Journal of Biomechanics 32 (1999) 1005-1012

[10] S. Hengsberger, A. Kulik, and PH. Zysset. Nanoindentation Discriminates the Elastic Properties of Individual Human Bone Lamellae Under

Dry and Physiological Conditions. Bone Vol. 30, No. 1.January 2002:178–184

[11] Y. N. Yeni and T. L. Norman. Fracture Toughness of Human Femoral Neck: Effect of Microstructure, Composition, and Age. Bone Vol. 26, No. 5 May 2000:499–504

[12] R.K.Nalla, J.J.Kruzic, J.H.Kinney, R.O.Ritchie. Effect of aging on the toughness of human cortical bone: evaluation by R-curves. Bone 35 (2004) 1240 – 1246

[13] Kuangshin Tai1, Ming Dao1, Subra Suresh, Ahmet Palazoglu And Christine Ortiz. Nanoscale heterogeneity promotes energy dissipation in bone. Nature Materials VOL 6 JUNE 2007

[14] Peter Zioupos a, Ulrich Hansen b, John D. Currey. 2008, Microcracking damage and the fracture process in relation to strain rate in human cortical bone tensile failure. Journal of Biomechanics, 41, 2932–2939

Development of nano-grained Calcium Hydroxyapatite using slip casting technique

Howa Begam, Abhijit Chanda and Biswanath Kundu

Abstract— The purpose of this study is to synthesize nano-grained Calcium Hydroxyapatite (HAp) through slip casting technique. For this, hydroxyapatite powders were synthesized using two methods, wet chemical method and Ammoniacal method. The as-prepared powders and calcined powders were characterized using XRD, FTIR, to study the phases of the powders. The hydroxyapatite powder calcined at 1000°C for 2hr was used to prepare 50 vol% slurry using DN40 (sodium polyacrylate) as dispersing agent. After slip casting, the green bodies were sintered at different temperatures, 1100, 1200, 1250 and 1300°C with 2hr soaking time. The sintered dense samples were characterized for physical, mechanical and biological behavior.

Key Words - Slip casting, Hydroxyapatite, Ammoniacal method

———————————————— ◆ ————————————————

1 INTRODUCTION

CALCIUM phosphate ceramics, mainly Hydroxyapatite (HAp), do not exhibit any cytotoxic effects, shows excellent biocompatibility with hard tissues and also with skin and muscle tissues, and is the most appropriate ceramic material for hard-tissue replacement. This is a bio-active material with wide range of applications ranging from bone replacement to controlled drug delivery system. Albee (1) reported the first successful medical application of calcium phosphate bioceramics in humans, and in 1975 Nery et al. (1) reported the first dental application of these ceramics in animals. Since then, lot of work has been done on the in-vivo performance of Hap (2,3). But the applications of pure HAp are limited due to its poor mechanical reliability. Though HAp is generally used for replacement of different types of bones, it can not be used for load bearing situations. In recent past, much work has been done towards the development of advanced processing techniques for achieving more functionally reliable bioceramic bodies. Refinement of grains (4), advancement in sintering process through microwave (5), doping with various ions (6-8) are some of the approaches adopted so far for improving the mechanical strength of this group of bio-ceramic materials. Due to high surface energy, nano-sized HAp powder results in better sinterability and therefore improves mechanical properties. For this reason, researchers have developed many HAp synthesis techniques, including sol-gel methods (8,9), co-precipitation (10), emulsion techniques(11), mechano-chemical methods (12), electrochemical deposition (13), and hydrothermal processes (14). Colloidal processing can also be a promising technique to achieve this objective.

In colloidal processing, particle dispersion is the most important factor, which has a strong influence on both the rheology and homogeneity of the suspensions. Various colloidal processes, such as slip casting (15), tape casting (16) and gel casting (17,18) etc. have been reported for preparation of HAp with different size and complicated shapes. It has been shown that anionic polyelectrolyte is very effective to obtain homogeneous distribution of HAp particles. M. Toriyama et al (18) investigated the dispersion behavior of mechanochemically synthesized hydroxyapatite. They showed that surface charge reversed as anionic polyelectrolytes were added in considerably high amount (3 wt%). It was further shown that polyacrylates were effective in stabilizing 10 vo1% suspensions while pourable casting slips were obtained at the maximum solid concentration of 21.4 vol %. Rao et al (15) studied the dispersibility of HAp in deionized water using sodium hexametaphosphate and ammonium polymethacrylate as dispersing agent.

In the present work, Hap powder was prepared by two methods- wet chemical and ammoniacal route method. These powders were characterized by different tools and slips were prepared with different concentration of dispersant to optimize the slip viscosity ideally suited for slip casting. Sinterability of these slip cast samples and their physical, mechanical and biological characterization have been performed.

2 MATERIALS AND METHOD

Hydroxyapatite powders were synthesized using two methods- wet chemical route and Ammoniacal route. In wet chemical method, G.R. calcium hydroxide ($Ca(OH)_2$) and orthophosphoric acid (H3PO4)(EMerck,India) were used. H_3PO_4 solution was added drop by drop in the $Ca(OH)_2$ solution in stirring condition. Synthesis was done at 80°C at a pH of 11-12. In ammoniacal method, A.R. grade Calcium Acetate (($CH_3COO)_2Ca.XH_2O$) and

L.R. grade (Ammonium phosphate Di basic) $((NH_4)_2HPO4$ NICE Chemicals Pvt. Ltd, India) were used. Aqueous solution of ammonia and Di Ammonium hydrogen orthophosphate were added drop by drop to the aqueous solution of Calcium Acetate with continuous stirring at 80°C. After 48 hours, the precipitate was filtered and washed continuously with distilled water to remove ammonia. The as-prepared powders were then calcined at 1000°C for 2hr. Samples prepared with wet chemical method were designated with the name URH while those prepared via Ammoniacal route were termed as ARH.

The calcined powders were planetary milled in ethanol medium with 50 rpm for 3hr to gain suitable particle size distribution and specific surface area. The particle size distribution was measured using laser diffraction technique in Mastersizer instrument (Malvern Instruments Ltd., UK) and the specific surface area was measured by BET method (Sorpty 1750, Italy). The phase composition of as prepared powder and calcined powder was analyzed by X-ray diffraction (XRD, Cu Kα1, Philips Analytical, PW1710, Netherlands). FTIR analysis was made in mid IR region (4000-400 cm-1) using KBr pellets in a Perkin-Elmer, Model 1615 (USA) instrument. The differential thermal analysis and thermo gravimetric analysis was done for as prepared powders using DTA/TGA instrument (NETZSCH, Germany),

An anionic polyelectrolyte with trade name Dispex N 40 (DN40) was used as deflocculating agent to prepare HAp slip. To optimize the amount of DN40 to be added, zeta potential of HAp- loaded slip was measured with different concentration of DN40 (0 - 5wt %) and with varying pH. It was performed in Zetasizer instrument (Malvern Instruments Ltd., UK). The slurry was planetary milled for 16 hours to obtain homogeneous solution. The rheological behavior of the aqueous suspensions of HAp powder (40-50 vol %) was studied with varying shear rate from 0 – 450/s using a cup-and cone arrangement (HAAKE, VT500 with sensor-5 mode) at room temperature with about 50% relative humidity.

HAp slurry with a maximum volumetric load of 50% was finally used for slip casting. It was done in a Plaster of Paris mould with 40-50 % porosity to prepare dense Hap. Sinterability of the dense HAp was studied in an electrical resistance furnace (Naskar Furnace, Kolkata, India) at different temperatures 1100, 1200, 1250 and 1300°C with 2 hrs dwelling time for each. The heating rate was kept constant for all the cases at 3°C/min. After cooling, the sintered samples were cut in different shapes for physical, mechanical and biological characterization.

The bulk density and porosity of the sintered samples were measured using water displacement method which is based on Archimedean principle. After taking the weight of the dried samples (D), they were soaked in water to get soaked weight (W). Weight of the samples in water-suspended condition (S) was also measured. Density, open and closed porosity were calculated using standard expressions as mentioned below.

Bulk density (B.D.) = D/(W-S) gm/cc, Apparent Porosity (A.P.)= (W-D)/(W-S) x100%
Total Porosity (T.P.) = (1-B.D./True Density) x100%, Closed Porosity (C.P.) = T.P.-A.P. True density of HAp was 3.167g/c.c. The phases of sintered HAp samples were identified by X-ray diffractometry (XRD, Cu Kα1, Philips Analytical, PW1710, Netherlands). Scans were recorded from diffraction angle (2θ) of 10° -80°, at a speed of 40/min with step size of 0.050. The surface morphology as well as the microstructure of the dense samples were observed to assess the influence of the sintering temperature on the grain size of hydrxyapatite, by using a Scanning Electron Microscope (S3400N, Hitachi).

Young's modulus and hardness was calculated from the nanoindentation technique (CSM, Switzerland), at a load of 100mN for both ARH and URH samples using standard Oliver and Pharr technique. Bi-axial flexural strength and compressive strength was also measured using a Universal testing machine (Instron 5500R, UK).

Cell adhesion behavior was studied for the HAp pellets at Shree Chitra Tirunal Institute for Medical Science and Technology, Kerala, India. L929 mouse fibroblast cells were seeded on the surface of conditioned HAp samples with minimal essential medium supplemented with 10% fetal bovine serum, 100 units/ml of penicillin and 100 µg/ml streptomycin, respectively and incubated at 37°C in humid atmosphere and 5 % CO_2 for 48 hours.

3   RESULTS AND DISCUSSION

Various results of physical, mechanical and biological studies are presented and discussed below. The XRD was performed on as prepared powder as well as the calcined powder. It has been found that as prepared URH powder consists of phase pure hydroxyapatite with the major peak at 31.8 degree (Fig 1a). When the same powder was calcined at 1000°C minor quantities of secondary phase viz. calcium oxide appeared (Fig. 1b).
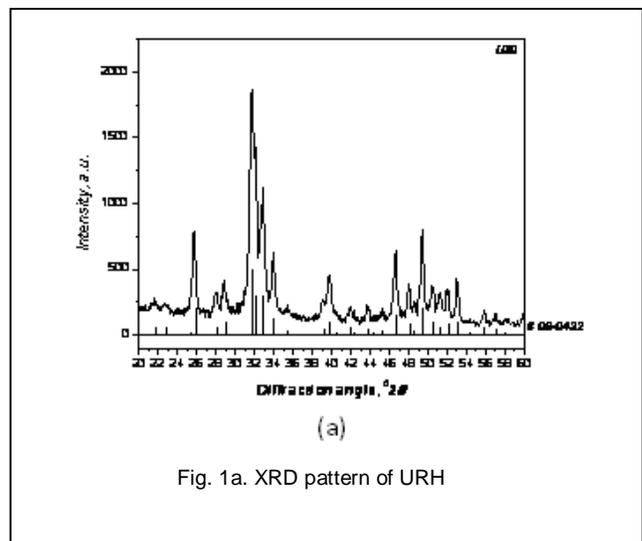
Fig. 1a. XRD pattern of URH

This may due to the degradation of Hydroxyapatite at that temperature and at that atmospheric pressure. After slip casting using this powder, the dense blocks were fired at 1100°C, 1200°C,1250°C, 1300°C and it was found that calcium oxide was retained as minor quantities up to the temperature as high as 1300°C.

Fig: 2a XRD pattern of ARH

Fig: 2a XRD pattern of ARH
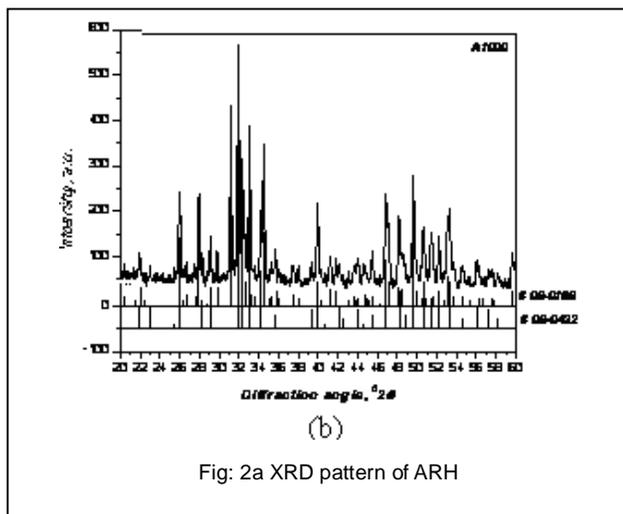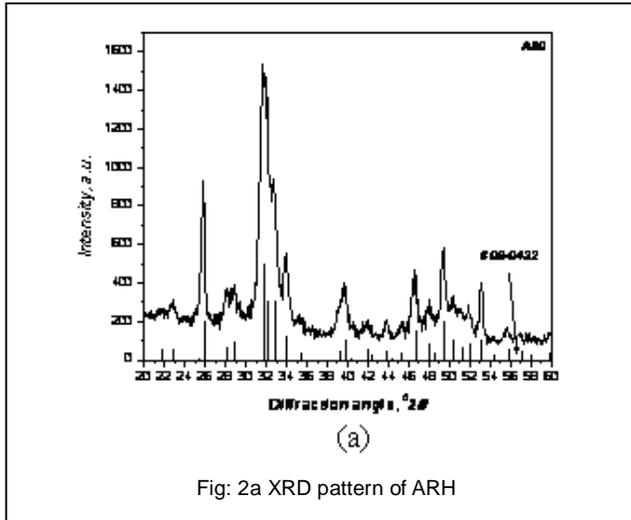
Similarly the as prepared ARH powder was composed of phase pure Hydroxyapatite (Fig 2a). When the same powder was fired for calcination at 1000°C some part of Hydroxyapatite decomposed to form another crystalline phase which is β-TCP (Fig 2b). It was found that percentage of β-TCP went on increasing from the temperature 1100-1300°C.

Average crystallite size is calculated from the XRD of as prepared powder as well as the sintered slip casted dense blocks sintered at different temperatures using by using Full Width at Half Maxima (FWHM) method. It was based on Scherrer's equation i.e., $D = 0.9 \lambda / \beta \cos\theta$ where, D=size of the crystal, λ=wavelength of X-ray used,

β=broadening of diffraction line at half of its maximum intensity in radians, θ =diffraction angle.

The average crystallite size of URH increases with increase in temperature upto 1200 °C. After this temperature the crystallite size decreases. For ARH the crystallite size went on increasing up to temperature of 1100°C for c Axis of Hydroxyapatite. It was not the case for the plane (300). The dimension of a axis went on increasing up to a temperature as high as 1250°C.

Percentage of crystallinity of powder and dense block were calculated by using the following relationship :

$X_C = (1 - V_{112/300} / I_{300})$

where, $V_{112/300}$ is intensity of hollow between the planes (112) and (300) reflections and I_{300} is the intensity of (300) reflection. It was found that for URH powder and sintered blocks percentage of crystallinity went on increasing up to temperature of 1200°c. After that it was decreasing up to 1300°c possibly due to presence of minor quantity of CaO.

Fig. 3a. FTIR spectra of As prepared and sintered URH powder
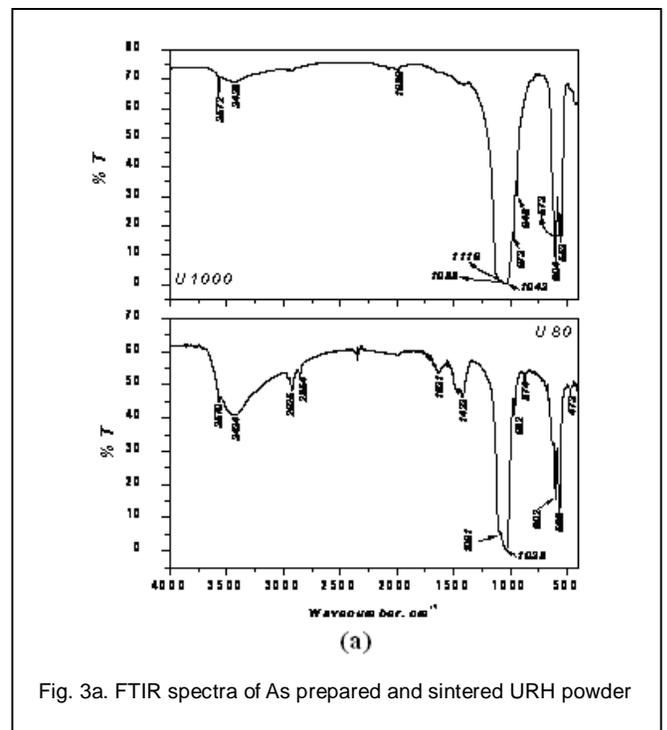
The crystallinity calculated from ARH1000 powder was found to be highest which drastically decreased at 1100°C and subsequently went on increasing upto 1250°C.

In FTIR plot for as prepared ARH, bulge shaped peak at 3436 was found which reveals the presence of hydroxyl group. In the A1000 (ARH sintered at 1000°C) plot this hydroxyl group peak is very short and tending to disappear. This might be due to the evaporation of the surface moisture of the powder for calcination. In A1000 the presence of peaks at 973, 963, and 947 reveals the presence of TCP. So from the FTIR plot too, evidence regarding the conversion of hydroxyapatite into TCP was obtained. The hydroxyl group was also present in as prepared URH like

ARH. In U1000 (URH sintered at 1000C) the hydroxyl group is trying to disappear this may due to the evaporation of the surface moisture of the powder after calcinations.
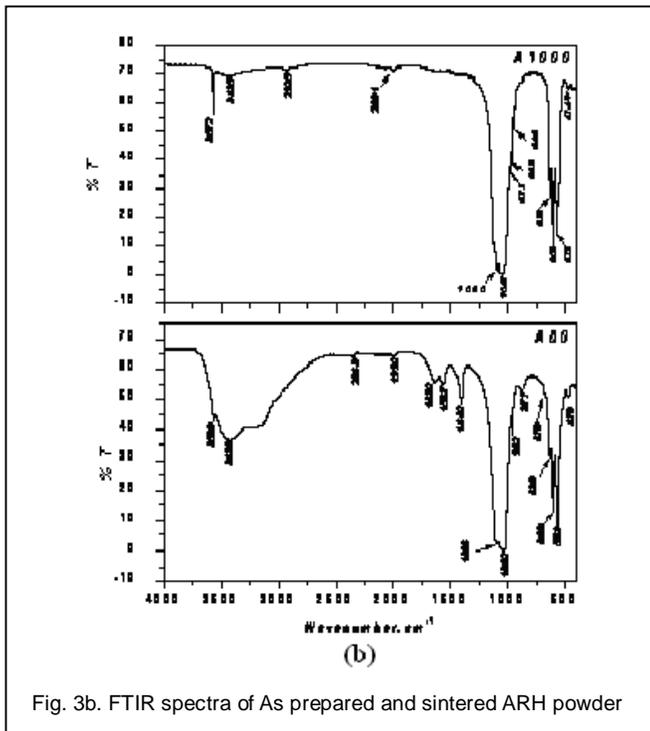
Fig. 3b. FTIR spectra of As prepared and sintered ARH powder

The surface morphology as well as the microstructure of the dense samples was observed to assess the influence the sintering temperature on the grain size of hydroxyapatite by using scanning electron microscope. It was observed that for both the type of Hap, samples sintered at 1100°C had homogeneous microstructure and grains were more or less uniform and fine. In high sintering temperature the microstructure changed and grain size increased for both the type of hydroxyapatite samples. Average grain sizes were 213 ± 64 nm, 438.92 ± 110.612 nm, 797.51 ± 156.3546 nm and 1.56 ± 0.288 micron respectively for URH samples sintered at 1100, 1200, 1250 and 1300°C. For ARH samples average grain sizes were 213 ± 53 nm, 851 ± 118 nm, 1073 ± 216 nm and 2136 ± 441nm for samples sintered at 1100, 1200, 1250 and 1300C respectively.

The specific surface areas in m2/gm for all samples were listed in the table.

TABLE:1

THE SPECIFIC SURFACE AREAS FOR ALL SAMPLES

Samples	Specific surface area, (m2/gm)
ARH(as prepared)	77.284
URH (as prepared)	40.843
ARH 1000	2.583
URH 1000	20.439

It reveals that calcined URH powders had higher specific surface area than ARH. Specific surface area played an important role for slurry stabilization and it was not possible to achieve high solid loading for higher specific surface area [19]. The specific surface area obtained was lower which made the slurry high solid content.

The median size of the URH before milling was 8.64 micron. Before milling the particle formed agglomerate. After milling the powder median size decreased and was subsequently was used for slip casting. The particle size distribution curve was a bimodal curve. We got two values for median size 1.61 and 4.92 micron. It was known that particle size around 2 microns is ideal for slip casting. The curve showed the presence of even smaller particles in the present case. In case of ARH before milling, the particle size distribution curve was a bimodal curve but after milling the curve had a fine Gaussian peak. The median size was 2.90 micron, not much different from the ideal one and hence suitable for slip casting.

Variation of surface charge with varying amount of percentage of dispersant for aqueous suspension of both URH1000 and ARH1000 particle were measured which is given in fig(4). It was found that with addition of dispersant, URH particles became highly deflocculated and the trend was continued up to 5wt%. For a particular aqueous suspension of powder particles it is known that the value of zeta potential anything between35 to 50 mV should have higher deflocculating properties. The same trend was observed for URH particles. From the plots it could be inferred that 4-5 wt% DN40 had the ability for idealized dispersion of URH1000 particles. In case of ARH, zeta potential values had a different behavior with addition of dispersant. In this case, the zeta potential values were found to be very high with the addition of DN40 from 1-4 wt %. So any quantities between these ranges may show high dispersion in the aqueous system of ARH1000 particles. After addition of 4% DN40 the particles started to agglomerate and continued up to 5wt%. So finally 4% of DN 40 was selected for effective defloculation.
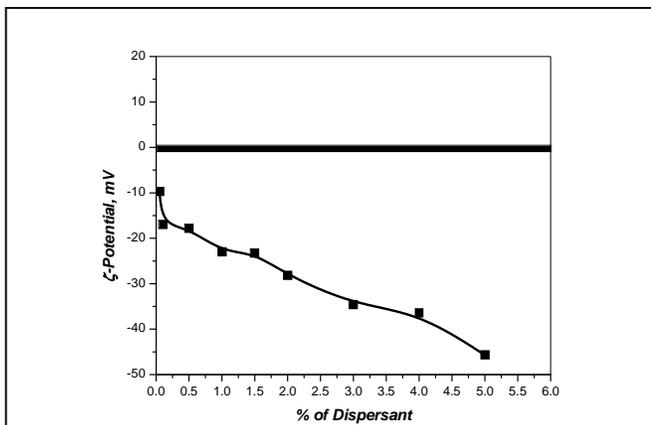
Fig. 4a. Surface charge with varying amount of DN40 of URH
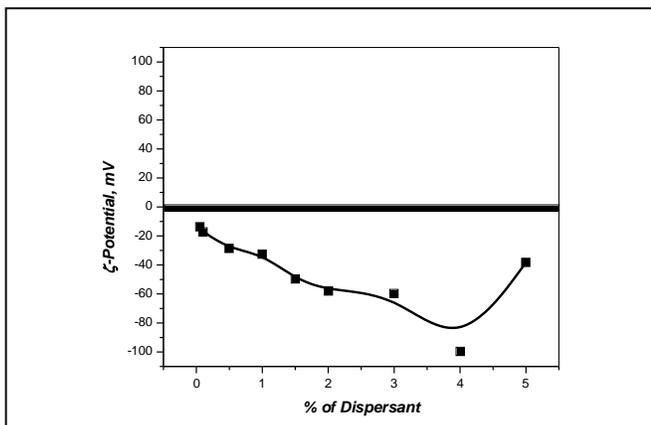
Fig. 4b. Surface charge with varying amount of DN40 of ARH

Changes of viscosity (expressed as mPa.-s) and shear stress for both URH and ARH particles with varying amounts of solid loading. It has been found that the slurries behaved like a Bingham plastic for the total range of shear rates. The behavior was more prominent for 50 vol% slurry where it was found to be almost a straight line having a y-axis intercept.

The bulk density was measured by water displacement method that employs Archimedes's principle. The figure (5) shows the variation of bulk density with sintering temperature. It was observed that bulk density increases with increase in sintering temperature for both URH and ARH samples and the bulk density was maximum at 1250°C as shown in the figure. For ARH samples sintered at 1250°C maximum density was obtained which was 97.56% of actual density. For URH samples we observed the same thing like ARH i.e. density increase with increase in sintering temperature. We got the maximum density 93.3% of actual density for URH1250. It was observed that URH and ARH samples sintered at 1100°C have more apparent porosity but the closed porosity is less compared to other samples. The apparent porosity

and closed porosity for ARH 1200, ARH1250 and ARH1300 were more or less same.
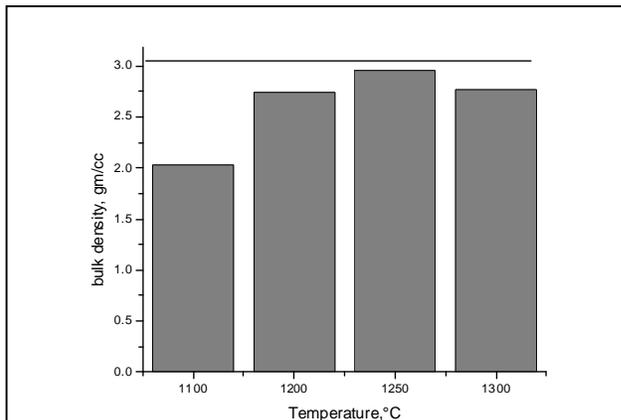
Fig. 5a.Sintering temperature versus bulk density of URH
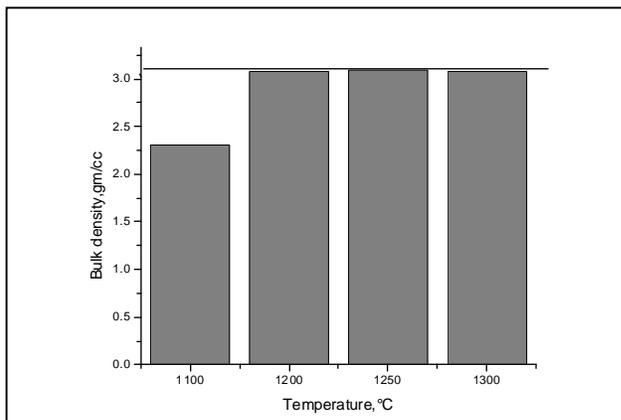
Fig. 5b.Sintering temperature versus bulk density of ARH

Vickers' hardness and nano indentation hardness was measured for both URH and ARH blocks sintered at 1100°C and 1250°C. It was found that URH1250 had more Vickers' hardness value (4.23GPa) than URH1100. Same results was observed for ARH 1250 i.e. Vickers' hardness value for ARH1250 was 4.14GPa which is better than ARH1100. The hardness obtained from nano-indentation showed more or less a similar trend like Vickers' hardness. The flexural strength was measured for both URH and ARH blocks sintered at 1100°C and 1250°C. The flexural strength of URH1250 and URH1100 were 94.01 MPa and 18.99 MPa respectively. For ARH1250 and ARH1100 the average flexural strength was obtained as 67.32 MPa and 38.04 MPa respectively. For flexural strength as well as hardness, the individual scatter was well within 5%. For both URH and ARH, at 1100C, the materials were found to be too porous. Drop in hardness or flexural strength for both URH and ARH at 1100°C might be attributed to this increased porosity. It may be

noted that the samples sintered at 1100°C with a soaking time of 2 hrs consisted of nano grain size but due to porosity their mechanical properties degraded. In case of compressive strength, however this was not followed in an identical manner in case of ARH. The exact reason behind it was not clear but in this case (1250°C) three samples were tested and the results showed large (more than 100%) scatter. The values varied widely from 13.82 MPa to 152.19 MPa while in 1100°C, with same number of specimens, the standard deviation was substantially lower, 22%. This might have caused some discrepancy in the general trend of lower mechanical property at 1100°C.

Fig. 6a.SEM micrograph of cellular adhesion URH blocks sintered at 1200°C/2hrs
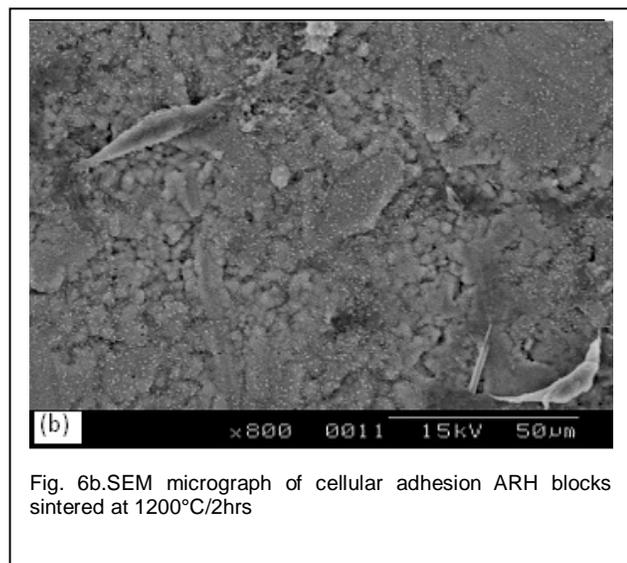
Fig. 6b.SEM micrograph of cellular adhesion ARH blocks sintered at 1200°C/2hrs

Cell-material interaction plays a very vital role in case of bioactive ceramics. In the present case, L929 mouse fibroblast cells were used to study the cell adhesion on the hydroxyapatite blocks sintered at 1250°C for 48 hrs. The time of the study was short due to some machine constraints. It was observed that preliminary fibroblast

cell adhesion on URH was better than ARH samples. Although in URH, the number of cells attached to the surface was not high, the primary signs of extending filopods clinging to micro-structural anchorage sites were observed. Such behavior was not found in ARH even after thorough scanning throughout the surfaces. It might be due to the presence of secondary phase (β-TCP) in ARH samples that did not promote cell attachment, cell growth and spreading capability like the other case. β-TCP are reported to affect the microenvironment of the cell culture [3] by leaching ions and changing the pH of the medium resulting in an unstable interface which prevent proper anchorage; attachment and adhesion in an acidic medium.

From the results discussed so far, it is evident that nano-sized hydroxyapatite was synthesized in a very simple and economical way of slip casting method without sacrificing the quality of the product. It has been observed that if suitable deflocculating process is adopted following thorough study of rheological behavior of the slip, casting of dense pellets are possible with consistent microstructure and compositional homogeneity. It is an accepted fact that grain refinement increases strength according to Hall and Petch [20,21] but porosity in the nano grained sample, can classically cause deterioration of mechanical strength. The present material was nano-grained with porosity as high as 30%. This was found for both URH and ARH HAp at 1100°C. The values were found not only from density measurement but also from image analysis. Furthermore the pores were found to be grossly interconnected which is ideal for in-vivo bony in growth and osteointegration.

Normally cortical bones have compressive strength in the order of 100-130MPa and hardness around 1-2 GPa. To verify a set of tests were conducted with human cadaveric bones (both male and female cortical bones) with same test set-ups as used here in case of HAp. The compressive strength values in axial direction were in the order of 150 MPa while hardness was less than 1 GPa. Here at 1100C, the maximum value of compressive strength we got for ARH was slightly less than 100 MPa, however hardness was comparatively high, around 3.11 GPa. So for load bearing implant this type of porous material might not be generally recommended right now from mechanical point of view. But this porosity would be advantageous for non-load bearing implants where cells could attach with high number of anchorage sites. This enhanced attachment could trigger other associated activities differentiation and proliferation.

4 CONCLUSION

Nano sized hydroxyapatite could be successfully synthesized using two co-precipitation methods. Submicron and/or nano ranges of grain size of Hydroxyapatite could be consistently formed by slip casting method using high deflocculated suspension of

URH & ARH. It was confirmed by the SEM, BET and particle size distribution study. Highly deflocculated suspension was prepared successfully for both the powders. Solid loading as high as 50 vol% could be achieved using an acrylic based anionic dispersant. This was correlated with the findings of assessment of zeta potential with varying amount of dispersant addition. The slip casted blocks were sintered at fairly low temperatures with > 95% density. URH 1250 showed highest values of hardness and Young's modulus as calculated from both Vickers indentation and nano indentation. Fibroblast cells adhesion on URH was better than ARH. Slip casting is very effective and efficient method to get fine grained hydroxyapatite.

REFERENCES

[1] R.Z. LeGeros, 40th Symposium on Basic Science of Ceramics, Convention Center, Osaka University, January 22–23,

[2] A.Banerjee, A. Bandyopadhyaya, S.Bose, W. M. Keck, Hydroxyapatite nanopowders: Synthesis, densification and cell–materials interaction, Materials Science and Engineering C 27 (2007) 729–735

[3] Annie John, H. K. Varma & T. V. Kumari, Surface Reactivity of Calcium Phosphate Based Ceramics in a Cell Culture System, Journal of biomaterial applications, V-18 (2003)

[4] Samar J. Kalita , Abhilasha Bhardwaj, Himesh A. Bhatt, Nanocrystalline calcium phosphate ceramics in biomedical engineering, Materials Science and Engineering C 27 (2007) 441–449

[5] I. Teoreanu , M. Preda , A. Melinescu, Synthesis and characterization of hydroxyapatite by microwave heating using Ca-SO4_2H2O and Ca(OH)2 as calcium source, J Mater Sci: Mater Med (2008) 19:517–523

[6] S. J. Kalita, D. Rokusek, S. Bose, H. L. Hosick, A. Bandyopadhyay, Effects of MgO-CaO-P2O5-Na2O-based additives on mechanical and biological properties of hydroxyapatite, 2004 Wiley Periodicals, Inc.

[7] S. J. Kalita, D. Rokusek, S. Bose,1 H. L. Hosick, A. Bandyopadhyay , CaO–P2O5–Na2O-based sintering additives for hydroxyapatite (HAp) ceramics, Biomaterials 25 (2004) 2331–2339

[8] A. Bandyopadhyay , Elizabeth A. Withey, J. Moore, S Bose, Influence of ZnO doping in calcium phosphate ceramics, Materials Science and Engineering C 27 (2007) 14–17

[9] A.H. Rajabi-Zamani, A. Behnamghader, A. Kazemzadeh Synthesis of nanocrystalline carbonated hydroxyapatite powder via nonalkoxide sol–gel method, , Materials Science and Engineering C (2008)

[10] Y. X. Pang and X. Bao, Influence of temperature, ripening time and calcination on the morphology and crystallinity of hydroxyapatite nanoparticles, J. Eur. Ceram. Soc.(2003) 1697.

[11] G. K. Lim, J. Wang, S. C. Ng, C. H. Chew and L. M. Gan, processing of hydroxyapatite via microemulsion and emulsion routes, Biomaterials18 (1997) 1433

[12] W. Kim, Q.W. Zhang, F. Saito, Mechanochemical synthesis of hydroxyapatite from Ca(OH)2–P2O5 and CaO–Ca(OH)2–P2O5 mixtures, J. Mater. Sci. 35 (2000) 5401–5405.

[13] L.Y. Huang, K.W. Xu, J. Lu, A study of the process and kinetics of electrochemical deposition and the hydrothermal synthesis of hydroxyapatite coatings, J. Mater. Sci. Mater. Med. 11 (2000) 667.

[14] Jae-Kil Han, Ho-Yeon Song, Fumio Saito, Byong-Taek Lee, Synthesis of high purity nano-sized hydroxyapatite powder by microwave-hydrothermal method, Materials Chemistry and Physics 99 (2006) 235–239

[15] R. Ramachandra Rao and Thandali S. Kannan , Dispersion and Slip Casting of Hydroxyapatite, J.Am.Ceram.Soc.84[8] 1710-16(2001)

[16] I.H. Arita, D.S. Wilkinson, V.M. Castazo, Synthesis and Processing of Hydroxyapatite Ceramic Tapes Casting with Controlled Porosity J. Mater. Sci. Mater. Med. 6 (1995) 19.

[17] S. Padilla, R. Garc´ıa-Carrodeguas, M. Vallet-Reg´ı, Hydroxyapatite suspensions as a precursors of pieces obtained by gelcasting method, J. Eur. Ceram. Soc. 24 (2004) 2223.

[18] M. Toriyama, Slip casting of mechanochemically synthesized hydroxyapatite, J. Mat. Sc. 30 (1995) 3216-3221

[19] F. Lelievre, D. Bernache-assollant, T. chartier, Influence of powder characteristics on the rheological behaviour of hydroxyapatite slurries, J. Mat. Sc.: materials in Medicine7 (1996) 489-494

[20] E.O.Hall, Proc. Phys. Soc., Ser. B, Vol. 64, pp. 747-753, (1951).

[21] N.J.Petch, J. Iron and Steel Institute, pp. 25-28, May 1953.

Application of Control Theory in the Efficient and Sustainable Forest Management

Md. Haider Ali Biswas, Munnujahan Ara, Md. Nazmul Haque, Md. Ashikur Rahman

Abstract— This paper focuses on the necessity of forest management using the model of control theory. Recent researches in mathematical biology as well as in life sciences closely depend on control theory. Various popular research papers have been received considerable attentions by engineers and research scholars due to the fact that it has been the central and challenging area of research for its wide range applications in the diverse fields. In this study we have briefly mentioned some of the fields in which these challenges are present, specially sustainable forest management is one of the warming issues in the present cencury. Our main objective in this paper is to investigate the scopes and applications of control theory in real life situation, specially the applications of control theory in the efficient and sustainable forest growth. A particular case of Sundarbans, the largest mangrove forest in the world is discussed with illustrative examples.

Index Terms— Control theory, Malthusian model, Logistic equation, Sundarban mangrove.

———————————————— ◆ ————————————————

1 INTRODUCTION

NOW-a-days the applications of Optimal Control (OC) theory have been increased surprisingly in almost all branches of modern science and engineering, specially the OC is playing significant, in some cases the dominant roles in the fields of aerospace engineering, medicine, agriculture and economics. See for examples [9], [14]. The theory of optimization continues to be an area of active research not only for the mathematicians but also for the engineers as an indication both of the inherent beauty of the subject and of its relevance to modern developments in science, industry and commerce. The optimal fuel landings of the space vehicle, the optimal strategies for the drug doses in the chemotherapy of infectious diseases along with all others are the fascinating challenges at present in the ongoing development of science and technology [28]. All the achievements in the field of optimal control theory are mostly due to Pontryagin. Although several authors made tremendous contributions to the further developments in this area, but *Pontryagin Maximum Principle* is still a milestone in OC theory. See for examples [6], [10], [16], [30] and references therein for details study on optimal control theory. However, we are mainly concerned in this paper to discuss the application of OC theory in the sustainable forest management. Many researches have been carried out over the years (see for examples [12], [17] and [37] for the recent development) emphasizing on the optimal strategies for an efficient and sustainable forest management due to the fact that forests are one of the best sources for saving lives of the human being from the poisonous greenhouse gases, from many natural disasters like cyclone and so on. European Tropical Forest Research Network (ETFRN) recently pointed out on the similar issues of forest management [20]. The people of the coastal region live depending on forests specially for a hunting, wood harvesting as well as collecting of honey (see for examples [23], [32]). Accordingly, in the economic point of view, forests help to enrich the national economy of a country. So steps should be taken so as to manage a sustainable and effective ecosystem where control theory is the essential tool. Different mathematical models were proposed for this purpose. Many studies have been undertaken to determine the optimal forest rotation length under different scenarios since the advent of the modern civilization. Some of these were focused on the optimal rotation length with the consideration of only timber value. Others searched for the optimal rotation length with the inclusion of both timber and nontimber benefits (See for examples, [27], [34], [36] and references within). These studies have provided important guidelines on how to manage the number of existing trees in the even-aged plantations. However, their applications in uneven-aged, or natural forests, are limited because age is no longer an appropriate variable under such circumstances. Also, in formulating a forest management plan/policy, particularly in the management of a large-scale (such as a regional or national scale) forest resource, it may be more relevant to determine how much timber should be harvested and what level of the forest stock should be maintained than to know when trees should be cut. However, in all the above mentioned models, the authors proposed the optimal managements of the forests only when the forests are full of trees. All those may bring a good result when the forests will be full of trees. In that case, the Malthusian model can give an optimal growth of trees after a certain time. In this study, we have proposed an alternative model discussing the Malthusian model and Logistic equation with illustrative example for the optimal controlling of the forest growth. A case study for survival of the existence of the largest mangrove forest in the world will be discussed for the sustainable management.

2 THE MALTHUSIAN MODEL

The Malthusian law of growth enunciated by Thomas

Robert Malthusian in his research documentary "*An essay on the principle of population*" and concluded that "Population when unchecked, increase, with a geometric ratio". See for details in [24]. Over the years, the Malthusian growth model is still the granddaddy of all population models, which is simply based on the famous exponential growth law. We begin this section with a simple derivation of the model.

Let t denote the time and $p(t)$ denote the number of individuals present at time t. In practice $p(t)$ is non negative integer. We assume that $p(t)$ is continuously differentiable.

The growth rate of a population is the rate at which population changes. If the population $p(t)$ at time t changes to $p(t+\delta t)$, the average per capita growth rate at time δt is

$$\frac{p(t+\delta t) - p(t)}{p(t)\delta(t)}$$

Taking limit $\delta t \to 0$, we get the instantaneous growth rate per capita at time δt as

$$\lim_{\delta t \to 0} \frac{p(t+\delta t) - p(t)}{p(t)\delta(t)} = \frac{p'(t)}{p(t)}$$

Now let,

b = intrinsic birth rate.

= The average number of offspring born per individuals per unit time.

d = intrinsic death rate.

= The fraction of individuals of the population dies per unit time.

$r = b - d$

= intrinsic growth rate of the population.

= Excess of birth over death per unit time per individuals.

Now, we consider a single species of population, the growth model is then described by

$$\frac{p'(t)}{p(t)} = r$$

$$\Rightarrow p'(t) = rp(t) \qquad (1)$$

with the initial population

$$p(t_0) = p_0 > 0$$

We have the mathematical model described the growth of single spaces population as

$$p'(t) = rp(t), \, p(t_0) = p_0 > 0 \qquad (2)$$

The solution of this equation is $p(t) = ce^{rt}$

For the initial condition $p(t_0) = ce^{rt_0} \quad \Rightarrow c = \frac{p_0}{e^{rt_0}}$

$$\therefore p(t) = p_0 e^{-rt_0} e^{rt}$$

$$\Rightarrow p(t) = p_0 e^{r(t-t_0)} \qquad (3)$$

which is known as exponential law of growth or Malthusian law of growth.

The behavior of the solutions depend upon the growth rate. Now we will discuss the solution for following several cases.
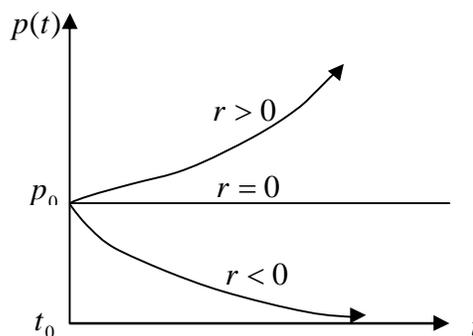
Fig. 1. Graph of the Malthusian model under 3 cases.

Case 01: For $r < 0$, we have

$$p(t) = \lim_{t \to \infty} p(t) = \lim_{t \to \infty} p_0 e^{r(t-t_0)} = 0$$

This implies extinction of population. That is if the growth rate is negative, in the long run the population will be extinct.

Case 02: For $r = 0$, we have

$$p(t) = \lim_{t \to \infty} p(t) = \lim_{t \to \infty} p_0 e^{r(t-t_0)}$$

$$= \lim_{t \to \infty} p_0 e^0 = p_0$$

This refers constant population at zero growth rates.

Case 03: For $r > 0$, we have

$$p(t) = \lim_{t \to \infty} p(t) = \lim_{t \to \infty} p_0 e^{r(t-t_0)} = \infty$$

This is the case of unlimited growth.

So the Fig. 1 shows that for a positive growth rate, the trees of a forest after a certain time can be managed in its full naturally balanced equilibrum state.

3 THE LOGISTIC EQUATION

A typical application of the logistic equation (see [15] for details) is a common model of population growth, originally due to Pierre-François Verhulst in 1838, where the rate of reproduction is proportional to:

- The existing population
- The amount of available resources

Let $N(t)$ be the population of trees. The Malthusian model assumes a rate of growth proportional to the population $N'(t) = aN(t)$. This gives the exponential growth law $N(t) = e^{at} N(0)$, which is only accurate for relatively small values of $N(t)$; overcrowding and competition for resources lower the rate of growth. A more realistic model assumes a steadily decreasing, eventually negative growth coefficient $a(N)$.

In Malthusian model we assume that the rate in which an organism will reproduce or die remain constant.

In case of many other populations, such as population which exhibits exponential growth for a limited period ultimately approaches to some steady state. The growth of any population in a restricted environment must eventually be limited because that individual member complete with each other for the limited living space, natural resources and food. This forces the growth rate to decline,

the model, $N'(t) = aN(t)$ may be modified to

$$N'(t) = Nr(N)$$ (4)

where $r(N)$ is some decreasing function of N satisfying $r'(N) < 0$ and the simplest function regarding $r(N)$, having this property, is $r(N) = a - bN$, $a, b > 0$. Thus the model with initial population $N(t_0) = N_0 > 0$

$$N'(t) = N(a - bN), \qquad N(t_0) = N_0$$
$$= (a - bN(t))N(t)$$

This is the well-known *logistic equation*. This model gives good results for bacteria populations, even for human population [15] but does not describe accurately phenomena such as forest growth because of the fact that, the inhibiting effects of new trees on the growth rate are negligible until these have reached a certain "adult" size. Thus, the rate should be a function not only of $N(t)$ but also of $N(t - h)$ for a suitable time delay $h > 0$, leading to the equation called the *delayed logistic equation*,

$$N'(t) = (a - bN(t - h))N(t)$$ (5)

Similar delay effects are observed in the influence of overcrowding in human populations, for an elementary exposition of logistic equations with and without delays equation (4) and (5) has two equilibrium solutions. One is $N(t) \equiv 0$, whereas the other is

$$N(t) \equiv N_e = \frac{a}{b}$$ (6)

Assume tree seeds are planted, and trees are logged with seeding and logging rates $u_0(t)$ and $u_1(t)$ respectively. Let k be time it takes a seed to become baby tree. Then the equation becomes

$$N'(t) = (a - bN(t - h))N(t) + cu_0(t - k) - u_1(t)$$ (7)

where the coefficient c for $(0 \le c \le 1)$ accounts for the function of seeds that actually result in a tree.

To start the equation we need to know the forest population in an interval of length h,

$$N(t) = N_0(t) \qquad (t - h \le t \le t_0)$$ (8)

Now in order to attain the equilibrium population N_e at a certain optimal time $t = \bar{t}$ we have the solution,

$$N(\bar{t}) = N_0(t)$$ (9)

and we say that the population is at equilibrium at $t = \bar{t}$ but not necessarily afterwards. If the population is to stay at equilibrium, the target condition must be

$$N(t) = N_e \qquad (\bar{t} - h \le t \le \bar{t})$$ (10)

This guarantees that $N(t) = N_e$ for all $t \ge \bar{t}$ if $cu_0(t - k) - u_1(t) = 0$ for $t \ge \bar{t}$. Target conditions of the form (9) are called Euclidean; those of the form (10) are called functional. To see the reason for this name, consider for instance the space $C[-h, 0]$ of continuous functions defined in the interval $-h \le t \le 0$. Given a function $y(t)$, denoted by $y_t(\tau)$ the section of $y(\tau)$ defined by

$$y_t(\tau) = y(t + \tau), \quad (-h \le \tau \le 0)$$ (11)

Then the target condition (10) can be written as an ordinary target condition in the space $C[-h, 0]$;

$$N_{\bar{t}}(\tau) = N_e$$ (12)

where N_e denotes the constant function. An optimal net profit problem is instance, to maximize the functional

$$J(t, u_1, u_2) = \alpha \int_0^t u_1(\tau) d\tau - \beta \int_0^t u_0(\tau) d\tau$$

with $\alpha, \beta \ge 0$ at some fixed time; the first term represents the profit from logging and the second, the cost of seeding. Clearly, $u_0(t), u_1(t)$ are nonnegative and it is reasonable to include upper bounds on both rates;

$$0 \le u_0(t) \le R, \quad 0 \le u_1(t) \le S$$ (13)

Straight maximization of the profit may result in destruction of the forest at time $\bar{t}$, thus we supplement the problem with a (functional) target condition, say

$$|N(t) - N_e| \le \varepsilon, \quad (\bar{t} - h \le t \le \bar{t})$$ (14)

N_e the equilibrium solution (6), and the terminate seeding at time $\bar{t} - k$. If the equilibrium position is stable, this means the forest population will stay near equilibrium after $\bar{t}$. Admissible controls for this problem are pairs $(u_0(t), u_1(t))$, where u_0 is defined in $-k \le t \le \bar{t}$ and u_1 is defined in $0 \le t \le \bar{t}$, and satisfying (13) in their respective intervals of definition.

The logistic equation discussed above is in dimensional form which have some free parameters, but sometimes the dimensional form is transformed into nondimensional form in order to reduce the parameters which gives good results in the numerical treatments. We present here the nondimensional form of the logistic equation,

$$\frac{d\eta}{d\tau} = \eta(1 - \eta), \text{ where } \tau = \frac{t}{t_*} \text{ and } \eta = \frac{N}{N_*}$$
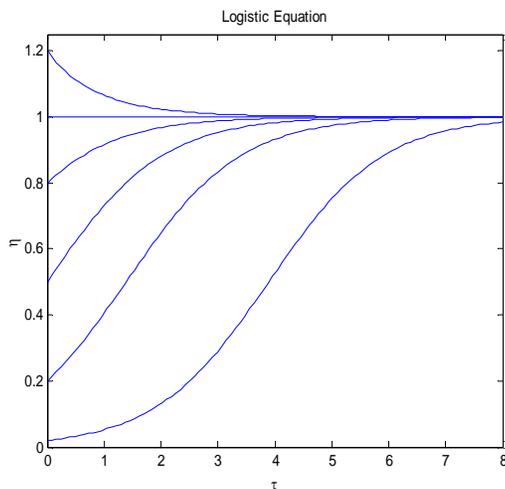
Fig. 2. The solution of nondimensional logistic equation

The numerical solution of the nondimensional logistic equation for the initial conditions 0.02, 0.2, 0.5, 0.8, 1 and 1.2 is shown in Fig. 2, where the lowest curve is the characteristic ´S-shape´ usually associated with the solution of the logistic equation. This sigmoid curve appears in many other models. See [4], [13] and [15] for details study on logistic equations and numerical applications.

Another growth model is described by the integrodifferential equation of the form,

$$N'(t) = \left(a - \left(\int_{-h}^{0} b(\tau) N(t+\tau) d\tau \right) \right) N(t) + cu_0(t-k) - u_1(t) \tag{15}$$

taking into account the inhibiting effects of new trees of all sizes on the growth rate. Our further research will be focused on that issue.

4 APPLICATION OF CONTROL THEORY

Numerous applications of control theory in the diverse fields have created this topic more challenging; specially an efficient and sustainable forest management is a crucial issue in the present world of climate change as well as the decay of ozone layer. In recent years, a number of economists and other experts have suggested sequestering carbon in forests to help mitigate the accumulation of greenhouse gases in the atmosphere. For more details studies see [1], [2], [21] and [22]. Forests currently store a substantial stock of carbon, amounting to 826 billion metric tons in trees and soil [11], and society can potentially remove carbon from the atmosphere by taking steps to increase this pool of carbon. These steps may include increasing the amount of carbon stored per hectare through management intensity or rotations ages (see for examples [21] and [37]) or increasing the area of land in forests. Carbon sequestration thus offers the promise of reducing the cost of greenhouse gas mitigation, which could lower the price of carbon and reduce global warming. Recent climate change is the global threat on the environment,

where sustainable forest management is the burning issue worldwide. For details servey on greenhouse effects and climate change issues see [2], [7] and [34].

Forests also provide a large variety of services such as: timber production, recreation and landscape, natural habitat for numerous species, protections of watersheds, and protections of villages from avalanches and landslides and buffering as well. Thus from a socio-economic point of view, the optimal management of forests must take these multiple services into account. All the issues mentioned above indicate that a sustainable forest management is a crying need where the optimal policy strategy is the key.

5 THE SUNDARBAN: A CASE STUDY

In this section, we will discuss the necessaity of forest management of a particular case of Sundarbans. The Sundarbans, situated both in Bangladesh and India is the world's largest delta, formed from the sediments brought down by three great rivers, the Ganges, the Brahmaputra and the Meghna which converge on the Bengal Basin. The forest consists of about 200 islands, separated by some 400 interconnected tidal rivers, creeks and canals. At 10,000 sq. km, it forms the largest mangrove forest in the world and it is the only mangrove tiger land on the earth. In 1947 the whole Sundarban mangroves were divided between India and Bangladesh (formerly East Pakistan), sharing 40% in India and 60% in Bangladesh. Although the two parts of the Sundarbans differ considerably in the nature and extent of investigations, conservations and managements due to belonging to the separate independent countries, the natural resources and beauties attract the global attentions simultaneously. As a results the Indian part of Sundarbans had been declared as a world heritage site in 1987 as *Sundarbans National Park* and Bangladesh part was declared as a world heritage site as *The Sundarbans* in 1997 by International Union for Conservation of Nature of UNESCO. The whole Sundarbans area was declared as Biosphere Reserve in 1989. See [5], [8], [25] and [32] for the details study about the history of Sundarbans and its local and global importance. However we are mainly concern to discuss the management about the Bangladesh part of Sundarbans. Presently the Sundarbans is one of the world's *new 7 wonders of nature* competitors' which will be declared at the end of 2011. The Sundarbans represents nearly half of the remaining forests of Bangladesh and is dominated by halophytic tree species such as **sundori** (*Heritiera fomes*) (from which Sundarbans derives its name), *gewa* (*Excoecaria agallocha*), *goran* (*Ceriops decandra*), *baen* (*Avicennia officinalis*), and *keora* (*Sonneratia apetala*). Sundarban is the habitat of many rare and endangered animals (*Batagur baska, Pelochelys bibroni, Chelonia mydas*), specially the Royal Bengal Tiger (*Panthera tigris*) and spotted deers. Javan rhino, wild buffalo, hog deer, and barking deer are now extinct from the area. The Sundarbans forest is the home to more than 400 tigers. However, the Royal Bengal Tiger is the king of all animals in the Sundarbans which have developed a unique characteristic of

swimming in the saline waters, and are world famous for their man-eating tendencies. Apart from the Royal Bengal Tiger; *Fishing Cats*, *Macaques*, *Wild Boar*, *C*ommon *Grey Mongoose*, *Fox*, *Jungle Cat*, *Flying Fox*, *Pangolin*, *Chital*, are also found in abundance in the Sundarbans. However, widespread hunting and forest depletion has reduced the tiger's range and numbers in Sundarbans. Natural disasters as well as climate change are also affecting this world heritage. In 2007, Cyclone CIDR hit directly on the coastal areas of Sundarbans and almost destroyed the whole forests and thus destroyed the conservation of the wild animals. However, proper management of the Sundarbans is crucial for the sustainable dwelling of wildlife and their prey. The major steps against the unplanned or illegal wood-cutting as well as for increasing the number of trees in the forest should be taken. According to the *Rio Convention* proposed by the United Nations Conference on Environment and Development (UNCED) Bangladesh has the obligation to accomplish the required of the convention which is 'timely, reliable and accurate information on forests and forest ecosystems is essential for public understanding'. In the meantime, to ensure the protection for wildlife habitat and the management of natural resources, three areas within the forest have been designated as Wildlife Sanctuaries: Sundarbans West (715 sq. km), Sundarbans South (370 sq. km), and Sundarbans East (312 sq. km) by Bangladesh Government. A long term plan named *Bangladesh Tiger Action Plan 2009-2017* has been taken by Bangladesh Government for the conservation of tigers and other wildlife recourses as a part of forest management. See for details [3], [5].

5.1 Importance of the Sundarbans

The Sundarbans ecosystem is unique in many respects. The area experiences a subtropical monsoon climate with the annual rainfall of about 1600-1800 mm and several cyclonic storms. The biodiversity includes about 350 species of vascular plants, 250 fishes and 300 birds, besides numerous species of phytoplankton, fungi, bacteria, zooplankton, benthic invertebrates, molluscs, reptiles, amphibians and mammals. Species composition and community structure vary east to west, and along the hydrological and salinity gradients. Large areas of the Sundarbans mangroves have been converted into paddy fields over the past two centuries, and more recently into shrimp farms. The Sundarbans has been extensively exploited for timber, fish, prawns and fodder.

The Sundarbans is the only largest mangrove forest in the world managed for commercial timber production and has had a history of scientific management since 1879. In Bangladesh it is now managed by the Sundarban West Forest Division and Sundarban East Forest Division of the Forest Department, divided into 20 sections each harvested in turn on a 20-year cycle, with the three peripheral wildlife sanctuaries on the coast. Early management consisted of revenue collection by enforcing simple felling rules, subsequent enforcement of which reduced the amount of over-cutting of the four main timber species. A wildlife conservation plan prepared under the joint sponsorship of the World Wildlife Fund and the U.S. National Zoological Park emphasised management

of the tiger and other wildlife as an integral part of sustainable forest and coastal management for both timber and the needs of the local population.

Approximately 2.5 million people lived in small villages surrounding the Sundarbans in 1981 which by 1991 had increased to 3 million. At nomination, some 35,330 people worked in the forest, 4,580 of whom collected timber and firewood, 1,350 collected honey and beeswax and 4,500 harvested the natural resources and hunted mainly deer, and 24,900 were fisherman and shrimp farmers. Today, the area provides a livelihood at some seasons of the year for an estimated 300,000 people. Some 4,500 people in Bangladesh are employed by contractors in the commercial logging of sundori and other timber, which is 45% of all that produced in state-owned forests. As well as construction timber they supply local
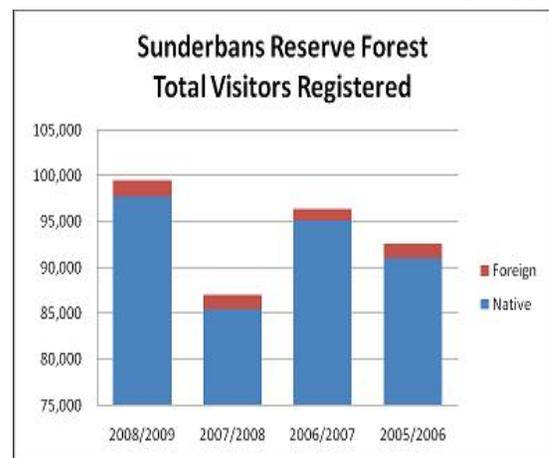
Fig. 3. The figure shows the statistics of total visitors visited the Sundarbans during 2005 to 2009 . Source: IPAC, 2009.

newsprint paper, match and board mills (see for examples [8], [32]).

Local people are also dependent on the forests and waterways for firewood, charcoal, timber for boats and furniture, poles for house-posts and rafters, nypa palm thatch for roofing, grass for matting reeds for fencing, shells and reptile skins, with deer, fish, crabs and shrimps taken for food. The season for collecting honey and wax is limited to ten weeks from April 1st. Thousands of people, with permits from the Forest Department, enter the forest for nests. Before Cyclone Sidr, which has destroyed the fishing industry, more than 10,000 fishermen from as far away as Chittagong camped along the coast for 3-4 months in winter before returning home at the start of the monsoon season in April, and as many or more local people fished year-round [32]. In 1986 the average annual catch was 2,500 tonnes.

The important sector based on Sundarbans is *Tourism Industry*. The Sundarbans may be more attractive to the visitors and tourists due to its natural attractive beauties as well as the rare wild animals which will play a significant role both in the national and global economy. In 1996, about 500 foreign tourists plus 5,000 local tourists visited the area, most in the South Wildlife Sanctuary and in recent years most nearly 100,000 local and about 1,500

foreign tourists per year visited the Sundarbans. Although there is no potential for mass tourism but limited eco-tourism from October to April or May is possible. Recent research (see the Fig. 3) on Sundarbans showed that the number of national and international visitors and tourists are increasing.The figure reflects some variability in visitor numbers over the last five years, with the highest numbers in the year 2008-2009. This indicates that efficient and sustainable management of Sundarbans and availability of logistic facilities are necessary for enriching the tourism industry with strong revenue earning.

6   SUNDARBAN FOREST MANAGEMENTS

The future of the Sundarbans will depend upon the sustainable management of freshwater resources as much as on the conservation of its biological resources. Considerable researches both in nationally and internationally have been carried out on the Sundarbans ecosystem and its wildlife (see for examples [18], [25], [26] and [33]). Since the last few decades several national (see [29]) and international (see [23], [32]) leading NGOs carried out different (short-terms and long-terms) research projects on Sundarbans emphasizing on the sustainable managements to meet the future challenges for the next generations. They also emphasized on Collaborative forest management (CFM) which is loosely defined as a working partnership between the key stakeholders in the management of a given forest—key stakeholders being local forest users and state forest departments, as well as parties such as local governments, civic groups and nongovernmental organisations, and the private sector [12].

The world heritage Sundarban is the source of natural beauty to the tourists as well as to the environmentalists, where the *Royal Bengal* tiger is one of the attractions to the visitors. However, the natural disasters like flood, cyclone etc. are very common issues taking place in Bangladesh specially in that region. All most every year the natural disasters hit on this forest and as a result, the forest loses its trees and animals and consequently its beauty. Sometimes the unplanned cutting of the trees makes the forest to be destroyed day by day. So it becomes necessary to protect the Mangrove from the unavoided situations by making an optimal design for sustainable management. The above mentioned model for example can be applied for such efficient growth of the trees in the Sundarban Mangrove area. A successful application of the Malthusian model for the sustainable management of Sundarban Mangrove may ensure the long-term survival with its full beauty within a certain time applied. For a better understanding of the model we present here a simple example.

6.1 Example

Suppose the population of forest of Sundarban $p(t)$ given by the exponential growth law. Growth rate r is positive. How much time it will take to increase p_0 to double, where p_0 is the initial trees of forest.

Solution: Let the required time is T.
The exponential law of growth is

$$p(t) = p_0 e^{r(t-t_0)} \tag{16}$$

For the given condition,

$$p(t_0 + T) = 2p(t_0)$$

$$\Rightarrow 2 = \frac{p(t_0+T)}{p(t_0)} \Rightarrow \frac{p_0 e^{r(t_0+T-t_0)}}{p_0 e^{r(t_0-t_0)}} = 2 \quad \text{[with the}$$

help of (16)]

$$\Rightarrow e^{rT} = 2 \Rightarrow rT = \log_e 2 \Rightarrow T = \frac{\log_e 2}{r}$$

After $T = \dfrac{\log_e 2}{r}$ times the trees in the forest will be double in compared to its initial state of the trees of the forest. This solution is shown in the Fig. 4 considering the growth rate $r = 100$.
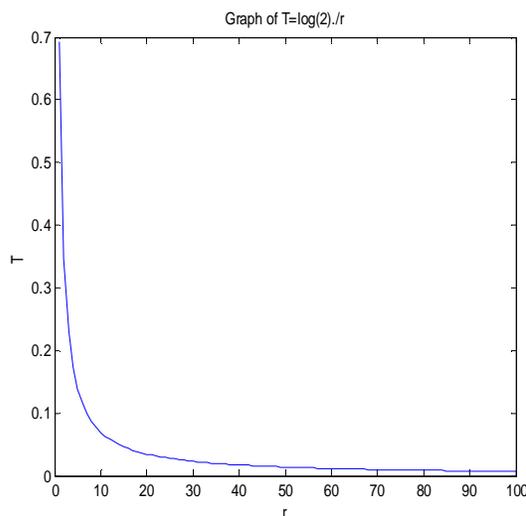
Fig. 4. The exponential growth of trees in the forest.

7   CONCLUSION

In Sustainable forest management is the warming issue due to the climate change and increasing the greenhouse gases. Optimal control strategy is the key to such conservation of forests. Among the different models for this optimal strategy Malthusian growth model is one which gives a good result where the logistic equation plays the vital role. This study investigates the application of control theory in the light of Malthusian model, which is used in controlling the forest growth and sustainable management. An illustrative example is presented taking into account the special case of the largest mangrove forest Sundarban. We claim that this study will play pioneer role in exploring further study in the field of control theory in agricultural and life sciences. The proposed model of controlling the optimal growth of the forests in

the largest mangrove in world can be applied rigorously which may help to increase the different trees in the forests destroyed by the natural disaster every year in the coastal region and thus can make a natural equilibrium. we believe many more interesting conclusions will be brought under the different assumptions by using the optimal control strategy, and we hope to focus on such management issues in our future research, also we welcome more researchers to take part in the field.

ACKNOWLEDGMENT

REFERENCES

[1] Adams, R. M., D. M. Adams, J. M. Callaway, C. C. Chang and B. A. McCarl. 1993. Sequestering Carbon on Agricultural Land: Social Cost and Impacts on Timber Markets. *Contemporary Policy Issues*. 11: 76-87.

[2] Adams, D.M.; Alig, R.J.; McCarl, B.A.; Callaway, J.M.; and Winnett, S. M. 1999. Minimum cost strategies for sequestering carbon in forests. *Land Economics*. 75(3) 360-374.

[3] Ahmad, I. U., Greenwood, C. J., Barlow, A. C. D., Islam, M. A., Hossain, A. N. M., Khan, M. M. H. and Smith, J. L. D. 2009. *Bangladesh Tiger Action Plan 2009-2017*. Bangladesh Forest Department, Ministry of Environment and Forests, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh.

[4] Allman, E. S. and Rhodes, J. A. 2003. Mathematical Models in Biology An Introduction. Cambridge University Press, New York.

[5] Altrell, D., Saket, M., Lyckebäck, L. and Piazza, M. *National Forest and Tree Resources Assessment 2005-2007, Bangladesh.* Implemented by Bangladesh Forest Department, Ministry of Environment and Forest (MoEF), Bangladesh Space Research and Remote Sensing Organization, Ministry of Defence, Bangladesh.

[6] Athans, M. and Falb, P. L. 1966. *Optimal Control*, McGraw-hill, New York.

[7] Backéus, S., Eriksson, L. O. and Garcia, F. 2005. Impact of climate change uncertainty on optimal forest management policies at stand level. Zerger, A. and Argent, R.M. (eds) *Proceeding on MODSIM 2005 International Congress on Modelling and Simulation.* Modelling and Simulation Society of Australia and New Zealand, December 2005.

[8] Bangladesh: Sundarbans Biodiversity Conservation Project, *ADB Validation Report.* Project Number: 30032, Reference Number: PCV: BAN 2008-37, Asian Development Bank, 2008.

[9] Biswas, M. H. A.; Huda, M. A.; Ara, M. and Rhaman, M. A. 2011. Optimal Control Theory and It's Applications in Aerospace Engineering. To be appeared in the *International Journal of Academic Research*, Vol. 3, No. 2, March 2011.

[10] Boltyyanskii, V. G. 1971. *Mathematical Methods of Optimal Control*. Holt, Rinehart and Winston, New York.

[11] Brown, S. 1998. Present and future role of forests in global climate change. In B. Goapl, P. S. Pathak, and K. G. Saxena (eds.), *Ecology Today: An Anthology of Contemporary Ecological Research*, International Scientific Publications, New Delhi, pp. 59-74.

[12] Carter, J. and Gronow, J. 2005. *Recent Experience in Collaborative Forest Management: A Review Paper*. CIFOR Occasional Paper No. 43. Center for International Forestry Research, Jakarta 10065, Indonesia.

[13] Chasnov, J. R. 2009. Mathematical Biology, Lecture Notes. Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong.

[14] Fattorini, H. O. 1999. *Infinite Dimensional Optimization and Control Theory.* Cambridge University Press, London.

[15] Gershenfeld, N. 1999. *The Nature of Mathematical Modeling*, Cambridge University Press, ISBN 978-0521-570954, Cambridge, UK.

[16] Glowinski, R. 1984. *Numerical Methods for Nonlinear Variational Problems* (2nd Edition), Springer-Verlag Publications, New York.

[17] Goetz, R. and Xabadia, A. 2003. Applications of Distributed Optimal Control in Economics – The Case of Forest Management. Working Paper, Economy Series E2003/37, Site A: Foundation Centre of Andalusian Studies, University of Girona, Spain.

[18] Gopal, B. and Chauhan, M. 2006. Biodiversity and its conservation in the Sundarban Mangrove Ecosystem. *Aquat. Sci. 68, 338–354.*

[19] Hoen, H. F. and Solberg, B. 1994. Potential and economic efficiency of carbon sequestration in forest biomass through silvicultural management. *Forest Science,* 40(3): 429-451.

[20] Holopainen, Jani and Marieke Wit (eds). (2008). *Financing Sustainable Forest Management.* Issue No. 49. European Tropical Forest Research Network (ETFRN), Tropenbos International, Wageningen, the Netherlands.

[21] Intergovernmental Panel on Climate Change (IPCC). 1996. Scientific-Technical Analysis of Impacts, Adaptations, and Mitigation of Climate Change. Report of Working Group II. Climate Change 1995: IPCC Second Assessment Report. Cambridge: Cambridge University Press.

[22] Intergovernmental Panel on Climate Change (IPCC). 2000. IPCC Special Report: Land Use, Land-Use Change, and Forestry. Cambridge: Cambridge University Press.

[23] Integrated Protected Area Co-Management (IPAC): Strengths, Weaknesses Opportunities, & Threats (SWOT) of Tourism in the Sundarbans Reserve Forest. USAID/Bangladesh, 2009.

[24] Malthus, T. J. 1798. *An Essay on the Principle of Population*. Chapter 1, p. 13 in Oxford World's Classics reprint.

[25] Manna, S., Chaudhuri, K., Bhattacharyya, S. and Bhattacharyya, M. 2010. Dynamics of Sundarban estuarine ecosystem: eutrophication induced threat to mangroves. *Saline Systems* 2010, 6:8.

[26] Musa, K. B. 2008. *Identifying Land Use Changes and It's Socioeconomic Impacts; A Case Study of Chakoria Sundarban in Bangladesh.* Master's Thesis, Department of Computer and Information Science (IDA), Linköping University, Linkoping, Sweden.

[27] Noack, F. 2008.*The bioeconomics of livestock rearing and timber production in the Caspian forest.* Diploma Thesis in the study program of Landscape Ecology and Nature Conservation, Institute of Botany and Landscape Ecology, Ernst Moritz Arndt University Greifswald.

[28] Pinch, E. R. 1993. *Optimal Control and the Calculus of Variations*, Oxford University Press Inc., New York.

[29] Policy Brief on "Environmental Policy", *CPD Task Force Report.* 2001. Centre for Policy Dialogue, Dhaka, Bangladesh.

[30] Pontrryagin, L.; Boltyyanskii, L. S.; Gamkrelidze, R. V. and Mishchenko, E. F. 1964. *The mathematical theory of optimal processes*. Pergamon press, Oxford.

[31] Stavins, R. 1999. The Costs of Carbon Sequestration: A Revealed Preference Approach. *American Economic Review*. 89(4): 994 - 1009.

[32] Sundarbans National Park, India and The Sundarbans, Bangladesh, 2008. "Natural site datasheet from WCMC". World Conservation Monitoring Centre. http://www.unep-wcmc.org/sites/wh/pdf/Sundarbans.

[33] Tuhin, G., Gupinath, B. and Sugata, H. 2003. Application of a 'bio-engineering'technique to protect Ghoramara Island (Bay of Bengal) from severe erosion. *Journal of Coastal Conservation 9: 171-178.*

[34] Van Kooten, G. Cornelius; Binkley, Clark S.; and Delcourt, Gregg. 1995. Effect of carbon taxes and subsidies on optimal forest rotation age and supply of carbon services. *American Journal of Agricultural Economics.* 77(5): 365-374.

[35] Wilkie, M. L., Holmgren, P. and Castaneda, F. 2003. *Sustainable forest management and the ecosystem approach: two concepts, one goal.* Forest Management Working Papers, Working Paper FM 25. Forest Resources Development Services, Forest Resources Division, FAO, Rome, Italy.

[36] Xabadia, A. and Goetz, R. U. 2008. *The Optimal Selective Logging Regime and the Faustmann Formula.* Barcelona Economics Working Paper Series, Working Paper No. 353, University of Girona, Spain.

[37] Xiao, J., Kang, W., Yin, S. and Zhai, H. 2010**.** An Analytical Optimal Strategy of the Forest Asset Dynamic Management under Stochastic Timber Price and Growth: A Portfolio Approach. *Scientific Research: Low Carbon Economy*, 1, 25-28, 2010.

Parameter Ranking and Reduction in Communication Systems

M.H. Azmol, M.H. Biswas, and A. Munnujahan

Abstract—Parameter reduction from experimental data is an important issue arising in many frequently encountered problems with different types of applications in communications engineering. However, the computational effort grows drastically with the number of parameters in such types of applications. This paper proposes a technique that reduces the performance parameters of a communication system based on eigenvalues of covariance matrix as well as providing a weighted rank of parameters by an approach called non-negative matrix factorization (NMF). The factorization of original matrix provides a weight metric that offers a means of ranking and selecting meaningful important parameters. The relative importance of each parameter is measured from the sequentially ordered eigenvalues. The main aims of this paper are to determine, identify and reduce the reasonable number of performance parameters which will reflect the best measurement system of a describing network scenario.

Index Terms—Eigenvalues, Parameter reduction, Non-negative matrix factorization

———————————————— ◆ ————————————————

1 INTRODUCTION

IN the modern era, recent technological developments of telecommunication devices, computer based instrumentation and electronic data storage have resulted in increasing quantities of data and in a consequence, overwhelm many of the classical data analysis tools available. Processing these huge amounts of data has created many challenges and new concerns of data representation with a few numbers of dimensions. Many researchers have been exploring in these fields and a large portion of these efforts are concerned with the challenging task of evaluating the performances of a network by measuring some quality of service (QoS) support parameters such as bandwidth, load, delay, jitter, error rate etc. However, measuring all of these variables together is very challenging due to the growing complexity of telecommunication systems as well as some constraints on information gathering devices. The collected data from these complex phenomena represent the integrated result of several interrelated variables since they act together. The original data becomes noisy, overlap and ambiguous. In most phenomena, some parameters may have little effect on model predictions, and the effects of some parameters can be highly correlated with the effects of others, making parameter estimation difficult. As a result, modelers often use simplified models that contain a reduced number of parameters or they estimate only a subset of the parameters in their complete model. Moreover, being ignorant experimenters we don't know which, let alone how many, axes and dimensions are important to measure. Even we often don't know which measurements best reflect the dynamics of our system. Generally, in many applications, the data set is gathered from a huge number of dimensions (variables). This *curse of the dimensionality* (variables) degrades the query efficiency and accuracy of a system. So, the question is raised, how can we reduce these dependent parameters as well as provide a compact system model which will provide fidelity near the significant level of QoS of the original system without much loss of information's. A suitable parameter reduction technique may address all of these issues. The problem of parameter reduction has received a broad attention in different areas of research such as computer vision [2], information retrieval [3], machine learning, pattern recognition etc. Moreover, it has also widespread application in other areas, ranging from ecological systems, power systems, production systems, chemical reactions and biochemical networks to wastewater treatment processes. There exist many parameter reduction techniques such as Principal Component Analysis [6], [7], Singular Value Decomposition [8], Feature Selection [4], Non-negative Matrix Factorization (NMF) [5], and Factor Analysis [9]. The main disadvantage of feature selection and factor analysis approaches are that some combination of parameters may give better results, but could be excluded because of parameters selected at earlier steps. In, principal components, there are some limitations on non-linear cases. We use NMF algorithms as it is simple-nonparametric method which reveals easily underlying structures in complex data set. By using, NMF algorithms of the original data matrix, we derive a weight metric to obtain their ranking. The relative importance of each parameter is obtained from the sequentially ordered eigenvalues of covariance matrix, which in turn expresses the attributive energy of the data. Eigenvalues quantifies the importance of each dimension by describing the variability of a data set. In particular, the measurement of variance along each dimension provides a means for comparing the relative importance of each variable. An implicit hope behind employing this method is that a small number of dimensions (i.e. less than the number of measurement types) will provide a reasonable characterization of the complete data set. In fact, categorizing eigenvalues in this manner,

a significant insight of the whole network can be obtained. For an illustration, in our network data matrix, we have explored six parameters of a WLAN network scenario and compare them. In general, we know that "throughput" and "network load" are the most important parameters for assessing the performance of a network. In our method, we have proved that these two parameters among six exhibit 94% characteristic of the whole network. In summary, we have proposed two measurement metrics in this paper: (i) a weighted metric for selecting the parameters that reveals important characteristics of a network (ii) and stipulated eigenvalues, measure the relative importance of each parameter. The primary objectives of our work are to determine, identify and reduce significant number of performance parameters in a systematic and representative way, so that a describing network scenario will characterize and reflect the best measurement system. The rest of this paper is organized as follows. The notations used throughout the rest of this paper are given in section 2. Related works are described in section 3. We outline the steps taken to collect the data in section 4. A brief overview of mathematical measurement technique of non-negative matrix factorization is discussed in section 5. In section 6, we discussed our experimental results. Concluding remarks are presented in sections 7.

2 BACKGORUND NOTATIONS

In order to facilitate discussion in subsequent sections we introduce relevant notation first. We use X is a m-by-n measurement matrix in which each column denotes the predicted variables and each row represents the sample observations. $\Re$ is a set of non-negative numbers, P and Q are the factorization of X of size m-by-k and k-by-n respectively. Y is the new transformation of X, $\|.\|$, is the Euclidean norm, λ_i and v_i are the corresponding eigenvalues and eigenvectors of covariance matrix X respectively. In our experiment data, we have taken six variables and 300 observations. So the size of X is *300 by 6*. Unless otherwise stated, all the vectors in this paper are column vectors.

3 RELATED WORKS

Our ideas behind parameter reduction in communication systems are motivated by a variety of past studies. Although to our knowledge, parameter reduction in a network using NMF and eigenvalues has not been previously applied directly. Author's in paper [11] and [12] present a framework for nonlinear systems analysis based upon controllability and observability covariance matrices. This covariance matrix is used for reduction of the nonlinear model. Sanjay L. et al. [13] introduced a new method of model reduction for nonlinear system with inputs and outputs. A new technique based on Hankel

singular values for reducing the parameter set of a fundamental model is introduced in paper [14]. The main drawback of this method is that the physical meaning of the parameters is lost during this procedure. There is a little prior work that is closely related to ours. Authors in paper [16] measured the performance of wireless networks by state space reduction. They propose the stochastic method based on the comparison of Markov chains by following three steps: first, choice of the state space of comparison, second, choice of the relation order on the chosen space and finally, construction of bounding models by applying aggregation. Paper [4] describes feature selection algorithm based on measuring similarity between features. A. Lakhina et al. in [17] analyzed the network flow and found that, in a network with over a hundred of origin to destination flows, these flows can be accurately modeled in time using a small number (10 or less) of independent components or dimensions. In [18] and [19] authors proposed different functional test areas to verify next generation broadband network. They classified the equipments into different groups by some scores such as MUST, SHOULD, MAY and CAN. This method can be used in order to compare any device of different vendors with each other but not their own devices, since all tests don't have the same value. Hence this method will not be as reliable as needed. Most of the existing reduction techniques based on model reduction and focus on to reduce the number of states, whereas relatively few techniques are available for reduction of the number of parameters in these models.

4 NETWORK ENVIRONMENT AND EXPERIMENTAL DATA

For network simulations and raw data of different parameters, we use OPNET modeler which is one of the most leading and powerful simulation tools for the analysis, planning, optimization and evaluating of communication networks, devices and protocols. The measurement platforms and environment in which we conduct our experiments (Fig-1) are described in this section. We use two WLAN fixed nodes separated by 80 meter distance in our test bed. We perform experiments with high resolution video conference as an application in IEEE 802.11b WLAN. The attributes of different performance parame-
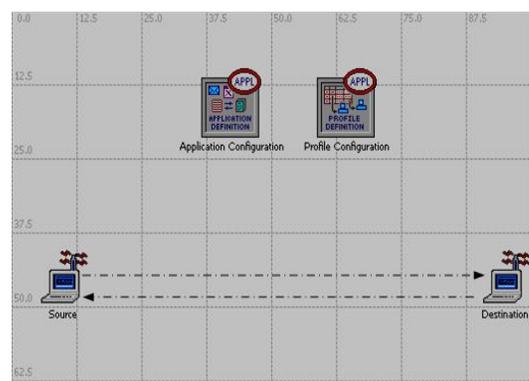
Fig. 1. WLAN network scenario.
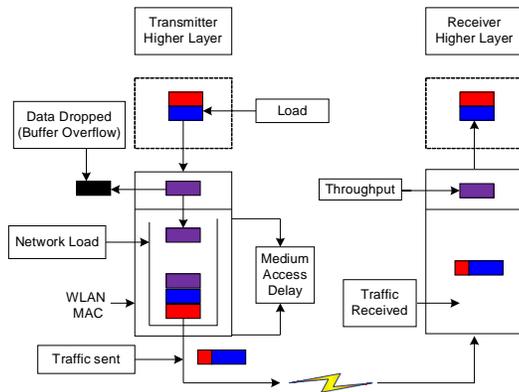
Fig. 2. Attributes of different parameters.

TABLE 1
SIMULATION PARAMETERS OF WLAN

Wireless LAN parameters	Values
Physical Characteristics	DSSS
Data Rate	11 Mbps
Channel Settings	Auto Assigned
Transmit Power	0.001w
Packet Reception-Power Threshold	-95
RTS Threshold	None
Fragmentation Threshold	None
CTS-to-Self Option	Enabled
Short Retry Limit	7
Long Retry Limit	4
Buffer Size	256000
Large Packet Processing	Drop
PCF Parameters	Disabled
HCF Parameters	Not Supported

TABLE 2
MEASUREMENT UNITS OF PARAMETERS

Collected parameters	Units
Data Dropped (Buffer Overflow)	bits/sec
Delay	sec
Load	bits/sec
Media Access Delay	sec
Network Load	bits/sec
Throughput	Bits/sec

ters are presented in Fig-2. We collect the global statistics of six parameters: data dropped (data overflow), delay,

load, media access delay, network load and throughput. The configuration of WLAN parameters and measurement unit of each parameter are shown in Table-I and Table-II.

5 MATHEMATICAL APPROACH

In this section, we discuss our mathematical approach. Suppose X is our original data set, where each row corresponds to a set of measurements from one particular trial (a single sample) and each column of X corresponds to all measurements of a particular type. In our data set X is an *m-by-n* matrix where *m = 300* and *n = 6*. Our first approach is to find some orthonormal matrix v such that $Xv = Y$, and $Y^T Y$ is a diagonal matrix. Here generally, v is the eigenvector of the covariance matrix $X^T X$ and Y is a new transformed presentation of X, derived from $Xv = Y$. Generally, the primary motivation behind our first approach is to de-correlate the data set, i.e. to remove second-order dependencies. Y is arranged with the stipulated eigenvalues $\lambda_1 > ... > \lambda_r$. Our second approach is to factorize the original data X so that we can find a weight matrix to make some rank of the variables in X. Here, we discuss the factorization technique.

5.1 Non-negative Matrix Factorization

In an article in *Nature*, Lee and Seung, 1999, [20] started a flurry of research and proposed the notion of non-negative matrix factorization algorithm. Prior to its publication several lesser known paper published and actually they deserve more credit for the factorization of non-negative matrix. Given a data matrix $X = [x_1, x_2, ..., x_n] \in \Re^{m \times n}$, each column of X is a sample vector and a positive integer $k < \{m, n\}$, non-negative matrix factorization aims to find two non-negative matrices $P = [p_{ij}] \in \Re^{m \times k}$ and $Q = [q_{ij}] \in \Re^{k \times n}$ which minimize the following function

$$f(P,Q) = \frac{1}{2} \|X - PQ\|_F^2 \tag{1}$$

where $\|.\|_F$ denotes the Frobenius norm. The Frobenius norm of a matrix $X = [x_{ij}]$ is defined by $\|X\|_F = \sqrt{\sum_{ij} X_{ij}^2}$.

The product PQ is called the non-negative matrix factorization of X, though X is not necessarily equal to the product of PQ. The product PQ is an approximate factorization of X of rank at most k. Usually, an appropriate decision on the value of k is critical, but the choice of k is often problem dependent. The objective function (1) is convex in both P and Q only, it is not convex in both

variables together. Therefore, to expect an efficient algorithm for finding global minima of $f(P,Q)$ is unrealistic. This important problem affects the numerical minimization of (1) since it includes the local minima due to the non convexity. Various minimization methods for the solution of (1) have been proposed in an effect to speed up the convergence. Lee and Seung's iterative multiplicative update algorithms have become a balance algorithm as follows:

$$P_{ij}^{t+1} \leftarrow P_{ij}^{t} \frac{\left(XQ^{T}\right)_{ij}}{\left(PQQ^{T}\right)_{ij}}$$

$$Q_{ij}^{t+1} \leftarrow Q_{ij}^{t} \frac{\left(X^{T}P\right)_{ij}}{\left(Q^{T}P^{T}P\right)_{ij}}$$

Multiplicative Update Algorithms:

Lee and Seung [20] proposed the prototype of multiplicative algorithm with the mean squared error objective function which is provided below:

P = rand (m, k)   % initialize P as random dense matrix

Q = rand (k, n)   % initialize Q as random dense matrix

for   i = 1: maxiter

$$Q = Q.*\left(P^{T}X\right)./\left(P^{T}PQ+10^{-9}\right);$$
$$P = P.*\left(XQ^{T}\right)./\left(PQQ^{T}+10^{-9}\right);$$

end

To avoid division by zero, 10^{-9} is added in each update rule.

5.2 Basic Geometric Review of NMF

NMF essentially tries to find two non negative matrices P and Q such that $X \approx PQ$. This approximation can be viewed column by column as $x_i \approx \sum_{i=1}^{k} p_i q^i$ where p_i denotes *i-th* column vector of P and q^i denotes *i-th* row of matrix Q. Thus each data vector x_i is approximated by a linear combination of the columns of P weighted by the components of Q. Therefore, P can be considered as basis

matrix which optimized the linear transformation of X. The outer product of PQ demonstrates how the rows of Q essentially specify the weights of each of the column-vector of P. If we look geometrically, NMF is a conical coordinate transformation. Graphical interpretation of NMF is shown in Figure-3. The two basis vectors P_1 and
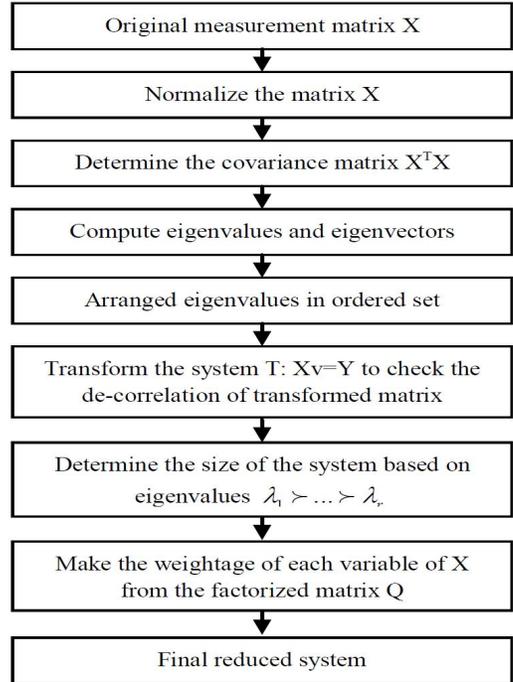
Fig. 4. Flowchart of reduction procedure.

TABLE 3
WEIGHT MATRIX Q

0.0552	0.1336	0.8836	0.1362	0.2894	0.3099
0.0734	0.1522	0.8868	0.1546	0.2737	0.2938

TABLE 4
EIGENVALUES AND THEIR PERCENTAGE

Eigenvalues	Percentage	Cumulative sum
4.6262	75.0049	0.7500
1.0827	18.6317	0.9364
0.2366	04.2039	0.9784
0.0528	01.4876	0.9933
0.0017	00.6498	0.9998
0.0002	00.0221	1.0000

Fig. 3. Geometric behavior of NMF.
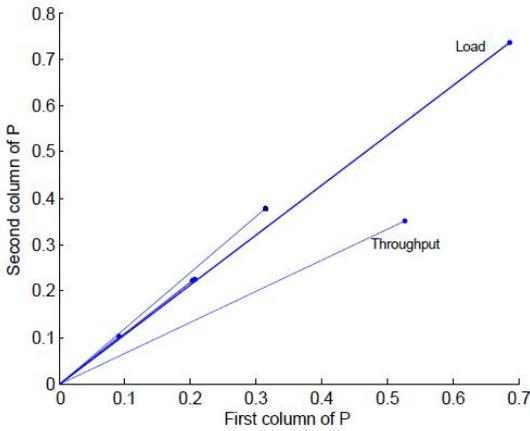
Fig. 5. Relative weight of variables in X w.r. to varianles in P.

P_2 describe a cone which encloses the approximation data set of X. Due to the non-negative constraints the points

inside this cone can be reconstructed through linear combination of these basis vectors: $x = (p_1, p_2) \cdot (q_1 \cdot q_2)^T$
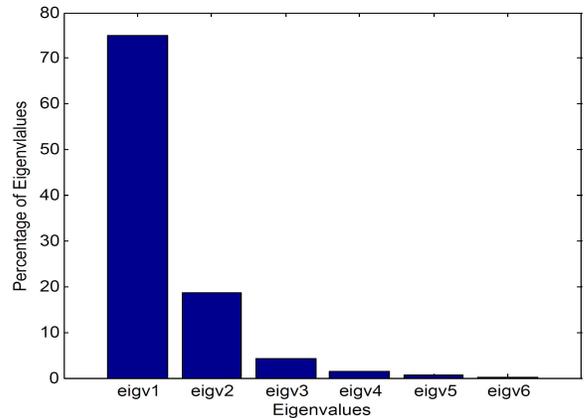
Fig. 6. Plot of eigenvalues versus their percentage.
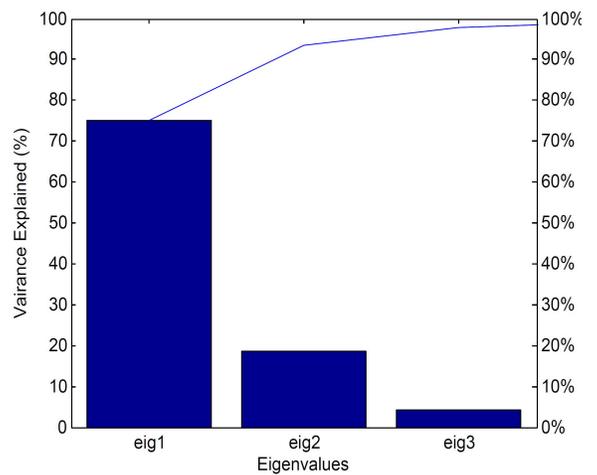
Fig. 7. Scree plot of the percent variability explained by eigenvalues.

6 EXPERIMENTAL RESULTS

Experimental results are discussed in this section. The flowchart of our experiment is shown in Fig 4. For our experiment, we use a multidimensional data of six variables collected from a WLAN network scenario. The sample size of our data is *300 by 6*. When we analyze our data set, an important and unmet challenge arises against different units or scales. In real world data set, variables are measured in different units and these discrepancies can distort the calculations also the analysis may lead inaccurate. We mitigate this problem by converting and normalizing all the values of data set by subtracting each column from the minimum value for moving towards center of the data and dividing by corresponding standard deviation of each variable for scaling. We factorize our original matrix $[X]_{300 \times 6}$ to $[P]_{300 \times 2}$ and $[Q]_{2 \times 6}$ so that $X \approx PQ$. Curious reader may ask why we chose the value of k is 2. This issue is explained in later. For the space constraints we show only the weight matrix Q in Table-III. The matrix Q represents the coefficients of the linear combinations of the original six variables in X that generate the transformed new variables in P i.e. rows of

Q provide the relative contributions of the six variables in X to produce the new variables in P. The third variable "load" in X (weight .8836) strongly influences the first predictor in P so we can consider it as the highest contributed variable in producing the matrix P. From the second row of Q the third variable "load" and the sixth variable "throughput" provide relatively strong weights. Since we have chosen "load" as the most contributed variable, we can select the sixth variable "throughput" as the second highest contributed variable. Finally, we can rank all the variables by providing weight in this manner. We visualize the relative contributions of the variables in the column space of P in Fig-5. The relative magnitude of the vectors provides the corresponding weights. The derived eigenvalues and their percentage of the total variability explained are given in Table IV. The bar plot Fig-6 shows the eigenvalues versus their explained variation. The first eigenvalue exhibits the maximum variation compare to others. We plot only three components in Fig-7 and it clearly shows that the first two components explain 94% energy of the data. Only two variables amongst six variables represent almost whole of the characteristic of original data matrix X. This is the reason why we

choose the values of k is 2.

7 CONCLUSION

In this paper, we have proposed a weight metric, derived from a data matrix of a network scenario to select lesser parameters that consequently leads to a parameter reduction. This method is easy to implement and the results are easily interpreted. The method gives a simple matrix manipulation and the ordered eigenvalues is an inherent property of the method itself. We emphasize the eigenvalues of the covariance matrix which believed to be important. Additionally, we provide the importance as well as quantifying contribution of the parameters by discounting those eigenvalues that are believed to be unimportant. The weighted metric and eigenvalues of the covariance matrix lead to parameter choices that appear finally parameter reduction. The primary benefit of our method is that, it quantifies the importance of each dimension for describing the variability of a data and provides a means for comparing the relative importance of each dimension. This method is not used just for WLAN but it is more general. From our experiment we have found that only two parameters "network load" and "throughput" among six, exhibit 94% percent contributions of whole network. So to evaluate the performance of a network, we emphasize more on these two parameters. We expect that, our approach will help to demystify the parameter reduction technique of a network and of great use in future.

ACKNOWLEDGMENT

REFERENCES

[1] IEEE 802.11 WG, Part 11: *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, Aug. 1999.

[2] K.V.R. Kanth, D. Agarwal, A.E. Abbadi, & A. Singh, "Dimensionality Reduction for Similarity Searching in Dynamic Databases," *ACM SIGMOD Conference Proceeding*, pp. 166-176, 1998.

[3] Ye J. Janardan, R. Li, "An Efficient Dimension Reduction Scheme for Image Compression and Retrieval," *ACM SIGKDD Conference Proceedings.* 2004.

[4] P. Mitra, C.A. Murthy, K. Sankar Pal, "Unsupervised Feature Selection Using Feature Similarity," *IEEE transaction on pattern analysis and machine intelligence.* vol. 24, no 3, march 2002.

[5] M. Berrya, M. Brownea, A.N. Langvilleb, V.P. Paucac, R.J. Plemmonsc, "Algorithms and Applications for Approximate Nonnegative Matrix Factorization," *Computational Statistics and Data Analysis* 52. pp. 155-173, 2007.

[6] J. Shlens, "A Tutorial on Principal Component Analysis", *Center for Neural Science*, New York City, NY 10003-6603, April 22, 2009.

[7] I.T. Jolliffe, "*Principal Component Analysis*," Springer-Verlag, New York, 2002.

[8] M.A. Hasan, "Low Rank Approximations with Applications to Principal Singular Component Learning Systems," *Proceedings of the 47th IEEE Conference on Decision and Control, Cancun Mexico*, Dec. 9-11, 2008.

[9] H.H. Harman, *Modern Factor Analysis*. 3rd Ed. Chicago: University of Chicago Press, 1976.

[10] Y. Jieping , "Generalized Low Rank Approximations of Matrices," *Proceedings of the 21st International Conference on Machine Learning.* 2004.

[11] B.C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," *IEEE Transactions on Automatic Control* 26 (1), 17-32, 1981.

[12] J. Hahn, T.F. Edgar, W. Marquardt, "Controllability and Observability Covariance Matrices for the Analysis and Order Reduction of Stable Nonlinear Systems," *Journal of Process Control* , 13, 115-127, 2003.

[13] S. Lall, J.E. Marsden, S. Glavaski, "Empirical Model Reduction of Controlled Nonlinear Systems," *14th IFAC World Congress*, Beijing, 1999.

[14] C. Sun, J. Hahn, "Reduction of Differential-Algebraic Equation Systems Via Projections and System Identification," *Journal of Process Control*, 15, 639-650, 2005.

[15] C.L. Sun, J. Hahn, "Parameter Reduction for Stable Dynamical Systems Based on Hankel Singular Values and Sensitivity Analysis," *Chem. Eng.Sci.*, 61(16), pp. 5393–5403, 2006.

[16] H. Castel-Taleb, L. Mokdad, "Performance Measure Bounds in Mobile Networks by State Space Reduction," *Proceedings of the 13th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems* (MASCOTS'05), 2005.

[17] A. Lakhina, K. Papagiannaki, M Crovela, C. Diot, Eric D. Kolaczyk, N. Taft, "Structural Analysis of Network Traffic Flows," *SIGMETRICS/Performance.* New York, NY, USA, June 12–16, 2004.

[18] Y. Nasr Harandi, M. Yaghoubi Waskasi, M. Pirhadi, M. Mirzabaghi, A. Iravani Tabrizipoor, " A Test Methodology for Testing Next Generation Broadband IP Access Services," *International Conference on Advanced Communication Technology* (ICACT 2007), Seoul, Korea, Feb. 2007.

[19] S.M.R. Sadri, M.Pirhadi, Y. N. Harandi, M.Y. Waskasi, A.I. Tabrizipoor, M. Mirzabaghi, " Test Strategy For DSL Broadband IP Access Services," *Highcapacity Optical Networks and Enabling Technologies* (HONET 2007), Dubai UAE, Nov. 2007

[20] D. Lee, H. Seung, "Algorithms for Non-Negative Matrix Factorization," *Adv. Neural Inform. Process. Systems 13*, pp. 556-562, 2001.

Trust Enhanced Authorization for Distributed Systems

Priyanka Dadhich, Dr. Kamlesh Dutta, Dr. M.C.Govil

Abstract— The trust –management approach to distributed system security is developed as an answer to the inadequacy of traditional authorization mechanism. The subjective concept of trust not only enables users to better understand the paradigm of pervasive computing, but also opens new direction of research for solving existing problems such as security [8], management of online communities or e-services lifecycle .This paper specifies research issues in the areas of authorization and trust in distributed environments involving mobile networks, pervasive and ubiquitous computing networks . We here discuss the notion of trusted computing and examine existing authorization mechanisms and their inadequacies. Next we define a logic program based languages and policies that facilitate the modeling process.To the end various approaches to trust enhanced security for overall authorization security in distributed systems are discussed.

Index Terms— Distributed, trust management, trusted computing, trust enhanced security, subjective trust

—————————————— ◆ ——————————————

1 INTRODUCTION

TRUST based security models have shown the potential to overcome the drawbacks of traditional security models by ensuring a higher level of trustworthiness of authorized entities and thus raising the security levels.

The paper lays emphasis on the design and do management of authorization policies for distributed applications and introduces the notion of trust enhanced authorization to improve security decision making. Section two confers design principles and architectural frameworks for distributed authorization. Section three examine existing authorization mechanisms with their inadequencies. Section four explains the trust –management approach as an answer to inadequacy of authorization mechanisms by exploring some trust management inference engines . Section five discusses authorization policies and languages for trust modeling. Section six concludes and section seven puts forward some future prospects to enhance authorization using trusted platforms in distributed applications.

Authorization in distributed system is called distributed authorization. It is much richer than authentication both in terms of types of privileges required, its nature and its degree of interaction between participating entities. In earlier times, considerableefforts have been spent on formalizing security protocols and access control schemes for general distributed systems that include aut-

----

- *F.A. Author is Research Scholar with the Department of Computer Science & Engineering ,National Institute of Technology, Hamirpur (India). E-mail: prynkmshr@ gmail.com.*
- *S.B. Author is Associate Professor with the Department of Computer Science & Engineering, National Institute of Technology,   Hamirpur(India) E-mail: kd@nitham.ac.in*
- *T.C. Author is professor with the Department of Coputer Engineering, Malviya National Institute of Technology,Jaipur and presently working as the Principal Govt. Women Engineering College ,Ajmer (India) E-mail:govilmc@yahoo.com*

hentication logic and access control calculus by Abadi et al [36,4], a logic language for authorization specifications proposed by Jajodia et al [10], an access control policy description language proposed by Kurlowski [8] and Levier et al [6]. But these models combining authorization and authentication did not approach trust directly but rather deal with trust in an indirect way for identifying security flaws in the existing security protocols.

2 DESIGN PRINCIPLES FOR DISTRIBUTED AUTHORIZATION(DA)

1. Designing of DA can only be accomplished by designing appropriate authorization attributes.
2. Designing should involve authorize information in the security service. Here security mechanisms are required to support these security service and the authorities involved in the management of the service[13].
3. Designing involves the locations where authorization checks can be made . These are:
   a. **CoarseLevel Check:**These determine whether access to the application is allowed or not.
   b. **Function Access check:** It is made on the type of function or operation being requested.
4. Designing of Distributed Authorization Service basically involves design of two distinct stages:
   a. **Administration Design Phase:** Involves design of facilities and services for the specification of authorization policies[15], updating and deleting of policies and their administration.
   b. **Runtime or Evaluation Phase:**It is concerned with the design of the use of these authorization policies in the evaluation of the access requests .

Authorization Architecture Frameworks

- Authorization Architecture (AA) should involve to locate the static and generic authorization information i.e. responsible for a collection of clients and server principals[11].
- Frameworks should involve the dynamic and specific authorization information to be located near the target enabling the target system authorities to be involved in their management.
- These specific and dynamic authorization information needs to be " pulled" at the time of the decision process.
- Authorization frameworks consists of components like administration component where the authorization policies[6] are entered and stored in one representation and a runtime evaluation component that stores the authorization rules at a different representation for runtime access[40].

3 INADEQUACIES WITH SECURITY MECHANISMS

One security mechanisms used in Operating System is the ACL(Access Control Lists). This ACL is a list describing which access rights a principal has on an object(resource). Foe eg: UNIX file system "permissions" mechanisms is essentially an ACL. But unfortunately these ACL 's are inadequate for distributed systems(DS). These are:

1. **Authentication :** In DS, identity of principal is not known but known in OS. Since authentication is accomplished by username/ password mechanisms so this simple password-based protocols are inadequate in network computing environments[4]. Here simple eavesdropping can destroy security.
2. **Delegation:** Delegation enables decentralization of administrative tasks. It is needful for scalability of DS. In DS, policy(or authorization)[15] are specified at the last step in the delegation chain(the entity enforcing policy) in form of an ACL. This leads to inconsistencies among locally specified sub-policies[45] .
3. **Expressibility and Extensibility:** ACL approach do not provide sufficient expressibility or extendibility[24] . Hence may security policy elements that are not directly expressible in ACL form should have to be hard-coded into applications. Hence whenever there is change in security policy it often requires reconfiguration, rebuilding and rewriting of applications.
4. **Local Trust Policy:** Since the number of administrative entities in a DS are very large so each entity is given a different (local) trust model to be used by different users and by other entities.

For example: System A may trust System B to authenticate its users correctly but system A do not trust system C but system B trust system C.

All above security mechanisms are insecure, inadequate and non-scalable authentication mechanisms that are used in conjunction with ACLs. All these unintuitiveness and problematic mechanisms are in use because of the lack of alternatives that suit to DS.

4 TRUST MANAGEMENT

The term 'trust management' was first introduced by Blaze et al [5] { role of trust management in security} . It is a unified approach specifying and interpreting security policies, credentials and relationships that allows direct authorization of security critical actions. These trust-management approach developed as an answer to the inadequacy of previous authorization mechanisms.

Trust Management system combines the notion of specifying security policy with the mechanisms for specifying security credentials. Credentials describe specific delegation of trust among public keys that bind keys to names, to perform specific tasks. These system supports delegation, policy specification, refinement at the different layers of a policy hierarchy. So, the system solves the consistency and scalability problems present in ACLs. Role of various components in Trust Management Architecture are:

1. **Trust Manager:** key component of proposed architecture that provides trust management services.
2. **Trust Inference Engine**: built on subjective logic primitives[30] .
3. **Trust Policy Base**: contains established trust relationships.
4. **Trust Update**: dynamically update the trust relationships in the trust base.
5. **Trust Decision**: provide trust decision from an owner host to requesting entities by preparing an itinerant computation.Trust decisions come from a set of trust based on initial set of trust relationships, recommended trust from others and observations of trust related actions over time[10].

Recommendation Protocol:These protocols are initiated by trust manager in the event of seeking trust information from its trusted entities about other unknown hosts[15]. This protocol helps to maintains a list of hosts (in its trust base) that are trusted for making recommendations.Recommendation is simply an attempt at communicating a party's reputation from one community context to another[20]. A poor recommendation may be detriment-mental to one's reputation and there is no separate term for "negative recommendation".
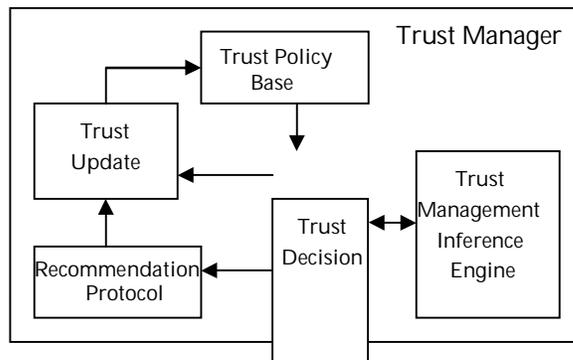
Fig.1 Trust Management Architecture

4.1 Trust Management Inference Engine(TMIE)

It is a separate system component that takes input, outputs a decision about whether compliance with policy is proven or not and if not proven then outputs some additional information detailing how to proceed.

TMIE avoid the need to resolve "identities" in an authorization decision. These engines express privileges and restrictions in a programming language allowing for increased flexibility and expressibility and standardization of modern, scalable security mechanisms.

4.2 Designing of Trust Management Inference Engine

- Design should lay principals for defining proof of compliance.
- There should be some language or notations to express the policies and credentials.

4.3 Tools designed to avoid inadequacies in Distributed Authorization

1. **PolicyMaker:** It was the first tool for processing signed request embodying the trust management. It addresses the authorization problem directly rather than handling the problem indirectly by authentication or access control. Credentials and policies of PoylicyMaker are fully programmable and so called "assertions".

    PolicyMaker is a trust management application that specifies what a public key is authorized to do[22]. PolicyMaker system is essentially a querry engine tool that can either built into applications or run as a daemon service.

2. **KeyNote:** KeyNote[4] [10]. It has same design principals as Policymaker. Keynote uses credentials that directly authorize actions instead of dividing the authorization task into authentication and access control as in PolicyMaker.

    In KeyNote, standardization and ease of integration is developed to give applications. So, KeyNote assign more responsibility to Trust management engine and less function to calling application. By fixing a specific and appropriate assertion language, KeyNote goes further than PolicyMaker toward facilitating efficiency, interoperability and widespread use of written credentials and policies.

3. **REFEREE:** This system fully supports programmability of assertions(i.e. policies and credentials) just like PolicyMaker. REFREE execution environment allows assertion programs to call each other as subroutines and to pass different arguments to different subroutines[18]. While PolicyMaker execution environment requires that each assertion program write anything that it wants to communicate , on a global blackboard i.e. seen by all other assertions. Refree system supports a more complicated form of inter-assertion communication than PolicyMaker.

4.4 Application Areas of Trust Management System

- In active networks
- In Mobile Code security
- In Access control Distributions

5 AUTHORIZATION POLICIES AND POLICY LANGUAGES

A fundamental objective of any authorization system is to enable , to represent and to evaluate a range of access policies that are relevant and required. These policies capture the authorization requirements of the distributed applications.

Policy languages are useful in separating out the policy representation from policy enforcement. Some languages given in [8,10,11]{authorization and trust enhanced security for DA}are solely dedicated for specifying authorization policies. Languages discussed in [6,7] are mathematical logic based, some are graph based and some languages[8,9,10] are programming based. A standard policy language is useful for interoperability between different systems and applications.

Policy language's such as [18, 19, 20]{trust magmt survey} make it possible to automatically determine whether certain credentials are sufficient for performing certain actions or not to authorize the trustee. One trust management framework called Sultan trust management include a language for describing trust and recommendation relationships in the system. Constraints can easily be attached to these relationships and through them , the relationships can be connected to the Ponder policylanguage[22]{trust mangmt survey}.

Sufficient flexible policy system provide the backbone for a trust management system. Tonti et al[21] compare three languages for policy representation and reasoning[23]{. KAoS[24, 25] , Rei[21] and Ponder [16]are used as the basic languages for sketching some general properties desirable in future work on policy semantics.

5.1 Features of Policy languages

- Policy Language should deal with expression and do structuring of complex and dynamic relationships.
- Languages should be simple enough to enable the administrators and policy setters to use the language in specifying their policies.
- Language should have significant expressive and analytical power to represent and evaluate a range of policies used in practical systems.

5.2 Advantages of Policy Languages

- Use of these languages helps the administrators to save time and money because they are not required to rewrite their policies in many different programming languages.
- Developers are not require to invent new policy languges and write code to support them, so time is saved for developers.

- If policy languages are standardized , there are good opportunities for emerging good tools for writing and managing policies for a policy language.

5.3 Authorization Policies

These policies can range from simple identity based to complex dynamic and collaboration policies[24][12]. Some commonly used access policies are:

- Identity based policies
- Group based policies
- Role based policies
- Delegation policies
- Static separation of duty policies
- Dynamic separation of duty and Chinese wall policies
- Joint action policies
- Collaboration access policies

6 APPROACHES TO TRUST ENHANCED SECURITY

Trust enhanced security services require some form of "trusted " authorities to establish and manage "trust" between the mutually suspicious entities A and B over an untrusted network.

For authorization services, we have trust management components, authorization policies and mechanisms. Though the term trust is being used around many decades in different disciplines, but in security the concept of trust came in late 1970's.

With the development of TCSEC(Trusted Computer System Evaluation Criteria)[26] , trust is used in the system's model , design and implementation for its correctness and security. Afterwards, came TCB(Trusted Computer Base)that encapsulates all the security relevant components i.e. both hardware and software that are necessary for enforcing security policies in a system. Trust is the firm belief in the competence of an entity to act dependently, securely and reliably within a specified context[27].

6.1 Trust Notions

Trust provides better understanding of security and privacy problems.

- It acts as centralized control in a system.
- It issues resources to build reputation.
- It performs separation of concern.

Trust records feedback about the security evaluations of other nodes. Trust management enables the trust system to track the behaviour of each node and make corresponding reactions to the tracked behaviours. Trust management can establish a set of effective rules to make a reliable analysis of certain suspicious nodes.

6.2 Concept of Trust Management

Trust Management focuses on designing languages, compliance checkers, identifying applications and building practical toolkits. Beth et al [20] is one of the earliest trust models for authentication in distributed system focusing on relationship modeling whereas Abadi et al [11] {same} presented a modal logic based trust model for modeling distributed authentication and access control. Blaze et al [12] proposed a new well-known trust management system. The common shortcoming of these models is that they did not address the trust based on behavior or past experiences dynamically. Lin and Varadharajan [13] developed trust model for agent security management, but this model did not taken into account security risks that itself trust model has brought. So, all above factors, lead to the research for development of trust enhanced security models for distributed systems.

Later, Kagal et al [23] { a trust based security system for ubiquitous and pervasive computing} presented an architecture based on trust management applicable to distributed system and towards pervasive computing environments. This trust based architecture has a security policy i.e. responsible for assigning credentials to entities, delegating trust to third parties and reasoning about user's access rights. Ngai and Lyu provide a public key authentication service based on a trust model to monitor malicious and colluding nodes. This model allows mobile nodes in distributed system to monitor and rate each other with an authentication metric. The trust value can be updated in conjunction with public key certification. Zhu et al attempt to establish a secure route from a source node S to a designated node D, and provide an approach to calculate the trust value by applying a delegation graph. The mapping between a delegation edge and an authenticated transition graph is used to compute the trust value based on the transitive property.

6.3 Trust Management Authority

Below architecture is "rule-based" and "event-based" architecture.Here rules are used to define the policy of the trust management authority and categorize events that may occur in transactions. This architecture is adaptable to various domains of service oriented applications.

For provision of security services, trusted authorities such as authorization server and authentication server are involved that provide complete trust. For example: if entities A and B trust the authorization server (AS), this server will perform functions of A and B correctly and honestly. This AS will keep the authorization policies securely , perform authorization checks correctly and ensure that software of AS is free from any malicious software.Trust management system such as[31,32,33,34,35] are designed to support specification, acquisition, revocation, degradation and evolution of trust according to some model. It is the unified approach for specifying and inter interpreting security policies, credentials and relationships that allow direct authorization of security-critical actions[31].

Examples as described above: Some automated trust management systems are: PolicyMaker[23], KeyNote[9], REFEREE[17] being delegated.
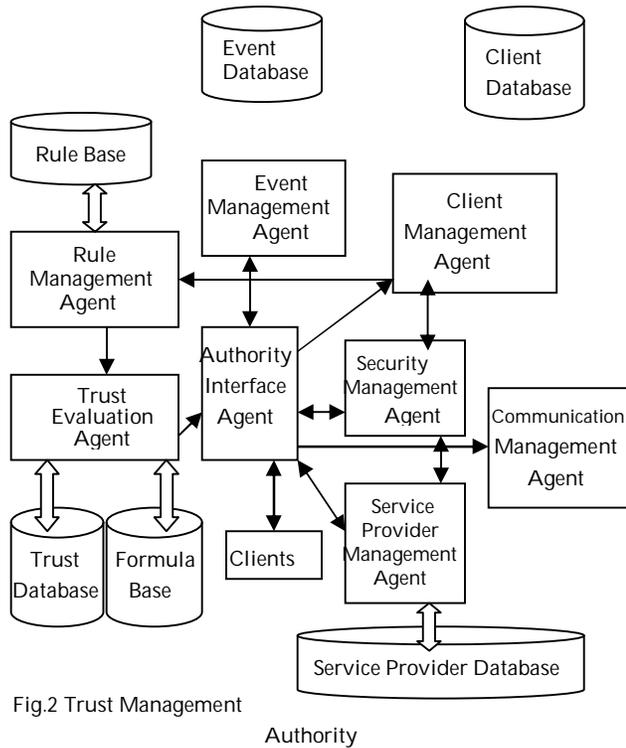
Fig.2 Trust Management

Authority

6.3 Hybrid Trust

It is a composite trust relationship formed by combining hard and soft trust.

1. **Hard Trust:** Denotes the trust relationships that can be derived from the underlying cryptography based security mechanisms such as digital certificates and cryptographic checksums.

These trust relationships indicate one agent host's belief in another in terms of authentication of the relevant host's identity (i.e. authentication trust)and the belief in the benevolence and competence of another host in producing good code (i.e. code trust) and the belief in the honesty and faithful and competent execution of the task requested by a visiting mobile agent i.e. migrating node or program called Execution Trust [3].

Benefits of hard trust Models:

- Enable trust to be extracted from the security mechanism: by extracting trust from security mechanisms, we are able to find actual trust requirements of the underlying security mechanisms that helps us to make more effective security decisions.

- Enable categorization of hard trust related security mechanisms: using the hard trust notion , we can determine a range of hard trust mechanisms that can process and manage the hard trust information and helps to build trust models that can work with security models effectively.

- Enable trust management and its integration with the underlying security mechanisms for enhancements of security performance with help of hard trust models. Trust management systems can be designed that helps to feedback the trust decisions back to the underlying security mechanisms for performance enhancements.

2. **Soft Trust**:Soft trust is based on trust relationships derived from localized and external observations of system entity behavior[1]. These trusts are obtained through social control mechanisms such as direct observations, recommendations or combination of both.

Many trust models [42,43] } are taken into soft trust models. Examples are subjective logic based trust model developed by Josang [30] and classical model of Beth et al [49].

Benefits of Soft Trust Models:

- Social control principles are extensively studied in soft trust models so as to do research and to develop counter measures for malicious behavior in general distributed systems [19,6,18].

- It gives linking between behavior and evidence through mapping .

- By help of trust management operations, these soft trust models can calculate dynamic trust valuations based on the opinion calculus which is used for flexible trust decision making based on the specified thresholds for different trust requirements, in the form of several trust enhanced security protocols[34].

- Through these trust management protocols , the operations of recommendation based trust update and the end of transaction trust update make the distributed trust management possible.

7   CONCLUSION

In this literature , we have addressed some research issues in areas of authorization and trust in distributed environments. Some key design principles, policy language's and mechanisms, are discussed for the development of distributed authorization service. Trust management authority and hybrid trust concepts are explored to outline an idea for enhancing security concerns in distributed systems.

8   FUTURE WORK

With the development of term TCPA[39](Trusted Computing Platform Alliance){authorization and trust enhanced security for distributed application}, currently known as TCG (Trusted Computing Group) lead to the discovery of trusted platform technology comprising of a hardware based subsystem devoted to maintaining trust and security between machines. With the help of the availability of

trusted platform[33] and its characteristics any two enti-ties that want to communicate with each other, has to go through trust determination phase before performing authorization at the beginning of the authorization process.

This above scheme can be extended to transfer autho-rization policies between two authorization server sys-tems in two different domains. We currently need to de-velop such a distributed Authorization service on trusted platforms[39]. Also need to develop an application i.e. showing secure access of its operations using trust en-hanced distributed authorization service[38]. Example of applications can be any military application, network management operations, healthcare applications or any e-commerce applications or any other[2].

REFERENCES

1. Heather, J., Hill, D.,:I'm Not Signing That! In Di-mitrakos, T., MArtinelli, f., eds.: proceedings of the 1st Int'l Wksp on Formal Aspects in security and trust (FAST 2003), Pisa, Itlay(2003)71-81.

2. Ishaya, T., Mundy, D. P.:Trust development and management in virtual communities. In Trust Management : 2nd international conference ,itrust 2004, Oxford 2004.

3. Rindeback, C., Gustavsson, R.,: Why Trust is hard-Challenges in e-mediated services. In: Pro-ceedings of the 7th Int'l wksp on Trust in Agent societies, New York, USA ,2004.

4. Lampson. B., Abadi, M., Burrows, M., Wobber, E.,: Authentication in distributed systems: theory and practice. ACM Trans. On Computer Science 10(4),1992.

5. Lampson. B., Abadi, M., Burrows, M., Plotkin, G.,: A calculus for access control in distributed systems. ACM Trans. On programming languag-es and systems 15(4),1993.

6. Jajodia, S., Samarati, P., Subrahmanian, V.S.: A logic language for expressing authorizations. In: Proc. IEEE Symp. On research in security and privacy, pp.31-42, 1997.

7. Maurer, U.,: Modelling a public key infrastruc-ture. In : Martella, G., Kurth, H., Montolivo, E., Bertino, E.(eds) ESORICS 1996. LNCS, vol 1146 Springer , 1996.

8. Levien, R., Aiken, A.,: Attack –resistant trust metrics for public key certification. In: Proceed-ings of 7th USENIX security Symposium , 1998.

9. M. Blaze, J.Feigenbaum, J. Ioannidis and A. Ke-romytis. The KeyNote Trust-Management .Work in Progress, http:// www.cis.upenn.edu/angelos/keynote.html.

10. M. Blaze, J.Feigenbaum, J.Lacy. Decentralized Trust Management. In Proc. of the 17th Sympo-sium on security and Privacy, pages 164-173. IEEE Computer Society Press, Los Alamitos,1996.

11. V. Varadharajan, C. Crall and J. Pa-to,"Authorization for Enterprise wide Distri-buted Systems" Proceedings of the IEEE Com-puter security Applications Conference, AC-SA'98,1998 USA.

12. M. Hitchens and V. Varadharajan, "ower: A lan-guage for Role Based Access Control" proceed-ings of Int'l Wksp on Policies for Distributed Sys-tems and Networks,UK , 2001pp 88-106.

13. S. Indrakanti, V. Varadharajan , M. Hitchens and R. Kumar,"Secure Authorizations for Web Ser-vices" Proceedings of the 17th IFIP Conference on Data and Applications Security, USA, 2003.

14. S. Jajodia, P. Samarati and V.S. Subrahmanian," A Logical Language for Expressing Authoriza-tions',Proceedings of the IEEE Symposium on Se-curity and Privacy, USA, 1997.

15. Y. Bai and V. Varadharajan ,' A logic for State transformations in Authorization Policies' Pro-ceedings of the IEEE Computer Security Founda-tions Wksp, USA, 1997.

16. N. Damianou, N. Dulay, E. Lupu and M. Slo-man,"the ponder Policy specification Language', proceedings of Int'l Wksp on Policies for Distri-buted systems and networks, UK, 2001, pp 88-106.

17. Chu, Y. H.., Feigenbaum, J., LaMacchia, B., Res-nick , P., Strauss, M.: REFEREE: Trust Manage-ment for web Applications. Computer Networks and ISDN systems 29 (1997) 953-964.

18. Blaze, M., Feigenbaum, J., Keromytis, A. D..: KeyNote: Trust management for public-key in-frastructures(position-paper) In: security proto-cols:6th Int'l Wksp , Cambridge, UK, April 1998. Proceedings. Volume LNCS Springer- Ver-lag(1998) 59-63.

19. T. Grandison and M. Sloman. A survey of trust in internet application. IEEE Communications Surveys, 2000.

20. T. Grandison and M. Sloman. Specifying and analyzing trust for Internet applications. In : pro-ceedings of 2nd IFIP Conference on e-commerce, e-business , e- government 13e2002, Lisbon, Por-tugal 2002.

21. Tonti, G., Bradshaw, J. M., Jeffers, R., Montanari, R., Suri, N., Uszok, A.: Semantic web languages for policy representation and reasoning: A com-parison of KAoS, Rei and Ponder. In: The Seman-tic Web – ISWC 2003. Vol LNCS 2870/2003.419-437.

22. Damianou , N., Dulay, N., Lupu, E., Sloman, M.: The Ponder policy specification language. In: Wksp on Policies for Distributed System and Networks HP Labs Bristol 29-31 Jan 2001. Vol 1995, 2001.

23. Kagal, L., Finin, T., Joshi, A., : a policy language for a pervasive computing environment. In pro-ceedings of tenth Knowledge Acquisition fpr knowledge-based system wksp,1995.

24. Uszok, A., Bradshaw, J. M., Jeffers, R. : KAoS: A Policy and domain services framework for grid computing and semantic web services. In: Trust Management : Second Int'l Conference, itrust 2004, Oxford, UK, March 29-April1, 2004. Proceedings. Volume LNCS 2995/(2004) 16-26.

25. Bradshaw, J. M..: KAoS: An open agent architecture supporting reuse, interoperability and extensibility . In : Proceedings of 10th Knowledge Acquisition for Knowledge-Based Systems Workshop(1995).

26. Dept. of Defense, " trusted Computer System Evaluation Criteria",(TCSEC), DoD5200.28 STD Dec. 198.

27. L. Kagal, T. Finin, A. Joshi. Trust based security in pervasive computing environments, Computer 34(2001) 154-157.

28. H. Zhu, F. Bao, R. H. Deng. Computing of trust in wireless networks, In: proceedings of IEEE 60 th Vehicular technology Conference 2004, pp 2621-2624.

29. T. Grandison and M. Sloman. A survey of trust in internet application. IEEE Communications Surveys, , 2000.

30. A. Josang . A logic for uncertain probabilities . Int'l journal of uncertainty, Fuzziness and knowledge based systems 2001.

31. G. Zacharia and P. Maes. Trust management through reputation mechanisms. Applied Artificial Intelligence,2000.

32. C. Castelfranchi and R. Falcone. Principles of trust for mas: cognitive anatomy , social importance and quantification. In Demazeau, y. (ed) proceedings of the 3rd int'l Conference on Multi-Agent systems, IEEE Computer Society, 1998.

33. C. Lin Trust Enhanced Security for MA, PhD thesis, Macquarie University, August 2006.

34. C. Iin V. VAradharajan, Y. Wang and V. Pruthi. Trust enhanced security for MA. In 7th int'l IEEE conference on e-commerce technology 2005, IEEE Computer Society Press 2005.

35. B. Yu and M. Singh . A Social mechanisms of reputation management in electronic communities. In M. Klusch and L. Kerschberg, editors, CIA-2000 Wksp on Cooperative Information Agents, 1860 of LNAI, Springer, 2000.

36. B. Lampson, M. Abadi, M. Burrows and E. Wobber, Authentication in distributed systems:theory and practice. ACM Transactions on Computer Systems, 1992.

37. TCPA " Trusted Computing Platform Alliance", Building a trust in the PC, jan 2000, http://www.trustedcomputing.org(now known as trusted Computing Grpup, https://www.trustedcomputinggroup.org/home).

38. V. Varadharajan, "trust enhanced authorization and its applications", 2005.

39. B. Balacheff et al.," trusted computing Platforms: TCPA Technology in context", Prentice-Hall, 2003.Ramdomly---

40. M. Burrows, M. Abadi, R. Needham, a Logic of authentication, In: proceedings of the 12 th ACM symposium on Operating Systems Principles, 1989.

41. D. L. Hoffman, T. P. Novak, M. Peralta, Building consumer trust online, Communications of the ACM 1999.

42. Abdul-Rahman , A., Hailes,S.,: A Distributed Trust Model . In Proceedings , ACM New Security paradigmsWorkshop'97, Cumbria, UK 1997.

43. Wagealla, W., Carbone, M., English, C., Terzis, S., Nixon. P.: A formal model on trust lifecycle management. In : wksp on formal Aspects of security and trust (FAST 2003)at FM 2003. VOL IIT TR-10/2003. IIT-CNR, Itlay 2003.

44. Resnick, P., Zeckhauser, R.,Friedman, E., Kuwabara, K.: reputation Systems. Communication of the ACM ,2000.

45. R Yahalom,B Klein and T Beth. Trust relationships in secure systems-a didtributed authentication perspective.Proceedings of IEEE Conference on Research in Security and Privacy,1993.

46. B.Lampson,M. Abadi,M. Burrows, and E.Wobber.Authentication in distributed systems:Theory and practice.ACm Transcations on Compuetr Systems,1992,10(4),pp.265-310.

47. Matt Blaze,Joan Feigenbaum,and Jack Lacy.Decentralised trust management .In Proceedings of the 1996 IEEE conference on security and privacy,Oakland,CA may,1996,pp.164-173

48. C.Lin,V.Varadharajan,"Trust Enhanced Security-A New Philosophy for Secure Collaboration of MobileAgents"Proceedings of the Workshop on Trusted Collaboration,part of Collaborate-Com2006,Atlanta,Georgia,USA.pp.17-20.

49. R. Yahalom, B. Klein ad T. Beth. Trust relationships in secure systems- a distributed authentication prospective. Proceedings of IEEE Conference on research in Security and Privacy, 1993.

Near State PWM Algorithm with Reduced Switching Frequency and Reduced Common Mode Voltage Variations for Vector Controlled Induction Motor Drive

K. Satyanarayana, J. Amarnath, A. Kailasa Rao

Abstract— In this paper a Near State Pulse Width Modulation (NSPWM) algorithm with reduced switching frequency is presented for vector controlled induction motor drives for reduced common mode voltage/currents. The proposed algorithm utilizes a group of three neighbor voltage vectors to construct the reference voltage space vector. In the proposed algorithm in each sector any one of the phases is clamped to either positive dc bus or negative dc bus. Hence, the proposed algorithm reduces the switching frequency and switching losses of the inverter. The simulation results illustrate that the proposed NSPWM algorithm results in reduced common mode voltage, has low switching frequency and has low switching losses of the inverter.

Index Terms—common mode voltage, induction motor drives, near state PWM, SVPWM, vector control.

———————————————— ◆ ————————————————

1 INTRODUCTION

THE vector control methods are widely used for the control of induction motor drives in high-performance applications due to its advantages [1]. The vector control algorithm gives a decoupling control between torque and flux so that the induction motor can be controlled as a separately excited dc motor. However, the classical vector control algorithm uses hysteresis controllers for the generation of gating signals to the voltage source inverter (VSI), which results in variable switching frequency operation. To achieve constant switching frequency operation of the inverter, the space vector pulse width modulation (SVPWM) algorithm [2] has been used for vector controlled induction motor drive. Among the various possible PWM algorithms SVPWM gives superior performance and gives reduced harmonic distortion [3]. The SVPWM algorithm divides the zero voltage vector time equally among the two zero voltage vectors.

To reduce the harmonic distortion, the zero voltage space vectors are used in SVPWM algorithm, which results in large common-mode voltage (CMV) variations. The poor CMV characteristics lead to prohibitive amount of common-mode current (CMC) in induction motors. In induction motor drive applications, this may lead to motor bearing failures, electromagnetic interference (EMI) noise, or interference with other electronic equipment in the vicinity [4]-[5]. Such problems have increased recently due to increasing PWM frequencies and faster switching times. The filters can be utilized to suppress the effect of the CMV from the source. However, these methods involve additional hardware, and thus, they significantly increase the drive cost and complexity. An alternative approach is to modify the pulse pattern of the standard PWM algorithm such that the CMV is substantially reduced from its source and its effects are mitigated at no

cost [9]-[11]. A novel remote state PWM algorithm is presented in [9] and a near state PWM (NSPWM) algorithm is presented in [10]. The detailed survey and analysis of various PWM algorithms has been carried out and proved that NSPWM algorithm gives superior performance when compared with the other PWM algorithms [11].

This paper presents space vector based NSPWM algorithm with reduced switching frequency for reduced CMV in vector controlled induction motor drives.

2 SVPWM ALGORITHM

Voltage source inverters (VSI) are utilized in many applications. The three-phase, two-level VSI has a simple structure and generates a low-frequency output voltage with controllable amplitude and frequency by programming high-frequency gating pulses. For a 3-phase, two-level VSI, there are eight possible voltage vectors, which can be represented as shown in Fig. 1. Among these voltage vectors, V_1 to V_6 vectors are known as active voltage vectors or active states and the remaining two vectors are known as zero states or zero voltage vectors. The reference voltage space vector or sample, which is as shown in Fig. 1 represents the corresponding to the desired value of the fundamental components for the output phase voltages. In the space vector approach this can be constructed in an average sense. The reference voltage vector (V_{ref}) is sampled at equal intervals of time, T_s referred to as sampling time period. Different voltage vectors that can be produced by the inverter are applied over different time durations with in a sampling time period such that the average vector produced over the sampling time period is

equal to the sampled value of the V_{ref}, both in terms of magnitude and angle. It has been established that the vec-
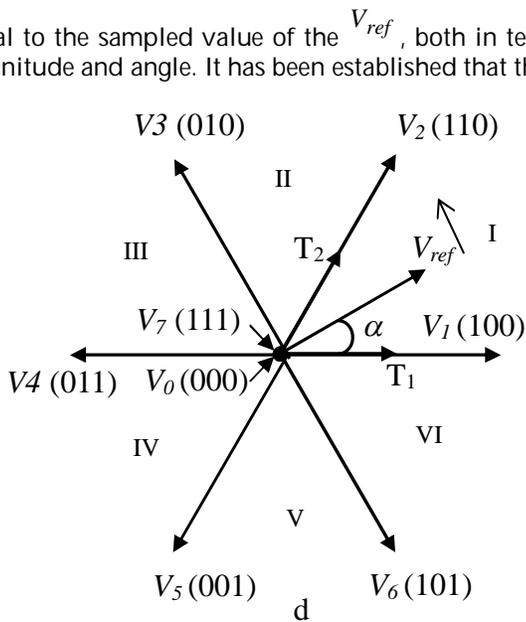
Fig. 1 Possible voltage space vectors and sector definition in SVPWM algorithm

tors to be used to generate any sample are the zero voltage vectors and the two active voltage vectors forming the boundary of the sector in which the sample lies. As all six sectors are symmetrical, the discussion is limited to the first sector only.

For the required reference voltage vector, the active and zero voltage vectors times can be calculated as in (1), (2) and (3).

$$T_1 = \frac{2\sqrt{3}}{\pi} M_i \sin(60^o - \alpha)T_s \qquad (1)$$

$$T_2 = \frac{2\sqrt{3}}{\pi} M_i \sin(\alpha)T_s \qquad (2)$$

$$T_z = T_s - T_1 - T_2 \qquad (3)$$

where M_i is the modulation index and defined as in [1]. In the SVPWM algorithm, the total zero voltage vector time is equally divided between V_0 and V_7 and distributed symmetrically at the start and end of the each sampling time period. Thus, SVPWM uses 0127-7210 in sector-I, 0327-7230 in sector-II and so on.

3 NSPWM ALGORITHM

The near state PWM (NSPWM) algorithm uses a group of three neighbor voltage vectors to construct the reference voltage vector. In order to reduce the common mode voltage variations, the proposed NSPWM algorithm did not use the zero voltage vectors. These three voltage vectors are selected such that the voltage vector closest to reference voltage vector and its two neighbors are utilized in each sector. Hence, the utilized voltage vectors are changed in every sector. As shown in Fig. 2, to apply the method, the voltage vector space is divided into six sectors. Here also, as all six sectors

are symmetrical, the discussion is limited to the first sector only. For the required reference voltage vector, the active voltage vectors (V_1, V_2 and V_6) times can be calculated as in (4), (5) and (6).

$$T_1 = \left\{ -1 + \frac{3}{\pi} M_i \cos(\alpha + \frac{\pi}{3}) + \frac{3\sqrt{3}}{\pi} M_i \sin(\alpha + \frac{\pi}{3}) \right\} T_s \qquad (4)$$

$$T_2 = \left\{ 1 - \frac{3}{\pi} M_i \cos(\alpha + \frac{\pi}{3}) - \frac{\sqrt{3}}{\pi} M_i \sin(\alpha + \frac{\pi}{3}) \right\} T_s \qquad (5)$$

$$T_6 = T_s - T_1 - T_2 \qquad (6)$$
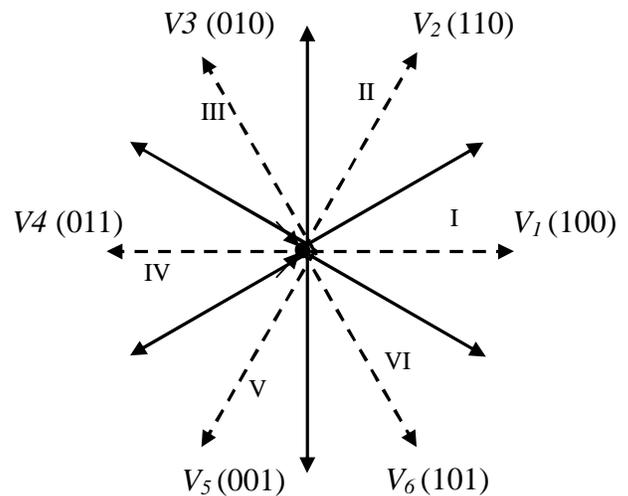
Fig. 2 Possible voltage space vectors and sector definition in NSPWM algorithm

But, in the NSPWM algorithm, the (4), (5) and (6) have a valid solution when the modulation index is varying between 0.61 and 0.906 [10]. In order to get minimum switching frequency and reduced common mode voltage the NSPWM algorithm uses 216-612 in sector-I, 321-123 in sector-II and so on.

The total number of commutations in SVPWM algorithm is three in a sampling time interval, where as the number of commutations in NSPWM algorithm is two. Moreover, the modulating waveform of NSPWM algorithm is similar to the DPWM1 waveform and hence any one of the phases is clamped to the positive or negative DC bus for utmost a total of 120^0 over a fundamental cycle. Hence, the switching losses of the associated inverter leg are eliminated. Hence, the switching frequency of the NSPWM algorithms is reduced by 33% compared with SVPWM algorithm.

4 NSPWM ALGORITHM BASED VECTOR CONTROLLED INDUCTION MOTOR DRIVE

In the vector controlled induction motor drive, a VSI is

supposed to drive the induction motor so that the slip fre-quency can be changed according to the particular require-ment. Assuming the rotor speed is measured, and then the slip speed is derived in the feed-forward manner. The block diagram of proposed vector controlled induction motor drive is as shown in Fig. 3. This shows how the rotor flux linkage position can be obtained by integrating the sum of rotor speed and actual speed. In the vector control scheme, to re-gulate λ_r and rotor speed to desired values are the two ob-jectives.

ate the desired rotor flux linkage and rotor speed are not directly related to these variables. So the alternative way is to regulate the rotor flux linkage and rotor speed through PI controllers and the outputs of these two controllers give out the reference values for the q- and d-axis stator currents in synchronous reference frame. Then the actual q- and d-axis stator currents are regulated to these two reference currents to get the stator voltages. Then these two-phase voltages are converted into three-phase voltages and given to the RPWM block, which generates the gating pulses to the VSI.
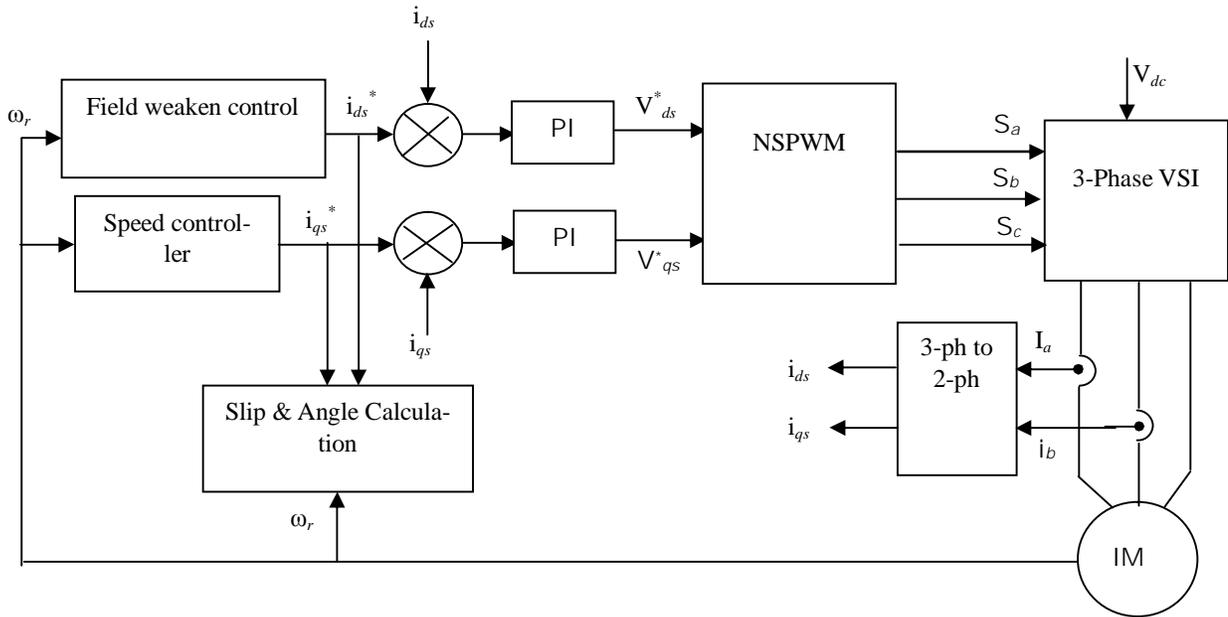
Fig. 3 block diagram of NSPWM based vector controlled I.M.drive

Apparently the stator voltages that are required to gener-

5 SIMULATION RESULTS AND DISCUSSIONS

To validate the proposed NSPWM algorithm, the numerical simulation studies have been carried out using MATLAB. For the simulation studies, the average switching frequency of the inverter is taken as 5 kHz. In order to maintain the constant average switching frequency, the switching fre-quency of SVPWM algorithm is taken as 5 kHz and that of NSPWM algorithm is 7.5 kHz. The induction motor used in this case study is a 4 kW, 400V, 1470 rpm, 4-pole, 50 Hz, 3-phase induction motor having the following parameters: R_s= 1.57Ω, R_r = 1.21Ω, L_s = 0.17H, L_r = 0.17H, L_m = 0.165 H and J = 0.089 Kg.m^2.

The simulation results of SVPWM algorithm based vector controlled induction motor drive are shown in Fig.4-Fig. 7.
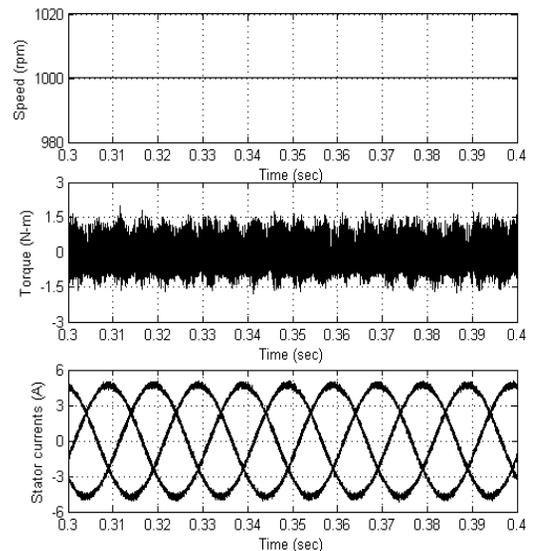
Fig. 4 Steady state plots of SVPWM algorithm based vector con-trolled induction motor drive
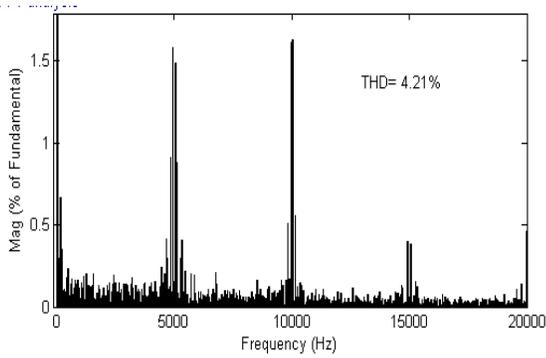
Fig. 5 Harmonic spectra of line current in SVPWM based vector controlled induction motor drive
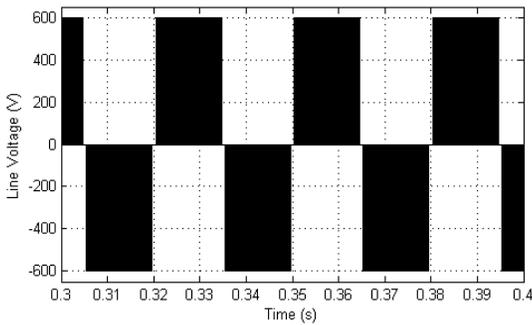
Fig. 6 Line voltage waveform of SVPWM algorithm based vector controlled induction motor drive
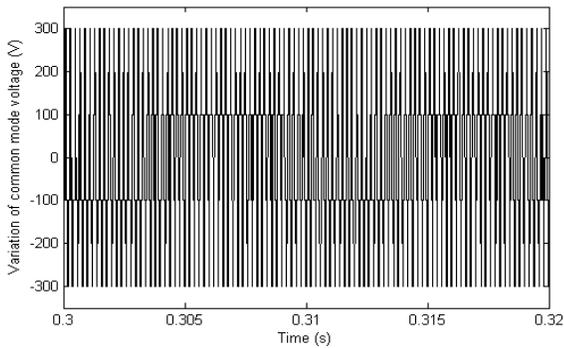
Fig. 7 Common mode voltage variations in SVPWM algorithm based vector controlled induction motor drive

Fig.4 shows the steady state plots of vector controlled induction motor drive. Fig. 5 shows the harmonic spectra of line current along wth the THD value.Fig. 6 shows the line voltage waveform of SVPWM algorithm based vector controlled induction motor drive in steady state and Fig. 7 shows the common mode voltage variations with the SVPWM algorithm, from which it can be observed that the common mode voltage is varying between $+0.5V_{dc}$ and $-0.5V_{dc}$. the simulation results of NSPWM algorithm based drive are shown from Fig. 8 to Fig. 11.
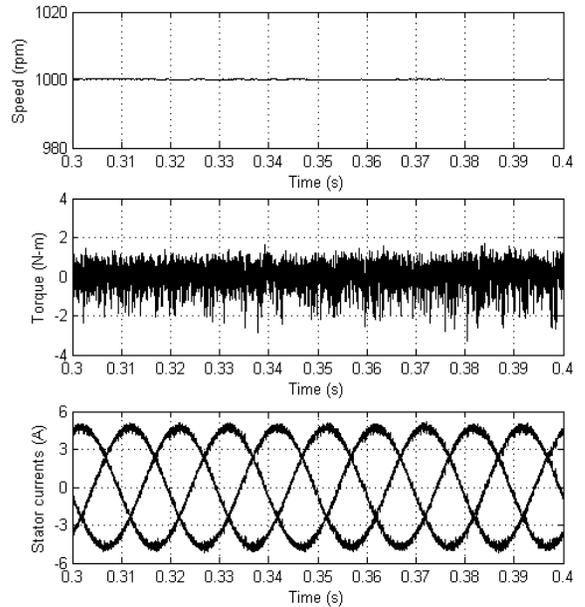
Fig. 8 Steady state plots of NSPWM algorithm based vector controlled induction motor drive
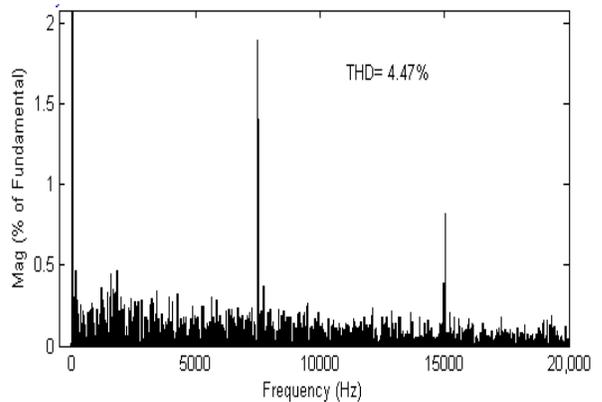
Fig. 9 Harmonic spectra of line current in NSPWM algorithm based vector controlled induction motor drive
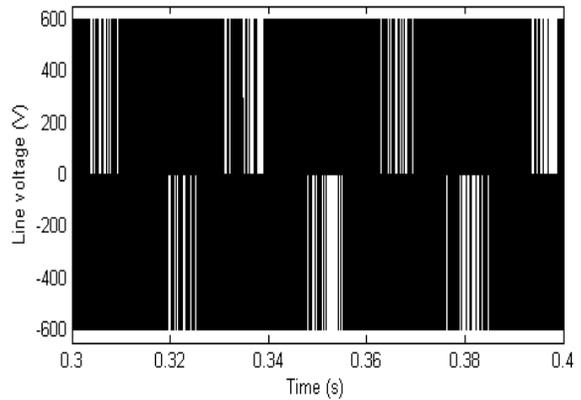
Fig. 10 Line voltagewaveform of NSPWM algorithm based vector controlled induction motor drive
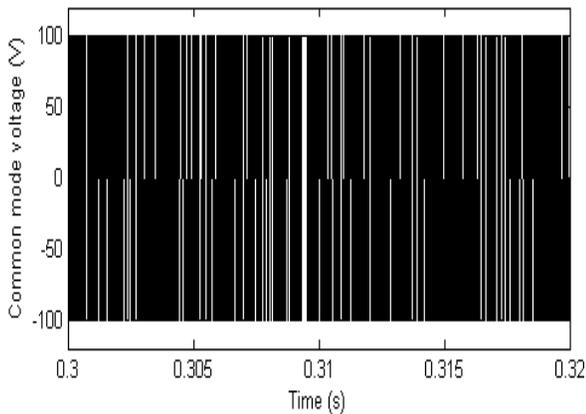
Fig. 11 Common mode voltage variations in NSPWM algorithm based vector controlled induction motor

From Fig. 5 and Fig. 9, it can be observed that the NSPWM algorithm gives more THD when compared with the SVPWm algorithm. Moreover, from Fig. 7 and Fig. 11, it can be observed that the NSPWM algorithm gives less common mode voltage variations when compared with the SVPWM algorithm.

Thus, from the simulation results, it can be observed that the proposed NSPWM algorithm reduces common mode voltage when compared with the SVPWM algorithm with slight increase in harmonic distortion. Moreover, from the line voltage waveforms, it can be observed that there are opposite pulses in line voltage of NSPWM algorithm. Also, as the proposed algorithm clams any one of the phases for a total period of 120 degrees over a fundamental cycle, it reduces the switching losses of the inverter and also the switching frequency of the inverter is 2/3 times to that of the SVPWM algorithm.

6 CONCLUSIONS

In this paper, a NSPWM algorithm is presented for vector controlled induction motor drives. to validate the proposed algorithm, simulation studies have been carried out and results are presented. From the results, it can be observed that the proposed NSPWM algorithm gives less common mode voltage variations when compared with the SVPWM algorithm with slight increase in harmonic distortion. Also, as the proposed NSPWM algorithm is a bus-clamping sequence, it reduces the switching losses by 33.33%. Hence, the switching frequency of the proposed NSPWM algorihm is also less when compared with the SVPWM algorithm.

REFERENCES

[1]   F. Blaschke "The principle of field orientation as applied to the new transvector closed loop control system for rotating-field machines," *Siemens Review*, 1972, pp 217-220.

[2]   Heinz Willi Vander Broeck, Hnas-Christoph Skudelny and Georg Viktor Stanke, "Analysis and realization of a pulsewidth modulator based on voltage space vectors" *IEEE Trans. Ind. Applicat.*, vol. 24, no. 1, Jan/Feb 1988, pp. 142-150.

[3]   Joachim Holtz, "Pulsewidth modulation – A survey" *IEEE Trans. Ind. Electron..*, vol. 39, no. 5, Dec 1992, pp. 410-420.

[4]   J.M. Erdman, R. J. Kerkman, D.W. Schlegel, and G. L. Skibinski, "Effect ofPWMinverters onACmotor bearing currents and shaft voltages," *IEEE Trans. Ind. Appl.*, vol. 32, no. 2, pp. 250–259, Mar./Apr. 1996.

[5]   G. L. Skibinski, R. J. Kerkman, and D. Schlegel, "EMI emissions of modern PWM AC drives," *IEEE Ind. Appl. Soc. Mag.*, vol. 5, no. 6, pp. 47–81, Nov./Dec. 1999.

[6]   Y. S. Lai and F. S. Shyu, "Optimal common-mode voltage reduction PWM technique for inverter control with consideration of the dead-time effects—Part I: Basic development," *IEEE Trans. Ind. Appl.*, vol. 40, no. 6, pp. 1605–1612, Nov./Dec. 2004.

[7]   Y. S. Lai, P. S. Chen,H.K. Lee, and J.Chou, "Optimal common-mode voltage reduction PWM technique for inverter control with consideration of the dead-time effects—Part II: Applications to IM drives with diode front end," *IEEE Trans. Ind. Appl.*, vol. 40, no. 6, pp. 1613–1620, Nov./Dec. 2004.

[8]   J. Zitzelberger andW. Hofmann, "Reduction of bearing currents in inverter fed drive applications by using sequentially positioned pulse modulation," *EPE J.*, vol. 14, no. 4, pp. 19–25, 2004.

[9]   M. Cacciato, A. Consoli, G. Scarcella, and A. Testa, "Reduction of common mode currents in PWM inverter motor drives" *IEEE Trans. Ind. Applic.* Vol.35, no.2, pp. 469-476, Mar/Apr, 1999.

[10] Emre Ün and A.M. Hava, "A near state PWM method with reduced switching frequency and reduced common mode voltage for three phase voltage source inverters" *IEEE*-IEMDC, pp. 235-240, May, 2007.

[11] A.M. Hava, and Emre Ün, "Performance Analysis of Reduced Common-Mode Voltage PWM Methods and Comparison with Standard PWM Methods for Three-Phase Voltage-Source Inverters" IEEE Trans. Power Electron., vol.24, no.1, pp. 241-252, Jan, 2009.

Teacher's awareness about the availability and use of technology for Visually Impaired: A Study

Prof. Madhuri Isave, Dr. Megha Uplane, Prof. Suresh Isave

Abstract -   This study focuses on, what is the use of technology for VI? What different types of technology available for VI? And how it's useful as a learning resource for VI? People with VI are gaining access to technology and assistive device designed to minimize the effect of their disabilities. Technological advances open up a new world for people with severe visual impairment. These advances give there greater participation and independence in all aspect of modern society. We have noted the technology affects many aspect of life for person with VI. This system is using different mean of input that is visual, auditory and tactile. These devices improve independent learning opportunity and enhanced participation in recreational and leisure time activities. Student with VI may need to learn to read and write by using different method. Braille is one of the best sources for VI, but today Braille is less popular. Why? The reason is that the unavailability of teacher who know how to teach Braille and another reason is that the increasing the availability of ICT. Advances of the technology have significant influence on the life of blind and low vision.

Index Terms: - Technology, Visually impaired, Teacher awareness

——————————————— ◆ ———————————————

1. Introduction

Advances in technology have significantly influenced in the blind and low vision individuals. Over the past 20th year improvement in computer has allowed for readily access to VI. Today a large percentage of students with VI spend over 80% of their school days in general educational classroom. Student with very severe visual impairment may need to learn read and write using different methods. Braille is a coded system of dots embossed on paper, so that individuals can feel a page of text. Braille is use for different type of reading such as Maths, and Music. Now seen fewer people are using Braille as a reading method today. First reason is that Braille method is slow. According to Tuttle and Ferrell (1995) reported that good Braille reader achieve a rate of only 100 words per minutes. Nolan (1967) found that average high school students who is blind reads even fewer words per minutes.

   Can you think of some other reason, Why Braille is less popular today? The first reason is that the teachers don't know how to use or teach the Braille and unavailability of the experts. Another reason is increasing availability of audio tape, immediate computerized print to voice translation difficulty of getting Braille version of books. Braille literacy has become focus of a great debate. Advanced technology is a reason for its unpopularity.

VIBUG (Visually Impaired Blind User Group), the Boston Computer Society are exchanging information to expand computer literacy among person with visual impairment. Gaining access of technology and the assistive device designed to minimize the effect of their disability. These exciting technological advances open up a new world for people with severe visual impairment. Technological aids categorized under three heads.

2.Technological Aids

CCTV - It can be used to enlarge the print found in printed texts and books.

Microcomputer – Using special word processing program can produce large print display that allow person with low vision.

Kurzweil Reader – One of the first computerized systems designed for people with visual impairments that translate print into synthesized speech.

Audiodescription – A technique in which trained narrators describe visual and nonverbal information during the pause in the audio or scripted dialogue of plays, films and TV shows by using FM transmission or extra sound track available on stereo TV.

Talking Books – A books available in auditory format.

Braille – A system of reading and writing that uses dot codes that are embossed on paper, developed by Louis Braille in 1929.

Perkins Brailler – It is a compact and portable machine that uses keys that, when pressed down, emboss special paper with the Braille code.

Braille Printer – A special designed Braille printer is attached to a micro computer, standard text can be translated into Braille, allowing teacher who does not know how to use Braille to produce Braille copies of handouts, tests, maps, charts and other class materials

Community Radio – It is a recent development in the technology.(Sakal Newspaper Published news on 23 March 2010) conducted exam of VI students with the help of community Radio. It is a great contribution of Vidyavani section – Advanced Educational lab for Blind of University of Pune, in the area of special education and especially for VI

3. Use of technology for VI

- Students with very severe visually impaired may need to learn read and write.
- Immediate computerized print to voice and voice to print translation of document.
- Many low vision students they can read a specially adapted version of the text.
- Greater and easier access to classroom material for student with severe visual impairment
- Benefits from improving their listening skills.
- Independent learning is possible.
- Increase the confidence level and minimize the effect of disability.
- Enhanced participation in recreational and leisure activities.

Availability of technology and assistive educational devices is important for VI person is no doubt, but now the question is raised in mind that, what is the duty of teacher? How they know about technology? Are they interested to get knowledge about technology? Will they are aware about technology or not?

To find out this curious view point the researcher took following study:

4. Method

The purpose of the present study to explores the teacher awareness about the technology and its use for VI. The main objective of the study was to find out the technology awareness among teacher of VI. This is a descriptive research and a survey method was used. All teachers who teach in special school and integrated school set up in Pune district. 23 teachers who teach at primary and secondary level from special school and integrated school were selected. A purposive sample method was used. To understand the awareness about technology, the researcher constructed the questionnaire. The researcher used open and close ended question to get maximum information from teacher.

5. Result and discussion

Table -1

Technology available in school

	Braille Printer	Screen reader	Graphical Embosser	Instant book reader	Karzweil Reader	CCTV TTI	Others
Responses	Yes	Yes	Yes	No	No	Yes	----

Table 1 revealed that Braille Printer, Screen reader, CCTV TTI, Graphical Embosser are available in school but Karzweil Reader, Instant book reader are not available in school. Out of these technological aids, no others aids are available in school. Teachers have not responded the others.

Table -2

Knowledge about Technology

Response →	Braille Printer	Screen Reader	Graphical Embosser	Instant book reader	Karzweil Reader	CCTV TTI
Yes	21	18	13	6	9	14
Percent	91%	78%	57%	26%	39%	61%
No	2	5	10	17	14	9
Percent	9%	22%	43%	74%	61%	39%

Table 2 shows that the teachers have very well knowledge about technology. As compare to others aids teachers have minimum knowledge about Instant book reader and Karzweil Reader because of unavailability of aids in school

Table - 3

Technology use in teaching

Response ↓	Braille Printer	Screen reader	Graphical Embosser	Instant book reader	Karzweil Reader	CCTV TTI
Yes	11	07	4	00	00	2
Percent	48%	30%	17%	00%	00%	9%
No	12	16	19	23	23	21
Percent	52%	70%	83%	100%	!00%	91%

Table 3 -indicates that a few teachers are use technology in their teaching. Only 9% teachers use CCTV for VI because the electricity problem and inconvenience about use.

Table -4

Provide training program about Technology

Response	Frequency	Percent
Yes	5	22%
No	18%	78%
Total	23	100%

Table 4 highlighted very high percent (78%) teachers said that no any training program arrange about the technology for them.

Table - 5

Need of training Program

Response	Frequency	Percent
Yes	23	100%
No	00	00%
Total	23	100

Table -5 indicates that (100%) all teacher have need a training program about technology and how its use. The program should organize for VI students.

Table -6

Need of training Program from which level

Response	Frequency	Percent
primary	22	96%
Secondary	1	4%
Higher secondary	00	00%
Total	23	100

Table -6 indicates that (96%) teachers are said that the training program should be arranged from the primary level.

6 Responses of the Teachers

1. Technology is really good and it is very useful for person who can not see.
2. Because of technology, student can learn independently without any supporter
3. For the easy interaction with normal children
4. Increase their reading, and listening skills Technology is really good and it is very useful for person who can not see.
5. Because of technology, student can learn independently without any supporter.
6. For the easy interaction with normal children
7. Increase their reading, and listening skills
8. Technology is not available in school
9. It is available but we do not have knowledge about technology.
10. Financially it is not affordable for school.

7. Conclusions

- Technology is available in school. Karzweil Reader and instant book reader is not available in school. Most of the teachers have the knowledge about technology.
- In case of the awareness about technology the study found that the teachers are better aware about the technology.
- All (100%) teachers wanted a special training program about technology, through which they will get sufficient knowledge

Referances

1. Husen,T. " The International Encyclopedia of Education, Research &Studies," Vol.IX, Pergaman, Oxford NewYork,1985
2. Gay, M. "Educational Research,"8th .edi. Pearson Merill Prentice hall ,2006
3. Gray,D. " Doing Research In Real World,"2nd edi. Sage Publication.2009
4. Rajput,J.S. "Encyclopedia of Indian Education,"Vol.I.(A-K),2002
5. Rajput,J.S. "Encyclopedia of Indian Education,"Vol.II.(A-K),2004

Privacy of Data, Preserving in Data Mining

Deepika Saxena

Abstract — Huge volume of detailed personal data is regularly collected and sharing of these data is proved to be beneficial for data mining application. Such data include shopping habits, criminal records,medical history, credit records etc .On one hand such data is an important asset to business organization and governments for decision making by analyzing it .On the other hand privacy regulations and other privacy concerns may prevent data owners from sharing information for data analysis. In order to share data while preserving privacy data owner must come up with a solution which achieves the dual goal of privacy preservation as well as accurate clustering result. Trying to give solution for this we implemented vector quantization approach piecewise on the datasets which segmentize each row of datasets and quantization approach is performed on each segment using K means which later are again united to form a transformed data set.Some experimental results are presented which tries to finds the optimum value of segment size and quantization parameter which gives optimum in the tradeoff between clustering utility and data privacy in the input dataset

Keywords—datamining, data privacy, preserving datamining, model, quantization approach, predictive information, ppdp, ppdm.

————————————— ◆ —————————————

1 INTRODUCTION

Data mining is a technique that deals with the extraction of hidden predictive information from large database. It uses sophisticated algorithms for the process of sorting through large amounts of data sets and picking out relevant information. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. With the amount of data doubling each year, more data is gathered and data mining is becoming an increasingly important tool to transform this data into information. Long process of research and product development evolved data mining. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this Evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

☞ Massive data collection

☞ Powerful multiprocessor computers

☞ Data mining algorithms

Scope of data mining

Data mining gets its name from the similarities between finding for important business information in a huge database for example, getting linked products in gigabytes of store scanner data and mining a mountain for a vein of valuable ore. These processes need either shifting through an large amount of material, or intelligently searching it to find exactly where the value resides. Data mining technology can produce new business opportunities by providing these features in databases of sufficient size and quality, automated prediction of trends and behaviors. The process of finding predictive information in large databases is automated by data mining. Questions that required extensive analysis traditionally can now be answered directly from the data quickly with data mining technique. A typical example is targeted marketing. It uses data on past promotional mailings to recognize the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.· Automated discovery of previously unknown patterns. Data mining tools analyze databases and recognize previously hidden patterns in one step. The analysis of retail sales data to recognize seemingly unrelated products that are often purchased together is an example of pattern discovery. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying data that are anomalous that could represent data entry keying errors. Data mining techniques can provide the features of automation on existing software and hardware platforms, and can be implemented on new systems as existing platforms are upgraded and new products developed. On high performance parallel processing systems when data mining tools are used, they can analyze huge databases in minutes. Users can automatically experiment with more models to understand complex data by using faster processing facility. High speed makes it possible for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

Applications of data mining

There is a rapidly growing body of successful applications in a wide range of areas as diverse as:

☞ Analysis of organic compounds, automatic abstracting, credit card fraud detection, financial forecasting, medical diagnosis etc. Some examples of applications (potential or actual) are:

☞A supermarket chain mines its customer transactions data to optimize targeting of high value customers

☞A credit card company can use its data warehouse of customer transactions for fraud detection

☞A major hotel chain can use survey databases to identify attributes of a 'high-value' prospect.

Applications can be divided into four main types:
 1 Classification
 2 Numerical prediction
 3 Association
 4 Clustering.

Data mining using labeled data (specially designated attribute) is called supervised learning. Classification and numerical prediction applications falls in supervised learning. Data mining which uses unlabeled data is termed as unsupervised learning and association and clustering falls in this category.

Data mining and Privacy

Data mining deals with large database which can contain sensitive information. It requires data preparation which can uncover information or patterns which may compromise confidentiality and privacy obligations. Advancement of efficient data mining technique has increased the disclosure risks of sensitive data. A common way for this to occur is through data aggregation. Data aggregation is when the data are accrued, possibly from various sources, and put together so that they can be analyzed. This is not data mining per se, but a result of the preparation of data before and for the purposes of the analysis. The threat to an individual's privacy comes into play when the data, once compiled, cause the data miner, or anyone who has access to the newly compiled data set, to be able to identify specific individuals, especially when originally the data were anonymous.

What data mining causes is social and ethical problem by revealing the data which should require privacy? Providing security to sensitive data against unauthorized access has been a long term goal for the database security research community and for the government statistical agencies. Hence, the security issue has become, recently, a much more important area of research in data mining. Therefore, in recent years, privacy-preserving data mining has been studied extensively. We will further see the research done in privacy area .In chapter 3 general survey of privacy preserving methods used in data mining is presented.

PRIVACY-PRESERVING DATA MINING

The recent work on PPDM has studied novel data mining techniques that do not require accessing sensitive information. The general idea of PPDM is to allow data mining from a modified version of the data that contains no sensitive information.

CENTRALIZED MODEL

In the centralized model, all data are owned by a single data publisher. The key issues are how to modify the data and how to recover the data mining result from the modified data. Answers often depend on data mining operations and algorithms. One common technique is randomization, by introducing random noise and swapping values in the data. The randomized data preserves aggregate properties (such as means and correlations) in the collection of records, but has little use when each record is examined individually. Another common technique is encryption. The data publisher transforms the original data into an encrypted form for data mining at an external party. Since the data mining results are in the encrypted form and since the data publisher is the only one who can decrypt the results, this approach is applicable only if the data publisher himself is the user of data mining results.

DISTRIBUTED MODEL

In the distributed model, multiple data publishers want to conduct a computation based on their private inputs, but no data publisher is willing to disclose its own output to anybody else. The privacy issue here is how to conduct such a computation while preserving the privacy of the inputs. This problem is known as the Secure Multiparty Computation (SMC) problem. The aim of SMC is to enable multiple parties to carry out distributed computing tasks in a secure manner with the assumption that some attackers, who possibly are the participating parties themselves, want to obtain extra information other than the output. SMC has two major requirements, privacy and correctness. The privacy requirement states that parties should learn their output and nothing else during and after the SMC process. The correctness requirement states that each party should receive its correct output without alteration by the attackers. Extensive research has been conducted on secure protocols for data mining tasks including association rule mining, classification analysis, and clustering analysis. Refer to for surveys on this distributed model of PPDM.

COMPARING PPDP AND PPDM

In many real life applications, the data publisher wants to *publish* some data, but has little or no interest in data mining results and algorithms. For example, a hospital may publish the patient data to a drug research institute; although willing to contribute its data to drug research, the hospital is not interested in and has no expertise in data mining algorithms because drug research is not its normal business. This privacy-preserving data publishing (PPDP) scenario differs from PPDM in several major ways. PPDP focuses on the data, not data mining results; therefore, published records should be meaningful when examined individually. This implies that randomization and encryption are inapplicable. PPDP seeks to anonymize the data by hiding the identity of individuals, not hiding sensitive data. The anonymized data is expected to be analyzed by traditional data mining techniques; therefore, no new data mining techniques are needed. We did not intend to dismiss the contribution of the randomization and encryption approaches. They are effective anonymization methods if the data records will not be examined individually.

CONFIDENCE BOUNDING

solution is possible. In particular, we iteratively disclose domain values in a top-down

manner by suppressing all domain values. In each iteration, we disclose the suppressed domain value to maximize some criterion taking into account both information gained and privacy lost. We evaluate this method on real life data sets. Several features make this approach practically useful:

☞ *No taxonomy required.* Suppression replaces a domain value with *?* without requiring

a taxonomy of values. This is a useful feature because most data do not have an

associated taxonomy, though taxonomies may exist in certain specialized domains.

☞ *Preserving the truthfulness of values.* The special value *?* represents the union," a

less precise but truthful representation, of suppressed domain values. This truthfulness

is useful for reasoning and explaining the classification model.

☞ *Subjective notion of privacy.* The data publisher has the exibility to her own notion of privacy using templates for sensitive inferences.

☞ *Excient computation.* It operates on simple but effective data structures to reduce

the need for accessing raw data records.

☞ *Anytime solution.* At any time, the user (the data publisher) can terminate the computation and have a table satisfying the privacy goal.

☞ *Extendibility.* Though we focus on categorical attributes and classification analysis,

this work can be easily extended to continuous attributes and other information utility

criteria.

CONCLUSIONS

Due to the wide use of the Internet and the trends of enterprise integration, one-stop service, simultaneous cooperation and competition, and outsourcing in both public and private sectors, data publishing has become a daily and routine activity of individuals, companies, organizations, government agencies. Privacy-preserving data publishing is a promising approach for data publishing without compromising individual privacy or disclosing sensitive information. In this thesis, we studied different types of linking attacks in the data publishing scenarios of single release, sequential release, and secure data integration . Our contributions can be summarized as follows:

☞ *Preserving Privacy and Information.* We considered the problem of protecting individual privacy while releasing person-specific data for classification modeling. We

chose classification analysis as the information requirement because the data quality

and usefulness can be objectively measured. Our proposed framework can easily adopt

other information requirement with a different selection criterion.

☞ *A Unied Privacy Notion.* We demeaned a new privacy notion, called privacy template in the form of *hX; Y; ki*, that unique anonymity template and confidentiality template.This unified notion is applicable to all data publishing scenarios studied in this thesis.

☞ *A Framework of Anonymization Algorithm.* Despite the data publishing scenarios are

very di®erent, we presented a framework of anonymization algorithm, Top-Down Re-

¯nement (TDR), to iteratively specialize the data from a general state into a special

state, guided by maximizing the information utility and minimizing the privacy speci-

¯city. This top-down approach serves a natural and e±cient structure for handling

categorical and continuous attributes and multiple privacy templates. Experiments

suggested that our TDR framework effectively preserves both information utility and

individual privacy and scales well for large data sets in different data publishing sce-

narios.

☞ *Extended Data Publishing Scenarios.* Most existing works considered the simplest data publishing scenario, that is, a single release from a single publisher. Such mechanisms are insufficient because they only protect the data up to the release or the cipient. Therefore, we also extended the privacy notion and anonymization framework

to other real life data publishing scenarios, including sequential release publishing and

Secure data integration.

REFERENCES:-

A. Benjamin C. M. Fung On PRIVACY-PRESERVING DATA PUBLISHING. Benjamin C. M. Fung 2007 SIMON FRASER UNIVERSITY Summer 2007.

[1] C. C. Aggarwal. On *k*-anonymity and the curse of dimensionality. In *Proc. of the 31st*

International Conference on Very Large Data Bases (VLDB), pages 901{909, Trond-

heim, Norway, 2005.

[2] C. C. Aggarwal, J. Pei, and B. Zhang. On privacy preservation against adversarial data

mining. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining, Philadelphia, PA, August 2006.

[3] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and

A. Zhu. Anonymizing tables. In *Proc. of the 10th International Conference on Database*

Theory (ICDT), pages 246{258, Edinburgh, UK, January 2005.

[4] R. Agrawal, A. Ev¯mievski, and R. Srikant. Information sharing across private

databases. In *Proc. of the 2003 ACM SIGMOD International Conference on Man-*

agement of Data, San Diego, CA, 2003.

[5] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of

items in large datasets. In *Proc. of the 1993 ACM SIGMOD*, pages 207{216, 1993.

[6] R. Agrawal and R. Srikant. Privacy preserving data mining. In *Proc. of the 2000 ACM*

SIGMOD International Conference on Management of Data, pages 439{450, Dallas,

Texas, May 2000.

[7] S. Agrawal and J. R. Haritsa. A framework for high-accuracy privacy-preserving mining.

In *Proc. of the 21st IEEE International Conference on Data Engineering (ICDE)*, pages

193{204, Tokyo, Japan, 2005.

[8] R. J. Bayardo and R. Agrawal. Data privacy through optimal k-anonymization. In

Proc. of the 21st IEEE International Conference on Data Engineering (ICDE), pages

217{228, Tokyo, Japan, 2005.

[9] L. Burnett, K. Barlow-Stewart, A. Pros, and H. Aizenberg. The gene trustee: A uni-

versal identi¯cation system that ensures privacy and con¯dentiality for human genetic

databases. *Journal of Law and Medicine*, 10:506{513, 2003.

[10] Business for Social Responsibility. BSR Report on Privacy, 1999. http://www.bsr.org/.

123

BIBLIOGRAPHY 124

[11] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms.

Communications of the ACM, 24(2):84{88, 1981.

[12] S. Chawathe, H. G. Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ull-

man, and J. Widom. The TSIMMIS Project: Integration of heterogeneous information

sources. In *16th Meeting of the Information Processing Society of Japan*, pages 7{18,

1994.

[13] C. Clifton. Using sample size to limit exposure to data mining. *Journal of Computer*

Security, 8(4):281{307, 2000.

[14] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu. Tools for privacy

preserving distributed data mining. *SIGKDD Explorations*, 4(2), December 2002.

[15] L. H. Cox. Suppression methodology and statistical disclosure control. *Journal of the*

American Statistics Association, Theory and Method Section, 75:377{385, 1980.

[16] T. Dalenius. Finding a needle in a haystack - or identifying anonymous census record.

Journal of O±cial Statistics, 2(3):329{336, 1986.

[17] U. Dayal and H. Y. Hwang. View de¯nition and generalization for database integration
in a multidatabase systems. *IEEE Transactions on Software Engineering*, 10(6):628{
645, 1984.
[18] A. Deutsch and Y. Papakonstantinou. Privacy in database publishing. In *ICDT*, 2005.
[19] W. Du, Y. S. Han, and S. Chen. Privacy-preserving multivariate statistical analysis:
Linear regression and classi¯cation. In *Proc. of the SIAM International Conference on*
Data Mining (SDM), Florida, 2004.
[20] W. Du and Z. Zhan. Building decision tree classi¯er on private data. In *Workshop*
on Privacy, Security, and Data Mining at the 2002 IEEE International Conference on
Data Mining, Maebashi City, Japan, December 2002.

Filtering Noise on two dimensional image Using
Fuzzy Logic Technique

Anita Pati, V. K. Singh, K. C. Mishra

1 . Abstract

This paper presents one simple and novel technique for removal of impulse noise from corrupted image data. The algorithm involves impulse detection followed by spatial filtering of the corrupted pixels. In this method the presence of impulse noise is detected by a simpler method called a fuzzy logic based technique (FLT). However, the filtering idea is to recover the healthy pixel by the help of neighboring pixels. Sometimes the loss of edges or presence of noise makes the image noisy or blurred in appearance. This fuzzy logic filter is presented through 5 stages. (1)A sliding moving window is constructed to check every pixel of the whole image.(2)Two conditional rule are applied according to the averaging value of the neighboring pixels. (3)some membership functions are generated to improve the intensity of pixel value so that it can be distinguishable.(4) A simpler function is developed for better reorganization and removal of noisy data. (5)The resulted matrix appears with suppression of less no of noise. It is shown that FLT is a preferred method of rejecting impulse noise both in terms of computational complexity and lower residual NSR(noise to signal ratio).

2. Introduction

An image acquired by optical or electronic means is usually degraded in the form of sensor noise, blur due to camera misfocus, relative object camera-motion and random atmospheric turbulence. In digital signal processing the filtering of such noises is one of the most important task. This is achieved by linear filtering technique. However, in many situations such as the presence of sharp edges and impulse noise, the performance of linear filter is poor. To overcome these loopholes , nonlinear method of filtering has been proposed. The most popular non-linear filter is the median filter. It is computationally efficient, but yields blurred and distorted outputs. Subsequently, a fast real time algorithm has been reported for median filtering of signal and images. In this method noise filtering based on their local mean and variance for both the additive and multiplicative cases has been suggested. Recently, it has been shown that the use of local statistics works better for removal of additive white and multiplicative noise . However, it is not suitable for the removal of impulse noise as it employs optimal linear approximations. Another effective algorithm of noise filtering which does not require image modeling for both the additive and multiplicative noise cases has recently been reported . Some statistical properties of median filter are analyzed and it is shown that the median filter can remove impulsive and gaussian white noise.

This filtering scheme is based on replacing the central pixel value by the general mean of all pixels inside a sliding window. As the probability of noise corruption increases, its performance decreases while that of the median filter remains constant.   Besides, if both positive and negative types of impulses are present, the performance of generalized mean filter is unsatisfactory. This filter is also not suitable for simultaneous removal of impulsive and non-impulsive noises. Its performance in presence of signal dependant noise is satisfactory. A novel class of non-linear filter for image processing known as order statistics filter has been reported. This filter is used for reduction of white noise, signal-dependent noise, and impulse noise. Another filter known as signal adaptive median filter has been developed which performs better than other non-linear adaptive filters for different .kinds of noises. The adaptive averaging filter proposed in   performs 'poorly' in the presence of impulsive noise and does not remove   noise   close   to   the   edges.   The   filtering

scheme proposed in cannot suppress the impulsive noise sufficiently, but can preserve the edge better than the mean filter. It is claimed that decision based order statistics filters can reduce both impulsive and non-impulsive noise and can also enhance blurred edges better than many other order statistics filter. A fuzzy operator has been presented in for enhancement of blurred and noisy images. A new approach to spatial adaptive image restoration, which employs minimum additional computational load compared to the direct techniques have been proposed. The use of wavelet coefficient presents a new method for adaptive restoration and yields very good edge preservation in the restored results. A novel algorithm for removing impulse noise from images is presented, in which the nature of filtering operation is conditioned on a state variable. The key of the algorithm is a classifier that indicates in the probability of impulse corruption by operating on the rank ordered differences within a sliding window. This technique significantly outperforms a number of well known techniques in presence of impulsive Gaussian and mixed type of noise. A reliable and efficient computational algorithm for restoring blurred and noisy images has been proposed . By using inverse filtering technique blurred images can be restored. In a recent publication Malladi and Seth Ian have proposed an unified approach for noise removal, image enhancement, and shape recovery. This approach relics on the level of set formulation of curves and surface motion, which leads to a class of PDE-hascd algorithm. Enhancement of medical images can be successfully achieved by this technique. Several adaptive least mean square (LMS) filters have been recently proposed for noise suppression from images. This algorithim is computationally efficient In a recent paper an adaptive order statistics noise filter is proposed for gamma corrupted image sequence. This technique estimates the weights of an adaptive order statistics estimator that adapts to the probability density function of the noise. This approach is quite successful in handling the signal dependent noise. Impulse noise can also removed using higher order statistics. But this method involves corruption of higher order statistics, which is computationally expensive Filtering of impulse noise is also performed using artificial neural network (ANN). It is reported that a single layer neural network accurately detects the impulse noise of varying amplitudes . However, the multi-layer ANN involves large number of' connecting weights, bias weights, and output non-linearity. This increases the training time and the computational complexity. The technique involves two steps:

- **Detection of impulse noise**

- **Spatial filtering of noise of corrupted pixels**

In Section 11, we present the details of the techniques employed for impulse detection and filtering operation. Section III describes the simulation. study of the proposed algorithms, which are applied to different images. Section IV deals with comparative performance obtained from the simulation results experimented under different noise conditions. Section V presents the concluding remarks.

3. Techniques Employed

Consider an original image Y corrupted by the impulse noise n_i . The resultant distorted image X, may be written as:

Xi $= Yi + n_i$;

The impulse noise n_i has been added with a probability P. Since P is less than 1, it is useful to filter only those pixels which are corrupted by noise.

This approach reduces the blurring of the signal. The first step in this approach is to detect the presence of an impulse at a pixel position. To achieve this, we consider a sliding window of small size and propose an algorithm to test whether the center pixel of the test window is corrupted or not. If the test pixel is found corrupted, then we replace it by suitable filtering otherwise, we slide the test window to test the next center pixel. The process is repeated until the entire image is covered .

3.1   Fuzzy   Logic   Based   Technique:

In this method, the detection of the presence of an impulse is based oil fuzzy logic. It is assumed that the image does not contain any sharp rise or fall in gayness at any of the pixel positions. The fuzzy detector is described below.

3.1.1 Noise Detection Process

Step 1: Consider a 3 x 3 test window x_T containing noisy pixels constructed from the corrupted image X ; as-

$$x_T = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,1} & X_{2,2} & X_{2,3} \\ X_{3,1} & X_{3,2} & X_{3,3} \end{bmatrix}$$

3.1.2 Pass the X_T through a mathematical manipulator (MM) and compute A s

i=1,...,4 defined as:

$$\Delta 1 = (X_{1,1} + X_{3,3})/2 - X_{2,2}$$
$$\Delta 2 = (X_{3,1} + X_{1,3})/2 - X_{2,2}$$
$$\Delta 3 = (X_{2,1} + X_{2,3})/2 - X_{2,2}$$
$$\Delta 4 = (X_{1,2} + X_{3,2})/2 - X_{2,2}$$

3.1.3 Pass all A is through two membership functions, $\mu_1(.)$ and $\mu_2(.)$ characterized us

$$\mu_1(x) = \begin{cases} 0 & x<a \\ (x-a)/(b-a) & a < x < b \\ 1 & x>b \end{cases}$$

$$\mu_2(x) = \begin{cases} 0 & x>-a \\ (x+a)/(a-b) & -b<x<-a \\ 1 & x<-b \end{cases}$$

These two membership functions are graphically tested and Let's denote $\mu i(\Delta j)$ by μ_{ij}.

3.1.4 The rule for the impulse detection in a pixel is

• If either Δl or Δ2 or Δ3 or Δ4 is very large then the test pixel $x_{2,2}$ is corrupted.

• If either Δl or Δ2 or Δ3 or Δ4   is very low then the test pixel x_{22} is corrupted.

Using the above two rules, the output O of the impulse detector may be written as

O= $\max\{\max(\mu 1j), \max(\mu 2j))$; where j=1..4

$$H(O)= \begin{cases} 1, & x>t \\ 0 & \text{otherwise, where t-threshold value} \end{cases}$$

As O is in non-binary form, it is passed through a hard limiter (H) defined above to give a discrete yes or no decision.

Computation of Eq^n.   takes more time while computing a number of comparison operations. To reduce this time,   the   Eqn.may   be   approximated   to   a   simpler   form   as   given   below:

$$O \qquad = \qquad \sum\sum \qquad \mu_{i,j} \qquad\qquad ;$$

O is then passed through the hard limiter defined in Eqn below to get the value of d, the desired decision regarding the presence of an impulse i.e.

$$d = H(O) \quad ;$$

If the value of d is 1 then impulse noise is present in the test pixel and filtering process is revoked. The scheme of the proposed impulse detector can be depicted through a diagram.

3.1.2 Noise Filtering Process:

Compute a term g=$(x_{1,1}+x_{2,2}+x_{3,3})$/3

The moving window $x_{i,j}$ is then shifted by one row/one column until all pixels of the corrupted image is covered. The impulse detection and filtering process is carried out for all the windows. The complete fuzzy logic based filtering scheme can be shown through the proposed impulse detector.
.
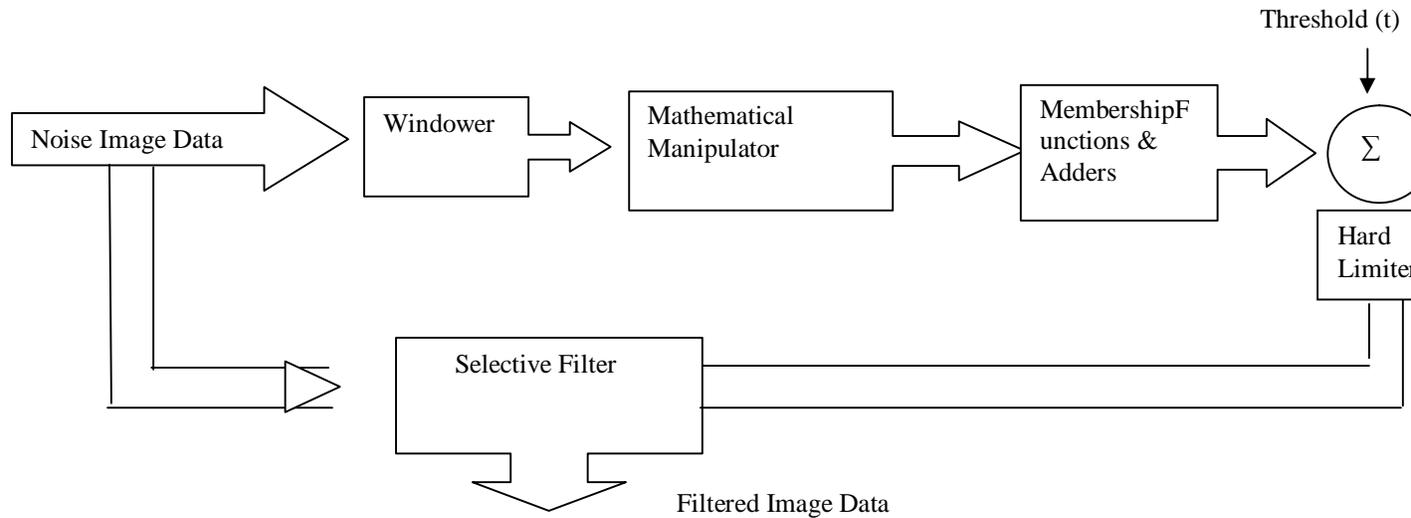
Fig 4 Fuzzy Logic Based Filtering Scheme

4.Simulation

To validate the efficiency of the proposed schemes, simulation study was carried out.
 Two standard images Clown and Lenna were selected for simulation. At first, thesetwo images were converted into their normalised form, so that the   the grav scale value of each pixel lies between 0 and 1   Each of these images was then distorted with impulse noise with its strength varying between 0,3   to   0.7. Probability of corruption was set between 5 to 30 percent. As the range of normalised pixel values was different for different images, the threshold values selected were different in each case.   Both impulse detection and spatial filtering were carried out on the corrupted images using both double derivative and fuzzy

logic based techniques. The noise to signal ratio (NSR) in dB was estimated using conventional formula. The residual NSR associated with the restored image was used as the index of efficiency of a particular method. The noise strength in each case was varied and the corresponding NSR in dB in the restored image was computed through the simulation using both the methods. The comparative performance has been made in the next section.

5.Conclusion

The following   represents the results of both Clown and Lenna images for a15 percent noise case. In each case, the original image, its corrupted version, the restored image obtained by DDT (Double Derivative method)and FLT (Fuzzy Logic Method) are shown. It may be observed that the restored Clown image exhibits a close similarity with its original image. Further, the restored image obtained from FLT is more closer to the original image compared to the image restored by DDT. The same findings are also observed in case of Lenna image. In the second part of the simulation, the percentage of noise is varied between 5 to 30 percent and in each case the residual NSR in dB is obtained by using both DDT(Double Derivative technique) and FLT(Fuzzy Logic Technique). The comparative performance of NSR present in restored images is shown in Fig.6. In general, it is observed that for both the images, the FLT outperforms the DDT in terms of residual NSR in dB, Further, to compare the computational complexity involved in both the methods, total number of operations per window is computed . In general, if the size of image is M x N pixels, the DDT requires (M-2) x (N-2) M x N windows to be processed both for impulse detection and spatial filtering for the entire image. Similarly, the FLT involves processing of (M-1) x (N-1) M x N windows. As the total number windows to be processed are of the same order in both DDT and FLT, a performance ratio E is defined as.

E = Total operations in DDT [Total operations in PUT =0.67/0.49= 1.36

Hence, it is observed that FLT is computationally efficient about 36 percent than DDT.
This paper proposes novel and efficient techniques of removing impulse noise contaminated with 2-D image, Both the methods involve two steps: impulse detection and spatial filtering. The simulation study has been performed on two standard images. Comparative performance study between two techniques has been made in terms of residual NSR and computational complexity. It is observed that FLT yields lower NSR in restored image compared to the DDT, which indicate that the first method possesses better noise filtering capability. Further,it is shown that the FLT involves less number of" I computations as compared to the DDT. Thus, it is concluded that the FLT is a preferred method of filtering the impulse noise from the iina c data both in terms of filtering potentiality and computational complexity.

References

(1)   A.K.Jain, Fundamentals of digital Image Processing, Englewood Cliff, N.J. Prentice Hall ,1989.
.

(2) E. Argyle., "Techniques for edge detection", In
Proceedings of the IEEE, vol. 59, pp. 285-286,
1971.

(3) Ayman A Aly and Abdallah A. Alshnnaway "an Edge

Detection and Filtering Mechanism of Two Dimensional Digital objects Based on Fuzzy inference, World Academy of Science Engineering and Technology 53,2009.
.
(4) J. L. Homer,"Optical Spatial Filtering with the Least mean Square Error Filter," Journal of Optical Society of America, 51(5): 553-558, May 1969.

(5)   T. S. Huang, G. J. Yang and G. Y. Tang, "A Fast Two-Dimensional Median Filtering Algorithm,'' IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-27(1): 13-18, February 1979.

6)   Peli and D. Malah, "A Study of Edge Detection on Algorithms " Computer Graphics and Image Processing, Vol. 20: 1-21 , 1982.

(7)   Chanda, B., Chaudhuri, B.B. and Dutta Majumdar, D., "A differentiation/enhancement edge detector and it's properties", IEEE Trans. On System, Man and Cybern. SMC- 15: 162-168, 1985.

(8)   Ziou, D. and S. Tabbone, "Edge detection techniques an overview," International Journal of Pattern Recognisation Image Analysis,vol.8:537-559,1998.

 (9)   Davis, L. S.,"Edge detection techniques", Computer Graphics Image Process., vol. 4: 248-270,1995.

(10)   .Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, " Robust Visual Method for assessing the Relative Performance of Edge Detection Algorithms," IEEE Trans. Pattern Analysis and Machine Intelligence,19(12): 1338-1359,1997.

(11)   Umbaugh, S., " Computer Imaging: Digital Image Processing and Analysis",India, Prentice Hall of India (2008).

Calculation of PID Controller Parameters for Unstable First Order Time Delay Systems

Hamideh Hamidian, Ali Akbar Jalali

Abstract— In this paper, a numerical approach for the fractional order proportional-integral-derivative controller (FO-PID) design for the unstable first order time delay system is proposed. The controller design is based on the system time delay. In order to obtain the relation between the controller parameters and the time delay, for several amounts of the plant time delay and the fractional derivative and integral orders, the ranges of stabilizing controller parameters are determined. First, for a typical time delay plant and the fractional order controller, the D-decomposition technique is used to plot the stability region(s). The controller derivative gain has been considered as one. By changing the fractional derivative and integral orders, a small amount in each stage, some ranges of proportional and integral gains are achieved which stabilize the system, independent of the fractional λ, μ orders. Therefore a set of different controllers for any specified time delay system is obtained. This trend for several various systems with different values of time delay has been done and the proportional and integral gains of the stabilizing controller have been calculated. Then we have fitted these values to the exponential functions and the proportional and integral gains have been obtained in terms of the system time delay. Using these relations, we can specify some ranges of the proportional and integral gains and obtain a set of stabilizing controllers for any given system with certain time delay. In these relations, fractional derivative and integral orders haven't part, and therefore can be applied to any fractional order controller design (for $0.1 \leq \lambda, \mu \leq 0.9$). Thus we have reached a numerical approach from the graphical D-decomposition method. In this method, there is freedom in choosing the values of λ and μ (they can fall in the range of [0.1, 0.9]), and there is no need to plot the stability boundaries and check the different regions to determine the stable one. This numerical method does not offer the complete set of the stabilizing controllers. Whenever the system time delay is more, the specified range of proportional and integral gains will be smaller. In other words, the extent of obtained stability region is inversely proportional to the system time delay. Finally, the introduced numerical approach is used for stabilizing an unstable first order time delay system.

Index Terms—Fractional order PID controller, numerical approach, time delay.

———————————————— ◆ ————————————————

1 INTRODUCTION

ALTHOUGH great advances have been achieved in the control science, the proportional-integral-derivative controller is still the most used industrial controller.

According to the Japan Electric Measuring Instrument Manufacturers' Association in 1989, PID controller is used in more than 90% of control loops [1], [2]. As an example for the the application of PID controllers in industry, slow industrial processes can be pointed, low percentage overshoot and small settling time can be obtained by using this controller [1]. Widespread application of the PID controller is due to the simple and implementable structure and its robust performance in the wide range of the working conditions [3], [4]. This controller provides feedback, it has the ability to eliminate steady-state offsets through integral action, and it can anticipate the future through derivative action. The mentioned benefits have caused widespread use of the PID controllers. The derivative action in the control loop will improve the damping, and therefore by accelerating the transient response, a higher proportional gain can be obtained. Precise attention must be paid to setting the derivative gain because it can amplify high-frequency noise. In this paper, for the fractional order PID controller design, the derivative gain (k_d) is set 1, that will result in design simplicity. Most available commercial PID controllers have a limitation on the de-

rivative gain [2]. During the past half century, many theoretical and industrial studies have been done in PID controller setting rules and stabilizing methods [3]. So far several different techniques have been proposed to obtain PID controller parameters and the research still continues to improve the system performance and increase the control quality. Ziegler and Nichols in 1942 proposed a method to set the PID controller parameters. Hagglund and Astrom in 1995, and Cheng- Ching in 1999, introduced other techniques [5]. By generalizing the derivative and integral orders, from the integer field to non-integer numbers, the fractional order PID controller is obtained. In fractional order PID controller design, there is more freedom in selecting the parameters and more flexibility in their setting . This is due to posse of choice -both integer and non-integer numbers- for integral and derivative orders. Therefore control requirements will be easier to comply [6], [7].

Before using the fractional order controllers in design, an introduction to fractional calculus is required. Over 300 years have passed since the fractional calculus has been introduced. The first time, calculus generalization to fractional, was proposed by Leibniz and Hopital for the first time and afterwards, the systematic studies in this field by many researchers such as Liouville (1832), Holmgren (1864) and Riemann (1953) were performed [8].

Fractional calculus is used in many fields such as electrical transmission losses systems and the analysis of the mechatronic systems. Some controller design techniques are based on the classic PID control theory generalization [7]. Due to the recent advances in the fractional calculus field and the emergence of fractance electrical element, the fractional order controller implementation has become more feasible [6], [9], [10]. Consequently, fractional order PID controller analysis and synthesis have received more attention [11], [12], [13], [14], [15], [16]. Results obtained from various articles published in this field, indicate that the fractional order PID controllers enhance the stability, performance and robustness of the feedback control system [6], [11], [12]. Maiti, Biswas and Konar [1] have significantly reduced the overshoot percentage, the rise and settling times, compared to classic PID controller, using the fractional order PID controller. Applying the fractional order PID controller ($PI^{\lambda}D^{\mu}$), the system dynamic characteristics can be adjusted better [17]. Many dynamic processes can be described by a first order time delay transfer function [18]. The need to control time delay processes can be found in different industries such as rolling mills. Varying time delay process control becomes difficult using classical control methods [19]. Simple formulas are available for setting the PID controller parameters for the stable first order time delay system, but when the system is unstable, the problem will be more difficult and therefore the unstable systems control requires more attention. Many attempts have been made in field of their stabilization [20], [21], [22], [23], [24]. So far, various design techniques have been suggested for the fractional order controller design [13], [14], [25], [26]. It has been shown that fractional order PID controllers have a better performance comparing to integer order ones, for both integer and fractional order control systems.

In the controller design for an unstable system, the most important design issue is stabilizing the closed-loop system [6]. As an example of previous research in stabilizing the unstable processes, we can point to De Paor and O'Malley research in 1989, which discussed unstable open loop system stabilization with a PID or PD controller [23]. Hamaci [3] has concluded that fractional order PID controller has a better response than classic one. In this paper, a numerical method is introduced to design the fractional order controllers for any unstable first order system with specified time delay.

2 THE FRACTIONAL ORDER PID CONTROLLER DESIGN

2.1 A Review to Design Methods

Hamamci and Koksal [4] have designed the fractional order PD controller to stabilize the integration time delay system, which result that stability region extent is reversed with the system time delay. Maiti, Biswas, and Konar, in 2008, could significantly reduce the overshoot percentage, the rise time, and the settling time by using fractional order PID controllers. They introduced PSO (particle swarm optimization) optimization technique for the fractional order PID controller design. In their me-

thod, the controller has been designed based on required maximum overshoot and the rise time. In the mentioned technique, the closed loop system characteristic equation is minimized in order to get an optimal set of the controller parameters [1]

One of the methods to obtain the complete set of stabilizing PID controllers is plotting the global stability regions in the (k_p, k_i, k_d) -space, which is called the D-decomposition technique [3], [4], [6], [8]. This technique is used in both fractional and integer order systems analysis and design.

Cheng and Hwang [6] has designed the fractional order proportional - derivative controller to stabilize the unstable first order time delay system and D- decomposition method has been used. The graphical D- decomposition technique results for such systems are simple.

The D- decomposition technique can be used for fractional order time delay systems and fractional order chaos systems. In this method, the stability region boundaries are obtained, which are described by real root boundary (RRB), infinite root boundary (IRB), and complex root boundary (CRB). By crossing these boundaries in the (k_p, k_i, k_d) -space, several regions will be achieved. By choosing an arbitrary point from each region and checking its stability, the region's stability is tested. If the selected point is stable, the region including that point would be stable, and if the selected point is not stable then the region would be unstable. By obtaining the stability boundaries and plotting the stability regions, a complete set of stabilizing fractional order controller parameters is obtained. The mentioned algorithm is simple and effective.

2.2 The D-decomposition Technique

In general, the characteristic equation of the fractional order closed loop system is defined as

$$P(s) = p_k s^{q_k} + p_{k-1} s^{q_{k-1}} + ... + p_1 s^{q_1} + p_0. \tag{1}$$

In P parameter space, the boundaries between stable and unstable regions are defined by three following parts:

Real root boundary (RRB): A real root crosses over the imaginary axis at s=0. Thus the real root boundary is obtained by setting s = 0 in (1). RRB is determined as $p_0 = 0$.

Complex root boundary (CRB): A pair of complex roots, cross over the imaginary axis at $s = j\omega$

Infinite root boundary (IRB): An infinite real root crosses over the imaginary axis at $s = j\infty$ Therefore IRB line is obtained by putting $p_k = 0$ in (1). First, by using the D-decomposition graphical method, stability boundaries and then stability region(s), are obtained. RRB and IRB lines are given by

$$RRB : k_i = 0 \tag{2}$$

$$IRB : k_d = \begin{cases} 0 & \mu > 1 \\ \infty & \mu \le 1 \end{cases}. \tag{3}$$

Then by writing k_p and k_i equations in terms of k_d, λ, and μ, the CRB curve equation is derived. The transfer function of the plant and fractional order PID controller and closed-loop system characteristic equations have been given in (4), (5), and (6), respectively. Fig. 1 shows the block diagram of the closed loop system.
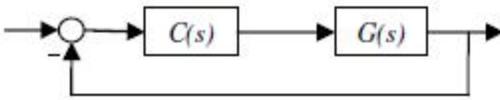
Fig. 1. The closed loop system block diagram.

$$G(s) = \frac{e^{-\theta s}}{\tau s - 1}, \tau = 1 \tag{4}$$

$$C(s) = k_p + \frac{k_i}{s^\lambda} + k_d s^\mu \tag{5}$$

$$P(s) = (\tau s - 1)s^\lambda + e^{-\theta s}(k_p s^\lambda + k_i + k_d s^{\lambda+\mu}) = 0 \tag{6}$$

CRB curve equations are given in (7).

$CRB : P(\omega; k) = 0$

$$CRB : \begin{cases} k_p = \dfrac{A(\omega)B(\omega) - C(\omega)D(\omega) + k_d(E(\omega)B(\omega) - F(\omega)D(\omega))}{G(\omega)B(\omega) - D(\omega)H(\omega)} \\ k_i = \dfrac{C(\omega)G(\omega) - A(\omega)H(\omega) + k_d(F(\omega)G(\omega) - E(\omega)H(\omega))}{G(\omega)B(\omega) - D(\omega)H(\omega)} \end{cases} \tag{7}$$

Where

$$A(\omega) = \left(\cos\frac{\lambda\pi}{2}\right)\omega^\lambda + \tau\left(\sin\frac{\lambda\pi}{2}\right)\omega^{(\lambda+1)} \tag{8a}$$

$$B(\omega) = (-\sin\omega\theta) \tag{8b}$$

$$C(\omega) = \left(\sin\frac{\lambda\pi}{2}\right)\omega^\lambda - \tau\left(\cos\frac{\lambda\pi}{2}\right)\omega^{(\lambda+1)} \tag{8c}$$

$$D(\omega) = (\cos\omega\theta) \tag{8d}$$

$$E(\omega) = -\omega^{(\lambda+\mu)}\left(\cos(\omega\theta).\cos\frac{(\lambda+\mu)\pi}{2} + \sin(\omega\theta).\sin\frac{(\lambda+\mu)\pi}{2}\right) \tag{8e}$$

$$F = \omega^{(\lambda+\mu)}\left(-\cos(\omega\theta).\sin\frac{(\lambda+\mu)\pi}{2} + \sin(\omega\theta).\cos\frac{(\lambda+\mu)\pi}{2}\right) \tag{8f}$$

$$G(\omega) = (\cos\omega\theta).\left(\cos\frac{\lambda\pi}{2}\right)\omega^\lambda + (\sin\omega\theta).\left(\sin\frac{\lambda\pi}{2}\right)\omega^\lambda \tag{8g}$$

$$H(\omega) = (\cos\omega\theta).\left(\sin\frac{\lambda\pi}{2}\right)\omega^\lambda - (\sin\omega\theta).\left(\cos\frac{\lambda\pi}{2}\right)\omega^\lambda \tag{8h}$$

By plotting IRB and RRB lines and CRB curve in the (k_p, k_i)-plane, for fixed k_d, λ, and μ, as ω runs from 0 to ∞, Stability regions are obtained [3], [4], [6], [8]. Assuming $k_d = 1$, the D-decomposition technique is used for various values of λ, μ and θ, and stability regions are obtained by arbitrary selected test points. In this paper, Nyquist stability test has been used to check the stability of the selected point of each region. The range of changes for λ and μ are considered between zero and 1.

3 OBTAINING THE CONTROLLER PARAMETERS BASED ON THE PLANT TIME DELAY

To determine the relations between the controller parameters with time delay system, the stability regions are determined using D-decomposition technique. In Fig. 2 the stable region of $\frac{e^{-0.5s}}{s-1}$ plant with $PI^{0.4}D^{0.7}$ controller is marked in gray. Notice that the small region that includes small proportional and integral gains (near the origin) is not a part of the stability regions.
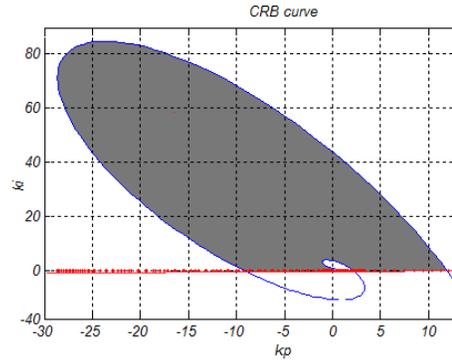
Fig. 2. The stability region for the $\frac{e^{-0.5s}}{s-1}$ plant and the $PI^{0.4}D^{0.7}$ controlle.

Two limited rectangles are selected from the stability region, one rectangle in the first quarter of the (k_p, k_i)-plane and another one in the second quarter of the plane. Fig. 3 shows two selected rectangles. Rectangle (1) which includes positive values of the proportional and integral gains is marked in light gray, and rectangular (2) which contains negative proportional gain and positive integral gain, is marked in dark gray. Two rectangles are chosen so that the minimum value of k_i is zero (means that rectangles have been placed on the RRB line) and the minimum value of k_p is close to zero as much as possible. Maximum value of the proportional gain in the second rectangle is selected near zero as much as possible.
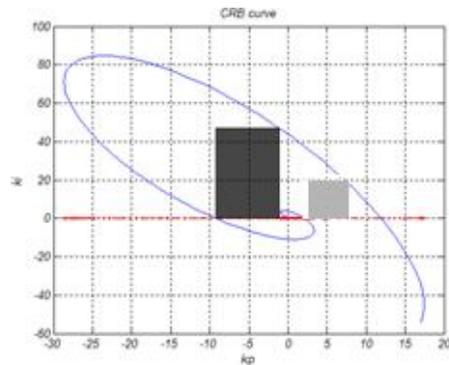
Fig. 3. Two selected rectangles, for the $\frac{e^{-0.5s}}{s-1}$ system, which is controlled by the $PI^{0.4}D^{0.7}$ controller.

Considering the mentioned criteria for selecting two stability rectangles, for various amounts of the system time delay, we have obtained the minimum and maximum values of k_p and maximum value of k_i. First, by considering different $PI^\lambda D^\mu$ controllers $(0 < \lambda, \mu < 1)$ for $\frac{e^{-0.1s}}{s-1}$

system, we choose two mentioned stability rectangles. In both rectangles, the maximum and minimum value of the proportional gain and the maximum value of the integral gain are shown with $k_{p_{max}}$, $k_{p_{min}}$ and $k_{i_{max}}$, respectively. By changing the fractional derivative and integral orders, the position of the rectangles on the (k_p, k_i)-plane will change, but in all, some values of k_p and k_i are the same.

This procedure is performed in two stages. First, rectangle 1 is considered. For $\theta = 0.1$, considering several controllers with different fractional orders, it can be observed that if the controller proportional gain is selected between 15 and 40 and the integral gain between zero and 25, for any fractional derivative and integral orders between 0.1 and 0.9, the controller will stabilize the system. By changing the value of the system time delay, this process is repeated and for each system, some ranges of k_p and k_i are determined, any arbitrary controller with these obtained parameters (where the fractional orders λ and μ are between 0.1 and 0.9) can stabilize the system. For some first order systems with the time delay θ, the obtained ranges of stabilizing controller parameters are given in Table 1. These k_p and k_i values belong to rectangle 1 within the stable region and have been obtained independently of the fractional derivative and integral orders. To study time delay effect on the range of controller parameters, in each step θ has been changed slightly (0.05).

To obtain the proportional and integral gains range independent of fractional derivative and integral orders, we consider different values of λ and μ, which change slightly in each step (0.05). By increasing the system time delay, the values of the minimum and maximum proportional gain and maximum selected value of the integral gain become smaller, Table 1 also confirms this reduction.

Table 1. The stabilizing parameters ranges in rectangle 1(for any arbitrary $PI^{\lambda}D^{\mu}$ controller $(0.1 \leq \lambda, \mu \leq 0.9)$)

θ	$k_{p_{min}}$	$k_{p_{max}}$	$k_{i_{max}}$
0.1	15	40	25
0.3	5	14	5
0.7	2	5	1.5
1	1.3	3.3	0.8
1.3	1.1	2.6	0.7
1.8	1	1.4	0.25

To obtain these ranges based on the system time delay, several different systems with $0.1 \leq \theta \leq 1$ have been considered. Different values of the time delay are considered from 0.1 up to 1, and the obtained minimum and maximum values of k_p and k_i are fitted to the exponential functions. In Fig. 4 the obtained proportional gain ranges (which will result the stabilizing controller, independently of the fractional derivative and integral orders) and also their fitting to the exponential functions are shown.

The upper curve is obtained from the maximum proportional gain values fitting, and the underlying curve is obtained from fitting the integral gain minimum values. These curves should not exceed $\left[k_{p_{min}}, k_{p_{max}} \right]$.
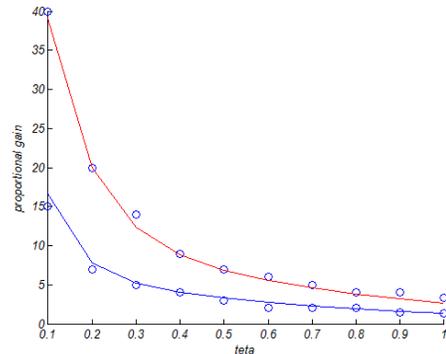
Fig. 4. The selected ranges for the proportional gain and the fitting results.

Maximum proportional gain in rectangle 1 is fitted to $a.e^{-b\theta} + c.e^{-d\theta}$, such that the resulting exponential function is close to the maximum value of k_p as much as possible (minimize the fitting error), also the fitting result should be located between the selected minimum and maximum proportional gain values. This fitting result is $76e^{-10.8\theta} + 16e^{-1.8\theta}$. Similarly, the minimum proportional gain in rectangle 1, is fitted to $45e^{-15\theta} + 8e^{-1.8\theta}$. The maximum integral gain which is selected from rectangle 1 is fitted to $60.1e^{-13\theta} + 8e^{-2.5\theta}$. According to these fitting results, for FO_PID controller design for the system with the time delay θ $(0.1 \leq \theta \leq 1)$, selecting the proportional gain from (9), the integral gain from (10), and the fractional orders from the given range in (11), the closed-loop system would be stable.

$$k_p = a.e^{-b\theta} + c.e^{-d\theta}, \begin{cases} 45 \leq a \leq 76 \\ 11 \leq b \leq 15 \\ 8 \leq c \leq 16 \\ d = 1.8 \end{cases} \quad (9)$$

$$k_i = a.e^{-b\theta} + c.e^{-d\theta}, \begin{cases} 0 \leq a \leq 60 \\ 13 \leq b \\ 0 \leq c \leq 8 \\ 3 \leq d \end{cases} \quad (10)$$

$$0.1 \leq \lambda, \mu \leq 0.9 \quad (11)$$

Therefore, for any unstable first order system with time delay $(0.1 \leq \theta \leq 1)$, the obtained exponential functions can be used to calculate the stabilizing controller parameters and to obtain a set of proportional and integral gains. These values can be used in any controller which its fractional derivative and integral orders are between 0.1 and 0.9, and therefore a wide set of the stabilizing controllers will be available. In this paper, the system time delay is considered smaller than or equal to the system time constant.

All obtained exponential functions, are only the functions of the system time delay, and are independent of the

fractional controller derivative and integral orders. Now, we explain the system time delay effect on the position of rectangle2 in the (k_p, k_i) –plane and like before, we consider different values of the time delay and the derivative order and the integral order. For several values of the system time delay $(\theta \le 2)$, the position of rectangle 2 has been obtained and Table (2) shows the minimum and maximum proportional gain and maximum integral gain in rectangle 2.

Table2. The stabilizing parameters range in rectangle 2, for some delay system.

θ	$k_{p_{min}}$	$k_{p_{max}}$	$k_{i_{max}}$
0.1	-45.7	-27	150
0.4	-10.5	-4.7	30
0.6	-6.7	-2.3	17
0.9	-4.3	-0.9	10
1.2	-3.1	-0.5	6.8
1.5	-2.4	-0.3	4.9
2	-1.9	0	3.1

Ranges which were obtained for the proportional and integral gains, are independent of λ and μ values, and for any λ and μ between [0.1, 0.9] the resultant controller will stabilize the closed loop system. When the system time delay increases, the minimum and maximum of k_p and maximum of k_i become smaller, as seen in Table 2. In controller design for the plant with specified delay, if the relations between stabilizing parameters and the plant time delay are given, using them a set of stabilizing controllers can be obtained. To get these relations, minimum and maximum of proportional gain and maximum of integral gain are fitted to the exponential functions of θ .The fitting results are given in Table 3.

Table3. The rectangle2 boundaries fitted to the exponential functions (independent of λ and μ)

$0.1 \le \theta \le 1 ; 0.1 \le \lambda, \mu \le 0.9$	
Fitting result of $k_{p_{min}}$	$-\left(98 e^{-12\theta} + 18.5 e^{-1.7\theta}\right)$
Fitting result of $k_{p_{max}}$	$-\left(48 e^{-6.3\theta} + 2.1 e^{-0.9\theta}\right)$
Fitting result of $k_{i_{max}}$	$310 e^{-12\theta} + 55 e^{-2\theta}$

Using the exponential functions which are obtained from fitting the boundaries of the two selected rectangles, various ranges of stabilizing controllers can be obtained, and for the system with specified delay some values of stabilizing proportional and integral gains can be easily obtained.

For varying time delay systems, these exponential func-

tions can be used to design the fractional order PID controller (with any arbitrary fractional derivative and integral order between 0.1 and 0.9). Since the obtained relations can be used in fractional controller design for $0.1 \le \lambda, \mu \le 0.9$, there are many alternatives in the controller choice. When the system time delay increases, the range of parameters would be smaller.

4 ILLUSTRATION

We consider an unstable first order plant which its transfer function is $\dfrac{e^{-0.34s}}{s-1}$. We use the introduced numerical method to obtain a set of stabilizing controller for this system. Some ranges of the proportional and integral gains are given in (12) .Here, both gains are positive.

$$k_{p_{min}} = 45 e^{-15\theta} + 8 e^{-1.8\theta} = 4.62 \qquad (12a)$$
$$k_{p_{max}} = 76 e^{-10.8\theta} + 16 e^{-1.8\theta} = 10.6 \qquad (12b)$$
$$k_{i_{min}} = 0 \qquad (12c)$$
$$k_{i_{max}} = 60.1 e^{-13\theta} + 8 e^{-2.5\theta} = 4.1 \qquad (12d)$$

Also, we can refer to Table 3 to calculate the controller parameters. According to this table, by choosing the proportional gain in the range [-12,-7.2], integral gain in the range [0, 33.1] and arbitrary λ and μ in the range [0.1, 0.9], a set of stabilizing controllers will be obtained.

5 CONCLUSION

In this paper, a numerical method has been proposed to design a fractional order PID controller for the unstable first order time delay system. In this method, some ranges of the proportional and integral gains are obtained based on the system time delay. If the proportional and the integral gains are selected from these ranges, the closed-loop system would be stable. The arbitrary fractional derivative and integral orders are selected from the range [0.1, 0.9]. The D-decomposition technique is used to derive the numerical relations. By introduced numerical method, without having to determine the stability boundaries and plotting them in the (k_p, k_i) -plane and checking the stability of all regions, a set of fractional order controller parameters for an unstable first order time delay system with transfer function $\dfrac{e^{-\theta s}}{s-1}$ is determined easily. In fractional controller design using the mentioned relations, derivative and integral orders can be chosen arbitrary numbers between 0.1 and 0.9. Although this approach does not get all the stabilizing controllers for the specified time delay system, but we have a simple design method because of simple calculations and the freedom to choose the fractional λ and μ orders. This method can also be used in varying time delay systems to obtain the proportional and integral gains as functions of the time delay, wherein using classical control methods will be difficult for these systems.

REFERENCES

[1] D. Maiti, S. Biswas, and A. Konar, "Design of a Fractional Order PID Controller Using Particle Swarm Optimization Technique," Proc. 2nd National Conference on Recent Trends in Information Systems (ReTIS-08).

[2] G. J. Silva, A. Datta, S. Bhattacharyya, "PID Controllers for Time-Delay Systems," Birkhauser Boston., pp. 161-190, 2005.

[3] S. E. Hamamci, "An Algorithm for Stabilization of Fractional-Order Time Delay Systems Using Fractional-Order PID Controllers," IEEE Transactions on Automatic Control, vol. 52, no. 10, pp. 1964-1969, 2007.

[4] S. E. Hamamci, M. Koksal, "Calculation of all stabilizing fractional-order PD controllers for integrating time delay systems," Computer and Math. Applic, vol. 59, no. 5, pp. 1621-1629, 2010.

[5] K. Bettou, A. Charef, "Control quality enhancement using fractional $PI^\lambda D^\mu$ controller," International Journal of Systems Science, vol. 40, no. 8, pp. 875-888, 2009.

[6] Y. C. Cheng, C. Hwang, "Stabilization of unstable first-order time-delay systems using fractional-order PD controllers," J. Chin. Inst. Eng., vol. 29, no. 2, pp. 241-249, 2006.

[7] R. Caponetto, L. Fortuna, and D. Porto, "Parameter tuning of a non integer order PID controller," In: Proceedings 0f 15th Int. Symp. on Mathematical Theory of Networks and Systems, Notre Dame, Indiana, 2002.

[8] S. E. Hamamci, "Stabilization using fractional-order PI and PID controllers," Nonlinear Dynamics, vol. 51, pp. 329-343, 2008.

[9] M. Nakagava, K. Sorimachi, "Basic Characteristics of a Fractance Device," IEICE Transactions Fundamentals, vol.E75-A, no. 12, pp. 1814-1818, 1992.

[10] P. Arena, R. Caponetto, L. Fortuna, and D. Porto, "Nonlinear Noninteger Order Circuits and Systems," World Scientific publishing Co., Inc., River Edge, NJ, USA, 2000.

[11] R. S. Barbosa, J. A. Machado, I. M. Ferreira, "Tuning of PID Controllers Based on Bode's Ideal Transfer Function," Nonlinear Dynamics, vol. 38, no. 1-4, pp. 305-321, 2004.

[12] A. Djouambi, A. Charef, and T. Bouktir, "Fractional Order Robust Control and $PI^\lambda D^\mu$ Controllers," WSEAS Transactions on Circuits and Systems, vol. 4, no. 8, pp. 850-857, 2005.

[13] I. Podlubny, "Fractional-order systems and $PI^\lambda D^\mu$ controllers," IEEE Trans, Autom. Control, vol. 44, no. 1, pp. 208-214, 1999.

[14] J. F. Leu, S. Y. Tsay, and C. Hwang, "Design of optimal fractional order PID controllers," J. Chinese Inst. Chem. Eng., vol. 33, no. 2, pp. 193-202, 2002.

[15] C. A. Monje, A. J. Calderon, B. M. Vinagre, Y. Q. Chen, and V. Feliu, "On Fractional PI Controllers: Some Tuning Rules for Robustness to Plant Uncertainties," Nonlinear Dynamics, vol. 38, no. 1-4, pp. 369-381, 2004.

[16] Z. B. Wang, Z. L. Wang, G. Y. Cao, and X. J. Zhu, "Digital Implementation of Fractional Order PID Controller and Its Application," Journal of Systems Engineering and Electronics, vol. 16, no. 1, pp. 116-122, 2005.

[17] I. Petras, "The fractional order controllers: methods for their synthesis and application," J. Electrical Engineering, vol. 50, no. 9-10, pp. 284-288, 1999.

[18] C. C. Valentine, M. Chidambaram, "PID controller of unstable time delay systems," Chem. Eng. Comm., vol. 162, pp. 63-74, 1997.

[19] D. M. Schneider, "Control of Processes with Time Delays," IEEE Transactions on Industry Applications, vol. 24, n0. 2, pp. 186-191, 1988.

[20] M. Kavdia, M. Chidambaram, "On-line controller tuning for unstable systems," Computer Chem. Eng., vol. 20, pp. 301-305, 1996.

[21] C. S. Jung, H. K. Song, and J. C. Hyun, "A direct synthesis tuning method of unstable first-order-plus-time-delay processes," J. Process Contr., vol. 9, no. 3, pp. 265-269, 1999.

[22] V. Venkatashankar, M. Chidambaram, "Design of P and PI controllers for unstable first-order plus time delay systems," Int. J. Control 60, pp. 137-144, 1994.

[23] A. M. De Paor, M. O'Malley, "Controllers of Ziegler-Nichols type for unstable process with time delay," Int. J. Control 49, pp. 1273-1284, 1989.

[24] T. Vinopraba, N. Sivakumaran, N. Selvaganesan, and S. Narayanan, "Stabilization Using Fractional-Order PID^α Controllers for First Order Time Delay System," International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009.

[25] JC. A. Monje, B. M. Vinagre, V. Feliu, and Y. Q. Chen, "On auto-tuning of fractional order $PI^\lambda D^\mu$ controllers," in Proc. FDA 2006 Fractional Derivatives and Appl., Porto, Jul. 19-21, 2006.

[26] C. Hwang, J. F. Leu, and S. Y. Tsay, "A note on time-domain simulation of feedback fractional-order systems," IEEE Trans. Auto. Control, vol. 47, no. 4, pp. 625-631, 2002.

Enhanced Mode of Extended Set of Target Fault Techniques in Single Stuck-at Fault for Fault Coverage in Benchmark Circuits

P. Amutha, C.Arunprasath

Abstract— Considering the full scan benchmark circuit, in which the undetectable single stuck-at faults, tends to cluster in certain areas. This indicates that certain areas may remain uncovered by a test set for single stuck-at faults. The extension to the set of target faults aimed at providing a better coverage of the circuit in the presence of undetectable single stuck-at fault. The extended set of target faults consists of double stuck-at faults that include an undetectable fault as one of their components. The other component is a detectable fault adjacent to the undetectable fault. Test sets that contain several different tests for each fault (n-detection test sets) are expected to increase the likelihood of detecting defects associated with the sites of target faults. This phenomenon is discerned from the gate level description of the circuit, and it is independent of layout parameters. In addition, the clustering is based on the gate level, and remains valid for any layout of the circuit. The fault simulation and test generation for the extended set of target faults is simulated using modelsim along with that the test set compaction is achieved by reseeding method.

Index Terms— Benchmark circuits, fault simulation, stuck-at faults, test quality, unresolved faults.

———————————————— ◆ ————————————————

1 INTRODUCTION

FAULT models used as targets for test generation are expected to guide the generation of tests. Thus, a test set generated for single stuck-at faults is expected to detect defects associated with the sites of stuck-at faults. Test sets that contain several different tests for each fault (n-detection test sets) are expected to increase the likelihood of detecting defects associated with the sites of target faults. When a single stuck-at fault is undetectable, it leaves an uncovered site in the circuit. As we demonstrate later, in benchmark circuits, undetectable single stuck-at faults tend to cluster in certain areas. This implies that certain areas of the circuit remain uncovered, or less covered than other areas, by a test set for single stuck-at faults. We demonstrate that undetectable single stuck-at faults in full-scan benchmark circuits tend to cluster in certain areas of the circuit. We consider the set F1 of uncollapsed single stuck-at faults. This allows us to define clustering, based on structural adjacencies between faults in the circuit, without the need to account for undetectable faults that are missing due to fault collapsing. The fault simulation and test generation experiment described in Section III to full-scan IS-CAS-89 benchmark circuits. We defined sets of double faults such that $|F2| \approx p|F1|$, for p = 1, 2, 4, and 8. We include in T1, every random test that detects a new fault from F1, when it is simulated. This test set does not detect all the detectable single stuck-at faults in the benchmark circuits considered. After targeting double faults and extending the test set into a new test set T2, we perform fault simulation of T2 to check whether any additional single faults are detected. This phenomenon is discerned from the gate level description of the circuit, and it is independent of

layout parameters. Specifically, undetectable single stuck-at faults are introduced by logic synthesis. In addition, our definition of clustering is based on the gate level, and remains valid for any layout of the circuit. To obtain a better estimate of coverage for the circuit in the presence of undetectable faults, and provide a target for improving this coverage, we propose to consider double stuck at faults that include an undetectable fault as one of their components. The other component is a detectable fault adjacent to the undetectable fault. The motivation for considering such double faults is two fold.

2 DETECTION OF A DOUBLE STUCK AT FAULT

The detection of a double stuck-at fault, which includes an undetectable stuck-at fault fi and an adjacent detectable fault fj, provides indirect coverage for the site of fi. Although fi is not detected when it is present alone since it is detected when another, adjacent fault is present. This improves the defect coverage around the site of fi. For example, bridging faults are often not targeted directly by tests used for manufacturing testing due to the difficulty in extracting and generating tests for potential bridges and the larger test set sizes needed. In such cases, one depends on the accidental detection of bridges by tests generated for single stuck-at faults. For bridges occurring in parts of the circuit with undetectable faults, the probability of accidental detection can be improved by adding tests for double stuck-at faults as suggested in this paper.
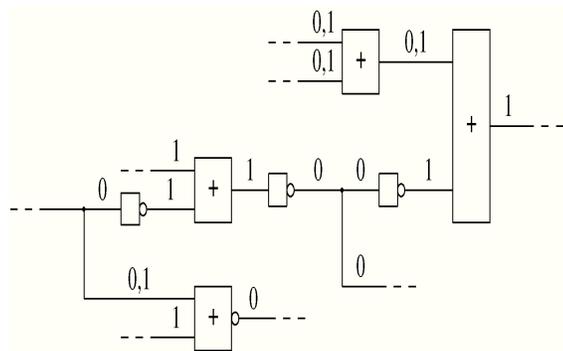
Fig.1. Sub circuit-1

Based on above, if an undetectable fault fi occurs in the circuit without being detected and a second fault fj occurs, a test set that detects fj when it is present alone may not detect the double fault that consists of fi and fj. Considering the double fault explicitly it is possible to ensure that fj will be detected when fi is present. Test sets that detect single stuck-at faults are known to detect large percentages of multiple stuck-at faults consequently, the requirement to detect double stuck-at faults is based on undetectable single stuck-at faults need not add a significant number of tests to a test set for single stuck at faults. Nevertheless, the added tests can be important in covering defects in the parts of the circuit with undetectable single stuck-at faults. It should be noted in this regard that even if a test set detects a fault fj , it is not guaranteed to detect a double fault that consists of fj and a second fault, fi. Such double faults will be targeted here when fi is an undetectable fault and f j is an adjacent detectable fault.

2.1 Clustering of Faults

In this section, we demonstrate that undetectable single stuck-at faults in full-scan benchmark circuits tend to cluster in certain areas of the circuit. We consider the set F1 of uncollapsed single stuck-at faults. This allows us to define clustering, based on structural adjacencies between faults in the circuit, without the need to account for undetectable faults that are missing due to fault collapsing. Let T1 be a given test set that detects all the detectable faults in F1. Suppose that T1 detects a subset D1 of F1. The set U1 = F1 − D1 consists of the undetectable faults in F1. We say that two faults fi and fj are adjacent, if one of the following conditions is satisfied. Let gi be the site of fi and let gj be the site of fj . For a gate G, gi is the input of G and gj is the output of G, For a gate G, gi and gj are inputs of G. gi is a fan-out stem and gj is one of its fan-out branches, or vice versa.For a fan-out stem g, gi and gj are fan-out branches of g. We apply the adjacency relation to pairs of faults in U1 in order to partition U1 into subsets S0, S1, Let U1 = {f0, f1, …, fm−1}. Initially, we set Si = {fi} for 0 ≤ i < m. We then repeat the following process in order to merge pairs of subsets that contain adjacent faults until no additional merging is possible. For every pair of subsets, Si1 and Si2 such that i1 < i2, we check whether Si1 and Si2 contain faults fi1 and fi2, respectively, such that fi1 and fi2 are adjacent. If so, we add the faults from Si2 to Si1, and remove Si2. We

computed subsets of adjacent undetectable faults for

TABLE 1

CLUSTRING OF UNDETECTABLE FAULTS

Circuits	Faults	Undet	Subsets	Subset size	
				Ave	Max
S1432	2820	26	11	2.36	6
S5378	10 590	120	16	7.50	30
S9234	18 468	118	52	21.50	46
S13207	26 358	298	46	6.48	42
S15850	31 694	789	76	10.38	73

full-scan ISCAS-89 benchmark circuits. The results are shown in Table I. Under column "Faults" we show the number of uncollapsed single stuck-at faults. Under column "Undet" we show the number of undetectable faults. Under column "Subsets" we show the number of subsets of adjacent undetectable faults. Under column "Subset Size" we show the average and maximum size of a subset of adjacent undetectable faults. From Table I, it can be seen that benchmark circuits have large subsets of adjacent undetectable faults, similar to the one shown in Fig. 1 based on s5378. This is the motivation for attempting to increase the coverage of subcircuits that contain undetectable faults by considering double faults.

3 EXTENDED SET OF TARGET FAULTS

Under this extended set, we define an extended set of target stuck-at faults that consists of double stuck-at faults. The goal of the extension is to provide a target for improving the coverage for sites of undetectable single stuck-at faults. We describe a particular way of selecting the double faults. Other approaches can be used instead to define a larger or smaller subset of double faults. As before, we consider the set F1 of uncollapsed single stuck-at faults, and a test set T1 that detects all the detectable faults in F1. We denote by D1 the subset of F1 that T1 detects. The set U1 = F1 − D1 consists of the undetectable faults in F1. We provide a target for additional coverage for the sites of the faults in U1 by using a set of double stuck-at faults, denoted by F2. To define the set of double faults F2, we use pairs of single stuck-at faults consisting of undetectable faults and detectable faults that are adjacent to them. By using detectable faults that are adjacent to undetectable faults we improve the coverage of areas of the circuit that contain undetectable faults. We avoid undetectable double faults as described later. We consider the faults in U1 one at a time. For every fi ∈ U1, we add double faults to F2 as follows. Let fi ∈ U1 be the fault gi stuck-at ai. We first mark gi and the lines that are adjacent to gi. We use two variables, adj (gj) and adj2 (gj) for every line gj. Initially, we set adj (gj) = 0 and adj2 (gj) = 0 for every line gj. We set adj

(gi) = 1. For every line gj, if gj is adjacent to gi, we set

Fig. 2. Sub circuit 2.

ble. If suppose that line gj carries an unspecified value

Fig. 3. Sub circuit 3.
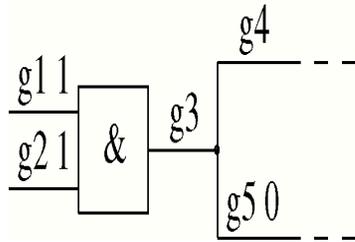
adj2 (gj) = 1. We use fi and the lines with adj2 (gj) = 1 to add faults to F2 as follows. For every fault fj ∈ D1, if fj is the fault gj stuck-at aj and adj2 (gj) = 1, we add the fault (fi, fj) or (fj, fi) to F2. We use (fi, fj) if i < j, or (fj, fi) if j < i. If the number of faults added to F2 based on fi is smaller than a constant N, we mark additional lines that are adjacent to the lines already marked. We then add additional faults based on fi using the newly marked lines. This is done as follows. For every line gj, if adj2 (gj) = 1, we set adj (gj) = 1 and adj2 (gj) = 0. This causes all the lines that were already used to define double faults based on fi to have adj (gj) = 1. For every line gj such that adj (gj) = 0, if gj is adjacent to a line gk such that adj (gk) = 1, we set adj2 (gj) = 1. We obtain a new set of lines with adj2(gj) = 1, based on which we add double faults to F2. We repeat this process until the number of faults added to F2 based on fi reaches N.

An undetectable fault gi stuck-at ai is marked with the value ai next to the line name. Our goal is to add N = 5 faults based on the undetectable fault g1 stuck-at 1. The first pass of marking adjacent lines results in adj2 (gj) = 1 for j = 2 and 3. Based on these lines we add to F2 the faults (g1/1, g2/0), (g1/1, g3/0) and (g1/1, g3/1), where g/a is the fault g stuck-at a. We do not add the fault (g1/1, g2/1) since both g1 stuck-at 1 and g2 stuck-at 1 are undetectable. Since the number of faults added to F2 is smaller than N = 5, we set adj (gj) = 1 for j = 2 and 3, and mark lines that are adjacent to them. We obtain adj2 (gj) = 1 for j = 4 and 5. Based on these lines we can add to F2 the faults (g1/1, g4/0), (g1/1, g4/1), and (g1/1, g5/1). When the number of faults added to F2 reaches N = 5, we stop adding faults based on g1 stuck-at 1. To avoid adding undetectable faults to F2, we extend the process as follows. Before starting to add faults based on fi ∈ U1, we find the forward implications of setting gi to the value ai. In the circuit with these implications, we trace the circuit backward from the outputs, and mark the lines that have x-paths to the outputs. An x-path is a path that has unspecified (x) values on all its lines. Suppose that the implications of gi = ai include a value aj on a line gj . Let fj be the fault gj stuck-at aj . The double fault that consists of fi and fj is undetectable, since fj does not affect the value of line gj in the faulty circuit that contains fi, and fi is undetectable. We do not add the fault with components fi and fj to F2. For example, in the circuit of Fig. 3, if g1 stuck-at 0 is undetectable, the double faults (g1/0, g5/0), (g1/0, g6/0), and (g1/0, g7/0) are undetecta-

when gi = ai is implied, but gj does not have an x-path to an output. This implies that the value of line gj cannot affect the output values in the presence of fi. Since fi is undetectable, the double fault consisting of fi and fj is undetectable.
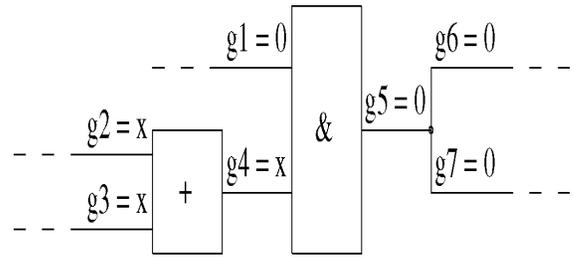
4 EXPERIMENTAL EVALUATIONS

We applied the fault simulation and test generation experimet to full-scan ISCAS-89 benchmark circuits. We defined sets of double faults such that $|F2| \approx p|F1|$, for p = 1, 2, 4, and 8. The variable p under experimental forms was used for keeping the number of double faults manageable. Since we include in F2 approximately N double faults for every single fault in U1, we obtain $|F2| \approx N|U1|$. To obtain $|F2| \approx N|U1| \approx p|F1|$, we used N = p · $|F1|$ / $|U1|$. We increase the number of faults in F2 in steps by increasing p, and use the test set T2 computed for the previous value as a starting point for the generation of additional tests. We used a simulation-based test generation process for double stuck-at faults. The test generation process is outlined next. Test generation is carried out for every fault (fi, fj) ∈ U2. If a test t is obtained for (fi, fj), t is added to the test set T2. All the faults in U2 are then simulated under t with fault dropping. A test t for a fault (fi, fj) ∈ U2 is computed as follows. We note that one of the components of (fi, fj) is an undetectable fault, and the other is a detectable fault. Let the detectable fault be fi and let the undetectable fault be fj. For fi, the test set T1 contains a test that detects it. Let the test be t. The simulation-based process we use modifies t into a test for the double fault (fi, fj) by complementing the bits of t one at a time. For a circuit with n inputs, let t = t(0)t(1)...t(n−1). Let I = {0, 1, ..., n−1}. In a step of the test generation process, we select an index k ∈ I randomly and remove it from I. We then compute the test ˆt = t(0)...t(k − 1)t(k)t(k + 1)...t(n − 1) by complementing the value of input k. We simulate fi and (fi, fj) under ˆt . If fi or (fi, fj) is detected, we set t = ˆt to accept the complemented value of input k. Otherwise, input k retains its previous value in t. If (fi, fj) is detected by t, we stop the modification of t and accept the test. Otherwise, we continue until I = ∅ . We start a new iteration by setting I = {0, 1, ..., n − 1} and repeating the process. This is done up to three times. The current test set is T1 when p = 1, or T2 generated for a lower value of p when p > 1. For example, for s38584, 142 tests in T1 leave 489 undetectable double faults for p = 1.

TABLE II

BRIDGING FAULT SIMULATION

Circuit	Test Set	Tests	Bridge
S1423	T1	26	83.53
S5378	T1	100	89.76
S5378	T2,p=1	101	89.90
S9234	T1	111	81.09
S9234	T2,p=1	113	81.53
S9234	T2,p=2	115	81.88
S9234	T2,p=4	132	83.33
S9234	T2,P=8	143	83.59
S13207	T1	235	87.97
S13207	T2,p=1	236	89.08
S15850	T1	97	86.39
S15850	T2,p=1	103	89.52
S15850	T2,p=2	104	89.72
S15850	T2,p=3	105	89.74

Test generation adds seven tests to T2 for a total of 155 tests, and detects 28 double faults. With p = 2, test generation adds three tests to T2 for a total of 158 tests, and detects 15 double faults. To illustrate that better coverage of the circuit is obtained due to the tests added to T2, we simulated nonfeedback fourway bridging faults under T1, and under the test sets T2 obtained with the various values of p. Simulation of bridging faults was used earlier to demonstrate the effectiveness of n-detection test sets for single stuck-at faults. A four-way bridging fault g/a/h is defined for a pair of lines g and h and a value a ∈ {0, 1}. In the presence of the fault, the value a on h dominates the value of g. The fault is detected by a test that sets h = a and detects the stuck-at fault on g. The model is referred to as four-way since it associates four faults with every pair of lines, differing in the dominating line and value. For every line g and every value a, we select ten four-way bridging faults randomly by selecting h randomly ten times without repetition. The results of bridging fault simulation are shown in Table III. The first row for every circuit corresponds to T1. Additional rows correspond to T2 with various values of p. The type of the test set is shown under column "Test Set." Under column "Tests" we show the number of tests in the test set. Under column "Bridge" we show the four-way bridging fault coverage. From Table III, it can be seen that adding tests for double stuck-at faults increases the four-way bridging fault coverage. Thus, the additional tests cover defects that are not covered by the single stuck-at test set.

If two patterns have no conflicting values in any bit position, we can say that the two patterns are compatible. If a cer-

tain position of a pattern is 1, and the corresponding position of another pattern is 0, and vice versa, that the two patterns are conflicting, in other words are incompatible. Combining the compatible test patterns will not ifluence the test coverage of F1 uncollapsed single stuck-at faults.

5 THE PRINCIPLE OF RESEEDING

In BIST, the reseeding is a widely-used method to test data compression. As CUT, we suppose a sequential circuit consisting of a combinational part and of n flip-flops, which form a scan chain of equal length. The TPG (Test Pattern Generator) circuit consists of a LFSR with n (n<m) flip-flop cells and an on-chip or off-chip ROM (Read Only Memory) used for storing seeds.

5.1 The method of test pattern grouping and

Combination

Usually, for all single stuck-at faults in a CUT, only about 1%~10% of bits in test patterns that generated from an ATPG tool need to specify logical value, while the others are in 'x' form. Just because test patterns have the characteristics of low specified bit density and arbitrary 0 and 1 distribution of 'x' bit, we find that a lot of test patterns are compatible.

```
P1: 0  x  1  1  x  1  1
P2: x  0  x  1  0  1  x
P3: 1  x  0  0  x  x  1
P4: x  0  x  x  1  x  x
P5: 0  x  x  x  1  1  0
P6: x  0  x  x  x  x  1
P7: 0  1  0  1  1  x  x
```

The example of test patterns

The whole test pattern set. Pattern p1 and p2 in Figure2 is an example of compatible patterns. Taking the test pattern set in figure2 for example, all test patterns can be transformed into a graph with 7 vertices, as shown in figure3, then through our-grouping-combination algorithm, 3 test pattern groups can be acquired, namely , {p1,p2,p6}, {p3, p4},and {p5,p7}. Therefore, pattern p1, p2 and p6 can merge into 0011011; pattern p3 and p4 can merge into 10001x1; pattern p5 and p7 can merge into 0101110 LFSR reseeding, the k value should be the minimum. It is known to all, graph-coloring problem is NP-complete [8], so we present a new heuristic algorithm based on Brelaz algorithm. The computational complexity of this algorithm is $2 \square (V)$, where the variable V is the number of patterns. After applying the heuristic algorithm, the whole test pattern set can be divided into multiple clusters, in each of which test patterns can be combined into one pattern, so we can encode multiple patterns by one LFSR seed. The algorithm for test patterns grouping and combination is explained as follows: create a conflict graph G = (V, E) for test patterns; for (every vertex v){ saturationDeg(v) = 0; uncoloredDeg(v) = deg(v); } place colors c1,c2,…,ck in an order; while (not all vertices in G are colored){ if (all uncolored vertices have the same combine the patterns that correspond to

the same color and calculate LFSR seeds.
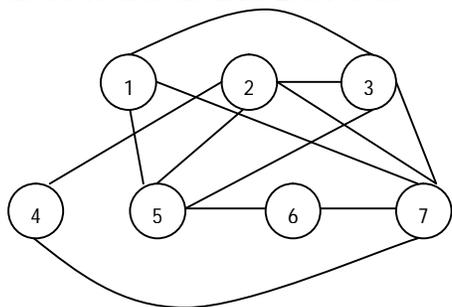
Fig.3. The graph of test patterns

5.2 Experimental results analysis

Experiments were performed on several full-scan versions of the largest ISCAS89 benchmark circuits. We used a LFSR, described with the verilog hardware description language, to generate 10000 pseudorandom patterns, which were able to detect the majority of single stuck-at faults, in each circuit. In order to attain 100% test coverage, we used ATPG tool to generate deterministic test patterns for the rest random-pattern-resistant faults. Our algorithm can successfully group and combine the compatible patterns and calculate corresponding seeds.In, one pattern can be encoded with one seed, in our method, one seed can encode multiple patterns. Table1 shows the quantity of the deterministic test patterns needed for the random-pattern-resistant faults and of the seeds which can be calculated using our method. We can draw a conclusion through comparison that our method has about 30% reduction of the seed number. The number of such faults varies with the circuit and with p. Overall; test generation for the faults in U2 adds tests to T1 in order to detect some of these faults. Although the number of tests is typically small, these tests are important due to the need to provide better coverage for areas of the circuit that may otherwise have reduced coverage. To illustrate that better coverage of the circuit is obtained due to the tests added to T2, we simulated nonfeedback fourway bridging faults [20], [21] under T1, and under the test sets T2 obtained with the various values of p. Simulation of bridging.

Table III.
Experimental results for ISCAS89

Circuit Name	Number of test patterns	Number of seeds after merging	Decreasing percentage
S1423	21	15	28.6%
S1488	18	12	33.3%
S1494	19	13	31.6%
S5378	43	30	30.2%
S9234	78	55	29.4%
S13207	69	50	27.5%
S15850	39	29	25.6%
S38584	57	36	36.8%

6 UNRESOLVED FAULTS

A test generation procedure may not be able to determine for every fault whether it is detectable or undetectable. Such a fault is said to be unresolved. The procedures described in the previous sections can treat unresolved faults as undetectable, and provide additional coverage for them. A possible byproduct of obtaining additional coverage for double faults based on an unresolved fault fi is that a test for fi would be found. We applied the fault simulation and test generation experiment to full-scan ITC-99 benchmark circuits with the same parameters as in Section IV, and with the following changes. To obtain the test set T1, we perform fault nsimulation with fault dropping of F1 under 100 000 random tests. We include in T1, every random test that detects a new fault from F1, when it is simulated. This test set does not detect all the detectable single stuck-at faults in the benchmark circuits considered. After targeting double faults and extending the test set into a new test set T2, we perform fault simulation of T2 to check whether any additional single faults are detected. As p is increased, we only define double faults based on single faults that are still undetected. It can be seen that the additional tests generated for covering sites of undetected single stuck-at faults also help detect additional detectable single faults. This is in addition to providing a better coverage for sites of faults that remain undetected. . Let the detectable fault be fi and let the undetectable fault be fj. For fi, the test set T1 contains a test that detects it. Let the test be t. The simulation-based process we use modifies t into a test for the double fault (fi, fj) by complementing the bits of t one at a time. For a circuit with n inputs, let $t = t(0)t(1)...t(n-1)$. Let $I = \{0, 1, ..., n-1\}$. In a step of the test generation process, we select an index $k \in I$ randomly and remove it from I. We then compute the test $\hat{t} = t(0)...t(k-1)t(k)t(k+1)...t(n-1)$ by complementing the value of input k. We simulate fi and (fi, fj) under $\hat{t}$. If fi or (fi, fj) is detected, we set $t = \hat{t}$ to accept the complemented value of input k. Otherwise, input k retains its previous value in t.

7 CONCLUSION

We indicated that undetectable single stuck-at faults in full-scan benchmark circuits are clustering in certain areas. We introduced an extended set of target faults based on double stuck-at faults, whose goal was to provide a target for improving the coverage of these areas. We demonstrated the enhanced mode of extended set of target fault techniques in single stuck at fault are generally provided with the compaction of test set by using the reseeding method.We presented experimental results of fault simulation and test generation in order to demonstrate the extent to which the coverage of areas with undetectable faults can be achieved that our enhanced reseeding method can significantly increase the ratio of test data compression.

REFERENCE

[1] S. M. Reddy, I. Pomeranz, and S. Kajihara, "Compact test sets for high defect coverage," IEEE Trans. Comput.-Aided Design, vol. 16, no. 8, pp. 923–930, Aug. 1997.
2003, pp. 1031–1040..

[2] P. Goel and B. C. Rosales, "Test generation and dynamic compaction of tests," in Proc. Test Conf., 1979, pp. 189–192.

[3] S. Kajihara, T. Sumioka, and K. Kinoshita, "Test generation for multiple faults based on parallel vector pair analysis," in Proc. Int. Conf. Comput.- Aided Design, 1993, pp.

[4] J.-S. Chang and C.-S. Lin, "Test set compaction for combinational circuits," in Proc. Asian Test Symp., 1992, pp. 20–25.

[5] H. Cox and J. Rajski, "A method of fault analysis for test generation and fault diagnosis," IEEE Trans. Comput.-Aided Design, vol. 7, no. 7, pp. 813–833, Jul. 1988.

[6] S. Kajihara, I. Pomeranz, K. Kinoshita, and S. M. Reddy, "Cost-effective generation of minimal test sets for stuck-at faults in combinational logic circuits," IEEE Trans. Comput.-Aided Design, vol. 14, no. 12, pp. 1496– 1504, Dec. 1995.

[7] I. Hamazaoglu and J. H. Patel, "Test set compaction algorithms for combinational circuits," in Proc. Int. Conf. Comput.-Aided Design, 1998,

[8] S. C. Ma, P. Franco, and E. J. McCluskey, "An experimental chip to evaluate test techniques experiment results," in Proc. Int. Test Conf., 1995, pp. 663–672.

 [9] M. Abramovici, M. A. Breuer, and A. D. Friedman, "Testing for single stuck faults," Digital Systems Testing and Testable Design. Piscataway,

NJ: IEEE, 1995, ch. 6, pp. 181–281...

[10] H. Tang, G. Chen, S. M. Reddy, C. Wang, J. Rajski, and I. Pomeranz, "Defect aware test patterns," in Proc. Design Autom. Test Eur. Conf., 2005, pp. 450–455.

[11] I. Pomeranz and S. M. Reddy, "Forming N-detection test sets without test generation," in Proc. ACM Trans. Design Autom., vol. 12. Apr. 2007, pp. 1–18.

[12] M. Abramovici and M. A. Breuer, "Multiple fault diagnosis in combinational

circuits based on an effect-cause analysis," IEEE Trans. Comput., vol. C-29, no. 6, pp. 451–460, Jun. 1980.

[13] Pomeranz, L. N. Reddy, and S. M. Reddy, "COMPACTEST: A method to generate compact test sets for combinational circuits," in Proc. Int. Test Conf., Oct. 1991, pp. 194–203.

A Study to Enhance Human-Resource Performance Efficiency for Minimizing Cost in Software Development Projects

Amrinder Kaur, Kamaljeet Singh

Abstract— Human resources in software development projects require a high level of individual intensity devoted to project tasks, which then is integrated collaboratively to complete the project. Human resources technical skills and implementation experience are key factors for project success and therefore must be allocated and managed judiciously. Despite of the efforts of organizations, problems related to cost escalation have been encountered regularly .Human resource efficiency and productivity shall be key for future success and sustainability of the businesses. Through this paper, the causes have been outlined in a study conducted on software development projects. The reasons outlined affect efficiency and productivity of employees in software development projects across various phases in software development viz. requirement, designing, coding/testing, implementation phase. And it is been devised that job satisfaction enhances when the job is more exploratory in nature.

Index Terms— human resource efficiency, software development phases, requirement phase, designing phase, coding phase, testing/verification phase.

———————————————— ◆ ————————————————

1 INTRODUCTION

Human resource efficiency

Efficiency is one of the very few sustainable advantages that creates a significant barrier to entry and at the same time ensures profitability for companies [1]. The efficiency is defined as the ratio between input and output, and the effectiveness is defined as the achieving level of the expected production output by a production system. In fact, efficiency and effectiveness represent different levels of performance, and there is no guarantee that both of them can be achieved simultaneously. However, an efficient organization must handle both of them well, and use the most efficient way to pursue maximum effectiveness.

Therefore, these benefits to efficiency can be categorized as tangible and intangible [2] The tangible benefits include the following:
• Reduced cost
• Improved productivity (i.e., amount of output produced per unit of input)
• Increased market share
• Savings in labor
• Increased consumer surplus (i.e., the accumulated difference between consumer demand and market price)
• Improved customer service quality
• Improved organizational efficiency
• Quicker response to customers
• Deeper knowledge and understanding of customers

On the other hand, the intangible benefits include:

• Improved decision-making ability
• Superior product quality
• Knowledge/information management and sharing
• Improved coordination/relationships with partners

In all, Contribution to the Company shall be

- **Effective management of the project**: Increase in the efficiency of human resource will lead to effective management of the project. Thus, it will lead to appropriate utilization of resources, and attainment of the desired deliverables in settled time frame and within the expected cost and quality.

- **Lesser redundancy and repetition of work**: effective utilization of resources will cut down the extra work and thus the energies can be synergized for work that is more productive.

- **Better customer image**: Effective project management will create better customer image and hence larger gains in future.

Software Development process

A **software development process**, also known as a **software development lifecycle**, is a structure imposed on the development of a software product [3]

The development process according to water fall model consists of a series of steps starting from requirement

collection, designing phase, coding phases, testing and verification phase followed with maintenance. The waterfall model shows a process, where developers are to follow these phases in order:

1. Requirements specification (Requirements analysis)

2. Software Design

3. Integration

4. Testing (or Validation)

5. Deployment (or Installation)

6. Maintenance

In a strict Waterfall model, after each phase is finished, it proceeds to the next one. Reviews may occur before moving to the next phase which allows for the possibility of changes (which may involve a formal change control process). Reviews may also be employed to ensure that the phase is indeed complete; the phase completion criteria are often referred to as a "gate" that the project must pass through to move to the next phase. Waterfall discourages revisiting and revising any prior phase once it's complete.[1]

Requirement Phase

The important task in creating a software product is extracting the requirements or requirements analysis. Customers typically have an abstract idea of what they want as an end result, but not what software should do. Incomplete, ambiguous, or even contradictory requirements are recognized by skilled and experienced software engineers at this point. Frequently demonstrating live code may help reduce the risk that the requirements are incorrect.

Once the general requirements are gathered from the client, an analysis of the scope of the development should be determined and clearly stated. This is often called a scope document.

Certain functionality may be out of scope of the project as a function of cost or as a result of unclear requirements at the start of development. If the development is done externally, this document can be considered a legal document so that if there are ever disputes, any ambiguity of what was promised to the client can be clarified.

Implementation, testing and documenting

Implementation is the part of the process where software engineers actually program the code for the project.

Software testing is an integral and important part of the software development process. This part of the process ensures that defects are recognized as early as possible.

Documenting the internal design of software for the purpose of future maintenance and enhancement is done throughout development. This may also include the writing of an API, be it external or internal. It is very important to document everything in the project.

Deployment and maintenance

Deployment starts after the code is appropriately tested, is approved for release and sold or otherwise distributed into a production environment.

Software Training and Support is important and a lot of developers fail to realize that. It would not matter how much time and planning a development team puts into creating software if nobody in an organization ends up using it. People are often resistant to change and avoid venturing into an unfamiliar area, so as a part of the deployment phase, it is very important to have training classes for new clients of your software.

Maintaining and enhancing software to cope with newly discovered problems or new requirements can take far more time than the initial development of the software. It may be necessary to add code that does not fit the original design to correct an unforeseen problem or it may be that a customer is requesting more functionality and code can be added to accommodate their requests. If the labor cost of the maintenance phase exceeds 25% of the prior-phases' labor cost, then it is likely that the overall quality of at least one prior phase is poor.[citation needed] In that case, management should consider the option of rebuilding the system (or portions) before maintenance cost is out of control.

Bug Tracking System tools are often deployed at this stage of the process to allow development teams to interface with customer/field teams testing the software to identify any real or perceived issues. These software tools, both open source and commercially licensed, provide a customizable process to acquire, review, acknowledge, and respond to reported issues. (Software maintenance).

Literature Review

Literature analysis shows that organizations usually adopt the process of business analysis in IT to enhance efficiency. Business analysis as a discipline has a heavy overlap with requirement analysis, but focuses on identifying the changes to an organization that are required for it to achieve strategic goals. These changes include changes to strategies, structures, policies, processes, and information systems.[3].Typically, business processes work to enhance project efficiency by reducing re-work and reducing the process length to contain the cost.

Efficiency improvements for the projects will have the main resource in form of human resource and there performance efficiency is crucial for the success of projects

In 1978, Harry pointed out that efficiency; effectiveness and productivity are three major parts of performance. In 1988, Fortuin placed the organizational goal in two categories: efficiency and effectiveness [1].

Thus, effective critical knowledge and skills may indeed create some frog-leaping of some economies as was predicted by [4] that:
"Fifty years from now—if not much sooner—the leadership in the world economy will
be moved to the countries and to the industries that have most systematically and most successfully raised knowledge-worker productivity."

Objective of the study

Enhancing efficiency to minimize cost in software development projects.

Methodology of the study

Research Design

The research design was exploratory till identification of software development parameters. Later it became descriptive when it came to evaluation of parameters in relation to the cost escalation.

Research Sample

The work had been the case of a software company viz. Vayam technologies Ltd. and four projects in various stages of software development. The respondents were the employees and teams working at various locations in delhi /NCR. Around 50 employees were the respondents for the study.

Data Collection tool

Data was collected through structured questionnaires. The questionnaires were designed by the help of Questionnaire Design and Survey Sampling [8] and project managers working at various projects in the software development phases in the company. The filled-in questionnaires were collected from the respondents in two weeks' time.

Data Presentation

The data collected from the questionnaire was checked for its reliability using standard deviation and standard error techniques. All the questions had their population mean within 95% of the confidence interval.

And hence the parameters for cost reduction with the help of efficiency improvement had following categories.

Project Clarity

This is regarding the goals/objectives to be achieved for the individual project.

Work done with structured plans
Commencement of work in the individual phases in the software development is done through careful planning.

Report preparation
Work at every phase is complied with reports for future reference and for further tracking by the project managers and team members.

Discussion with other team members
This is necessary for bottlenecking, innovation and clarity. At the same time it can work to maintain the job satisfaction of the team members.

Monitoring and training
As the name suggests monitoring of work flow at appropriate stages is significant for tracking of project within the timelines .Training helps the employees to update their skills and knowledge.

Job satisfaction

Job satisfaction describes how content an individual is with his or her job. The happier people are within their job, the more satisfied they are said to be. Job satisfaction is not the same as motivation, although it is clearly linked. Job design aims to enhance job satisfaction and performance; methods include job rotation, job enlargement and job enrichment. Other influences on satisfaction include the management style and culture, employee involvement, empowerment and autonomous work position. [6]

Communication problem with customers and team mates
Ninety percent of the common problems within teams are poor communication skills or a lack of communication skills, and a lack of understanding or appreciation of each team member.
Findings

Based on the parameters identified above, employee's responses were quantified. And the table 1 below represents the summary of the data on the basis of responses of fifty respondents.

Factors	Requirement Phase	Designing Phase	Coding Phase	Testing/Verification phase
Project Clarity	50%	80%	90%	90%

Work is done with structured Plans	80%	80%	80%	70%
Report Preparation	100%	100%	40%	70%
Discussion with other team members	80%	90%	90%	70%
Monitoring and Training	60%	60%	40%	70%
Job satisfaction	90%	70%	80%	80%
Communication problems with customer or team mates	80%	60%	10%	20%

In the table, the columns are the percentages of respondents who believe that they do hold the corresponding factor in the row.

Analysis of the collected data

The above result is head wise analyzed as follows:

Requirement Phase

The conclusions of requirement phase are as follows:

- Team members were clear about the project goals and Objectives.

- According to the results, if respondents face problems to carry out certain work then only 30 percent of respondents try to solve problems on their own. Remaining either keep pleading for details or talk to seniors for direction. Hence, there has to be more self-reliance in the team to carry out their work. This needs suitable experience, regular mentoring and training in the candidates. Moreover, the candidates selected for requirement phase should have experience in dealing with customers, should have leadership skills and should be responsible to carry out their task. Definitely, seniors need to be consulted because the time factor and cost minimization are important. However, referring the seniors for

every mundane task leads to more dependency, chaos and increases time to accomplish a task. This is because the seniors are more costly resources and they must put their energy and time to relatively more important work.

- The team / respondents need to be fairly trained, have leadership and problem solving traits, and customer-negotiation skills to cut down the cost.

- According to the results, around 60 percent of respondents need more training and monitoring to handle their jobs. However, around 20 percent of respondents sometimes feel that they need training and monitoring. Hence, requirement phase needs improvisation in training and mentoring to candidates so that they can better equip themselves with the task.

- According to the results, around 80 percent of respondents get communication problems with the customer while gathering the data. However, only 20 percent never feel such problems. Hence, the respondents / team need to be clearer about their project goals. The team needs to be assertive, to have good problem-solving skills. In addition, the team needs to have a fair amount of expertise in gathering and extracting information from customers to accomplish project goals.

Designing Phase

The conclusions of designing phase are as follows:

- According to the results, around 50 percent of respondents believe that they talk to seniors for direction if they meet some team members (from requirement phase) who are reluctant to divulge necessary details. Around 30 percent of respondents try to get details from team members of requirement phase assertively. Remaining 10 percent keep pleading for details. These results clearly show that interaction among the teams is less, which results in discrepancies, chaos, time lapse and inefficiency. This can be dealt by having fixed procedure form to transfer information from requirement to coding phase, which will make the work flow smoother.

- Designing and its techniques need to be defined very clearly to increase efficiency.

- The team needs to be given more training and mentoring to accomplish job.

Coding Phase

The conclusions of coding phase are as follows:

- There is a need to be little more impetus to prepare coding process report, which will help to reduce redundancy and increase efficiency of the team/respondents.

- Respondents need training in automation and optimization tools so that they can enhance their efficiency to many folds.

- Respondents require leadership and problem-solving skills to solve their problems on their own instead of relying on other team members and seniors for direction. Relying on others decreases efficiency and results in time lapse, which leads to under utilization of resources.

- According to the results, around 40 percent of respondents need more training and monitoring to handle their jobs in a better way. Moreover, around 50 percent of respondents sometimes feel that they need training and monitoring. However, remaining 10 percent do not think the same. Hence, this phase also needs improvisation in training and mentoring to candidates so that they can better equip themselves with the task.

Testing/Verification Phase

The conclusions of testing/verification phase are as follows:

- Respondents need leadership and problem-solving skills to solve their problems on their own instead of relying on other team members and seniors for direction. Relying on others decreases efficiency and results in time lapse, which leads to under utilization of resources.

- Respondents require more training and mentoring as results show that around 70 percent of respondents feel they need it.

- Respondents must be given training in automation tools so that they can enhance their efficiency to many folds.

- According to results, only 60 percent of respondents think that testing/verification issues are regularly scrutinized and casual analysis is done for testing defects. Therefore, there is a need to improve this.

Comparison of Phases

Comparison of requirement, designing, coding and testing/verification phases is as follows:

So according to the table:

- **Project Clarity**

  Percentage is highest in coding phase as, work is done through set procedures /processes and the team becomes clearer with project requisites, objectives and goals till it reaches the coding phase. On the other hand, project clarity is lowest in requirement phase as the requirement phase is beginning of the project. Hence, this phase needs special focus on this arena. Ambiguity of project leads to increased cost in the later phases. This is because software bugs are carried over to next phases, which results in redundancy, chaos and wastage of time and resources.

  > **Note:** As per the industry data, fixing a bug in the requirement phase costs 1 Unit and fixing it in the user site at production level costs 100 Units. Hence, Project clarity at requirement phase is imperative and very significant

- **Work is Done with Structured Plans**

  This factor is virtually same in all phases but marginally different for testing/verification phase. The reason for the difference is that it is the last phase and therefore, team has more confidence to handle the work without structured plans. The difference could also be due to sheer lethargy nature of the team. However, this factor can be checked for variance in performance if encountered at this stage.

- **Report Preparation**

  The percentage is lowest in coding phase. The reason for the difference is that it is the phase of settled process/procedures and hence, team has more confidence to handle the work without structured plans. The difference could also be due to sheer lethargy. However, this factor can be checked for variance in performance if encountered at this stage.

- **Discussion with Other Team Members**

  This factor is almost same in all phases.

- **Monitoring and Training**:

The need for monitoring and training is highest for coding phase. This can be attributed to the fact that coding phase has the highest technical portion, which gets updated with time. Hence, the mentoring and training definitely will enhance efficiency in this phase.

- **Job Satisfaction**

This factor is virtually same in all phases but highest in requirement phase. This is due to the reason that this phase requires interaction with customer and hence, no set process/procedures are followed. Newer challenges make work interesting, which results in the highest job satisfaction. This factor is an interesting observation for reference.

- **Communication Problems with Customer or Team Members**

This factor is highest in requirement phase, as this phase requires interaction with customer. Therefore, the team confronts discrepancies and problems in this phase. On the other hand, this factor is lowest in coding phase because this phase has set procedures. This factor is for reference and can be used to increase efficiency at different phases.

Conclusions:

- At requirement phase the project clarity is minimum, which depicts the arena for more work to enhance efficiency.

- Employees in requirement phase have more problems but more job satisfaction which shows that human interactions enhance the happiness quotient for better efficiencies.

- More automation and optimization procedures reduce the communication troubles within the teams.

Hence, software organizations can use this study in software development projects to enhance human resource efficiencies at various stages.

Limitation of study

The study was conducted on the parameters defined and identified after discussions with the software engineers, academicians, and subject experts.

Scope for further study

Parameters defined above can be individually researched to further outline the efficiency and productivity factors.

References

Citations
1. Measuring Human Resource Efficiency and Effectiveness
Human Resource Professionals Are Best Positioned to Significantly Impact an Organizational Effectiveness and Efficiency as They Interact with Everyone in the Company

2. Enhancing Human Resource Productivity Using Information and Communication Technologies: Opportunities and Challenges for Tanzania1
By Faustin Kamuzora, Mzumbe University, Tanzania, Pg 7-8

3.
http://en.wikipedia.org/wiki/Software_development_process

4. Drucker, P. (1999). *Management challenges for the 21st century*. New York: Harper Business. Pg-158.
5. http://en.wikipedia.org/wiki/Business_analysis
6. http://en.wikipedia.org/wiki/Job_satisfaction

Reference

7. Statistics for Management by Richard I.levin and David S.Rubin.Page no. 360-375.

8. Questionnaire Design and Survey Sampling by Professor Hossein Arsham.Link:University of Baltimore.

Improving Diffusion Power of AES Rijndael with 8x8 MDS Matrix

R.Elumalai, Dr.A.R.Reddy

Abstract— AES Rijndael is a block cipher developed by NIST as the Advanced Encryption Standard (AES) replacing DES and published as FIPS 197 in November 2001 [5] to address the threatened key size of Data Encryption Standard (DES). AES-Rijndael was developed by Joan Daemen and Vincent Rijmen, Rijndael [4, 5] and was selected from five finalists. Advancement in computation speed every day puts lots of pressure on AES and AES may not with stand attack for longer time. This work focuses on improving security of an encryption algorithm, beyond AES. Though there are various techniques available to enhance the security, an attempt is made to improve the diffusion strength of an algorithm. For enhancing the diffusion power AES Rijndael in MixColumn operation the branch number of MDS matrix is raised from 5 to 9 using a new 8X8 MDS matrix with trade off of speed [8, 9] and implemented on R8C microcontroller.

Index Terms— diffusion, MDS matrix, AES Rijndael, security, encryption standard, R8C, microcontroller.

———————————————— ◆ ————————————————

I. INTRODUCTION

THE AES Rijndael algorithm basically consists of four byte oriented transformation for encryption and inverse transformation for decryption process over number of rounds depending on plain text size and key length namely [1,2],

1) Byte substitution (S-box) a non linear operation, operating on each of the State bytes independently.

2) Shifting rows (Row transformation) is obtained by shifting row of states cylindrically.

3) Mix Column transformation, the columns of the State are considered as polynomials over GF(28) and multiplied modulo $X^4 + 1$ with a fixed polynomial c(x), given by
c(x) = '03' x^3 + '01' x^2 + '01' x + '02'
The inverse of MixColumn is similar to MixColumn. Every column is transformed by multiplying it with a specific multiplication polynomial d(x), given by
d(x) = '0B' x^3 + '0D' x^2 + '09' x + '0E' .

4) Add round key, a Round Key is applied to the State by a simple bitwise EXOR. The Round Key is derived from the Cipher Key by means of the key schedule. The Round Key length is equal to the block length.

The round transformation in C pseudo can be written as [1]
 Round(State,RoundKey)
{
ByteSub(State);
ShiftRow(State);
MixColumn(State);
AddRoundKey(State,RoundKey);
}

The final round of the cipher is slightly different. It is defined by:
FinalRound (State,RoundKey)
{
ByteSub(State) ;
ShiftRow(State) ;
AddRoundKey(State, roundkey);
}

In AES Rijndael confusion and diffusion are obtained by non-linear S-Box operation and by the linear mixing layer over rounds respectively.

II. DIFFUSION IN AES RIJNDAEL

The linear mixing layer guarantees high diffusion over multiple rounds. Rijndael in his proposal approved by NIST replacing DES in the 2001 proposed MixColumn which operates on space of 4-byte to 4-byte linear transformations according to the following criteria[1,2]:
1. Invertibility;
2. Linearity in GF(2);
3. Relevant diffusion power;
4. Speed on 8-bit processors;
5. Symmetry;
6. Simplicity of description.
Criteria 2, 5 and 6 have lead to the choice of polynomial multiplication modulo x4+1. Criteria 1, 3 and 4 impose conditions on the coefficients. Criterion 4 imposes that the coefficients have small values, in order of preference '00', '01', '02', '03'…The value '00' implies no processing at all, for '01' no multiplication needs to be executed, '02' can be implemented using xtime and '03' can be implemented using xtime and an additional EXOR. The criterion 3 induces more complicated conditions on the coefficients.

In Mix Column, the columns of the State are considered as polynomials over GF (2^8) and multiplied modulo $x^4 + 1$ with a fixed polynomial c(x) [1,2]. The Mix Column transformation operates independently on every column of the state and treats each sub state of the column as term of a(x) in the operating equation b(x)=c(x)$\otimes$a(x), where c(x)= '03'X^3+'01'X^2+'01'X+'02'.This polynomial is co-prime to (X^4 + 1) For example, in the figure1. a(x) is $a_{0,j}X^3$+$a_{i,j}X^2$+$a_{2,j}$X+a_3j and it is used as multiplicand of operation.
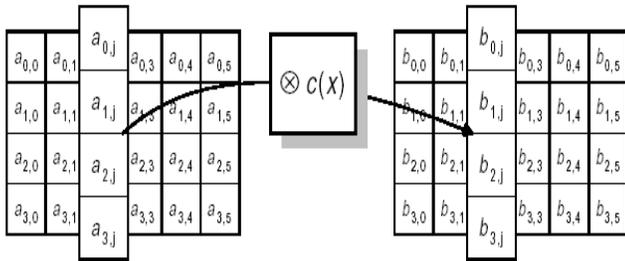

Figure1. MixColumn operation

This can be represented as

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Where c(x) has been represented by a circular matrix of 4x4.

Pseudo C code for Mixcolumn operation with Rijndael circular matrix is [10]

MixColumn (S)

{for (c=0 to3)

 Mixcolumn (s_c)

}

Mixcolumn (col)

    { copycolumn (col, t) //t is temporary column

     Col_0← (0x02) ● t_0 ⊕ (0x03) ● t_1 ⊕ t_2 ⊕ t_3

     Col_1← t_0 ⊕ (0x02) ● t_1 ⊕ (0x03) ● t_2 ⊕ t_3

     Col_2← t_0 ⊕ t_1 ⊕ (0x02) ● t_2 ⊕ (0x03) ● t_3

     Col_3← (0x03) ● t_0 ⊕ t_1 ⊕ t_2 ⊕ (0x02) ● t_3

    }

Inverse of MixColumn in Rijndael uses a MixColumn transformation with different polynomial i.e.

   c(x) = '03' x^3 + '01' x^2 + '01' x + '02' for MixColumn and

   d(x) = '0B' x^3 + '0D' x^2 + '09' x + '0E' for inverse MixColumn

The Branch Number of a linear transformation is a measure of its diffusion power. Let F be a linear transformation acting on byte vectors and let the byte weight of a vector be the number of nonzero bytes [1,2]. The byte weight of a vector is denoted by W (a).
Definition: The branch number of a linear transformation F is $Min_{a\neq 0}$ (W ((a) + W (F (a)))

A non-zero byte is called an active byte. For MixColumn it can be seen that if a state is applied with a single active byte, the output can have at most 4 active bytes, as MixColumn acts on the columns independently. Hence, the upper bound for the branch number is 5. The coefficients have been chosen in such a way that the upper bound is reached. If the branch number is 5, a difference in 1 input (or output) byte propagates to all 4 output (or input) bytes, a 2-byte input (or output) difference to at least 3 output (or input) bytes. Moreover, a linear relation between input and output bits involves bits from at least 5 different bytes from input and output.

III. OBJECTIVE OF THE PAPER

The Mixcolumn transformation of Rijndael so constructed has the property that the upper bound for the Branch number, which is 5, is reached. But the Inverse Mixcolumn transformation is not the same as the Mixcolumn transformation i.e. Mixcolumn does not have self inverse. This has led to degradation in performance on 8-bit processors for the Inverse cipher because it uses Inverse Mix Column and a modified Key schedule. Since the cipher and its inverse use two different transformations, a circuit that implements Rijndael does not automatically support the computation of the inverse of Rijndael and consumes more hardware when implemented on FPGA.

This work involves two important parameter for improving the diffusion strength of AES.

1. To find 8x8 matrix which has branch number 9 greater than 5 which is the case in AES Rijndael so that diffusion strength of the algorithm increases?
2. The matrix should be self inverse so that same matrix can be used for inverse MixColumn operation, which decreases the complexity of circuit and occupies less silicon area when implemented on hardware.

IV. MDS MATRIX

In AES linear transformations in the form of mappings based on Maximum Distance Separable (MDS) codes are used to achieve diffusion. A linear code over Galois field GF(2^p) is denoted as an (n, k, d)-code, where n is the symbol length of the encoded message, k is the symbol length of the original message, and d is the minimal symbol distance between any two encoded messages[10].

Theorem 1: An (n, k, d)-code with generation matrix G = [I | C] is MDS if, and only if, every square sub matrix of C is nonsingular.

An (n, k, d)-code is MDS if d = n − k + 1. A (2k, k, k+1)-code with generation matrix G = [I | C], where C is a k×k matrix and I is an identity matrix, determines an MDS mapping from the input X to the output Y through matrix multiplication over a Galois field as follows:

$$f_M: X \rightarrow Y = C \cdot X \quad (1)$$

where

$$X = \begin{bmatrix} x_{k-1} \\ \vdots \\ x_0 \end{bmatrix} \quad Y = \begin{bmatrix} Y_{k-1} \\ \vdots \\ Y_0 \end{bmatrix} \quad C = \begin{bmatrix} C_{k-1,k-1} & \cdot & \cdot & \cdot & C_{k-1,0} \\ \vdots & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ C_{0,k-1} & \cdot & & \cdot & C_{0,0} \end{bmatrix}$$

Each entry in X, Y, and C is an element in $GF(2^p)$. For a linear transformation, the branch number is defined as the minimum number of nonzero elements in the input and output when the input elements are not all zero.

There are different types of matrices which exhibit this property we will discuss two important types one used in Rijndael and other used in the work.

1. Circulant matrices: Given k elements α_0,…., α_{k-1}, a circulant matrix A is constructed with each entry $A_{i,j} = \alpha_{(i+j)}$ mod k. The probability that a circulant matrix is suitable for an MDS mapping C is much higher than that of a normal square matrix.

    The Matrix used in Rijndael is a circulant MDS Matrix. In Rijndael, substitution permutation network (SPN) uses optimal non-involution MDS mappings. When an SPN uses a non-involution MDS mapping optimized performance only for encryption, the inverse MDS mapping used in decryption has a higher complexity

2. Hadamard matrices: Given k elements α_0, . . ., α_{k-1}, a Hadamard matrix A is constructed with each entry $A_{i,j} = \alpha_{i \oplus j}$. Each Hadamard matrix A over a finite field has the following properties: $A^2 = \gamma \bullet I$ where γ is a constant. When $\gamma = 1$, A is an involution matrix. An involution MDS mapping is required by an involution SPN. When used in an SPN, the involution MDS mapping produces equally optimized performance for both encryption and decryption.

V. MATHEMATICAL MODEL

The matrix with high branch number and involution characteristics is found by applying Brute force method after arriving at required polynomial. The obtained polynomial H= had(01x, 03x, 04x,05x,06x,08x,0Bx,07x)

Using this polynomial 8×8 matrix is constructed. This matrix is checked for the involution property. This matrix is constructed based upon the types explained above. Then the MDS property of the matrix is calculated. i.e. an (n, k, d)-code is MDS if d = n − k + 1.This can be done by checking the branch number of the transformation. The input with one or two active byte column is multiplied with the matrix and the output column is checked, if the total number of active bytes including input and output bytes is equal to 9 then it satisfies the property of MDS. Finally this matrix is found out and verified that it satisfies the involution property and the MDS property with (16, 8, 9) code.

The linear transformation matrix is $GF(2^8) \rightarrow GF(2^8)$ is a linear mapping based on the [16,8,9] MDS code with generator matrix G_H=[IH],
where H=had(01x, 03x, 04x,05x,06x,08x,0Bx,07x) is the polynomial found using brute force technique. The Hadamard matrix with involution property based on above polynomial shown below;

$$H = \begin{bmatrix} 01 & 03 & 04 & 05 & 06 & 08 & 0B & 07 \\ 03 & 01 & 05 & 04 & 08 & 06 & 07 & 0B \\ 04 & 05 & 01 & 03 & 0B & 07 & 06 & 08 \\ 05 & 04 & 03 & 01 & 07 & 0B & 08 & 06 \\ 06 & 08 & 0B & 07 & 01 & 03 & 04 & 05 \\ 08 & 06 & 07 & 0B & 03 & 01 & 05 & 04 \\ 0B & 07 & 06 & 08 & 04 & 05 & 01 & 03 \\ 07 & 0B & 08 & 06 & 05 & 04 & 03 & 01 \end{bmatrix}$$

A simple inspection shows that matrix H is symmetric and unitary. Therefore it is an involution transformation. It may be verified that this transformation has the branch number equal to 9 and also satisfies the criterion 3.

In Rijndael the state matrix to each function is considered as a 4×4 matrix. Here inside the MixColumn function the state matrix is converted into an 8×2 matrix and multiplied with the MDS matrix. Then the resulting 8×2 matrix is converted into a 4×4 matrix and passed to the next function.

The following Example describes about the multiplication of the state matrix with the above MDS matrix.
Let the state matrix input to the MixColumn state be

$$\begin{bmatrix} 01 & 01 & 01 & 01 \\ 02 & 02 & 02 & 02 \\ 03 & 03 & 03 & 03 \\ 04 & 04 & 04 & 04 \end{bmatrix}$$

In MixColumn transformation the state matrix has to be multiplied with the standard matrix generated by the polynomial. In this scheme the multiplication of the matrix is performed with new 8x8 matrix generated by brute force

technique which satisfies the involution property as shown below;

$$H = \begin{bmatrix} 01 & 03 & 04 & 05 & 06 & 08 & 0B & 07 \\ 03 & 01 & 05 & 04 & 08 & 06 & 07 & 0B \\ 04 & 05 & 01 & 03 & 0B & 07 & 06 & 08 \\ 05 & 04 & 03 & 01 & 07 & 0B & 08 & 06 \\ 06 & 08 & 0B & 07 & 01 & 03 & 04 & 05 \\ 08 & 06 & 07 & 0B & 03 & 01 & 05 & 04 \\ 0B & 07 & 06 & 08 & 04 & 05 & 01 & 03 \\ 07 & 0B & 08 & 06 & 05 & 04 & 03 & 01 \end{bmatrix} X \begin{bmatrix} 01 & 01 \\ 02 & 02 \\ 03 & 03 \\ 04 & 04 \\ 01 & 01 \\ 02 & 02 \\ 03 & 03 \\ 04 & 04 \end{bmatrix}$$

$$= \begin{bmatrix} 08 & 08 \\ 3F & 3F \\ 2E & 2E \\ 1D & 1D \\ 08 & 08 \\ 3F & 3F \\ 2E & 2E \\ 1D & 1D \end{bmatrix}$$

The state matrix after the Mix Column transformation is

$$\begin{bmatrix} 08 & 08 & 08 & 08 \\ 3F & 3F & 3F & 3F \\ 2E & 2E & 2E & 2E \\ 1D & 1D & 1D & 1D \end{bmatrix}$$

The difference between the diffusion powers of the Mix column step of Rijndael algorithm and the new proposed Mix column are explained in the Figure 2 and 3. At the end of first round, active bytes are 8 in case of 8x8 MDS, whereas it is 4 in case of 4x4 MDS. This shows an encryption algorithm designed using 8x8 MDS will provide more security compared to 4x4 MDS. However, this is achieved at the cost of additional computation.
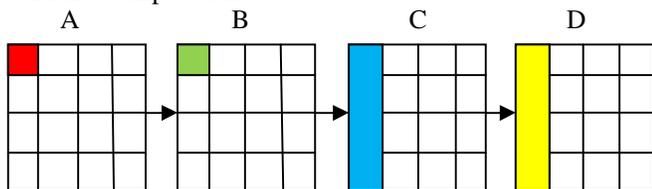

Figure 2 - Diffusion states for encryption in the AES in the first round. A – S-BOX, B – Shift Row, C – MixColumn and D – ADDRoundKey
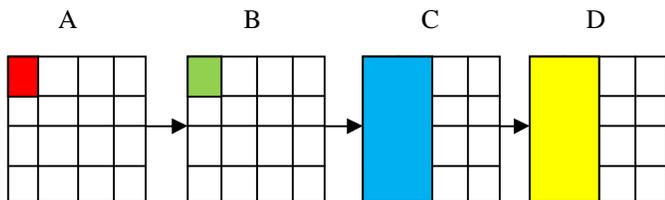

Figure 3 - Diffusion states for revised encryption with 8x8 matrix in the AES in the first round. A – S-BOX, B – Shift Row, C – MixColumn and D – ADDRoundKey.

VI. IMPLENTATION ON MICROCONTROLLER

A. Renesas R8C microntroller

The R8C/Tiny Series of single-chip microcomputers was developed for embedded applications by Renesas. The R8C/Tiny Series supports instructions tailored for the C language, with frequently used instructions implemented in one-byte op-code. It thus allows development of efficient programs with reduced memory requirements when using either assembly language or C. Furthermore, some instructions can be executed in a single clock cycle, enabling fast arithmetic processing.

R8C has features like CPU core operating at 20MHz, on chip ROM,RAM, and data Flash, programmable I/O ports, 9 interrupts with 7 priority levels, 14 bit watch dog timer, 3 timers, 4 UARTs, synchronous communication port, I^2C bus, LIN module, USB, 10 channel 10 bit ADC and 2 comparators which makes it most preferred industrial application microcontroller.

For implementation on R8C microcontroller, Renesas High performance Embedded workshop V.4.07.01 with simulator version 4.1.04.00 which is provided by Renesas was chosen for convenience. It provides integrated development environment composed of compiler and simulator also. Every cycle number and code size output depends on embedded workshop.

B. Simulation result

Code was run on the Renesas High performance embedded workshop V.4.07.01 with simulator version 4.1.04.00 with test vector from Brian Gladman's technical paper [11]. The implementation was optimized many times.

Module	Cycle	Code (Byte)
Precomputation	2178	1649
ByteSub	115	32
ShiftRow	51	256
MixColumn	201	134
AddRoundKey	127	48
branch	19	12
Total	8276	2135
(round0-10)	5917	784

Figure 4 – Simulation results for 10 round AES.

The total number includes consideration of the number of rounds. In this implementation, the message block and key length are 128-bit each. Therefore 10 rounds constitute one encryption procedure. In the 10 rounds of encryption procedure, precomputation is executed just once and other modules have all different execution times as in Cycle column of the Table2. Whereas, code length weight factor is irrelevant with cycle number weight factor because some modules are

reused every time while the other modules are not. For example, AddRoudKey module executed at round0, round1-9 and round10 but MixColumn executed only at round1-9, which makes the different code weight factor 3 and 1 respectively.

Module	Cycle	Code (Byte)
Precomputation	1	1
ByteSub	10	2
ShiftRow	10	2
MixColumn	9	1
AddRoundKey	11	3
branch	1	1

Table2. Weight factor of each module

With weight factor total number of cycle and code are computed by equation below.

Total cycle number =

$\sum$Cycle number(i) * weight factor(i)

Total code size =

$\sum$Code number(i) * weight factor(i)

Where ' i' is every module

Module	Cycle	Code (Byte)
Precomputation	2178	1649
ByteSub	115	32
ShiftRow	51	256
MixColumn	243	149
AddRoundKey	127	48
branch	19	12
Total	8376	2195
(round0-10)	6171	824

Figure 4 – Simulation results of revised 10 round AES with 8x8 MixColumn..
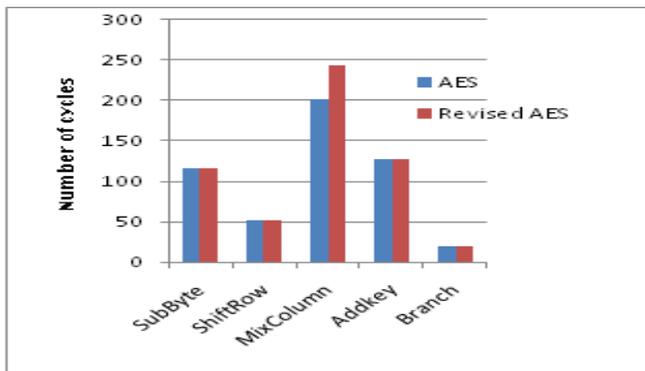
Figure 5 – Comparison of number of cycle consumed by two schemes.
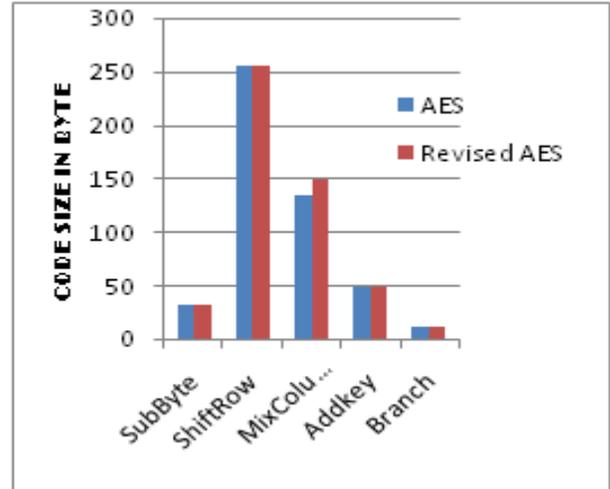

Figure 5 – Comparison of number of code consumed by two schemes.

VII.   CONCLUSION

The MixColumn module is still takes much part of cycle number. This module is also the critical part for register scheduling because it need 8 multiplication with 8 different sub state at the same time while keeping their initial state.

The result shows the number of cycle required is 20.08% more and code area consumed is 11.19% more in the case of revised AES with 8x8 MixColumn compared to AES Rijndael. The increase in cycle and code memory is the trade off for the increase in diffusion strength which increases the security of the algorithm.

REFERENCES

[1]   Daemen and V. Rijmen, AES Proposal: Rijndael (Version 2). NIST AES

[2]   NIST, Advanced Encryption Standard (AES), (FIP PUB 197), November 26, 2001

[3]   G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[4]   K. Ohkuma, H. Muratani, F. Sano, and S, Kawamura, "The Block Cipher Hiero-crypt", Workshop on Selected Areas in Cryptography. SAC 2000, Lecture Notes in Computer Science 2012, Springer-Verlag, pp. 72-88, 2001.

[5]   P. Barreto and V. Rijmen, "The Anubis Block Cipher", NESSIE Algorithm Submission 2000, available on: www.cosic.esat.kuleuven.ac,be/nessie.

[6]   P. Barreto and V. Rijmen, "The Khazad Legacy-Level Block Cipher", NESSIE Algorithm Submission, 2000, available on: www.cosic.esat,kuleuven.ac.be/nessie

[7]   A. Rudra, P.K. Dubey, C.S. Jutla, V, Kumar, J. R. Rao, and P. Rohatgi, "Efficient Rijndael Encryption Implementation with Composite Field Arithmetic", Cryptographic Hardware and Embedded Systems - CHES 2001, Lecture Notes in Computer Science 2162, Springer-Verlag, pp. 171-184, 2001

[8]   Lu Xiao and Howard M. Heys "Hardware Design and Analysis of Block Cipher Components"

[9]   Aarti Singh "Study of MDS Matrix used in Twofish AES algorithm and its VHDL implementation" M.Tech thesis.

[10] Behrouz A.Forouzan "Cryptography and network security " TATA-Mcgraw hill publication 2007 edition.

[11] A Specification for Rijndael, the AES Algorithm v3.3, Brian Gladman, May 2002

[12] P. Barreto and V. Rijmen, "The Khazad Legacy-Level Block Cipher", NESSIE Algorithm Submission, 2000, available on: www.cosic.esat.kuleuven.ac.be/nessie.

[13] R. Anderson, E. Biham, and L. Knudsen, "Serpent: a Proposal for the Advanced Encryption Standard", AES Algorithm Submission, available on: www.cl.cam.ac.uk/¢rja14/serpent.html

[14] A. Youssef, S. Mister, and S. Tavares, "On the Design of Linear Transformations for Substitution-Permutation Encryption Networks", Workshop on Selected Areas in Cryptography - SAC '97, Ottawa, 1997.

AUTHORS PROFILE

Dr.A.R.Reddy is professor in the department of Department of Electronics and Communication Engineering, at Madanapalli Institute of Science and Technology, Madanapalli, Andrapradesh, India. He received   M.Tech and Ph.D from IIT Kharagpur, India. His field of interest is Cryptogharaphy, Embedded system and VLSI.

R.Elumalai is Ph.d student in Department of Electronics and Communication Engineering, at Vinayaka Mission University, Salem, Tamilnadu, India. He received M.Tech from University Visvesvaraya College of Engineering, Bangalore University. India. His field of interest are Embedded system, VLSI and cryptography.

Design of HDLC Controller Using VHDL

K.Sakthidasan,   Mohammed Mahommed

Abstract- In this paper, we explore a High-level Data link control published by   International   Standards   Organization (ISO). HDLC is one of the most enduring and fundamental standards in Communications. HDLC is in itself a group   of   several protocols or rules for transmitting data between network points. The HDLC protocol also manages the flow or pace at which the data is sent. The data is organized into a unit called a frame. HDLC controllers are devices, which executes the HDLC protocol. Some of the key operations of the HDLC protocol implemented are handling bit oriented protocol structure and formatting data as per the packet switching protocol, it includes Transmitting and receiving the packet data serially and providing the data transparency through zero insertion and deletion. This controller generates and detects flags that indicate the HDLC status. The device contains a full duplex transceiver, with independent receive and transmit sections for bit-level HDLC protocol operations. The design is completely Synchronous, with separate clock inputs for receive and transmit allowing the two sections to operate asynchronously. These operations have been implemented using VHDL.

Index Terms- HDLC, VHDL, Bit-level HDLC protocol operations.

1. INTRODUCTION

1.1 Overview of HDLC:

The high level data link control (HDLC) protocol Defined by the ISO provides a transparent transmission Service at the data link layer of the ISO reference model. Many protocol suites use an HDLC link layer, including X.25, The IP point-to-point protocol (PPT) and SNA. It has been so widely implemented because it supports half duplex and full duplex communication lines, point-to-point (peer to peer) and multipoint networks, and switched or non switched channels. The procedures outlined in HDLC are designed to permit synchronous, code-transparent data transmission. The HDLC Protocol is a general purpose protocol, which Operates AT the data link layer of the OSI reference model. The protocol uses the services of a physical layer, to provide communications path between the transmitter and receiver. The users of the HDLC service provide PDUs, which are encapsulated to form data link layer frames. These frames are separated by HDLC "flags" and are modified by "Zero bit insertion" to guarantee transparency. Each piece of data is encapsulated in an HDLC frame by adding a trailer and a header. The header contains an HDLC address and an HDLC control field. The trailer is found at the end of the frame, and contains a Cyclic Redundancy Check (CRC), which detects any errors, which may occur during transmission. The frames are separated by HDLC flag sequences that are transmitted between each frame and whenever there is no data to be transmitted.

1.1.1 HDLC Frame Structure

HDLC uses the term "frame" to indicate an entity of data (or a protocol data unit) transmitted from one station to another. Figure below is a graphical representation of a HDLC frame with an information field.

F	A	C	I	FCS	F
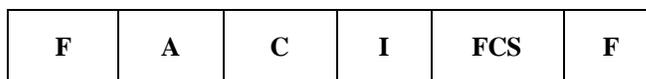
Fig 1- HDLC Frame Structure

Field Name	Size(in bits)
Flag Field (F)	8 bits
Address Field (A)	8 bits
Control Field (C)	8 or 16 bits
Information Field (I)	Variable; Not used in some frames
Frame Check Sequence (FCS)	16 or 32 bits
Closing Flag Field (F)	8 bits

The Flag Field

Every frame on the link must begin and end with a flag sequence field (F). Stations attached to the data link must continually listen for a flag sequence. The flag sequence is 01111110. Flags are continuously

transmitted on the link between frames to keep the link active. Two other bit sequences are used in HDLC as signals for the stations on the link. These two bit sequences are:

- Seven 1's, but less than 15 signals an abort signal. The stations on the link know there is a problem on the link.
- 15 or more 1's indicate that the channel is in an idle state.

HDLC is a code-transparent protocol. It does not rely on a specific code for interpretation of line control. This means that if a bit at position N in an octet has a specific meaning, regardless of the other bits in the same octet. If an octet has a bit sequence of 01111110, but is not a flag field, HLDC uses a technique called bit-stuffing to differentiate this bit sequence from a flag field.

Once the transmitter detects that it is sending 5 consecutive 1's, in inserts a 0 bit to prevent a flag sequence occurring. At the receiving end, the receiving station inspects the incoming frame. If it detects 5 consecutive 1's it looks at the next bit. If it is a 0, it pulls it out. If it is a 1, it looks at the 8th bit. If the 8th bit is a 0, it knows an abort or idle signal has been sent. It then proceeds to inspect the following bits to determine appropriate action. This is the manner in which HDLC achieves code-transparency. HDLC is not concerned with any specific bit code inside the data stream. It is only concerned with keeping flags unique.
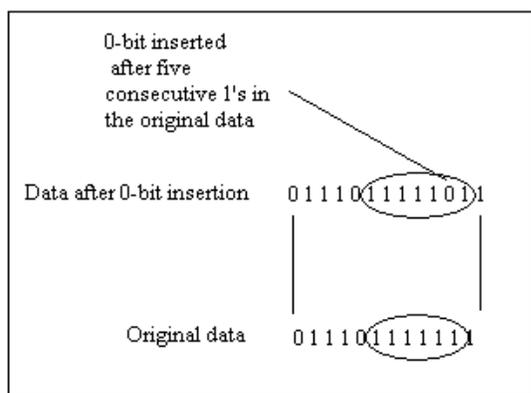
Fig 2 -Bit Insertion

0   1
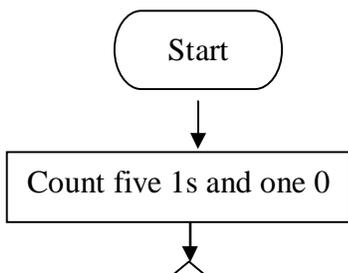
0   1

Fig 3 -Flowchart for Bit Deletion

The Address Field:

The address field (A) identifies the primary or secondary stations involvement in the frame transmission or reception. Each station on the link has a unique address. In an unbalanced configuration, the A field in both commands and responses refer to the secondary station. In a balanced configuration, the command frame contains the destination station address and the response frame has the sending station's address.

The Control Field:

HDLC uses the control field(C) to determine how to control the communications process. This field contains the commands; responses and sequences numbers used to maintain the data flow accountability of the link, define the functions of the frame and initiate the logic to control the movement of traffic between sending and receiving stations. There three control field formats:
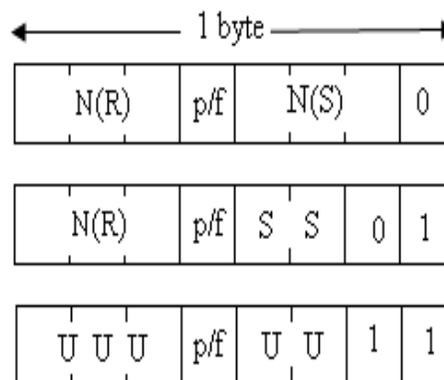
Fig 4 -Control Field Format

Information Transfer Format
The frame is used to transmit end-user data between two devices.

Supervisory Format
The control field performs control functions such as acknowledgment of frames, requests for re-transmission, and requests for temporary suspension of frames being transmitted. Its use depends on the operational mode being used.

Unnumbered Format
This control field format is also used for control purposes. It is used to perform link initialization, link disconnection and other link control functions. The 5th bit position in the control field is called the poll/final bit, or p/f bit. It can only be recognized when it is set to 1. If it is set to 0, it is ignored. The poll/final bit is used to provide dialogue between the primary station and secondary station. The primary station uses P=1 to acquire a status response from the secondary station. The P bit signifies a poll. The secondary station responds to the P bit by transmitting a data or status frame to the primary station with the P/F bit set to F=1. The F bit can also be used to signal the end of a transmission from the secondary station under Normal Response Mode.

The Information Field
This field is not always in a HDLC frame. It is only present when the Information Transfer Format is being used in the control field. The information field contains the actually data the sender is transmitting to the receiver.

The Frame Check Sequence Field
This field contains a 16 bit, or 32 bit cyclic redundancy check. It is used for error detection. CRC-16 = $x16 + x15 + x2 + 1$ is used.

2. Overview of VHDL:

VHDL is the VHSIC Hardware Description Language. VHSIC is an abbreviation for Very High Speed Integrated Circuit. It can describe the behavior and structure of electronic systems, but is particularly suited as a language to describe the structure and behavior of digital electronic hardware designs, such as ASICs and FPGAs as well as conventional digital circuits. VHDL is a notation, and is precisely and completely defined by the Language Reference Manual (LRM). This sets VHDL apart from other hardware description languages, which are to some extent defined in an adhoc way by the behavior of tools that use them. VHDL is an international standard, regulated by the IEEE. The definition of the language is non-proprietary. VHDL is not an information model, a database schema, a simulator, a toolset or a methodology! However, a methodology and a toolset are essential for the effective use of VHDL. Simulation and synthesis are the two main kinds of tools, which operate on the VHDL language. The Language Reference Manual does not define a simulator, but unambiguously defines what each simulator must do with each part of the language. VHDL does not constrain the user to one style of description. VHDL allows designs to be described using any methodology — top down, bottom up or middle out! VHDL can be used to describe hardware at the gate level or in a more abstract way. Successful high-level design requires a language, a tool set and a suitable methodology. VHDL is the language; you choose the tools, and the methodology.

2.1 Simulation

Top-down design first describes a system at a very high level of abstraction, like a specification. Designers simulate and debug the system at this very high level before refining it into smaller components. The method describes each component at a high level and debugs it alone and with the other components in the system. The design continues to be refined and debugged until it is complete down to its lowest building block. Mixed-level design occurs when some components are at a more detailed level of description than others. The advantage of the top-down design methodology is that engineers can discover and correct system problems early in the design cycle. Engineers can concentrate on overall design issues such as system requirements and timing. The tedious task of gate-level design can be left to synthesis tool. The bottom line is reduced cost and faster time to manufacturing.

2.2 VHDL design flow:

```
┌─────────────────────────────┐
│        Specification        │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│          Algorithm          │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│           Coding            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│         Simulation          │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│          Synthesis          │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│       Place and Route       │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│            FPGA             │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│            ASIC             │
└─────────────────────────────┘
```
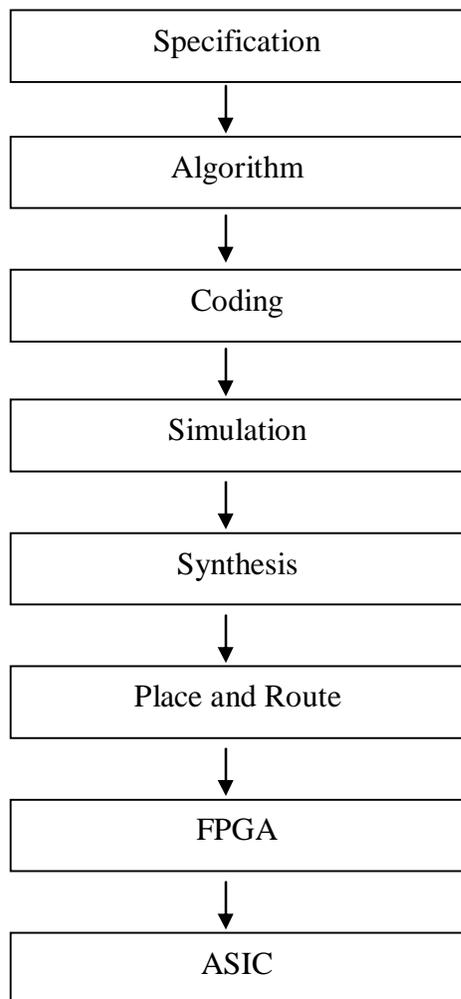
Fig 5- Design Flow

2.3 Need for a HDLC controller

As mentioned already, HDLC protocol is used widely in many applications. In many situations, the space occupied is a main criterion. Further, compatibility and processing speed pose a serious problem. In order to overcome these difficulties, the solution has been to bring out the protocol converter in the form of a handy controller. The controller can be implemented in two ways. As an embedded application or as a VLSI chip. The data transferred in the overall communication chip is purely digital and also at a very high data rate. Hence the more apt solution is to use the VLSI chip to construct the HDLC controller.

Further, in many applications the processing would need a central controller to monitor the entire process but the space constraint would dictate the need for minimal space occupation. In such a situation, the use of a system there becomes impossible. There arises the need for a controller. Replacement of the system is achieved through such efficient processors.

Transmitter :

The transmitter portion of the HDLC core will begin to transmit when the user's external logic asserts the TX_DATA_VALID signal. The transmitter will respond by asserting the TX_LOAD signal to load the first byte of the packet. The timing diagram assumes that IDLE_SEL is tied to a '1' and the transmitter is generating continuous '1' bits between frames. If IDLE_SEL is set to a '0', the number of clocks from the assertion of TX_DATA_VALID to TX_LOAD will vary from 5 to 12. Before the transmitter can begin to send data serially, it must send an opening flag (7E). Immediately after the flag is sent, the first byte is clocked out of the input shift register. Once a transmit frame has begun, the user is required to make sure that data is available for each subsequent requested byte. The transmitter will continue to request data by asserting TX_LOAD until the user supplies a TX_EOF signal. This informs the transmitter that the last byte is on the data bus. The transmitter then appends a 16- or 32-bit Frame Checking Sequence (FCS) to the transmitted data. After the FCS is sent, a closing flag (7E) byte is appended to mark the end of the frame.

Receiver :

The receiver clocks serial HDLC frames in continuously through the RXD pin. When an opening flag is recognized, the receiver locks to all subsequent octet bytes. The user informs the receiver of the ability to store the frame by asserting the RX_SPACE_AVAILABLE input. The receiver informs the user that a data byte is available by asserting the RX_READY signal. The receiver indicates the beginning of the frame by asserting the RX_SOF signal. Bytes will continue being passed to the user until the receiver recognizes the closing flag. At this point, the last byte of the FCS sequence will be passed to the user coincident with the RX_EOF signal. It must be stressed that the core does not contain the additional pipeline registers to "swallow" the 2 or 4 bytes of FCS, and these will therefore be passed on to the user. If this is undesirable, the corresponding pipeline should

be added externally to keep these bytes from passing on as part of the received frame. After the reception of the frame has completed, the receiver will pass a byte of

status information to the user by placing the status on the receive data bus and asserting the RX_STATUS signal.
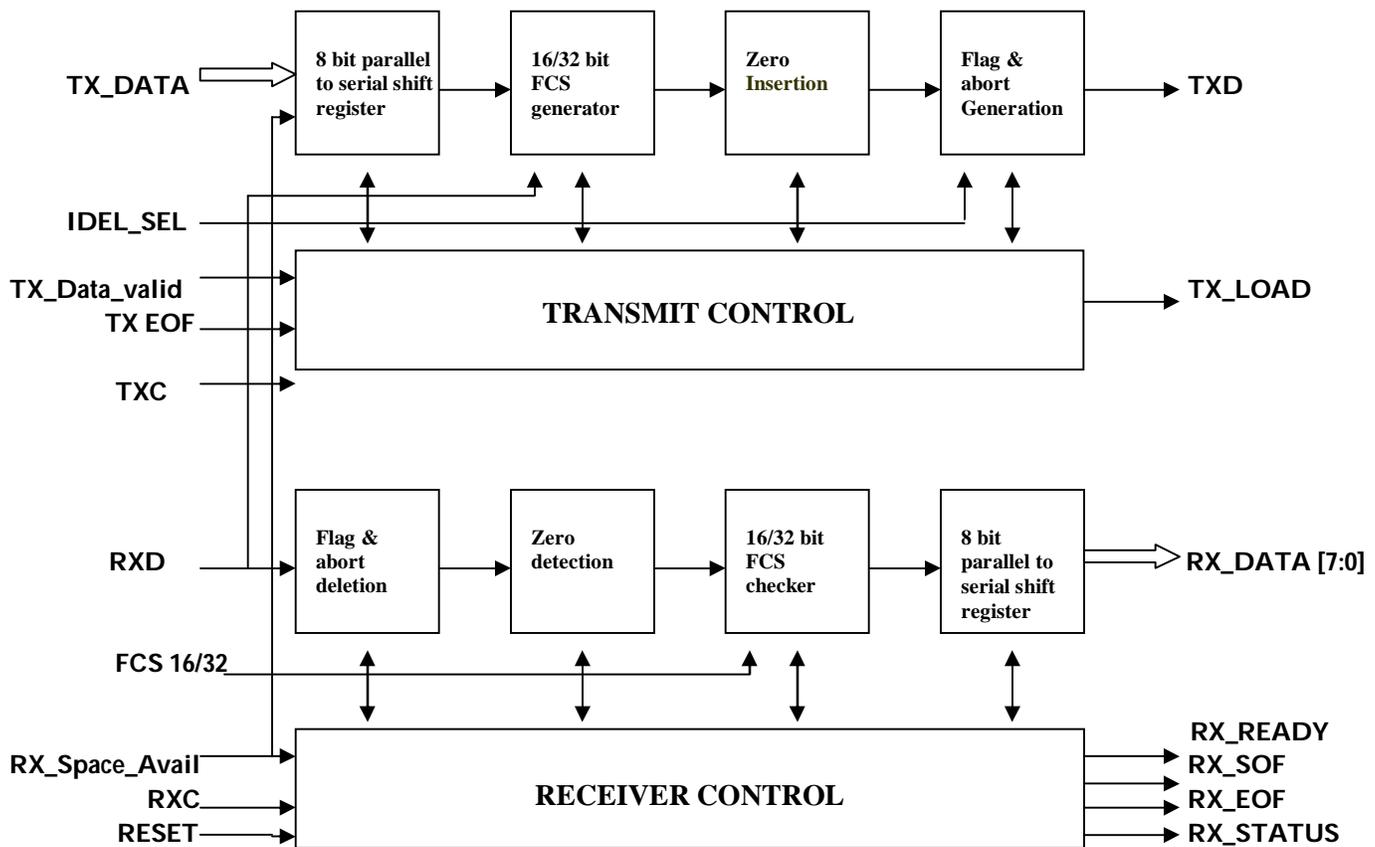
Fig 6- HDLC Controller Block Diagram

memory hierarchy to manage fast access to RAM on and off the chip . A comparison between the implemented HDLC

3. TESTING PROCESS USING VHDL:

The HDLC transceiver was designed and implemented using VHDL codes. The codes were simulated by ModelSim by forcing the clock and values to other input pins and monitoring the output waveforms. The reason for using Virtex FPGA is its various built-in features that help the designer throughout the process of design.This family provides a broad capability for chip-to-chip communications through programmable support forthe latest I/O standards, digital Delay-Locked Loops (DLLs) for clock signal synchronization on the FPGA and on the board, and a

transceiver in this research and some of the major existing products is presented in Table 3.The waveforms have been included below.
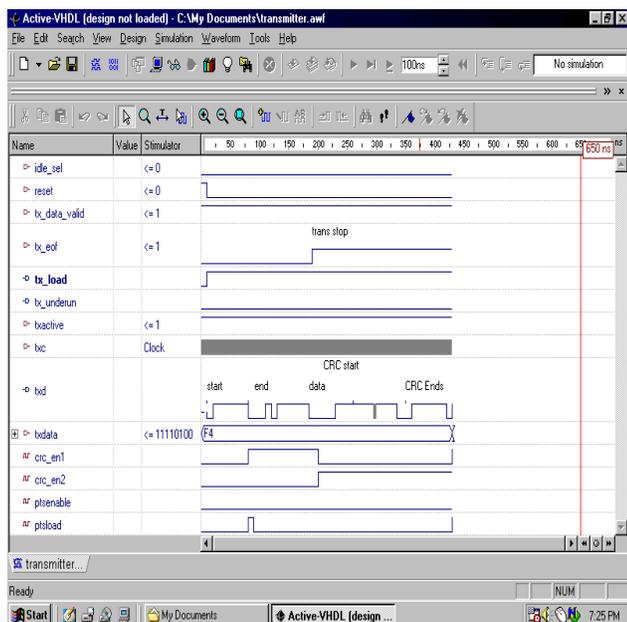
**Fig 7- Waveform for Transmitter simulated output**
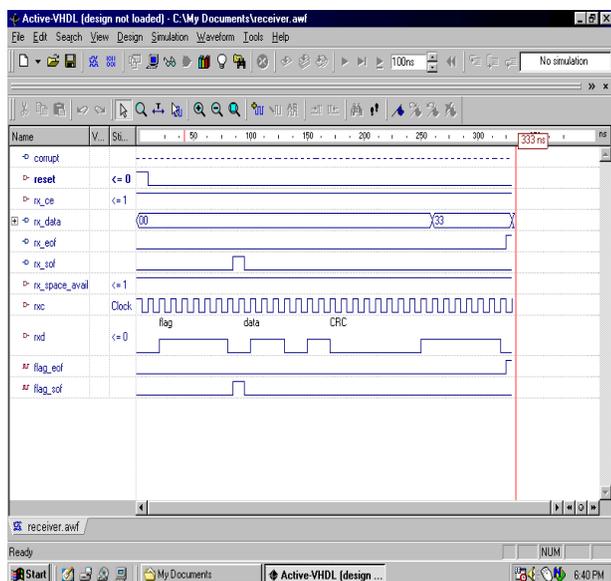
Fig 8-Waveform for Receiver simulated output:

4. CONCLUSION

HDLC Controllers are deployed within a number of different data networking applications. Being built on the capabilities of the Virtex family and due to its features and cost effectiveness. Hence it is uniquely poised to penetrate the ASSP marketplace. An FPGA based HDLC controller solution with efficient partitioning of hardware and software functions provides the necessary scalability and flexibility to handle all these applications and allows for tracking of new standards.

Thus through this project it is clear that to reduce the cost of design and to increase the probability of the IC design, FPGA design helps the designer in many ways. Thus it reduces the time consumption in designing as well as in testing the design. The coding written in VHDL language was checked using ALTERA simulator. The top level coding written for this paper is tested along with random stimulator values and the results obtained were up to the designer's expectation. The future enhancement in this paper is to implement this controller in real time application and also to accommodate more number of channels in the same design. Thus this controller can be used to provide communication by accommodating more number of channels using PCM highway technique.

5. FUTURE ENHANCEMENTS:

This paper can be used as Multi-channel HDLC controller with few modifications in the VHDL coding. The main difference single channel and multi channel HDLC controller is that a multiplexer and a demultiplexer are used in the transmitter and receiver respectively. The main advantage of this project is error correction using CRC-16 polynomial and through this controller we can achieve error free communication. Thus this project can be used in MODEMS where error detection part is performed by this HDLC controller.

Since, HDLC forms the basic standard for other protocols like LAPB, LAP, LLC, LAN etc., this controller can be used to generate frame for any of these protocols with slight modification in the coding part. This HDLC controller can also support ISDN frame format and thus the flexibility in this controller is very high.

REFERENCES:

[1] Guozheng Li Nanlin Tan State Key Lab. of Rail Traffic Control & Safety, Beijing Jiaotong Univ., Beijing, China " Design and Implementation of HDLC Protocol and Manchester Encoding Based on FPGA in Train Communication Network",

Information and Computing (ICIC), 2010 Third International Conference

[2] Lu, Y., Z. Wang, L. Qiao and B. Huanq, 2002. "Design and implementation of multi-channel high speed HDLC data processor," IEEE International Conference on Communications, Circuits and Systems, and West
Sino Expositions, 2: 1471-1475

[3] Amendola, A.M. Benso, A. Corno, F. Impagliazzo, L. Marmo, P. Prinetto, P. Rebaudengo, M. Sonza Reorda, M. CRIS, Napoli "Fault behavior observation of a microprocessor system through a VHDL simulation-based fault injection experiment Design Automation Conference, 1996, with EURO-VHDL '96 and Exhibition, Proceedings EURO-DAC '96, European ,16-20 Sep 1996

[4] Arshak, K. Jafer, E. McDonagh, D. Ibala, C.S. Univ. of Limerick, Limerick, "Modelling and simulation of wireless sensor system for health monitoring using HDL and Simulink mixed environment" Computers & Digital Techniques, IET Sept. 2007

[5] Gheorghiu, V., S. Kameda, T. Takagi, K. Tsubouchi and F. Adachi, 2008. "Implementation of frequency domain equalizer for single carrier transmission," In Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM '08.

[6] Jun Wang; Wenhao Zhang; Yuxi Zhang; Wei Wu; Weiguang Chang; Sch. of Electron. & Inf. Eng., Beihang Univ. (BUAA), Beijing, China "Design and implementation of HDLC procedures based on FPGA" , Anti- ounterfeiting, Security, and Identification in Communication, 2009. ASID 2009. 3rd International Conference, 20-22 Aug. 2009

[7] Meng, X. and V. Chaudary, 2009. "Boosting data throughput for sequence database similarity searches on FPGAs using an adaptive buffering scheme," Parallel Computing, 35(1): 1-11."

A Distributed Administration Based Approach for Detecting and Preventing Attacks on Mobile Ad Hoc Networks

Himadri Nath Saha , Prof. (Dr.) Debika Bhattacharyya , Prof.(Dr.) P. K. Banerjee

Abstract - Certain security attacks specific to Mobile Ad Hoc Networks (MANETs) such as black hole attacks, gray hole attacks and blackmail attacks and also flooding attacks are lethal in terms of hampering availability of network service. In this paper, we propose a protocol for detecting flooding, black hole, gray hole and blackmail attacks and taking measures against the nodes committing them. Our scheme is based on a concept of an underlying backbone network of administrator nodes that we assume to be trustworthy and honest throughout. These administrators have greater transmission and reception range than the general nodes in the MANET and have the power to take corrective actions on the basis of the reports sent by the other nodes. The association of these administrator nodes is dynamically increased to ensure better network coverage by upgrading certain general nodes to become administrators subject to certain constraints such as the transmission and reception range and the performance over a sufficiently large period of time. We have modeled a possible life cycle for a general node in the network and have shown how our protocol unlike the existing ones is resilient and conservative while taking actions against any node emphasizing that an honest node should not be penalized by mistake. We give an elaborate description of the procedures and how they lead to detection of the attacks.

Keywords: Black hole attack, Blackmail attack, MANET,Gray hole attack, Flooding, Watch Node, Adminstrator.

———————————— ◆ ————————————

1. INTRODUCTION

THE security of communication in ad hoc wireless networks is very important and at the same time is much more challenging than it is for structured networks. Security attacks on MANETs can be broadly classified into *active* and *passive* attacks. In passive attacks, the malicious nodes attempt to obtain information form the network without disrupting the network operations. On the other hand, active attacks hamper network operations and can be carried out by nodes that are external or internal to the network. Internal attacks are harder to tackle as the nodes carrying them out are already accepted as a part of the network and are associated with other nodes in the network through already established trust relationships. We are concerned about such internal nodes that carry out active attacks like flooding, black hole, gray hole and blackmail attacks after establishing themselves in the network. Before we proceed to deal with the

detection and prevention of these attacks, it is important to thoroughly understand these attacks and their characteristics.

Flooding attack: In flooding attack, a malicious node sends a huge number of junk packets to a node to keep it busy with an aim to prevent it from participating in other activities in the network. This can lead to an obvious disruption of network availability as the nodes communicating with the victim will not be attended. Apart from this threat, other complicacies can also be generated as follows:

- Two malicious nodes can cooperate to carry out an attack where one floods an honest node in their vicinity while the other carries out a packet dropping (black hole) attack thereby preventing the honest node from detecting the black hole attack being carried out by the other malicious node.

- In critical situations, where a node comes up and is waiting for receiving its identity from the other existing nodes, a malicious node can flood this node or the neighbors to delay the acceptance of the new node in the network.

These are two small examples out of many possible which justifies the need for a protocol which detects and takes action against the nodes trying to flood the network.

Black hole attack: In black hole attack, a malicious node upon receiving a route request packet from a node replies by sending a false routing reply to the sending node to misguide it to send the data to it and then it drops the data packet. This is the simplest way a black hole attack can be carried out and it is trivially easy to detect the node that is dropping all packets and consequently isolate it in the network. Let us look into a more complex scenario. In a situation where a group of nodes cooperate to create a black hole where the data packet is transferred and retransferred within the black hole until it runs out of its *time to live* (TTL) and gets eventually dropped without causing the node dropping it to be blamed anyhow. This is how cooperative black hole attack is carried out and it is increasingly challenging to detect the chain of nodes responsible and take corrective measures.

Gray hole attack: In a gray hole attack, the malicious nodes are harder to be detected as they selectively drop packets. Such a malicious node can pretend to be honest over a time in the network observing a pattern in the traffic flow. For instance, say after a node comes into the network and an existing node receives the request for identity from the new node, then it is expected that the existing node will reply with an identification data to the new node. At this moment, a malicious node can drop packets from any nodes meant for the new node thereby preventing or delaying participation of the new node in the network.

Blackmail attacks: In a blackmail attack, or more effectively a cooperative blackmail attack, malicious nodes complain against an honest node to make other nodes that need to send data to believe that routing through the victim is harmful. Such attacks can prevent senders from choosing the best route to the destination thereby hampering efficiency and throughput in the network.

Having discussed the threat areas we now introduce our approach to fight against these attacks. The crude definition of MANET calls for a cluster of mobile nodes with equal or different computing power but with equal status inter-communicating without taking the aid of any central authority whatsoever. An ideal MANET as per the basic definition incorporates only peer to peer communication.

In other words such networks are structure less.

We however have been deeply influenced by the concept of using a logical structure over on infrastructure less network as in [1].

We emphasize on deploying of MANET for a definite purpose such as military activities, fighting disaster in calamity struck areas and so on. So it is not unjust to assume that there exists a logical authority in the form of an individual or a group who sets up the communication network for a purpose and that they will always be honest. On the basis of this assumption we mark these nodes as administrator nodes that place themselves in positions so as to ensure maximum network coverage. They have large transmission and reception power than the general nodes that participate in the network activities. We make the general nodes go through four phases in their lives in the network, namely, WHITE, GRAY, BLACK and BLUE. As we have said earlier, our scheme takes special care to avoid rash decisions taken on nodes to prevent honest nodes getting misjudged as malicious. A WHITE node is one which is honest with a high probability and is the default phase of a new node in the network. A node which is under suspect is made GRAY and a GRAY node which does not improve its behavior is made a BLACK node and its isolation is effected. A WHITE node that has transmission and reception powers comparable to that of the administrators and has been honest for a sufficiently large amount of time can be upgraded to the status of BLUE nodes. We have given the general nodes the power to watch the activities of other nodes and judge independently whether other nodes in its vicinity are carrying out malicious activities. On detecting such activities, the general nodes can send out complains to the administrators who have the authority to take corrective measures on the basis of the received complains.
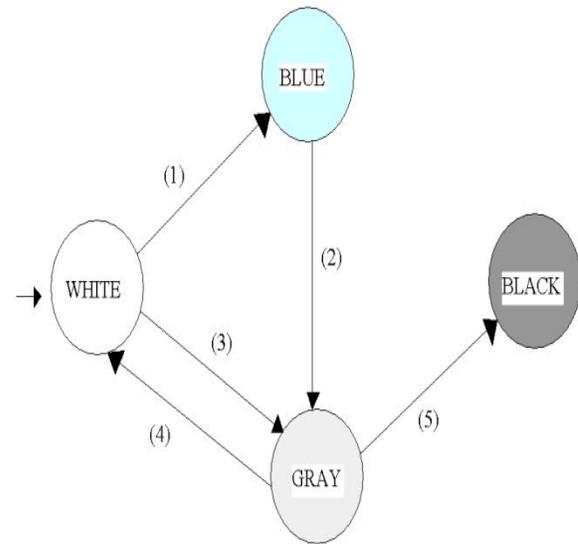

Fig 1. State diagram representing the phases in the life cycle of a general node in the network

A node is WHITE when it comes into the network. It can make transition (1) provided its hardware capabilities are as strong as the administrators and it has shown good behavior over a sufficient amount of time. However, we are very conservative about giving unlimited power to such a node. One of the existing administrators has to accept it as a BLUE node after which it can take corrective measures regarding security in the network. However any discrepancy of decision making if found out with any other initial administrators, the BLUE node is penalized by making the transition (2). A white node under suspicion is made to undergo transition (3). If a gray node's behavior continues to be suspicious then it is made to undergo transition (5). However if the GRAY node shows good behavior over a sufficiently long period of time, there is a high chance that the suspicion was erroneous and hence transition (4) is effected.

The rest of the paper proceeds as follows:

In section 2, we discuss related works in this area and the concepts which have influenced our work. Following this, we have presented the methodology of our work in section 3 and in section 4 we have concluded citing the scopes for further works in this area.

The improvements we have sought in our scheme are efficiency and robustness related apart from our concern towards honest nodes not having to face penalty due to their behavior during abnormal traffic loads or movements in the network.

2. Related works

Agrawal et. al. [1] have proposed a protocol on the basis of the logical backbone network over infrastructure less ad hoc networks. Their protocol also assumes that the strong nodes that form the backbone are honest. They have utilized the concept of sending data in terms of small equal sized packets rather than a continuous stream. According to their scheme however, regular nodes are not capable of monitoring the activities of other nodes which according to them strengthens the chances of black hole attacks. We contradict on this issue and our scheme allows general nodes to monitor the activities of other nodes. The decisive actions however can be taken only by the administrator nodes. Thus we have distributed the monitoring work among all the nodes which adds more generality to our protocol. From [2], we get the concept of guard nodes to monitor the activities of other nodes within their range. Razak, Furnell, Clarke and Brooke [3] describes a two tier Intrusion Detection System using a friend approach which is capable of minimizing the impact of colluding blackmail attacks in the system. Another algorithm [4] is not robust with respect to cooperative malicious nodes in the network. [5] presents another compute intensive algorithm which cannot tackle gray hole attacks. In [6] we get a similar concept of backbone network that we have incorporated in our scheme.

3. Methodology

For our convenience, we have visualized the MANET area on X-Y two dimensional plane where each node is aware of the coordinates (x,y) they are currently placed in. Moreover, we have assumed that the nodes have enough hardware capabilities to judge the coordinates of a node from which they are able to receive packet in a single hop. Next, we give procedures that the general nodes and the administrators execute followed by descriptions about them.

List of Terms

GEN_SEND_DATA()	A procedure used by general nodes while sending data
P	Packets
ACK	Acknowledgements
ALERT	An alert signal to caution the administrators of a possible black or gray hole attack
GEN_WATCH_NODES()	A procedure used by the general nodes to monitor the activities of other nodes in its vicinity

SND	Original sender of a packet
PREV	The node from which the packet has arrived
NEXT	The node to which the packet goes on the next hop
DST	The final destination of the packet
τ	Timestamp of the last noticed packet between a sender destination pair
f	Frequency of packet transmission between a sender destination pair
f_{max}	The threshold frequency beyond which we conclude chances of flooding attack
T_{safe}	The safe time limit beyond which packets transmitted between same sender destination pair are not subjected to suspicion of flooding attack
b	One flag bit to give suspected nodes a chance before registering complain

	against them
TTL	Time to live of a packet
TIMEOUT	A clock based time out event
DOUBT_COUNT_BLACK	A counter to store the doubts that a node is carrying out black hole attack
DOUBT_COUNT_GRAY	A counter to store the doubts that a node is carrying out gray hole attack
DOUBT_TOLERANCE	The threshold limit beyond which complain is registered against a node
L	The complain list
INIT_GEN_NODE()	A procedure which describes how a general node comes up and becomes a part of the network
PENALIZE(node)	A procedure to effect transitions as shown in Fig. 1
DETECT_BLACK_GRAY()	A procedure incorporated by the administrator nodes to detect cooperative black and gray hole attacks
ACT_ADMIN_NODE()	A procedure which describes the active

	life of an administrator node
ACT_GEN_NODE()	A procedure which describes the active life of a general node
ADMIN_SEND_DATA()	A procedure used by an administrator while sending data

Procedures for a General Node

GEN_SEND_DATA()

1. Decompose the message to be sent into small and equal sized packets, say P[1], P[2], ..., P[N].

2. Select next hop destination using routing protocol.

3. Set P[0] to a random nonce.

4. Send P[0] first, set timer and wait for ACK.

5. If ACK received within timeout, then repeat step 3 with P[1] and so on till P[N] and then go to step 7.

6. If timeout occurs, then possibility of black or gray hole attack.

7. Send ALERT in Broadcast mode.

8. Return.

In line 3 of the above procedure, the random nonce is used to verify whether at present the discovered route to the destination is proper. We believe that the initial stage of sending data is more critical in terms of becoming victims of black or gray hole attacks. However whenever we sense failure of packet delivery we alert the administrators in the neighborhood who subsequently trigger a detection procedure.

GEN_WATCH_NODES()

1. Start timer.

2. If there is packet being sensed, then

3. Capture packet P at time t.

4. Get Header information <SND, PREV, NEXT, DST>

5. If there is no record of the form <SOURCE, DESTINATION, TIMESTAMP, FREQUENCY> as <PREV, NEXT, τ, f> in the audit file, then add a record by making τ=t and f=1.

6. Else if f<f_{max}, then

7. t-τ ≤ T_{safe} →f=f+1

8. T_{safe} < t-τ ≤ 2T_{safe} and f > 0 → f=f-1 (slow retreat)

9. t-τ > 2T_{safe} and f > 0 → f=[f/2] (fast retreat)

10. Else if f ≥ f_{max}, then

11. If bit b=1, then append PREV into complain list L and reset b to 0.

12. Else if bit b=0 then set b=1 and refresh f to 0.

13. End if.

14. Update last entry of this packet in the audit file.

15. End if.

16. If TTL of P > 1, then

17. Start timer

18. If within TIMEOUT a packet from NEXT meant for some node, say x, does not arrive, then DOUBT_COUNT_BLACK is increased by 1 for node NEXT and packet dropping record is saved.

19. If for a node J, DOUBT_COUNT_BLACK exceeds DOUBT_TOLERANCE, then

20. If bit b=1 then, append J to complain list L, reset DOUBT_COUNT_BLACK to 0 and bit b to 0.

21. Else set b to 1 and reset DOUBT_COUNT_BLACK to 0.

22. End if.

23. If packet from NEXT meant for x had been dropped earlier then, double DOUBT_COUNT_GRAY.

24. If for a node J, DOUBT_COUNT_GRAY exceeds DOUBT_TOLERANCE, then

25. If bit b=1 then, append J to complain list L, reset DOUBT_COUNT_GRAY to 0, bit b to 0 and remove corresponding packet dropping record.

26. Else set b to 1 and reset DOUBT_COUNT_GRAY to 0 and remove corresponding packet dropping record.

27. End if.

28. Go back to step 1.

In the above procedure, lines 5 to 15 deal with the detection of a flooding attack. If the frequency of packet sending crosses a threshold between the same sender destination pair, then we start doubting a flooding attack. However, a sudden burst of data can be there at some instant. So such a behavior might not be caused by a flooding attack. Questioning the practicality of a flooding attack, we have concluded that a node with the aim of flooding the network will show such a behavior frequently if not continuously. So we give such a node a second chance with the help of the flag bit b. After the second chance, we make sure that the node makes a complain about the suspected node which will be attended to by the administrators. Lines 16 to rest deal with black and gray hole attacks. Whenever there is a packet being dropped whose TTL has not shrunk to 0, we start suspecting a black hole or gray hole somewhere. For black hole number of packets getting dropped will be much higher than for gray hole attacks. So we increase the suspicions of a black hole, that is, DOUBT_COUNT_BLACK linearly while we increase suspicions of a gray hole, that is DOUBT_COUNT_GRAY exponentially. Complains are registered once either of the doubt counts exceed a threshold DOUBT_TOLERANCE.

ACT_GEN_NODE()

1. Set timer.

2. If there is data to be sent, then create thread GEN_SEND_DATA() and execute it.

3. If there is an incoming packet to be routed, then use routing protocol to find next hop destination H and forward the packet to H.

4. Create thread GEN_WATCH_NODES() and execute it.

5. If NOT TIMEOUT, then go to step 2.

6. Create complain messages about the nodes in the complain list L.

7. Broadcast complain message.

8. Refresh timer and go back to step 1.

Here, we take care that we do not congest the network with complain messages. So we wait for a time over which we accumulate the complains in a list L and then send the complains together.

INIT_GEN_NODE()

1. Ascertain the current coordinates, say (x,y).

2. Create packet for IDENTITY_REQUEST and broadcast packet.

3. Set timer.

4. If IDENTITY_REPLY received, then go to step 7.

5. If NOT TIMEOUT, then go to step 4.

6. Move to a random (x′,y′) and repeat from step 2.

7. ACT_GEN_NODE()

(*Comment:* In this paper, we have not worked on dynamic identity allocation algorithms and we assume that all nodes possess valid identities obtained correctly and efficiently).

Procedures for an administrator node

PENALIZE(node)

1. If node is in WHITE list, then move its entry to GRAY list.

2. Else if node is in GRAY list, then move its entry to BLACK list.

3. Else if node is in BLUE list, then move its entry to GRAY list.

4. Else if node is in BLACK list, IGNORE.

5. End if

6. Broadcast updated message to other administrator nodes.

The above procedure makes use of the state diagram and its conditional transitions as presented in Fig. 1.

DETECT_BLACK_GRAY()

1. Note the SENDER and the DESTINATION of the failed packet.

2. Create more packets with the same SENDER DESTINATION pair each consisting of random nonce to aid in the detection

3. Send such a packet and store the hops.

4. If at any stage, a node is found to route the same packet twice in a circle, that node is subjected to PENALIZE(node) then and there.

5. Else if a single node is dropping packets, then a similar procedure as in GEN_WATCH_NODES() is incorporated and if criteria for attack are satisfied then PENALIZE(node) is effected.

In line 4, the idea we have presented is that for a cooperative black hole or gray hole attack,

more than one node form a closed loop in which they go on forwarding packets until TTL becomes 0. So a cycle detection not leading to packet dropping then and there is surely a malicious activity and needs to be taken care of immediately. The rest is similar to monitoring of the nodes by the general nodes.

ADMIN_SEND_DATA()

1. Decompose the message to be sent into small and equal sized packets, say P[1], P[2], ..., P[N].

2. Select next hop destination using routing protocol.

3. Set P[0] to a random nonce.

4. Send P[0] first, set timer and wait for ACK.

5. If ACK received within timeout, then repeat step 3 with P[1] and so on till P[n] and then go to step 8.

6. If timeout occurs, then possibility of black or gray hole attack.

7. Send a self ALERT to DETECT_BLACK_GRAY() thread.

8. Return.

The above procedure is similar to the one used by general nodes for sending data with the difference that it triggers its own action taking procedures rather than sending out complains.

ACT_ADMIN_NODE()

1. If there is data to be sent, then create thread ADMIN_SEND_DATA() and execute it.

2. If there is an incoming packet to be routed, then use routing protocol to find next hop destination H and forward the packet to H.

3. Probe on the forwarded packets to ensure end to end delivery. If not, then trigger DETECT_BLACK_GRAY() thread.

4. If there is an incoming complain list containing a complain about a node, say v, from u, then

5. If u is in WHITE list, then complain[v] is increased by 2.

6. Else if u is in GRAY list, then complain[v] is decreased by 1.

7. Else if u is in BLACK list, the ignore the complain.

8. Else if u is in BLUE list, then perform self detection to tally the results.

9. If Decisions match PENALIZE(v).

10. Else PENALIZE(u) keeping provisions for BLACKMAIL_DOUBT exceeding DOUBT_TOLERANCE..

11. End if

12. End if

13. If complain[v] exceeds threshold, then PENALIZE(v) and refresh complain[v].

14. End if.

15. Go back to step 1.

Lines 5 to 8 deal with giving priorities to the complaints from other nodes. If the node complaining is WHITE then higher priority is given to the complain than if the node

complaining is GRAY. We have not written other procedures that an administrator performs that are not related to the security aspects we have dealt with in our scheme. It is worth mentioning that like the general nodes, the administrators also run a similar monitoring over other nodes. The difference is that an administrator can watch over a huge area while the general nodes cannot. Then naturally a question might arise that what is then the need for the general nodes to watch other nodes' activities? The answer to this is however trivially simple. As the area of observation is large, so the number of nodes to be watched is much larger for an administrator. Consequently the process is slower. So, the general nodes' monitoring over other nodes increases the optimality of our scheme.

4. Conclusion and future work

We conclude by mentioning that we have worked upon the basic observation that the malicious nodes will continue showing malicious activities. Therefore we can differentiate between them and the nodes that are forced under situation to show behaviors which can be suspected to be malicious. Our protocol stands out among the rest in its conservative approach and avoidance to taking rash decisions. Moreover we keep provisions of nodes suspected before up to the GRAY level to be cleared of the suspicions in case improved behavior is noticed. These are extra points that we have mentioned about the protocol and which does not appear in the procedures. This is a deliberate attempt as we want to focus on the detection aspects in the procedures. We have tried to avoid clogging the procedures with supplementary features.

It is worth mentioning that there is scope of further research in this area, some of which are:

- The protocol can be further extended to incorporate the detection of worm hole attack as well

- Research on optimality analysis of our protocol is welcome.

5. References

[1] P. Agrawal, R.K Ghosh, S.K. Das, "Cooperative black and gray hole attacks in mobile ad hoc networks", ICUIMC'08.

[2] I. Khalil, S. Bagchi, N.B. Shroff, "MOBIWORP: Mitigation of the Wormhole attack in mobile multihop wireless network", Elsevier, 2007.

[3] S.A Razak, S. Furnell, N. Clarke, P. Brooke, "A Two-Tier Intrusion Detection System for Mobile Ad Hoc Networks – A Friend Approach", Springer-Verlag Berlin Heidelberg 2006.

[4] H. Deng, Wei Li, D.P Agrawal, "Routing Security in Wireless Ad Hoc Network", IEEE Communications Magazine, vol. 40, pp. 70-75, 2002.

[5] S. Ramaswamy, H. Fu, M. Sreekanaradhya, J. Dixon, K. Nygard, "Prevention of Cooperative Blackhole Attack in Wireless Ad Hoc Networks", Proceedings of 2003 International Conference on Wireless Networks, Las Vegas, Nevada, USA, pp. 570-575.

[6] I. Rubin, A. Behzad, R. Zhang, H. Luo, E. Caballero, "TBONE: A Mobile-Backbone Protocol for Ad Hoc Wireless Networks", Proceedings of IEEE Aerospace Conference, 2002, vol. 6, pp. 2727-2740.

[7] A.A Cardenas, T. Roosta, S. Sastry, "Rethinking security properties, threat models,

and the design space in sensor networks: A case study in SCADA systems", Elsevier, 2009.

[8] P.R Kumar, L.L Xie, "Ad Hoc Wireless Networks: From Theory to Protocols", Ad Hoc Wireless Networking, Kluwer Academic Publishers, 2004, pp. 175-195.

[9] W. Lou, Y. Fang, "A Survey of Wireless Security in Mobile Ad Hoc Networks: Challenges and Available solutions", Ad Hoc Wireless Networking, Kluwer Academic Publishers, 2004, pp. 319-364.

[10] C.S.R. Murthy, B.S. Manoj, "Ad Hoc Wireless Networks", Pearson, 2004.

[11] J. Schiller, "Mobile Communications", Pearson, 2003.

Color Image Segmentation – An Approach

S.Pradeesh Hosea, S. Ranichandra, T.K.P.Rajagopal

Abstract— The literature on the color image segmentation is very large and it has been delimited to review some important literature to trace the core issues. On the basis of the identified issues, objectives were drawn to prosecute a fresh study in the color image segmentation. This Literature review helps researcher to understand various techniques, themes, methodologies, approaches and controversies so for applied for color image segmentation. The algorithm combining color and texture information for the segmentation of color images. The algorithm uses maximum likelihood classification combined with a certainty based fusion criterion. It was validated using mosaics of real color textures presented in Color and texture fusion: application to aerial image segmentation.

Index Terms— color images, segmentation, image processing.

———————————————— ◆ ————————————————

1 INTRODUCTION

H D. Cheng, X. H. Jiang, Y. Sun, Jingli Wang in 2001, described the concept of monochrome image segmentation approaches operating in different color spaces such as histogram thresholding, characteristic feature clustering, edge detection, region-based methods, fuzzy techniques, neural networks in Color image segmentation [2].

In 2001, F.Kurugollu, B.Sankur and A.E. Harmanci proposed the techniques of multiband image segmentation based on segmentation of subsets of bands using multithresholding followed by the fusion of the resulting segmentation "channels". For color images the band subsets are chosen as the RB, RG and BG pairs, whose two-dimensional histograms are processed via a peak-picking algorithm to effect multithresholding is present in Color image segmentation using histogram multithresholding and fusion [3].

The method of simple and efficient implementation of Lloyd's k-means clustering algorithm, which we call the filtering algorithm. This algorithm is easy to implement, requiring a kd-tree as the only major data structure is present in an efficient k-means clustering algorithm. It was proposed by T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y Wu in 2002[4].

Chi zhang, P. Wang in 2002, described the concept based on K-means algorithm in HSI space and has the advantage over those based on the RGB space. Both the hue and the intensity components are fully utilized. In the process of hue clustering, the special cyclic property of the hue component is taken into consideration is present in A New Method of Color Image Segmentation Based on Intensity and Hue Clustering[5].

The method of a color image segmentation system that performs color, clustering in a color space followed by color region segmentation in the image domain. The region segmentation algorithm merges clusters in the image domain based on color similarity and spatial adjacency is present in Color Image Segmentation in the Color and Spatial Domains. It was proposed by Tie Qi Chen', Yi L. Murphey', Robert Karlsen and Grant Gerhartd in 2002[6].

Faguo Yang and Tianzi Jiang in 2003, described the concept of a novel pixon-based adaptive scale method for image segmentation. The key idea of our approach is that a pixon-based image model is combined with a Markov random field (MRF) model under a Bayesian framework is present in Pixon-Based Image Segmentation With Markov Random Fields.[7].

The method to split colox information is the image to be segmented. Hence, this is a blind colour image segmentation method. It consists of four subsystems: preprocessing, cluster detection, cluster fusion and postprocessing is present in A four-stage system for blind colour image segmentation. It was proposed by Ezequiel López-Rubio, José Muñoz-Pérez, José Antonio Gómez-Ruiz in 2003[8].

Dmitriy Fradkin, Ilya Muchnik in 2004, described the concept to constructing hierarchical classifiers using cluster analysis and suggests new methods and improvements in each of these approaches. We also suggest a new method for constructing features that improve classification accuracy is present in A Study of K-Means Clustering for Improving Classification Accuracy of Multi-Class SVM [9].

Cheolha Pedro Lee in 2005, described the concept based on the statistics of image intensity where the statistical information is represented as a mixture of probability density functions defined in a multi-dimensional image intensity space. Depending on the method to estimate the mixture density functions, three active contour models are proposed: unsupervised multi-dimensional histogram method, half-supervised multivariate Gaussian mixture

density method, and supervised multivariate Gaussian mixture density method is present in Robust Image Segmentation using Active Contours [10].

In 2007, Chris Vutsinas described the concept of Image Segmentation: K-Means and EM Algorithms. In this method, two algorithms for image segmentation are studied. K-means and an Expectation Maximization algorithm are each considered for their speed, complexity, and utility. Implementation of each algorithm is then discussed [11].

Ahmed REKIK, Mourad Zribi, Ahmed Ben Hamida, Mohammed Benjelloun in 2007, described the concept of Image analysis, usually, refers to a process of images provided by a computer in order to find the objects within the image. . It consists of subdividing an image into its constituent parts as well as extracting them, is present in the Review of satellite image segmentation for an optimal fusion system based on the edge and region approaches [12].

In 2008, Milind M. Mushrifand Ajoy K. Ray introduced the method of a new color image segmentation algorithm using the concept of histon, based on Rough-set theory, is presented in Color image segmentation: Rough-set theoretic approach. The histon is an encrustation of histogram such that the elements in the histon are the set of all the pixels that can be classified as possibly belonging to the same segment. In rough-set theoretic sense, the histogram correlates with the lower approximation and the histon correlates with upper approximation [13].

The concept of Fusion of multispectral image with a hyperspectral image generates a composite image which preserves the spatial quality from the high resolution (MS) data and the spectral characteristics from the hyperspectral data , is presented in Performance analysis of high-resolution and hyperspectral data fusion for classification and linear feature extraction. It was proposed by Shashi Dobhal in 2008[14].

Sheng-xian Tu, Su Zhang, Ya-zhu Chen, Chang-yan Xiao and Lei Zhang in 2008, a new hierarchical approach called bintree energy segmentation was presented for color image segmentation. The image features are extracted by adaptive clustering on multi-channel data at each level and used as the criteria to dynamically select the best chromatic channel, where the segmentation is carried out. In this approach, an extended direct energy computation method based on the Chan-Vese model was proposed to segment the selected channel, and the segmentation outputs are then fused with other channels into new images, from which a new channel with better features is selected for the second round segmentation. This procedure is repeated until the preset condition is met. Finally, a binary segmentation tree is formed, in which each leaf represents a class of objects with a dis-

tinctive color [15].

A novel method of colour image segmentation based on fuzzy homogeneity and data fusion techniques is presented. The general idea of mass function estimation in the Dempsteri-Shafer evidence theory of the histogram is extended to the homogeneity domain. The fuzzy homogeneity vector is used to determine the fuzzy region in each primitive colour, whereas, the evidence theory is employed to merge different data sources in order to increase the quality of the information and to obtain an optimal segmented image. Segmentation results from the proposed method are validated and the classification accuracy for the test data available is evaluated, and then a comparative study versus existing techniques is presented. It was described by Salim Ben Chaabane, Mouniri Sayadi, Farhat Fnaiech and Eric Brassart in 2009[16].

Fahimeh Salimi, Mohammad T. Sadeghi in 2009, introduced a new histogram based lip segmentation technique is proposed considering local kernel histograms in different illumination invariant colour spaces. The histogram is computed in local areas using two Gaussian kernels; one in the colour space and the other in the spatial domain. Using the estimated histogram, the posterior probability associated to non-lip class is then computed for each pixel. This process is performed considering different colour spaces. A weighted averaging method is then used for fusing the posterior probability values. As the result a new score is obtained which is used for labeling the pixels as lip or non-lip. The advantage of the proposed method is that the segmentation process is totally unsupervised [17].

In 2009, Damir Krstinic, Darko Stipanicev, Toni Jakovcevic described a pixel level analysis and segmentation of smoke colored pixels for the automated forest fire detection. Variations in the smoke color tones, environmental illumination, atmospheric conditions and low quality of the images of wide outdoor area make smoke detection a complex task. In order to find an efficient combination of a color space and pixel level smoke segmentation algorithm, several color space transformations are evaluated by measuring separability between smoke and non-smoke classes of pixels [18].

The concept of a new color thresholding method for detecting and tracking multiple faces in video sequence. The proposed method calculates the color centroids of image in RGB color space and segments the centroids region to get ideal binary image at first. Then analyze the facial features structure character of wait-face region to fix face region. The novel contribution of this paper is creating the color triangle from RGB color space and analyzing the character of centroids region for color segmenting. It was proposed by Jun Zhang, Qieshi Zhang, and Jinglu Hu in 2009[19].

REFERENCES

[1] M. -P. Dubuisson-Jollyand A. Gupta, Color and texture fusion: application to aerial image segmentation and GIS updating, Volume 18, Issue 10, July 2000, Pages 823-832.

[2] H. D. Cheng, X. H. Jiang, Y. Sun, Jingli Wang, Color image segmentation: advances and prospects Vol. 34, No. 12. (December 2001), pp. 2259-2281.

[3] F. Kurugollu, B. Sankur and A. E. Harmanci (2001) Color image segmentation using histogram multithresholding and fusion Volume 19, Issue 13, 1 November 2001, Pages 915-928

[4] T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, IEEE Trans. Pattern Analysis and Machine Intelligence, 24 (2002), 881-892.

[5] Chi zhang, P. Wang, A New Method of Color Image Segmentation Based on Intensity and Hue Clustering, volume 3 2002.

[6] Tie Qi Chen', Yi L. Murphey', Robert Karlsen 2, and Grant Gerhartd were present the method of Color Image Segmentation in the Color and Spatial Domains, 2002.

[7] Faguo Yang and Tianzi Jiang, Pixon-Based Image Segmentation, With Markov Random Fields, VOL. 12, NO. 12, DECEMBER 2003.

[8] Ezequiel López-Rubio, José Muñoz-Pérez, José Antonio Gómez-Ruiz. A four-stage system for blind colour image segmentation. Volume 10, Number 2/2003 Pages127-137.

[9] Dmitriy Fradkin, Ilya Muchnik, A Study of K-Means Clustering for Improving Classification Accuracy of Multi-Class SVM, DIMACS Technical Report 2004-02 April 2004.

[10] Cheolha Pedro Lee, Robust Image Segmentation using Active Contours: Level Set Approaches. 2005.

[11] Chris Vutsinas, Image Segmentation: K-Means and EM Algorithms, 2007.

[12] Ahmed REKIK, Mourad Zribi, Ahmed Ben Hamida, Mohammed Benjelloun (2007) Review of satellite image segmentation for an optimal fusion system based on the edge and region approaches.Vol.7 No.10, October 2007.

[13] Milind M. Mushrifand Ajoy K. Ray Color image segmentation: Rough-set theoretic approach Volume 29, Issue 4, 1 March 2008, Pages 483-493.

[14] Shashi Dobhal, Performance analysis of high-resolution and hyperspectral data fusion for classification and linear feature extraction. January, 2008.

[15] Sheng-xian Tu, Su Zhang, Ya-zhu Chen, Chang-yan Xiao and Lei Zhang, A bintree energy approach for colour image segmentation using adaptive channel selection, in February 2008.

[16] Salim Ben Chaabane, Mounir Sayadi, Farhat Fnaiech and Eric Brassart, Colour Image Segmentation Using Homogeneity Method and Data Fusion Techniques from Hindawi Publishing Corporation in may 2009.

[17] Fahimeh Salimi, Mohammad T. Sadeghi, Decision Level Fusion of Colour Histogram Based Classifiers for Clustering of Mouth Area Images from International Conference on Digital Image Processing in 2009.

[18] Damir Krstinic, Darko Stipanicev, Toni Jakovcevic, HISTOGRAM-BASED SMOKE SEGMENTATION IN FOREST FIRE DETECTION SYSTEM, ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2009, Vol.38, No.3.

[19] Jun Zhang, Qieshi Zhang, and Jinglu Hu, RGB Color Centroids Segmentation (CCS) for Face Detection, ICGST-GVIP Journal, ISSN 1687-398X, Volume (9), Issue (II), April 2009.

Integer Programming Model for Integrated Planning of Solid Waste Management in Jaipur

Archana Gupta, D. C. Sharma

Abstract--Rajasthan's economic growth has stimulated urbanization, but lack of commensurate investment in urban infrastructure and services has resulted in an overall deterioration of urban quality of life. Jaipur is the capital of the largest state of country, having an inefficient, outdated and unscientific waste management system. Jaipur socio-economic development potential relies on sound environmental management so that tourism may grow and become sustainable. This paper attempts to assess the existing state of municipal solid waste management (MSWM) in Jaipur city with the aim of identifying the main obstacles to its efficient and prospects for improvisation of solid waste management system in the city.

Keywords- RDF-refuse derive fuel, MSWM- municipal solid waste management, Jmc- Jaipur municipal -corporation, MT- metric tons.

1. INTRODUCTION

Maharaja Jai Singh second founded Jaipur city in the year 1728.It is a uniquely planned city with rich architectural heritage which attract to foreign tourist and known as the "Pink city". But environment degradation and lack of basic amenities threaten its viability. Solid waste management system of a city has vital role on quality of life. As per 2001 census the population of Jaipur city was 23 lacks approx. and as per Jmc's figure it will increase by the rate of 2.8% per annum so the present population of Jaipur in 2010 is 29 ,25,663.It is observed that bigger the size of the city (population and density wise) , greater is the quantity of waste generated. As per information received from various sources, Jaipur city generates 1000-1100 MT of waste per day. There by usual everyday practice maximum solid waste goes in to dumping sites without any recycling and separation process. On an average more than 60% of the population disposed off their solid waste on streets, open spaces, drains, lanes, and storm water drainage. The system has resulted unhygienic conditions in the city.

So presently approx. 600-650 Ton/day of waste is crudely dumped at mathuradaspura dumping sites and 400-450 Tons/day at sevapura dumping sites without segregation and reuse.

Presently we have only one RDF plant based on BOOT basis by ms.Grasim India Ltd. At Lagadiyawas, one bio-medical treatment plant at Khorarupadi on PPP run by Instromedics India private Ltd. And two open dumping spaces located at mathuradaspura and Sevapura sites. As per gathered information approximate 15% of total waste is used in RDF plant without segregation but it is not in everyday practice. The quantity of bio-medical waste expected to be generated 4 to 4.5MT in which approx.800kg.glassware has to be disinfected chemically, and only 0.15 MT is autoclaving and about 2.8 MT burned and disposed off at the site of plant. In Jaipur proper sanitary landfill has not been developed but one is under construction process at Lagadiyawas. Not proper plannining of collection and transportation method followed. Some time the collected waste of some wards goes to transfer station some time it will go to any dumping sites either sevapura or mathuradaspura.

So Proper planning and scheduling must be required in the present situation and we have to think with increasing emphasis on resource recovery of useful commodities from the waste (recycling), utilization of waste as a substitute for fuel in power plant , composting and other

advance technologies. In this paper our aim is to suggest the better way to improve the present system by implementing some changes in collection and transportation method, by applying segregation method, and empirically establish the existence of economics of scale for vermicular compost plant, recycling plant, mechanical compost plant, incinerator and one more sanitary landfill. So proposed model is used as an effective tool for efficient management when planner has made decision regarding political, social, and other practical aspects of the problem which are difficult to quantify. In our scientific optimization model we have total two transfer stations. MSW from a specific region should be firstly collected at different nodes, and by choosing suitable roots must be sent to a transfer station, then there it will sorted out by workers and allocated to various waste treatment plants like mechanical compost plant, vermicular compost plant RDF plant, recycling, incinerator or landfills. Biomedical waste should be collected from all hospitals, dispensaries, clinics and diagnostic centers and directly sent to bio- medical treatment plant. With these versatile programming techniques, different waste diversion rates under multiple policy scenario and complex uncertainties could be analyzed which will be very helpful to reduce quantity of disposal waste in a better economic and more sustainable way.

The current practice MSWM in all urban centers of the country is biased towards achieving 100% collection and its subsequent disposal with partial or no treatment/processing. Therefore an urgent need to shift the paradigm from open cycle to close cycle of waste. (given in fig 1.)
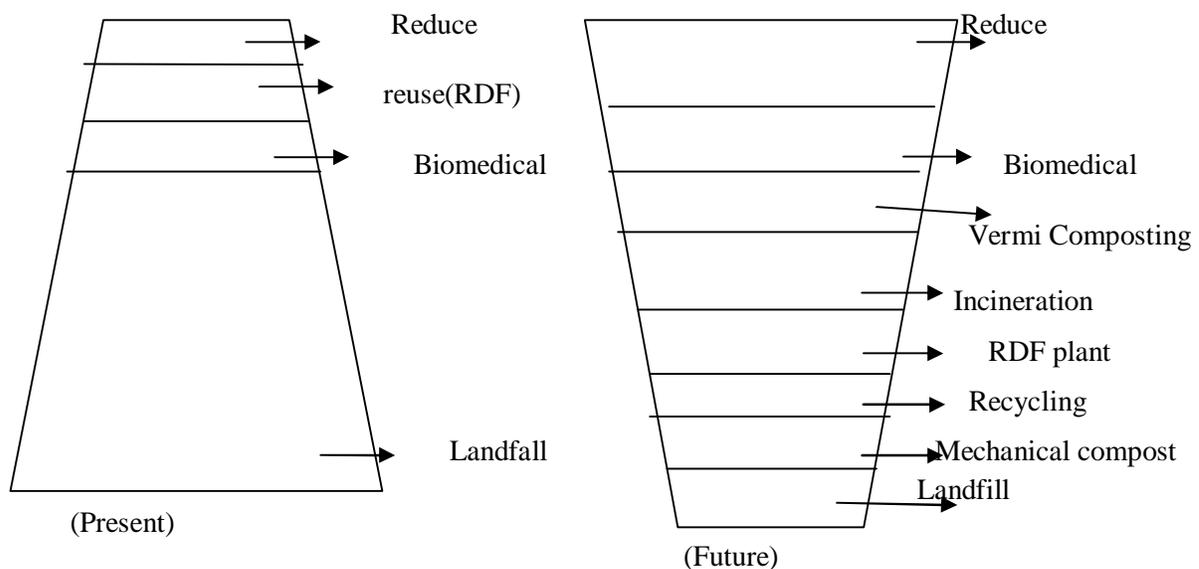
(Fig. 1) In Jaipur, change in strategy in waste management hierarchy

Model

This section describes the mathematical formulation of the integer linear programming model for integrated waste management planning. The model has been built upon the following assumptions.

1. Waste nodes are located at the centers of waste generating areas.

2. Waste separation is done at transfer stations.

3. All the proposed plants in the model are situated near landfill sites so that transportation cost of inert material transported from these plants to landfill is negligible.

Variables:-

X_{it1}, X_{jt2}, i=1, 2…5, j=1…4:- respectively total number of trips made by dumpers to carry waste from node i to transfer station at t1 and from node j to transfer station at t2.

x_{it1}, x_{jt2}, , i=1,2…5, j=1,…4:- respectively total number of dumpers used everyday to carry waste from node i to transfer station at t1 and from node j to transfer station at t2.

Y_{t1r}, Y_{t1m}, Y_{t1s}, $Y_{t1\ell 1}$, Y_{t2n}, Y_{t2v}, $Y_{t2\ell 2}$:- respectively total number of trips made by dumpers to carry waste from transfer station at t1 to R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, landfill at ℓ1 and from transfer station at t2 to incinerator at n, vermicular compost plant at v, landfill at ℓ2.

y_{t1r}, y_{t1m}, y_{t1s}, $y_{t1\ell 1}$, y_{t2n}, y_{t2v}, $y_{t2\ell 2}$:- respectively total number of dumpers used everyday to carry waste from transfer station at t1 to R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, landfill at ℓ1 and from transfer station at t2 to incinerator at n, vermicular compost plant at v, landfill at ℓ2.

$K_{r\ell 1}$, $K_{m\ell 1}$, $K_{s\ell 1}$, $K_{n\ell 2}$, $K_{v\ell 2}$:- respectively total number of trips made by dumpers to carry waste from R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, to a landfill at ℓ1 and from an incinerator at n, vermicular compost plant at v, to a landfill at ℓ2.

$k_{r\ell 1}$, $k_{m\ell 1}$, $k_{s\ell 1}$, $k_{n\ell 2}$, $k_{v\ell 2}$:- respectively total number of dumpers used everyday to carry waste from R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, to a landfill at ℓ1 and from an incinerator at n, vermicular compost plant at v, to a landfill at ℓ2.

P_o: - number of trips made by special vehicles which collect the waste from all hospitals, dispensaries, clinics, and diagnostic centers of city to carry biomedical waste to biomedical treatment plant.

p_o: - number of special vehicles used to carry biomedical waste from all hospitals, dispensaries, clinics, and diagnostic centers of city to biomedical treatment plant.

Z_r, Z_m, Z_s, Z_n, Z_v, Z_o, $Z_{\ell 1}$, $Z_{\ell 2}$:- 0-1 variables indicate respectively, the presence of RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o and landfills at ℓ1 and ℓ2.

W_r, W_m, W_s, W_n, W_v, W_o, $W_{\ell 1}$, $W_{\ell 2}$:- amount of waste transported everyday respectively, to a RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o and landfills at ℓ1 and ℓ2.

D_1 , D_2 : -respectively total numbers of dumpers used and special vehicle used for biomedical waste everyday.

Input Data Parameter:-

a_{it1}, a_{jt2}, i=1, 2…5, j=1…4:- expected number of trips respectively made by dumpers per day to carry waste from node i to transfer station at t1 and from node j to transfer station at t2.

b_{t1r}, b_{t1m}, b_{t1s}, $b_{t1\ell 1}$, b_{t2n}, b_{t2v}, $b_{t2\ell 2}$:- expected number of trips respectively made by dumpers per day to carry waste from transfer station at t1 to R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, landfill at ℓ1 and from transfer station at t2 to incinerator at n, vermicular compost plant at v, landfill at ℓ2.

$\lambda_{r\ell 1}$, $\lambda_{m\ell 1}$, $\lambda_{s\ell 1}$, $\lambda_{n\ell 2}$, $\lambda_{v\ell 2}$:- expected number of trips respectively made by dumpers per day to carry waste from R.D.F. plant at r, mechanical compost plant at m, recycling plant at s, to a landfill at ℓ1 and from an incinerator at n, vermicular compost plant at v, to a landfill at ℓ2.

α:- capacity (in tons) of a dumper.

C_{it1}, C_{jt2}, i=1, 2…5, j=1…4:- respectively transportation cost from node i to transfer station at t1 and from node j to transfer station at t2.

d_{t1r}, d_{t1m}, d_{t1s}, $d_{t1\ell1}$, d_{t2n}, d_{t2v}, $d_{t2\ell2}$:- *respectively transportation cost from transfer station at t1 to R.D.F.* plant at r, mechanical compost plant at m, recycling plant at s, and a landfill at ℓ1 and from transfer station at t2 to incinerator at n, vermicular compost plant at v, and a landfill at ℓ2.

$h_{o:}$ transportation cost of special vehicles which collect the waste from all hospitals, dispensaries, clinics, and diagnostic centers of city.

C_r, C_m, C_s, C_n, C_v, C_o :- revenue respectively per unit of waste from a RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o.

f_1, f_2 :-respectively cost of buying dumpers and special vehicle for bio medical waste.

ψ_1, ψ_2 :- are respectively total amount of waste at transfer stations at t1 and t2.

ρ_r, ρ_m, ρ_s, ρ_n, ρ_v, ρ_o :- fraction of unrecovered waste respectively at RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o.

Q_r, Q_m, Q_s, Q_n, Q_v, Q_o , $Q_{\ell1}$, $Q_{\ell2}$:-capacity per day respectively for a RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o and landfills at ℓ1 and ℓ2.

∂_r, ∂_m, ∂_s, ∂_n, ∂_v, ∂_o , $\partial_{\ell1}, \partial_{\ell2}$:-respectively fixed cost incurred in opening a RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o and landfills at ℓ1 and ℓ2.

ξ_r, ξ_m, ξ_s, ξ_n, ξ_v, ξ_o , $\xi_{\ell1}, \xi_{\ell2}$:-respectively variable cost incurred in handling a RDF plant at r, mechanical compost plant at m, a recycling plant at s, an incinerator at n, vermicular compost plant at v, bio-medical treatment plant at o and landfills at ℓ1 and ℓ2.

Objective function

Objective function represents the overall daily management cost.

The first component (F_1) gives the total of transportation cost i.e. Transportation cost from nodes i=1,…5 to transfer station at t1 + transportation cost from nodes j=1,…4 to transfer station at t2 +transportation cost from transfer stations at t1 and t2 to either at any treatment plant or to a landfill sites at ℓ1and ℓ2 respectively + transportation cost of special vehicles used to collect biomedical waste. The second component (F_2) gives the investment and handling expenses. The third component (F_3) gives the total cost for buying all dumpers and special vehicles for bio-medical waste. The fourth component (B) gives the benefits at plants owing to the production at different plant.

$$F_1(X,Y)=\sum_{i=1}^{5}C_{it1}\alpha X_{it1}+\sum_{j=1}^{4}C_{jt2}\alpha X_{jt2}+\sum_{t1r}d_{t1r}\alpha Y_{t1r}+\sum_{t1m}d_{t1m}\alpha Y_{t1m}+\sum_{t1s}d_{t1s}\alpha Y_{t1s}$$
$$+\sum_{t1\ell1}d_{t1\ell1}\alpha Y_{t1\ell1}+\sum_{t2n}d_{t2n}\alpha Y_{t2n}+\sum_{t2v}d_{t2v}\alpha Y_{t2v}+\sum_{t2\ell2}d_{t2\ell2}\alpha Y_{t2\ell2}+\sum_{o}h_o\alpha P_o$$

[1]

$$F_2(Z,w)=\sum_{r}\left(\delta_r Z_r+\xi_r w_r\right)+\sum_{m}\left(\delta_m Z_m+\xi_m w_m\right)+\sum_{s}\left(\delta_s Z_s+\xi_s w_s\right)$$
$$+\sum_{n}\left(\delta_n Z_n+\xi_n w_n\right)+\sum_{v}\left(\delta_v Z_v+\xi_v w_v\right)+\sum_{\ell1}\left(\delta_{\ell1}Z_{\ell1}+\xi_{\ell1}w_{\ell1}\right)+$$
$$\sum_{\ell2}\left(\delta_{\ell2}Z_{\ell2}+\xi_{\ell2}w_{\ell2}\right)+\sum_{o}\left(\delta_o Z_o+\xi_o w_o\right)$$

[2]

$$F_3(x,y,k)=\sum f_l(D_1)+\sum f_2(D_2)$$

[3]

$$B(w)=\sum_{r}C_r(1-\rho_r)w_r+\sum_{m}C_m(1-\rho_m)w_m+\sum_{s}C_s(1-\rho_s)w_s$$
$$+\sum_{n}C_n(1-\rho_n)w_n+\sum_{v}C_v(1-\rho_v)w_v+\sum_{o}C_o(1-\rho_o)w_o$$

[4]

So the objective function F to be minimizes is

$$\mathbf{F = F_1 + F_2 + F_3 - B}$$ [5]

Mass balance constraints:-

Total waste moved from each waste collection points i=1,…5 and j=1,…4 should at least be equal to

the total amount of waste at that point.

$$\sum_{i=1}^{5} \alpha X_{it1} + \sum_{j=1}^{4} \alpha X_{jt2} \geq \Psi_{t1} + \Psi_{t2}$$

[6]

In constraints (7) and (8) amount of waste carried from transfer stations at t1 and t2, to different plants should at least be equal to the amount of waste found at that point.

$$\sum_{t1r} \alpha Y_{t1r} + \sum_{t1m} \alpha Y_{t1m} + \sum_{t1s} \alpha Y_{t1s} + \sum_{t1\ell 1} \alpha Y_{t1\ell 1} \geq dt1$$

[7]

$$\sum_{t2r} \alpha Y_{t2r} + \sum_{t2m} \alpha Y_{t1m} + \sum_{t2s} \alpha Y_{t2s} + \sum_{t2\ell 2} \alpha Y_{t2\ell 2} \geq dt2$$

[8]

In constraints (9) – (13) amount of waste carried away from every plant to landfills ℓ1and ℓ2 should at least be equal to amount of waste found at that point.

$$\rho_r w_r \leq \sum \alpha K_{r\ell 1}$$

[9]

$$\rho_m w_m \leq \sum \alpha K_{m\ell 1}$$

[10]

$$\rho_s w_s \leq \sum \alpha K_{s\ell 1}$$

[11]

$$\rho_n w_n \leq \sum \alpha K_{n\ell 2}$$

[12]

$$\rho_v w_v \leq \sum \alpha K_{v\ell 2}$$

[13]

In constraint (14) amount of waste collected from all hospitals , dispensaries, clinics, and diagnostic centers should at least be equal to the amount of waste found at that point.

$$\rho_o w_o \leq \sum \alpha P_o$$

[14]

Capacity limitation constraint:-

In constraint (15)-(20) the maximum capacity for processing plants are accounted means amount of waste taken to different plants should not exceed the plant capacities. In constraints (21) and (22) same thing is done for sanitary landfills ℓ1and ℓ2.

$W_r \quad \leq \quad Q_r \; Z_r$ [15]

$W_m \quad \leq \quad Q_m Z_m$ [16]

$W_s \quad \leq \quad Q_s \; Z_s$ [17]

$W_n \quad \leq \quad Q_n Z_n$ [18]

$W_v \quad \leq \quad Q_v \; Z_v$ [19]

$W_o \quad \leq \quad Q_o Z_o$ [20]

$W_{\ell 1} \quad \leq \quad Q_{\ell 1} \; Z_{\ell 1}$ [21]

$W_{\ell 2} \quad \leq \quad Q_{\ell 2} Z_{\ell 2}$ [22]

Technical constraints:-

Constraints (23)-(34) means that, once the flow to either plant or sanitary landfill is positive that plant or landfill must exist. In constraint (35) same thing is done for bio-medical waste.

$\alpha\, Y_{t1rg} \leq \quad Q_r\, Z_r$ [23]

$\alpha\, Y_{t1mg} \leq \quad Q_m\, Z_m$ [24]

$\alpha\, Y_{t1sg} \leq \quad Q_s\, Z_s$ [25]

$\alpha\, Y_{t1\ell 1g} \leq \quad Q_{\ell 1}\, Z_{\ell 1}$ [26]

$\alpha\, Y_{t2ng} \leq \quad Q_n\, Z_n$ [27]

$\alpha\, Y_{t2vg} \leq \quad Q_v\, Z_v$ [28]

$\alpha\, Y_{t2\ell 2g} \leq \quad Q_{\ell 2}\, Z_{\ell 2}$ [29]

$\alpha\, K_{r\ell 1g} \leq \quad Q_{\ell 1}\, Z_{\ell 1}$ [30]

$\alpha\, K_{m\ell 1} \leq \quad Q_{\ell 1}\, Z_{\ell 1}$ [31]

$\alpha\, K_{s\ell 1g} \leq \quad Q_{\ell 1}\, Z_{\ell 1}$ [32]

$\alpha\, K_{n\ell 2g} \leq \quad Q_{\ell 2}\, Z_{\ell 2}$ [33]

$\alpha\, K_{v\ell 2g} \leq \quad Q_{\ell 2}\, Z_{\ell 2}$ [34]

$\alpha\, P_{og} \leq \quad Q_o\, Z_o$ [35]

Variable conditions:-
All the variables are integers and positive.

i.e., X_{it1}, X_{jt2}, are integer, i=1,…5, j=1,…4 and ≥ 0

$Y_{t1r},\ Y_{t1m},\ Y_{t1s},\ Y_{t1\ell 1},\ Y_{t2n},\ Y_{t2v},\ Y_{t2\ell 2}$ are integers

and $\quad \geq 0$

Variables in (36)-(43) are defined as Boolean. These are

used to determine existence of either plant or a landfill.

$Z_r \quad \epsilon \quad \{0, 1\}$ [36]

$Z_m \quad \epsilon \quad \{0, 1\}$ [37]

$Z_s \quad \epsilon \quad \{0, 1\}$ [38]

$Z_n \quad \epsilon \quad \{0, 1\}$ [39]

$Z_v \quad \epsilon \quad \{0, 1\}$ [40]

$Z_o \quad \epsilon \quad \{0, 1\}$ [41]

$Z_{\ell 1} \quad \epsilon \quad \{0, 1\}$ [42]

$Z_{\ell 2} \quad \epsilon \quad \{0, 1\}$ [43]

Definition:-
In definition (44)-(57) which were already mentioned in the beginning gives the expected number of trips made by dumpers per day from waste nodes i=1,…,5 and j= 1,…,4 to transfer stations t1 and t2, and from transfer stations at t1 and t2 to different plants and landfills are given.

$X_{it1} \quad = \quad a_{it1}\, x_{it1} \qquad I = 1,…, 5$ [44]

$X_{jt2} \quad = \quad a_{jt2}\, x_{jt2} \qquad I = 1,…, 4$ [45]

$Y_{t1r} \quad = \quad b_{t1r}\, y_{t1r}$ [46]

Y_{t1m} = $b_{t1m} \, y_{t1m}$ [47]

Y_{t1s} = $b_{t1s} \, y_{t1s}$ [48]

Y_{t2n} = $b_{t2n} \, y_{t2n}$ [49]

Y_{t2v} = $b_{t2v} \, y_{t2v}$ [50]

$Y_{t1\ell1}$ = $b_{t1\ell1} \, y_{t1\ell1}$ [51]

$Y_{t2\ell2}$ = $b_{t2\ell2} \, y_{t2\ell2}$ [52]

$K_{r\ell1}$ = $\lambda_{r\ell1} \, k_{r\ell1}$ [53]

$K_{m\ell1}$ = $\lambda_{m\ell1} \, k_{m\ell1}$ [54]

$K_{s\ell1}$ = $\lambda_{s\ell1} \, k_{s\ell1}$ [55]

$K_{n\ell2}$ = $\lambda_{n\ell2} \, k_{n\ell2}$ [56]

$K_{v\ell2}$ = $\lambda_{v\ell2} \, k_{v\ell2}$ [57]

Definitions (58)-(62) indicates the amount of waste transported from transfer stations at t1 and t2 to processing plants, while definitions (63) and (64) gives the amount of waste transfer to landfills at ℓ1and ℓ2.equation (65) gives amount of waste collected from all hospitals, dispensaries, clinics and diagnostic centers and send to bio-medical treatment plant

W_r = $\sum \alpha \, Y_{t1r}$ [58]

W_m = $\sum \alpha \, Y_{t1m}$ [59]

W_s = $\sum \alpha \, Y_{t1s}$ [60]

W_n = $\sum \alpha \, Y_{t2n}$ [61]

W_v = $\sum \alpha \, Y_{t2v}$ [62]

$W_{\ell1}$ = $\sum \alpha \, Y_{t1\ell1}$ [63]

$W_{\ell2}$ = $\sum \alpha \, Y_{t2\ell2}$ [64]

W_o = $\sum \alpha \, P_o$ [65]

Equations (66)-(67) indicates the amount of waste disposed of in a sanitary landfills ℓ1and ℓ2, everyday. Equation (68) gives total amount of waste collected from all waste sources per day. (This excluded waste generated from plants.)

$$\phi_{\ell1} = W_{\ell1} + \sum_r \alpha \, K_{r\ell1} + \sum_m \alpha \, K_{m\ell1} + \sum_s \alpha \, K_{s\ell1} \quad [66]$$

$$\phi_{\ell2} = W_{\ell2} + \sum_n \alpha \, K_{n21} + \sum_v \alpha \, K_{v\ell2} \quad [67]$$

$$W = \sum_r W_r + \sum_m W_m + \sum_s W_s + \sum_n W_n + \sum_v W_v + \sum_{\ell1} W_{\ell1} + \sum_{\ell2} W_{\ell2} + \sum_o W_o$$
[68]

Equation (69) gives the total number of dumpers and Special vehicles used per day are determined.

$$D = \sum_{i=1}^{5} x_{it1} + \sum_{j=1}^{4} x_{jt2} + \sum_{t1r} y_{t1r} + \sum_{t1m} y_{t1m} + \sum_{t1s} y_{t1s} + \sum_{t1\ell1} y_{t1\ell1}$$
$$+ \sum_{t2n} y_{t2n} + \sum_{t2v} y_{t2v} + \sum_{t2\ell2} y_{t2\ell2} + \sum_o p_o$$
[69]

Model application:-

Jaipur is divided in eight Zones having 77 wards.Zonewise quantity of waste generated per day in Jaipur is given in figure (3). In our model we have two transfer stations, First at Delhi road near meena petrol pump (t1) and second at Zalana (t2) and two landfill sites, first at Lagadiyawas (ℓ₁) and second at Sevapura(ℓ₂). RDF plant (r) , mechanical compost plant (m), and recycling plant (s), are located at Lagadiyawas site land and Incinerator (n), and vermicular compost plant (v), are located at Sevapura site land, and one bio-medical treatment plant located at khorarupadi.City is divided by north east

corner to south west corner in such a way that waste from Hawamahal (east), Hawamahal (west), Motidugri zone, Amer zone, and Civil line zone(18-22, 41-43, 63 wards) will be collected at nodes i=1,…5 respectively and then transfer to the transfer station at (t1). Waste from Saganer zone, Mansarovar zone, Vidyadharnagar zone, and Civil line(11-14 wards) will be collected at nodes j=1,…4 respectively and then transfer to the transfer station at (t2).At transfer station waste are segregated by workers and then transfer to the different plants which is shown in figure (2). Collection cost from house to house collection to nodes i=1…5 and j=1…4 will not be incorporated in the model. Biomedical waste shall be segregated into containers/bags at the point of generation treated and disposed in accordance with category given in the following table1.

Biomedical waste is handled without any adverse effect to human health and the environmental. The following precautions should be taken by every person/occupier of an institution who involved in handling of biomedical waste.

1. Bio-medical waste shall not be mixed with other waste.
2. No untreated bio-medical waste shall be kept beyond a period of 48 hours.
3. Not withstanding any thing contained in the motor vehicle act, 1988, on rules there under untreated bio-Medical waste shall be transported only in such vehicle
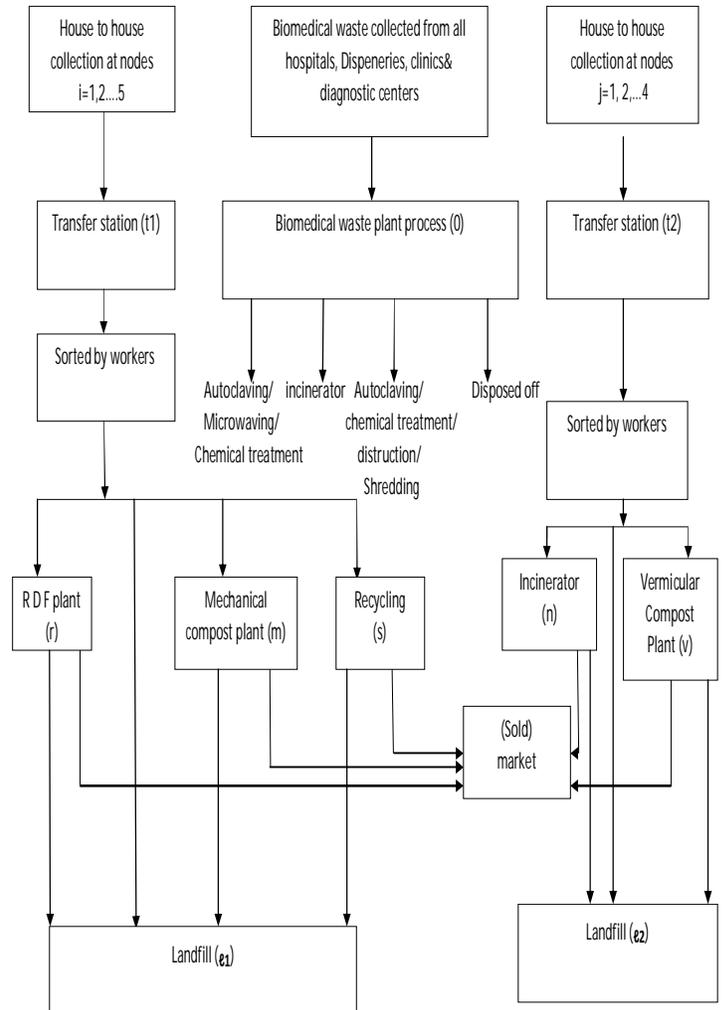
(Fig.2)Waste flow chart

as may be authorized for the purpose by the competent authority as specified by the government.

4. Deep burial shall be an option available only in towns with population less than five lakes and in rural areas.

5. Chlorinated plastics shall not be incinerated.
6. Mutilation/shredding must be such so as to prevent unauthorized reuse.
7.Liquid waste generated from laboratory, wasting, cleaning, also liquid chemical waste should be disinfected by chemical treatment and discharge into drains.

Table:1`.Container color coding, type of container Used for waste category and treatment option for Disposal of bio-Medical wastes.

Container color coding	Type of container-Waste category	Treatment option
Yellow	Plastic bags-human anatomical waste, animal waste, microbiology and biotechnology waste, solid waste items contaminated with blood.	Incineration/deep burial
Red	Disinfected container/plastic bag- microbiology and biotechnology waste, solid waste generated from disposal items other than the sharps such as tunings, catheters, intravenous sets etc.	Autoclaving/ microwaving/ chemical treatment
Blue/white translucent	Plastic bags/puncture proof-waste sharps like needles, syringes, scalpels, blades, glass etc., and also waste generated from disposal items other than waste sharps.	Autoclaving/ microwaving/ chemical treatment and destruction/sh redding
Black	Plastic bags- discarded medicines and cytotoxic drugs, ash from incineration of any biomedical waste, chemicals used in production of biological, chemicals used in disinfection, as insecticides, etc.	Disposed in secured landfill

Zone wise Quantity of waste generated in Jaipur shown by the following chart.(Fig.3)
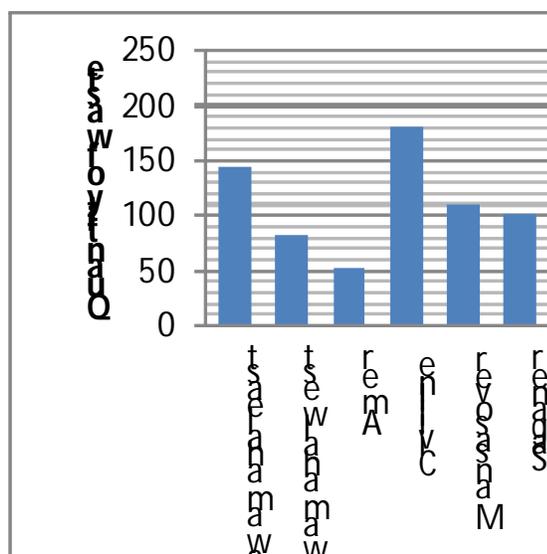
Conclusion and future development:-

The proposed model is a good starting point upon which future variation can be built. The existing status of waste management and the littered streets all over the city Cleary speak about the poor environmental health of the city. The aforesaid policies, if implemented have the potential to bring an improvement in the SWM system in the Jaipur city. A careful attention has been paid to provide a proper characterization of the system, as regards waste composition, heating value, material recovery and possible treatment which are mentioned in the proposed model. In such a way, it would be possible to determine optimal sequence of interventions, (building of new plants) over a given time horizon, capable of optimally deriving the MSW management system from the Present configuration to a final one.

References:-

1. P. costi, R. Minciardi, M. Robba, M. Ravati, R. Sacile, *An environmentally sustainable decision model for urban solid waste management* , waste management 24,(2004), 277-295.

2. Y. Sun , G. H. Huang, Y. P. Li, *ICQSWM: An inexact chance- constrainedquartic solid waste management model*, Resources, conservation and recycling 54,(2010), 641-657.

3. Gordon H. Huang, D. Linton, Julian Scott Yeomans, Reena Yoogalingam, *Policy planning under uncertainty: efficient starting populations for simulation- optimization methods applied to municipal solid waste management.* Journal of Environmental management 77, (2005) , 22-34.

4. Ministry of environment and forests notification, New Delhi, 20th July 1998.

5. Minciardi,R.,Paolucci,M.,Robba,M.,Sacile,R., *A multi objective approach for solid waste management.*IEMSS 2002 Congress, Lugano, Switzerland, June 24-27, (2002) , 205-210.

6. Michael K. Nganda, *Mathematical Models in municipal solid waste management.*P.hd. Thesis, Chalmers University of Technology and Gotebarg University, Sweden. (2007).

Surveillance Robot For Tracking Multiple Moving Targets

S.Pratheepa, Dr.Purushothaman Srinivasan

Abstract -- Object tracking is a challenging task in spite of all sophisticated methods that have been developed. The major challenge is to keep track of the object of a particular choice. In this work, a new video moving object-tracking method is proposed. The segmentation of the video is done by contextual clustering. Clustering is an important method in data analysis because of its ability to 'discover' the inherent features in the data. The fundamental concept in clustering techniques is to group a given set of objects into subsets according to properties associated with each object, so that the members in each individual subset share some similar properly defined features. A multitarget human tracking is attempted.

Index Terms-Contextual Segmentation,Clustering,Tracking.

———————————— ◆ ————————————

1. INTRODUCTION

Surveillance is the monitoring of behavior. Systems surveillance is the process of monitoring the behavior of people, objects or processes within systems for conformity to expected or desired norms in trusted systems for security or social control. Intelligent visual surveillance systems deal with the real time monitoring of persistent and transient objects within a specific environment. The primary aims of these systems are to provide an automatic interpretation of scenes and to understand and predict the actions and interactions of the observed objects based on the information acquired by sensors. The main stages of processing in an intelligent visual surveillance system are: moving object detection and recognition, tracking, behavioral analysis and retrieval. These stages involve the topics of machine vision, pattern analysis, artificial intelligence and data management.

The technological evolution of video-based surveillance systems started with analogue closed circuit television (CCTV) systems. These systems consist of a number of cameras located in a multiple remote location and connected to a set of monitors, usually placed in a single control room, via switches (a video matrix). Conventional CCTV cameras generally use a digital charge coupled device (CCD) to capture images. The digital image is then converted into an analogue composite video signal, which is connected to the CCTV matrix, monitors and recording equipment, generally via coaxial cables. The digital to analogue conversion does cause some picture degradation and the analogue signal is susceptible to noise. It is possible to have CCTV digital systems by taking advantage of the initial

digital format of the captured images and by using high performance computers.

The technological improvement provided by these systems has led to the development of semi-automatic systems, known as second generation surveillance systems. Most of the research in second generation surveillance systems is based on the creation of algorithms for automatic real-time detection events aiding the user to recognize the events. The increasing demand for security by society leads to a growing need for surveillance activities in many environments. The demand for remote monitoring for safety and security purposes has received particular attention, especially in the following areas. Transport applications include airports, maritime environments, railways, underground, and motorways to survey traffic. Public places such as banks, supermarkets, homes, department stores and parking lots use such surveillance systems. Remote surveillance of human activities such as attendance at football matches or other activities and surveillance to obtain certain quality control in many industrial processes, surveillance in forensic applications and remote surveillance in military applications are in use.

Recent events, including major terrorist attacks, have led to an increased demand for security in society. This in turn has forced governments to make personal and assess security a priority in their policies. This has resulted in the deployment of large CCTV systems.

Surveillance systems created for commercial purposes differ from surveillance systems created in the academic world, where

commercial systems tend to use specific-purpose hardware and an increasing use of networks of digital intelligent cameras. The common processing tasks that these systems perform are intrusion and motion detection and detection of packages. Research in academia tends to improve image processing tasks by generating more accurate and robust algorithms in object detection and recognition, tracking, human activity recognition, database and tracking performance evaluation tools. A review of human body and movement detection, tracking and also human activity recognition is presented. Other research currently carried out is based on the study of new solutions for video communication in distributed surveillance systems. The creation of a distributed automatic surveillance system by developing multi-camera or multi-sensor surveillance systems, and fusion of information obtained across cameras, or by creating an integrated system is also an active area of research.

2. Schematic Diagram

An user who is moving in and out of library system
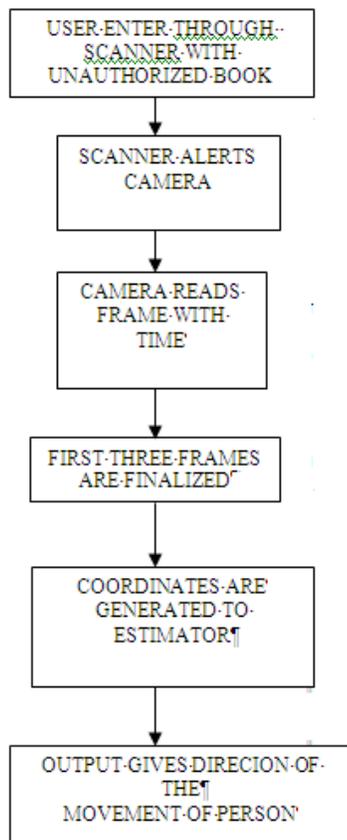
The sequence of implementation of this work starts with initialization of Radio Frequency Identification Technology (RFID) scanner at the library entrance. Whenever a person coming out of the library does not register his/ her books at the books issuing counter, and whenever, he/ she is trying to pass through the RFID scanner, the scanner alerts the video camera which is in focus at the entrance. The video camera covers a range of area at the entrance and records the actions. Inbuilt algorithm passes each frame to software. The software, segments the frame using contextual clustering algorithm. The segmented image contains the human being distinctly visible. As the frame covers certain are of the entrance, the position of the human in the frame is given with respect to the left hand side of the frame. In our calculation, the number of columns away from the left position of the frame are suitably scaled based on the area of the room and the position of the human is recorded based on coordinates. The coordinates of the first two frames are given to an estimator. The estimator estimates the next position the object. This estimation may not be exact as the position of the human in the third video frame. To overcome this drawback, back-propagation algorithm has been used to bring close the estimation accuracy. When more than one human is present, then separate procedure has to be followed. The segmented frame has to be coordinated with the input frame and corresponding color information of the human are noted. Incase , both the human are wearing the dress in same color, then there should be some difference in the height or width of the persons in the image.

3. Contextual Segmentation

Segmentation refers to the process of partitioning a digital image into multiple regions (sets of pixels). The goal of segmentation is to simplify and /or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly

different with respect to the same characteristics. Several general-purpose algorithms and techniques have been developed for image segmentation.

4. Experimental Setup

Hardware Requirements are any Processor above 500 MHz., Ram 128Mb, Hard Disk 10 Gb., Compact Disk 650 Mb, Input device Standard Keyboard and Mouse./ Yashika Camera, Output device VGA and High Resolution Monitor are used Software Requirements namely operating system Windows Xp, Image Processing, Front / Back End and Matlab 7 is used for implementing the work.

5. Results And Discussion
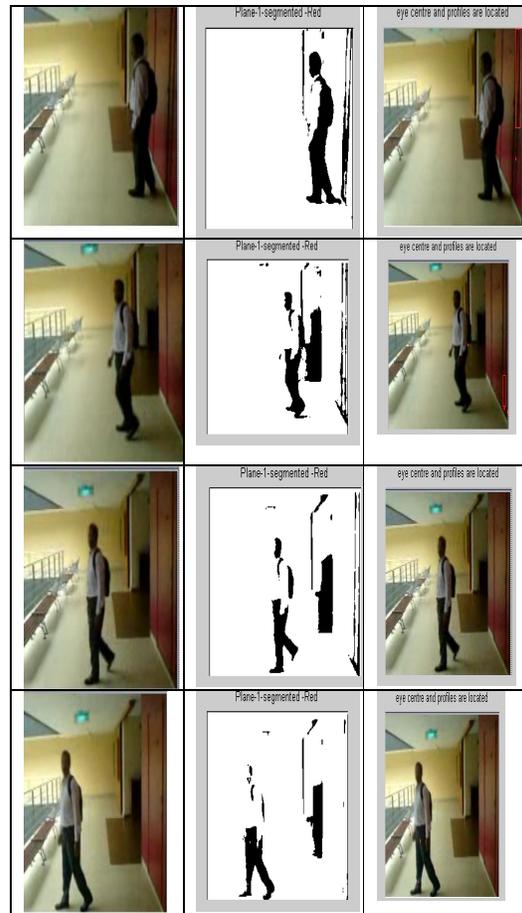
Figure 1 original segmented and identified images

6. Conclusion

Video tracking is an important process in tracking objects. It involves various image processing concepts. In this work, the acquired video has been separated into frames and segmented by using contextual clustering method. The features of the segmented image is further processed by the imfeature properties of the matlab. The imfeature provides 24 properties. In this work, two important properties are used to process the features of the segmented image for highlighting the presence of the human .

References

[1] Matlab Image processing Tool box
[2]Digital Image Processing, Rafael C Gonzales, Second Edition,Pearson Education
[3] Robotics, K.S.Fu, R.C.Gonzales, MeGraw-Hill
[4] I. Haritaoglu, D. Harwood, L.S. Davis, "W4: Real-Time surveillance of people and their activities". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 809-830, 2000.

[5]Siebel, N.; Maybank, S., "Fusion of Multiple Tracking Algorithms for Robust People Tracking," ECCV 2002, pp.373–387, 2002

[6]Spengler, M., Schiele, B.: Towards robust multi-cue integration for visual tracking,Machine Vision and Applications 14 (2003) 50–58

[7]Zhao, T.; Nevatia, R., "Tracking Multiple Humans in Complex Situations," PAMI 2004, pp1208–1221, 2004

[8] Y. L. Tian and A. Hampapur, "Robust Salient Motion Detection with Complex Background for Real-time Video Surveillance," IEEE Computer Society Workshop on Motion and Video Computing, Breckenridge, Colorado, Jan. 5-6, 2005.

[9] Y. Zhuang, W. Wang and R.Z. Xing, "Target Tracking in Colored Image Sequence Using Weighted Color Histogram Based Particle Filter". Proceedings of the IEEE International Conference on Robotics and Biomimetics, pp. 1488-1493, 2006.

[10] W. M. Hu, H. Min, T.N. Tan, etc, "Principal axis-based correspondence between multiple cameras for people tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, vol. 28, no. 4, pp.663-671, 2006.

Sea Surface Simulation for SAR Remote Sensing Based on the Fractal Model

Ding Guo, Xingfa Gu, Tao Yu, Xiaoyin Li, Jingjun Zheng, Hui Xu

Abstract— Based on the fractal ocean surface model, electromagnetic scattering model under Kirchhoff Approximation and the raw signal simulation procedure of dynamic scene based on time domain, the sea surface of the SAR remote sensing has been simulated. The images of the wave and complex fractal sea surface are in accordance with the hydrodynamic modulation, the tilt modulation and the velocity bunching modulation. The simulation has been developed in the Matlab programming language.

Index Terms— sea fractal surface, multi-polarization, scattering matrix, synthetic aperture radar, electromagnetic wave.

————————————— ◆ —————————————

1. INTRODUCTION

THE concept "fractal" is popularly was used in the world after the book "The fractal geometry nature" published in 1982. The fractal geometry is a simple tool of describing complicate world and mechanism and then many researchers paid attention to it.

Numerical simulation sea SAR surface is one of the foci of research now. Signification of applying "fractal" to the researching field lying to:

1. Researches such as multiple scattering and sea clutter et al., are in favor of improving performance of the radar system and communication.

2. It can not only promote the development of characteristics of sea surface, the application of SAR measurement of environment but also convenient for the management of navigation et al.

3. It can be used to explain physical phenomenon of sea surface, for example, the hydrodynamic evolvement of ocean wave, the air-sea power exchanging and the analyzing the ocean current of sea et al,.

Theoretical facts

Recently, F.Berizzi et al. presented sea fractal surface models, the 1-D and 2-D sea fractal model. We find: whether one dimensional or two dimensional sea Fractal surface model, the surface becomes rough when one of b and S increases while another of them

is invariable. So both b and S are effect on the roughness of the sea fractal surface.

$$\rho(\xi) = \frac{1 - b^{2(s-2)}}{1 - b^{2(s-2)N_f}}$$

$$\sum_{n=0}^{N_f - 1} b^{2(s-2)n} \cos\left[\frac{2\pi}{\Lambda_0} b^n \xi\right] \quad (1)$$

In order to study the polarization effects on scattering coefficient and Radar Cross Section, exploiting Huygen's principle, Kirchhoff approximation and the model of sea fractal surface presented by F.Berizzi et al ., we derived the scattering fields, scattering coefficient, RCS and Poynting vector of 2-D sea fractal surface with finite conductivity illuminated by arbitrary polarization wave under the condition that the shadowing effect and multiple scattering are neglected. Our results coincide with those of other literatures. Meanwhile, receiving signals in different position are simulated. The result shows that depolarization effects in any position can be neglected though we derived out the expression of depolarization, while the cross polarization can't neglected.

Exploiting the scattering matrix which is worked out and calculated full polarimetric radar cross section (RCS) in the third part. The numerical results show: Normal backscattering RCS of sea fractal surface is degraded in exponential form with

increasing of incident angles in HH polarization and VV polarization. When θ3 is a certain invariable value, the maximum RCS appears at θ2=θ1and RCS decreases when │θ2-θ1│increases. We also find that values of the radar cross section calculated according to the formula which is presented by F.Berizzi et al. are greater than those of our results because it is assumed that the conductivity of the sea is infinite and then there is no loss when wave is reflected on the sea surface. In fact, however, the conductivity of sea surface is finite. Comparing theoretic result with numerical result, one can find that whether the relation between σhh and σvv or the relation between σhv and σvh is decided by the relative position between receiving antenna and transmit antenna. Critical angles(θc1 and θc2) are derived in theory.

In the last part, numerically simulating the co-polarization signature with different sea fractal surface parameters in the experiments, we conclude: Sea fractal surface roughness has no effect on polarized parameters. The fundamental spatial wavelengths of the ocean wave have no effect on orient angle shift but have effect on ellipticity while other sea fractal Surface parameters are invariable. The radar incident angle only effects on the orient angle shift. The orient angle shift becomes zero when the radar incident angle is larger than a certain value. We find that the effects of sea fractal parameters on the ellipse parameters are virtually caused by mean slope of the sea surface.

I. Simulation and results

In this section, the 2-D ocean surface has been simulated according to (R. Garello et al.1993, Berizzi et al. 2002, Nunziata, F. et al.2008). It is shown in figure.
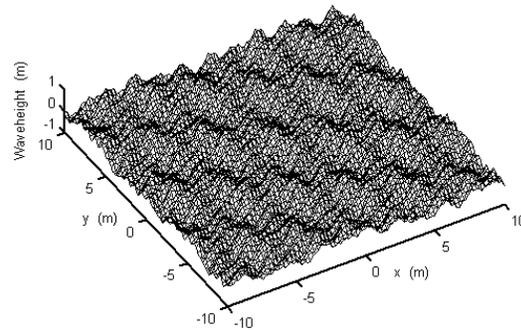

Fig 1 Simulated ocean suface

The RCS of the ocean surface is

$$\sigma_{qj} = 4\pi r^2 \left| S_{qj} \right|^2, \quad q=h_s, v_s, \quad j=h_i, v_i$$ (Ruck G

T et al. 1970), $\sigma_{h_s h_i}$ and $\sigma_{h_s v_i}$ are the depolarization parameters, where the radar wavelength is

$$\lambda = 0.23m, \quad \varepsilon_r = 72.1 + 60.979i.$$

The normalized RCS of the 2-D ocean surface

when $\theta_1 = 10°$, $\theta_3 = 0°$ is shown in figure 3. We

can find that the maximum RCS at $\theta_2 = \theta_1$ and

RCS increases with θ_2 varies from $-90°$ to $90°$,

$\theta_2 = \theta_1$ while decrease with the increasing of θ_2

after the peak value. It is agreed with literatures(Jakov V. et al. 1998, Ericl.Thorsos,1990). Rcs0 is the RCS when the electrical conductivity of ocean water is infinite.
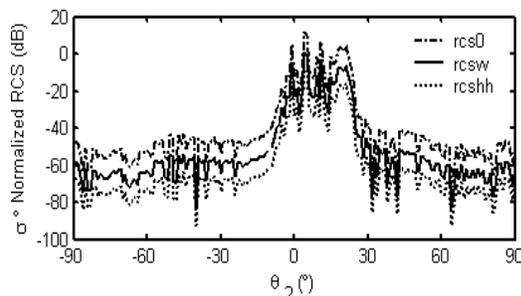

Fig 2 Normalized RCS with VV and HH polarization when $\theta_1 = 10°$, $\theta_3 = 0°$, θ_2 varies from $-90°$ to $90°$.

When the $\theta_1 = 30°, \theta_2 = 30°$, θ_3 varies from

$0°$ to $180°$. Figure shows the normalized radar cross section for HH and VV polarization. And the normalized radar cross section for HV and VH polarization.
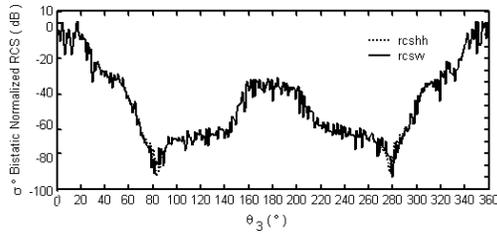

Fig 3 Normalized RCS with VV and HH polarization when $\theta_1 = 30°$, $\theta_2 = 30°$, θ_3 varies from $0°$ to $360°$.
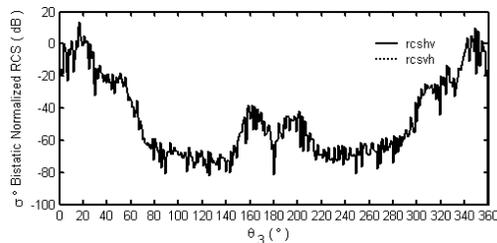

Fig 4 Normalized RCS with HV and VH polarization when $\theta_1 = 30°$, $\theta_2 = 30°$, θ_3 varies from $0°$ to $360°$.

In the first sea surface simulation case a single 60 m wavelength azimuth travelling long wave is simulated and the noisy SAR intensity image is shown in Fig. 5(a). To appreciate the results an azimuth transect (see white dotted line in Fig. 5(a)) is made in the noise-free SAR image and referred to the corresponding long wave, see Fig. 5(b), where are plotted the first 200 pixels. Since an azimuth travelling wave has been simulated, it can be experienced that the SAR imaging process is strongly non-linear in this case as clearly shown in Fig. 5(b). In fact analyzing the plots of Fig. 5(b) it is possible to recognize the non-linear effect of VB. It can be also evaluated the C parameter (section II) which, in this case, is equal to 1 witnessing a strongly non-linear imaging process. It can also be experienced that in this case Rt(·) is equal to zero.
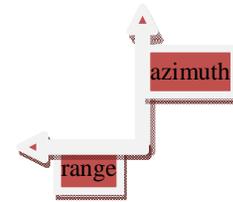

Fig 5(a)
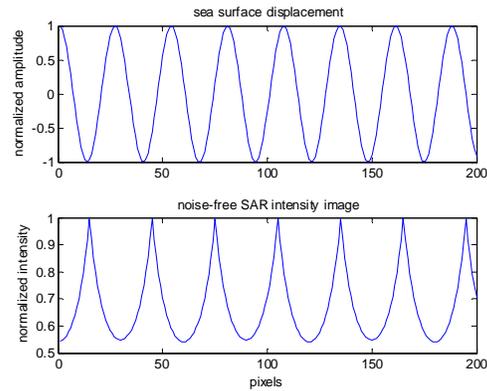
Fig 5(b)

As final cases, a broader spreading function is considered, described by 30 components. In particular the noisy SAR intensity images shown in Figs. 8 and 9 are relevant to a 100 m sea peak wavelength, range travelling, and azimuth travelling. The SAR images clearly show that a broaden spreading function has been employed. Once again it can be evaluated the C parameter, making reference to the peak wavelength and direction, recognizing that VB is a linear process in the first case and highly non-linear in the second one. And it can also be appreciated that the degree of non-linearity of the SAR imaging process decreases increasing the wind speed, as expected for fully-developed wind-seas.
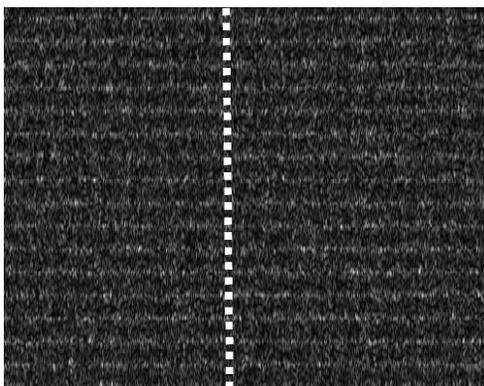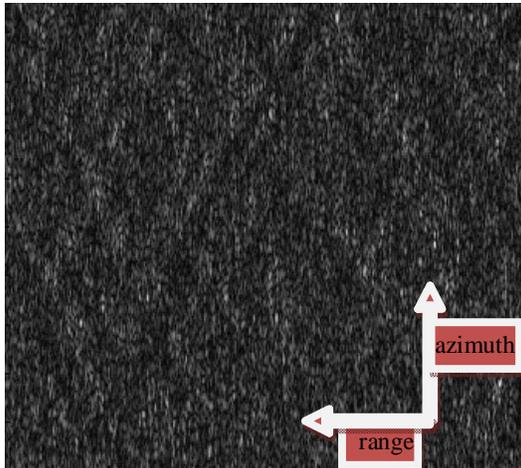
Fig 6
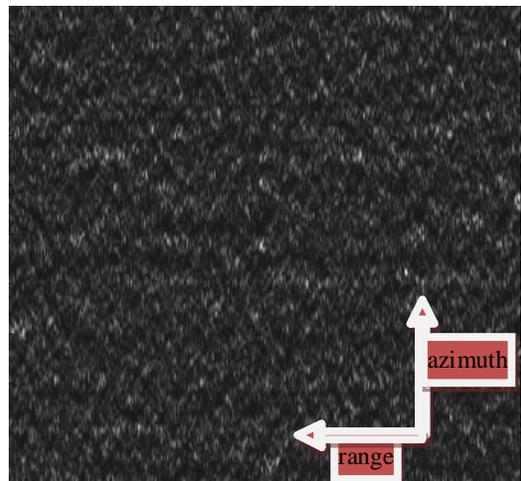
Fig 7

II. Conclusion facts

Using the 2-D ocean surface model, Huygen's principle, Kirchhoff approximation and the neglected multiple scattering, we simulated the full polarization RCS of the ocean surface scattering fields.

The numerical results show that the full polarization scattering model coincides with other literatures. And the other conclusions as bellow:

1) The normalized radar cross section of cross polarization with significant information can't be neglected.

2) When θ_1 , θ_2 are determined and θ_3 varies from $0°$ to $360°$,the relationship among θ_1 , θ_2 and θ_3 has been known. It can be used to predict the critical angle θ_3 .

3) It's obviously that the effects of depolarization can be neglected.

REFERENCES

[1]Kennaugh, E.M. Effects of the type of polarization on echo characteristics. In Rep.389, Antenna Lab., Ohio State Univ., Columbus, 1951.9.

[2]Jakob J.Van Zyl, Zebker. Howard A., and Charles Elachi. Imaging radar polarization signatures: Theory and observation. Radio Science, 1987, 22 (2) :

[3]R. Garello, S. Proust and B. Chapron, "2D Ocean Surface SAR Images Simulation: A Statistical Approach," *Proc. OCEANS'93*, vol. 3, pp. 7-12, 1993.

A.K.Fung Scattering and depolarization on reflection of EM waves from a rough surface. In Proc. IEEE (Communications), 1966.395-396.

[4]Gaspar R.Valenzuela. Depolarizaiton of EM Waves by Slightly Rough Surfaces. IEEE Trans. Antennas Propagation, 1967, 15 (4) :552-557.

[5]F. Berizzi and E. Dalle Mese. Scattering coefficient evaluation from a two-dimensional sea fractal surface. IEEE Trans. Antennas Propagation,2002, 50 (4): 426-434.

[6]Nunziata, F., Gambardella, A. and Migliaccio, M. An educational SAR sea surfacewaves simulator. International Journal of Remote Sensing, vol. 29, No. 11, 10 June 2008, 3051-3066

USENET: InternetNews Software, Security – Needs and Goals

Monika Saxena, Praneet Saurabh, Bhupendra Verma

Abstract — Usenet is a distributed bulletin board system, built as a logical network on top of other networks and connections. By design, messages resemble standard Internet electronic mail messages as defined in RFC822. The Usenet message format is described in RFC1036. This defines some additional headers. It also limits the values of some of the standard headers as well as giving some of them special semantics. Newsgroups are the classification system of Usenet. The required Newsgroups header specifies where a message, or article, should be filed upon reception. In addition to InterNetNews, there are two major Usenet packages available for UNIX sites. All three shares several common implementation details. USENET at first was built with effectively no security. There was limited auditing even to detect abuse, let alone prevent it. Over time abusers came, and this meant in many cases that "privileged" functions had to be in some places either shut down or "put on manual" at great administrative cost to admins. In some cases, actual security using digital signature was applied, for newgroup messages (pgpverify), moderated groups (pgpmoose) and NoCem. PGP was commonly used because it is a widely distributed standalone program capable of doing digital signature.

Keywords – Usenet, bulletin board, internet news software, newsgroups, internetnews architecture, usenet security

. ——————————— ◆ ———————————

1 INTRODUCTION

Usenet is a distributed bulletin board system, built as a logical network on top of other networks and connections. By design, messages resemble standard Internet electronic mail messages as defined in RFC822 [Crocker82]. The Usenet message format is described in RFC1036 . This defines some additional headers. It also limits the values of some of the standard headers as well as giving some of them special semantics. Newsgroups are the classification system of Usenet. [Adams87] The required Newsgroups header specifies where a message, or article, should be filed upon reception. Sites are free to carry whatever [transliteral] groups they want. Most sites carry the core set of so –called ``mainstream' groups.There are currently about 730 of these groups, and one or two new ones is created every week. Messages generated at a site are sent to the site's ``neighbors'' who process them and relay them to their neighbors, and so on. Sites can be interconnected -- indeed, on the Internet, this is quite common. See Figure
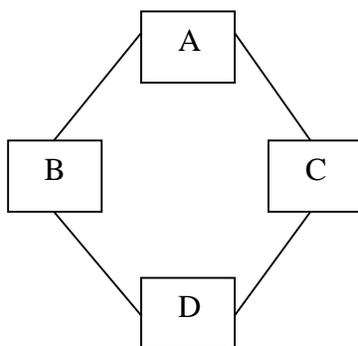

Figure 1: Small Usenet topology (all links are two -way).

The Path header is used to prevent message loops. For example, an article written at A could get sent to B , D , C , and then back to A. Before propagating an article, a site pretends its own name to the Path header. Before propagating an article to a site, the receiving host checks to make sure that the site that would receive the article does not appear in the Path line. For example, when the article arrived at site C , the Path would contain A!B!D , so site C would know not to send the article to A. Sites also keep a record of the Message -ID's of all articles they currently have. If D receives an article from B , it will reject the article if C offers it later. For self -protection, most sites keep a record of recent articles that they no longer have. This is very useful when another site dumps a (usually quite large) batch of old news back out to Usenet. For the past few years, the amount of data generated by Usenet sites has been doubling every year. A site that receives all the mainstream groups is receiving over 17 megabytes a day spread out over 11,000 articles . [Adams92] About 20% of the data is article headers, and while all of them must be scanned only half of it is must be processed by the Usenet software. The number of sites participating in Usenet has been growing almost as quickly. Based on articles his site receives and survey data sent in by participating sites, Brian Reid estimates that there are 36,000 sites with 1.4 million participants [Reid91] . A ``sendsys'' message to the ``inet'' distribution in June of 1989 received about 200 replies in the first twenty -four hours. A year later, nearly 700 replies were received. (Sendsys is a special article that asks all receiving sites to send back an email message, usually without human intervention; by convention, inet is

primarily the set of sites on the Internet.) [Adams87] The NNTP protocol is defined in Internet RFC 977 [Kantor86] published in February, 1986. This was

accompanied by the general public release of a reference implementation, also called ``nntp.'' This has been the only NNTP implementation that is generally available to UNIX sites. [Adams87]

Usenet Software

In addition to Internet News, there are two major Usenet packages available for UNIX sites. All three share several common implementation details. A newsgroup name such as comp.foo is mapped to a directory comp/foo within a global spool directory. An article posted to a group is assigned a unique increasing number based on a file called the *active* file. If an article is posted to multiple groups, links are used so that only one copy of the data is kept. A *sys* file contains patterns describing what newsgroups the site wishes to receive, and how articles should be propagated. In most cases, this means that a record of the article is written to a ``batchfile'' that is processed off-line to do the actual sending. The first Usenet package is called B News, also known as B2.11. The B news model is very simple: the program *rnews* is run to process each incoming article. Locking is used to make sure that only one *rnews* process tries to update the active file and history database. At one site that received over 15,000 articles per day, the locking would often fail so that 10 to 100 duplicates were not uncommon. Because each article is handled by a separate process, it is impossible to pre-calculate or cache any useful data. More importantly, file I/O had become a major bottleneck. A site that feeds 10 other sites does over 150,000 open/append/close operations on its batchfiles. It is generally agreed that B news cannot keep up with current Usenet volume; it is no longer being maintained, and its author has said more then once that the software should be considered ``dead.'' C News gets much better performance then B news by processing articles in batches [Collyer87] . The *relaynews* program is run several times a day to process all the articles that have been received since the last run. Since only one *relaynews* program is running, it is not necessary to do fine-grain locking of any of the supporting data files. More importantly, it can keep the entire active and sys file in memory. It can also use buffered I/O on its batchfiles, reducing the amount of system calls by one or two orders of magnitude. An alpha version of C News was released in October, 1987. Within four years it surpassed B news in popularity,

and there are now more sites running C News then ever ran B news. From the beginning, the NNTP reference implementation was layered on top of the existing Usenet software: an article received from a remote NNTP peer was written to a temporary file and the appropriate *rnews* or *relaynews* program processed it. In order to avoid processing an article the system already has, it first does a lookup on the history database to see if the article exists. It soon became apparent that invoking *relaynews* for every article lost all of C News's speed gain, so the NNTP package was changed to write a set of articles into a batch, and offer the batch to *relaynews.* When articles arrive faster then *relaynews* can process them, they must be spooled. If two sites (B and C in the previous examples) both offer a third site (D) the same article at the ``same time'' then an extra copy will be spooled, only to be rejected when it is processed, wasting disk space; this problem multiplies as the number of incoming sites increases.To alleviate this problem, most sites run Paul Vixie's *msgidd* , a daemon that keeps a memory -resident list of article Message -ID's offered within the last 24 hours. The NNTP server is modified so that it tells this daemon all of the articles that it is handing to Usenet and queries the daemon before telling the remote site that it needs the article. This is not a perfect solution -- if the first, spooled, copy of the article is lost or corrupted, the site will likely never be offered the article after the *msgidd* cache entry has expired. Going further, *msgidd* is work - around for a problem inherent in the current software architecture.

Other problems, while not as severe, lead to the conclusion that a new implementation of Usenet is needed for Internet sites. For example:

- Since all articles are spooled, *relaynews* cannot tell the NNTP server the ultimate disposition of the article, and the server cannot tell its peer at the other end of the wire. This hides transmission problems. For example, a site tracing the communication has no way of finding out an article was rejected because the remote site does not receive that particular set of newsgroups.

- The NNTP reference implementation is showing signs of age. Maintaining the server is becoming a maintenance nightmare; over one-tenth of its 6,800 lines are *#ifdef* related.

- All articles are written to disk at extra time. Disks are getting bigger, but not faster, while CPU's, memory, and networks are.

InterNetNews architecture

There are four key programs in the InterNetNews package :

- *Innd* is the principal news server for incoming newsfeeds;
- *innxmit* reads a file identifying articles and offers them to another site;
- *ctlinnd* sends control commands to *innd* ;
- *nnrpd* is an NNTP server oriented for newsreaders.

Of these programs, the most important is *innd*. We first mention enough of its architecture to give a context for the other programs, and then discuss its design and features in more detail at the end of this section.
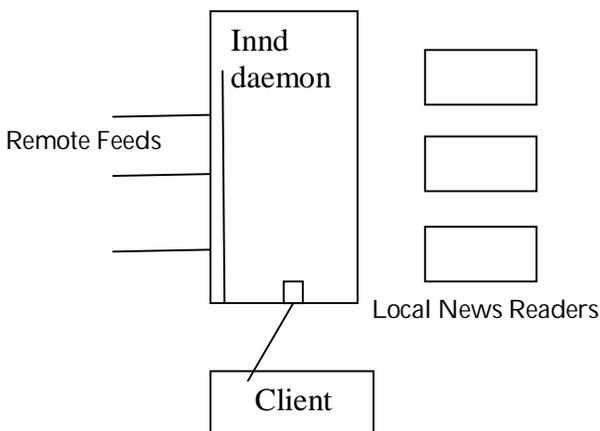

Figure 2 : Innd architecture

Innd is a single daemon that receives all incoming articles, files them, and queues them up for propagation. It waits for connections on the NNTP port. When a connection is received, a *getpeername* (2) is done. If the host is not in an access file, then an *nnrpd* process is spawned with the connection tied to its standard input and output. (Unlike other implementations, no single INN program implements the entire NNTP protocol.) It is worth noting that *nnrpd* is only about 3,500 lines of code, and 20% of them are for the ``POST'' command, used to verify the headers in a user's article. *Nnrpd* provides all NNTP commands except for ``IHAVE'' and an incomplete version of ``NEWNEWS''. On the other hand, it does provide extensions for pattern -matching on an article header and listing exactly what articles are in a group. The NNTP protocol seems to be a good example of the

UNIX philosophy: it is small, general, and powerful and can be implemented in a very small program. Articles are usually forwarded by having *innd* record the article in a ``batchfile'' which is processed by another program. For Internet sites, the *innxmit* program is used to offer articles to the host specified on its command line. The input to *innxmit* is a set of lines containing a pathname to the article and its Message -ID. Since the Message -ID is in the batchfile, *innxmit* does not have to open the article and scan it before offering the article to the remote site. This can give significant savings if the remote site already has a percentage of the articles. Until recently, *innxmit* used *writev* to send its data to the remote host. At start -up it filled a three - element *struct* iovec array with the following elements:

```
[0] { ".", 1 };
[1] { placeholder };
[2] { "\r\n" }
```

To write a line, the *placeholder* was filled in with a pointer to the buffer, and its length, and a single *writev* was done, starting from either element zero or one. While this implementation was clever, and simpler then what was done elsewhere, it was not very fast. *Innxmit* now uses a 16k buffer and only does a *write* when the next line would not fit. This is also consistent with ideas used throughout the rest of INN: use the *read* and *write* system calls, referencing the data out of large buffers while avoiding the copying commonly done by the standard I/O library. The *ctlinnd* program is used to tell the *innd* server to perform special tasks. It does this by communicating over a UNIX -domain datagram socket. The socket is behind a firewall directory that is mode 770, so that only members of the news administration group can send messages to it. It is a very small program that parses the first parameter in its command line and converts it to an internal command identifier. This identifier and the remaining parameters are sent to *innd* which processes the command, and sends back a formatted reply. For example, the following commands stops the server from accepting any new connections, adds a newsgroup, and then tells it to recompute the list of hosts that are authorized newsfeeds: ctlinnd pause "Clearing out log files" ctlinnd newgroup comp.sources.unix m vixie@pa.dec.com

```
ctlinnd reload newsfeeds "Added OSF feed"
ctlinnd go ""
```

The text arguments are sent to *syslog* (8) for audit purposes.

The most commonly -used *ctlinnd* command is ``flush.'' This directs the server to close the batchfile that is open for a site, and is typically used as follows:

```
mv batchfile batchfile.work
ctlinnd flush sitename
innxmit sitename batchfile.work
```

The flush command points out another difference between INN and other Usenet software. The B News *inews* program needed no external locking. Files were opened and closed for a very short window, the time needed to process one article. The C News *relaynews* could be running for a longer period of time. The only way to get access to a batchfile is to either lock the entire news system, which is overkill for the desired task, or to rename the file and wait until the original name shows up again. The INN approach is more efficient and conceptually cleaner.

USENET SECURITY -- NEEDS

USENET at first was built with effectively no security. Anybody, anywhere could introduce any article which could do anything. There was limited auditing even to detect abuse, let alone prevent it. Over time abusers came, and this meant in many cases that "privileged" functions had to be in some places either shut down or "put on manual" at great administrative cost to admins. In some cases, actual security using digital signature was applied, for newgroup messages (pgpverify), moderated groups (pgpmoose) and NoCem. PGP was commonly used because it is a widely distributed standalone program capable of doing digital signature.

Authority on USENET

USENET has no government. It is an anarchy -- the absence of government -- but this does not mean total chaos. It has rules, guidelines, traditions, movements and principles of governance, if not government. USENET is, in spite of its public nature, a privately owned network. It is a cooperative, owned by the owners of the sites that are on it. Nobody gets on the network or uses it without the permission, directly or indirectly of these owners. In these site owners lies all the authority on USENET. This makes sense, as anything on USENET involves storing or changing data on the site owner's machines. Those files are theirs. Of course, having each individual site owner privately administer all aspects of the net on their machine would never work. There are over a few hundred thousand machines on the net, serving millions of users. So means to delegate administration have been found. As noted, the first way to delegate it was to simply let anybody do anything. In fact, at first anybody could create a new newsgroup just by typing a new name. In the past there were not many malicious users, so the system worked.

Today we have malicious users. Both spammers and the like who abuse for imagined gains, and plain sociopaths like trollers and crackers who abuse the net or people on it for the sake of abusing it. Barring malice, in the past we still had politics -- different groups wanting different things. To solve this various anarchic and pseudo-democratic systems evolved to develop group consensus or a measurement of group will, and everybody agreed, without force, to go along with the group will where it was important. One example was the newsgroup voting system. This works because in fact to get anything done in a co-op like USENET, you need the almost unanimous consent of the site owners. Any site owner is free to not participate in any group, hierarchy or other activity. So you must keep them all happy if you want to do something netwide. [Postel82] While total unanimity is hard, near-unanimity, won through compromise, has actually worked better than might be expected. This is true in part because almost all of us are drilled from childhood to accept the democratic principle and accept things the majority wants so that we can get our way later when we agree with the majority.

What needs to be secured?

Security on USENET amounts to the question, "Should anybody and everybody be able to perform this action?" If the answer is yes, you need no security. If no, you need some security to divide those who you do wish to perform the action from those who you don't. Security of course has a cost, so sometimes you're willing to accept letting anybody perform some action if the risk of that is less than the bother of security. When the net was smaller, and there were few malicious people about, security wasn't necessary simply because even though anybody **could** do certain things, they tended not to. [Postel82]

Now on USENET, the only "action" is the posting of an article.

However, this gets broken down based on what the headers of the article do, and in particular the Control header on control messages. So while "post an article" is not the unit you secure, you are interested in "post a cancel" or "post an article in a moderated newsgroup."

Here is a list of the actions on USENET I believe most people would prefer not be available to anybody and everybody. As such, we must address how to secure them.

Arbitrary users should not be able to:

1. Cancel an article they were not involved in originating
2. Post an article identified as coming from somebody else against that person's will.
3. Post to a moderated group without the moderator's approval
4. Violate a newsgroup's policies on crossposting, article size, mime-types, etc.
5. Machine post vast volumes of articles to one or many groups
6. Create a group
7. Create or replace a named article with a specified purpose
8. Change or set the status of a group
9. Issue a sendsys in the name of another, or possibly any sendsys at all
10. Delete a group
11. Modify the headers or body of another's message
12. Block the propagation of another's message by hijacking its message-id
13. Issue a checkgroups or change a group description

While there is some fine debate about some, and in some cases these rules may vary in some hierarchies (for example alt might allow any party to create a group) I think that for the mainstream of USENET, ideally most people would prefer these functions were not entirely open. [Reid91]

Who can be trusted

If not all parties can be trusted to perform these actions, who can or should be trusted? Well, that varies from action to action. In some cases, like the cancel message, everybody agrees the original poster of a message should be trusted, and most agree the administrators of the equipment used to insert the posting into the net should be trusted as well. Many others wish to pick specific 3rd parties and trust them, to deal with abuse. For other functions it's more political. The actions themselves require subjective judgement and must be performed by individuals or groups who win the trust of the machine owners who in turn grant it. It turns out that the vast majority of people on USENET can be trusted, at at least given the benefit of the doubt, with their trust revoked only after it is abused. That's how the net used to work, but there was no way to revoke the trust when people started abusing. The answers as to who people want to trust to perform these actions are varied and many. The underlying security system has to allow people to create the various structures of trust and enabling that they desire.

System Goals

Delegation
It's vital that all trust be easily delegated. Site admins don't want to worry about the administrative problems of newsgroups or other small subsets of the net. They want to examine the big picture at best, and sometimes no picture at all. Most would like to just leave things to work, and only deal with problems if problems occur that are big enough to reach their attention. The delegation itself must be secure.

Synchronization
In addition, secure delegation is the only way that cooperation on USENET can work. If each site tunes its own parameters for a newsgroup independently, either by deliberate will or more commonly just by accident, then the group starts being useful to none. All sites have to be reasonably in sync -- the near unanimity -- about how most components of the net work. You can't have a third of the sites thinking a group is moderated with one moderator, another third naming a different moderator and the other third thinking the group is unmoderated. The group becomes damaged for everybody.

Extensibility
While I've listed common things we want to secure today, it is certain that other things will come up. New control messages or headers. New newsgroup policies and ways

to run newsgroups. Good software design insists that there be one system -- since this is complex enough as it is

-- and that it be general and extensible, just like the other facets of USENET are.

Security

Good security is of course secure. Once you have a general security system, you don't leave things unsecured without an explicit reason for doing so. Sadly, when you have malicious attackers, if you close one security hole they just move to the next. You must close all the holes they will try. This is usually not totally possible, but one tries to get as close as possible to the goal as one can.

Anarchic

Unlike most systems, a security system for USENET has to factor in its anarchic nature. Authority remains with site admins, you can never prevent that. However, the system must allow people to work together, to cooperate and compromise as they see fit. Security becomes a tradeoff between the burdens of complexity of security and cooperation and the risks of insecurity. [Reid91]

Conclusions and Comparisons

The InterNetNews architecture works. Profiling a production installation for 24 hours showed that *open* (2) accounted for 10% of the run time. Since the server only does one *open* (2) per article, it is not clear if any other performance tuning is needed. The profiling overhead accounted for 5% of the run-time. Several optimizations are available because there is only one process, and because it is always running. For example, avoiding duplicates is an integral part of the server. If a second site offers an article while a first site is sending it, the NNTP code will put the channel to ``sleep'' for a short while before replying to the second offer. This is usually enough time to have the first site finish sending the article, reducing the number of duplicates from hundreds to nearly none, with no external programs. Since the server is always running, the system has a much smoother performance curve. As a result, it ``feels'' much faster to users. Another unexpected benefit is that articles are accepted or rejected synchronously. A user can post

an article, and by the time their posting agent has returned, it has been written to the spool directory and queued for remote transfer. If there is a problem such as having an illegal newsgroup specified, the user founds out immediately. The design of the server seems to be very good, split into abstractions that are very independent. For example, sites have no knowledge of incoming NNTP connections. Using callbacks lets each portion of the server safely do I/O without worrying that it might affect other parts. Much of the Usenet processing becomes trivial when serialized, such as access to the history file. The design has also led to a fairly small program: it is under 13,000 lines, and about 20% of them are comments. This compares favorable to the 7,400 lines in the equivalent C News program and the 7,600 lines in the NNTP reference implementation. [Reid91]

References

1]. [Crocker82] David H. Crocker, *Standard for the Format of ARPA Internet Text Messages* , Request For Comments 822, Marina del Rey, CA: Information Sciences Institute, 1982.

2.][Adams87] Rick Adams, Mark Horton, *Standard for Interchange of USENET Messages* , Request For Comments 1036, Marina del Rey, CA: Information Sciences Institute, 1987.

3] [Adams92] Rick Adams, *Total traffic through uunet for thelast2weeks*Usenetmessage<1992Apr8.193050.8963@uunet.uu.net> in *news.lists* , April, 1992.

4] [Collyer87] Geoff Collyer and Henry Spencer, *News Need not be Slow* , Usenix Winter Conference, 1987.

5] [Kantor86] Brian Kantor, Phil Lapsley, *Network News Transfer Protocol*: A Proposed Standard for the Stream - Based Transmission of News , Request for Comments 977, Marina del Rey, CA: Information Sciences Institute, 1986.

6] [Postel82] Jonathan B. Postel, *Simple Mail Transfer Protocol* , Request For Comments 821, Marina del Rey, CA: Information Sciences Institute, 1982.

7] [Reid91] Brian Reid, *Usenet Readership Summary* Report for May 91 , Usenet message<1991Jun2.141124.12753 @pa.dec.com

Preparation, Verification and Finding Out of the Critical Current of Thin Sample of YBCO Compounds

Shaikh Md. Rubayiat Tousif, Shaiyek Md. Buland Taslim

Abstract— This paper is all about superconductors. It deals with the current state of high temperature superconductors (HTS), the application of these materials and possible breakthroughs in the field. It also provides information about how to synthesize YBCO compounds and explains the material from structural point of view. It describes the method of verifying the prepared material is a superconductor or not by observing Meisner effect at 77K. Finally, it describes the technique of finding out the critical current by finding out the resistances of a thin prepared sample of YBCO for temperatures between 77 and 300K.

Keywords—Applications, Critical current, Resistance measuremetns, Structure, Synthesis

——————————————— ◆ ———————————————

1 INTRODUCTION

The discovery of superconductivity above liquid-nitrogen temperature in cuprate materials (.High-Temperature Superconductors. .HTS.) raised an unprecedented scientific euphoria16 and challenged research in a class of complicated compounds which otherwise would have been encountered on the classical research route of systematic investigation with gradual increase of materials complexity only in a far future. The plethora of preparational degrees of freedom, the inherent tendency towards inhomogeneities and defects, in combination with the very short SC coherence length of the order of the dimensions of the crystallographic unit cell did not allow easy progress in the preparation of these materials. Nevertheless, after enormous preparation efforts HTS arrived meanwhile at a comparatively mature materials quality that allows now a clearer experimental insight in the intrinsic physics which is still awaiting a satisfactory theoretical explanation. The present situation of HTS materials science resembles in many aspects the history of semiconductors half a century ago. The new dimension in the development of HTS materials, in particular in comparison with the case of silicon, is that HTS are multi-element compounds based on complicated sequences of oxide layers. In addition to the impurity problem due to undesired additional elements, which gave early semiconductor research a hard time in establishing reproducible materials properties, intrinsic local stoichiometry defects arise in HTS from the insertion of cations in the wrong layer and defects of the oxygen sublattice. As additional requirement for the optimization of the SC properties, the oxygen content has to be adjusted in a compound-specific off-stochiometric ratio, but nevertheless with a spatially homogeneous microscopic distribution of the resulting oxygen vacancies or interstitials. Today, reproducible preparation techniques for a number of HTS material species are available which provide a first materials basis for applications. As a stroke of good fortune, the optimization of these materials with respect to their SC properties seems to be in accord with the efforts to improve their stability in technical environments in spite of the only metastable chemical nature of these substances under such conditions.

This article talks about various aspects of the high temperature superconductor (YBCO compounds). It includes the current state of superconductivity, application of these HTS materials and the reason behind their behabiour. Most importantly it describes experimental procedures of systhezing and finally finding electrical properties of an already prepared material.

2 APPLICATIONS

Besides the scientific interest, the search for applications has always been a driving force for superconductor materials science. Right from the discovery, it had been envisioned that SC coils with high persistent current might be used to generate strong magnetic fields. However, in the first generation of SC materials (.*type-I*.) superconductivity was easily suppressed by magnetic fields: The agnetic self-field generated by the injected current prevented high-field as well as high-current applications. A first step towards this goal was the discovery of *type-II* superconductors where the magnetic penetration depth λ exceeds the SC coherence length ξ. This enables a coexistence of superconductivity and magnetic fields, which are allowed to penetrate into the SC bulk in the quantized form of vortices. The concomitant substan-

---

- *Shaiyek Md. Buland Taslim is currently pursuing masters degree program in electric power engineering in Royal Institute of Techolgy (KTH) Sweden, PH-01913531992. E-mail: buland_taslim@yahoo.com*
- *Shaikh Md. Rubayiat Tousif is working as a Lecturer in American International University-Bangladesh (AIUB), Bangladesh, PH-01913531993. E- mail: tousif@aiub.edu*

tial reduction of the loss of SC condensation energy that has to be paid for magnetic field penetration facilitates the survival of superconductivity even in strong magnetic fields, at least up to a certain critical field H_{c2} where the SC state no longer survives the .vortex swiss cheesing.. The last ingredient required for technically applicable ".*hard*" superconductors was the discovery and engineering of pinning centers which fix penetrated magnetic flux and prevent its Lorentz force driven flow through the superconductor that otherwise generates power dissipation.

Today, NbTi and Nb_3Sn conductors are the basis of a billion Euro SC wire industry which delivers magnets that cannot be realized by means of conventional metal wire conductors, e. g., for Magnetic Resonance Imaging (MRI) systems and High-Energy Physics (HEP) particle accelerators. The enormously high critical fields $H_{c2} \sim 100$ T of HTS indicate their potential for extremely high-field applications. However, HTS vortex physics has turned out to be much more complex than what had been known from classical superconductors. This implies strong restrictions for high-field, high-temperature HTS magnet hopes[8]. Nevertheless, in spite of earlier concerns about the ceramic nature of HTS, flexible[9] HTS-based conductors are steadily progressing towards applications where a substantial size decrease justifies the cryogenic efforts. HTS current leads are just being introduced worldwide in HEP accelerators to transport kA-sized feed currents at a substantially reduced heat leakage from a liquid-nitrogen (.LN_2.) temperature region to LHe cooled SC NbTi coil systems.

Nb-based SC rf-cavities represent another recent technological progress of HEP accelerators: An extremely high quality factor provides here a much better transfer of acceleration energy to the particle bunches than in conventional cavities[11]. Miniaturized microwave filters, e. g., for mobile phone base stations, are at present the most advanced HTS electronics application: The low loss of thin film HTS resonator stripes with a typical size 50 μm * 1 cm allows a complex coupling of a large number of such resonators on a chip which enables filters with sharp frequency cut-offs.

Josephson junctions, well-defined weak links of SC regions, can be coupled to .Superconducting QUantum Interferometric Devices. (.SQUIDs.), magnetic flux detectors with quantum accuracy that are the most sensitive magnetic field detectors presently available. SQUIDs based on Nb/AlO_x/Nb Josephson junctions achieve today at LHe temperature a magnetic noise floor $\sim$ 1 fT/$\sqrt{}$ Hz which enable diagnostically relevant magnetic detection of human brain signals (.magnetoencephalography, .MEG.). HTS SQUIDs at liquid nitrogen operation have approached this magnetic sensitivity within one order of magnitude and are already in commercial use for the nondestructive evaluation (NDE) of defects in complex computer chips and aircrafts.

In the 1970s and 1980s, IBM as well as a Japanese consortium including Fujitsu, Hitachi, and NEC tested in large projects the fast switching of Josephson junctions from the SC to the normal state with respect to a post-semiconductor computer generation. Unfortunately, the switching from the normal to the SC state turned out to limit the practical performance to several GHz instead of the theoretical $\sim$1 THz[12]. Meanwhile, new device concepts based on the transport of single magnetic flux quanta reestablished the feasibility of THz operation. The hottest topic of present Josephson circuit investigations is the realization of quantum computing with "Qubits" encoded by the SC wave function around μm-sized loops containing single or even half flux quanta. At present, among all demonstrated Qubit realizations a SC electronics implementation appears to have the largest potential of upscalability to the size of several kQubit, which is required for first real applications: The lithographic requirements of $\sim$ 1 μm minimum feature size are already common practice in present semiconductor circuits.

For all these applications of superconducitvty, the necessity of cryogenics is at least a psychological burden. Nevertheless, with the present progress of small cryocoolers SC devices may evolve within foreseeable future to push-button black-box machines that may be one day as common practice as nowadays vacuum tube devices in ordinary living rooms.

3 SYSNTHESIS

Relatively pure YBCO was first synthesized by heating a mixture of the metal carbonates at temperatures between 1000 to 1300 K.

$$4\ BaCO_3 + Y_2(CO_3)_3 + 6\ CuCO_3 + (1/2-x)\ O_2 \rightarrow 2\ YBa_2Cu_3O_{7-x} + 13\ CO_2$$
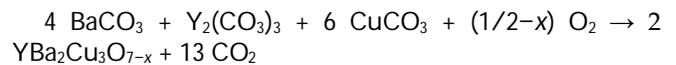
Modern syntheses of YBCO use the corresponding oxides and nitrates.

The superconducting properties of $YBa_2Cu_3O_{7-x}$ are sensitive to the value of x, its oxygen content. Only those materials with $0 \leq x \leq 0.65$ are superconducting below T_c, and when $x \sim 0.07$ the material superconducts at the highest temperature of 95 K, or in highest magnetic fields: 120 T for B perpendicular and 250 T for B parallel to the CuO_2 planes.

In addition to being sensitive to the stoichiometry of oxygen, the properties of YBCO are influenced by the crystallization methods used. Care must be taken to sinter YBCO. YBCO is a crystalline material, and the best superconductive properties are obtained when crystal grain boundaries are aligned by careful control of annealing and quenching temperature rates.

Numerous other methods to synthesize YBCO have developed since its discovery by Wu and his coworkers, such as chemical vapor deposition (CVD), sol-gel, and aerosol methods. These alternative methods, however, still require careful sintering to produce a quality product.

However, new possibilities have been opened since the discovery that trifluoroacetic acid (TFA), a source of fluo-

rine, prevents the formation of the undesired barium carbonate ($BaCO_3$). Routes such as CSD (chemical solution deposition) have opened a wide range of possibilities, particularly in the preparation of long length YBCO tapes. This route lowers the temperature necessary to get the correct phase to around 700 °C. This, and the lack of dependence on vacuum, makes this method a very promising way to get scalable YBCO tapes.

3.1 Sample Fabrication

First of all a suitable recipe must be found out to prepare YBCO compound in the laboratory. The recipe should follow the basic synthesis process described in the previous section.

After finding a suitable recipe and ingredients. The appropriate materials are mixed up. It is better to use the (ceramic) mortar and pistil to mix the powders. Enough powders are used to make about two, 1 cm³ pellets.

Next, the powders are annealed in air. Then the powders are placed into the alumina crucible and the Fisher Muffle oven is used. Metal "tongs" gloves and eye protection are taken when the crucible is inserted into the hot oven.

After the annealing period, a black, superconducting YBCO powder is obtained. Next, Powder is pressed into an, approximately, 1 cm diameter pellet. This is accomplished with the hydraulic "press" in Chemistry.

After powder is pressed into one or two pellets, the pellets must be annealed in oxygen. This final anneal is accomplished in the Lindberg tube furnace. This anneal is an easy, since this furnace is controlled electronically, and may be programmed for an entire temperature/time sequence.

After the fabrication of the sample, the Meisner effect is verified by levitating a magnet.
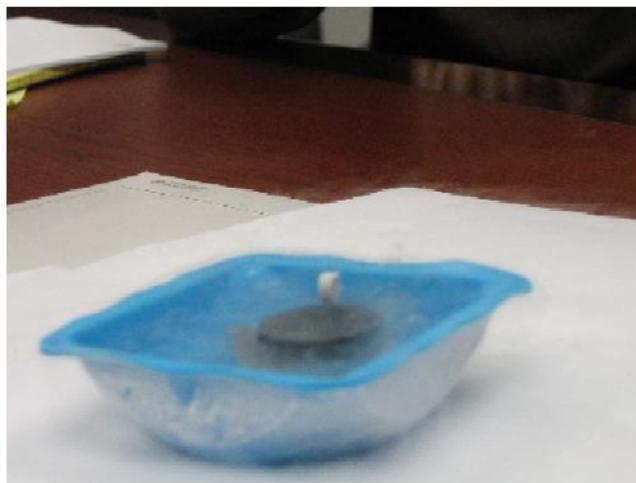

Fig.1 Demonstration of levitation of magnet by superconducting pellet. The black disk in the blue tray of liquid nitrogen at 77 K (-320.8º F). The object floating in the mist is a small permanent magnet.

Necessary laboratory preparation must be taken dur-

ing the whole experimental procedure.

4 STRUCTURE OF YBCO COMPOUNDS

YBCO crystallises in a defect perovskite structure consisting of layers. The boundary of each layer is defined by planes of square planar CuO_4 units sharing 4 vertices. The planes can some times be slightly puckered. Perpendicular to these CuO_2 planes is CuO_4 ribbons sharing 2 vertices. The yttrium atoms are found between the CuO_2 planes, while the barium atoms are found between the CuO_4 ribbons and the CuO_2 planes. This structural feature is illustrated in the figure.
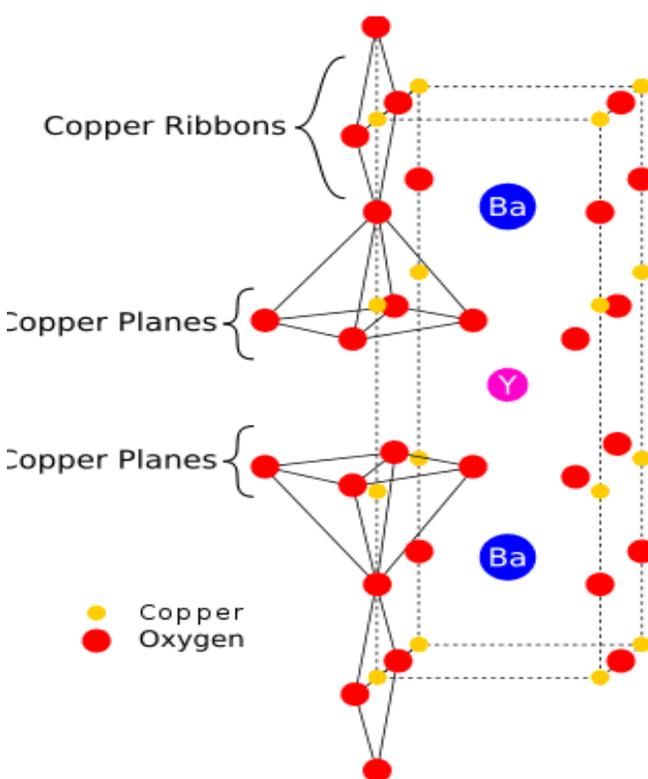

Fig.2 Structure of YBCO compound

Although $YBa_2Cu_3O_7$ is a well-defined chemical compound with a specific structure and stoichiometry, materials with less than seven oxygen atoms per formula unit are non-stoichiometric compounds. The structure of these materials depends on the oxygen content. This non-stoichiometry is denoted by the $YBa_2Cu_3O_{7-x}$ in the chemical formula. When $x = 1$, the O(1) sites in the Cu(1) layer are vacant and the structure is tetragonal. The tetragonal form of YBCO is insulating and does not superconduct. Increasing the oxygen content slightly causes more of the O(1) sites to become occupied. For $x < 0.65$, Cu-O chains along the b-axis of the crystal are formed. Elongation of the b-axis changes the structure to orthorhombic, with lattice parameters of $a = 3.82$, $b = 3.89$, and $c = 11.68$ Å. Optimum superconducting properties occur when $x \sim$

0.07 and all of the O(1) sites are occupied with few vacancies.

In experiments where other elements are substituted at the Cu and Ba sites evidence has shown that conduction occurs in the Cu(2)O planes while the Cu(1)O(1) chains act as charge reservoirs, which provide carriers to the CuO planes. However, this model fails to address superconductivity in the homologue Pr123 (praseodymium instead of yttrium). This (conduction in the Copper planes) confines conductivity to the *a-b* planes and a large anisotropy in transport properties is observed. Along the *c*-axis, normal conductivity is 10 times smaller than in the *a-b* plane. For other cuprates in the same general class, the anisotropy is even greater and inter-plane transport is highly restricted.

Furthermore, the superconducting length scales show similar anisotropy, in both penetration depth ($\lambda_{ab} \approx$ 150 nm $\lambda_c \approx$ 800 nm) and coherence length, ($\xi_{ab} \approx$ 2 nm, $\xi_c \approx$ 0.4 nm). Although the coherence length in the *a-b* plane is 5 times greater than that along the *c*-axis it is quite small compared to classic superconductors such as niobium (where $\xi \approx$ 40 nm). This modest coherence length means that the superconducting state is more susceptible to local disruptions from interfaces or defects on the order of a single unit cell, such as the boundary between twinned crystal domains. This sensitivity to small defects complicates fabricating devices with YBCO, and the material is also sensitive to degradation from humidity.


Fig. 3 Crystal structure of $YBa_2Cu_3O_7$ ("YBCO"). The presence of the CuO chains introduces an Ortho-rhombic distortion of the unit cell.
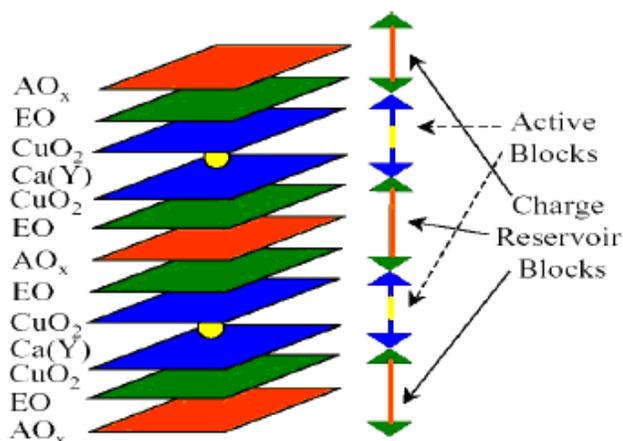
Fig. 4 General structure of a cuprate HTS.

5 RESISTANCE MEASUREMENTS

The idea was to measure the resistance of a YBCO sample from room temperature down to the boiling temperature of liquid nitrogen (77 K). This may be accomplished with two different sample holders. The choice of method depends on the availability of the equipment. The can be many methods for this purpose but here only two methods are talked about. Method 1 uses the "helium dipper" while method 2 uses the RMC closed-cycle helium refrigerator. The basic steps for either method are the same: (1) the YBCO thin film sample is mounted on a copper block whose temperature may be varied, (2) a temperature sensor mounted on the copper block is used to determine film temperature, and (3) four spring-loaded, pressure contacts are pushed against the sample and connected to a circuit that will enable to measure the sample resistance. Then the sample temperature was varied and the resulting resistances were recorded.

5.1 4-Probe Resistance Cryogenic Dipper

This holder uses a diode temperature sensor, similar to the one in the Janis model DT liquid nitrogen cryostat. If a temperature controller is available, it could be used to monitor the temperature. If not, a 10 μA current source can be used and the diode voltages are measured.
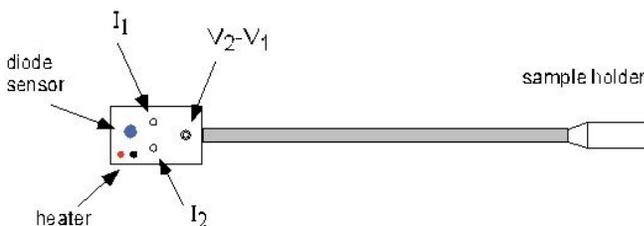

Fig.5 4-Probe Dipper

The "sleeve" must be removed and the sample holder must be inspected. An Ohmmeter is used to determine how the contacts are wired to the BNC connectors at the

top of the "dipper."2

The nuts are loosened that hold the contact assembly down and the contacts are lifted so that there is sufficient space to slide the thin-film YBCO sample between the contacts and the copper holder. An insulator is being sandwiched between YBCO sample and the copper holder. Using tweezers and a small screw-driver, the YBCO sample under the contact assembly is positioned. The four contacts are made sure to be used are placed squarely on the YBCO film. Gradually the nuts are tighten so that the contacts are forced down against the sample.3

Final adjustment of the contact pressure should be made while monitoring the resistance between all pairs of contacts with an Ohmmeter. With good contacts, the resistance between any two is made no more than a few hundred ohms. The contacts and reposition the film are raised if necessary to achieve this.
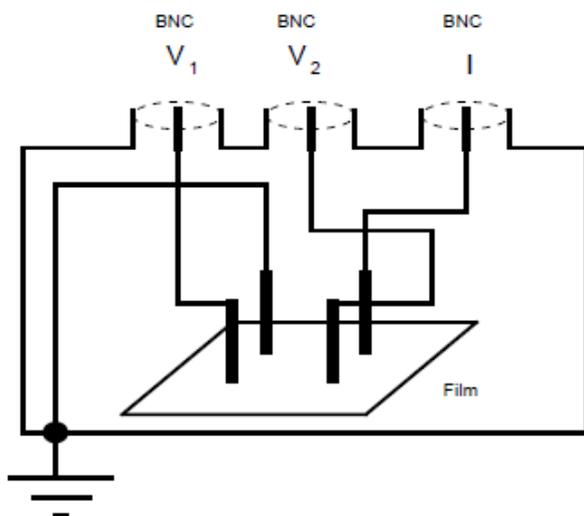
Fig.6 Wiring diagram for dipper.

A lock-in amplifier is used to perform the standard 4-probe resistance measurement on the film. Also a large ballast resistor ($R_B > 10$ kΩ) in series with the oscillator output of the lock-in to establish a constant current of (roughly) 10 µA.4 Send this current through the film using two of the contacts is used. The voltage drop (with the lock-in) across the other two film contacts is measured.
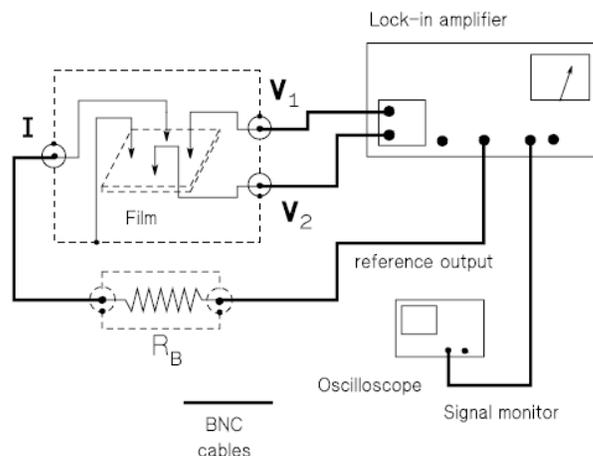

Fig.7 Wiring diagram to measure film resistance using a lock-in amplifier.

While monitoring the film resistance with the lock-in, the "sleeve" must be carefully slide back over the sample holder. Constant attention must be paid to the lock-in reading as the dipper is moved subsequently. It is easy for the contacts to degrade when the sample holder is jarred.

The diode temperature sensor on the dipper is used to measure sample temperature. This is done by hooking up either of the Lake Shore temperature controllers (if available) or by using a 10 µA current source while measuring the diode voltage. As the sample temperature is lowered periodically the lock-in voltage (proportional to the resistance of the YBCO sample) is recorded and temperature (or diode voltage, from which temperature may be determined). If available, a p. c. with a 2-channel A/D board or an X-Y recorder can be used to simultaneously record both of these voltages.

The top flange from the helium storage dewar is removed and replace it with the dipper/flange.

The pressure relief valve on the helium storage dewar must be opened. As the is dipper slowly lowered into the dewar the cryogenic fluid is boiled, generating gas pressure that must be released.

The dipper is slowly lowered into the cryostat, constantly monitoring the lock-in voltage and temperature. Ideally a continuous stream of data (X-Y recorder or digitized data) is obtained. The sporadic data are also recorded.

From the obtained data, film resistance versus temperature down to 77 K for a measurement current of 10 µA can be plotted. The onset temperature, Tc, and the R=0 temperature can be determined too.
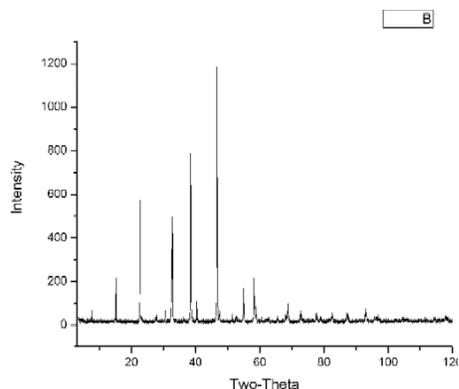
Fig. 8. X-Ray Diffraction pattern of the powder YBCO sample.

Figure 8 shows the X-Ray diffraction pattern of the powder YBCO sample. The YBCO compound prepared following the procedure described in the topic sample fabrication if passed through the X-Ray diffraction tube will give a reading as figure 8.
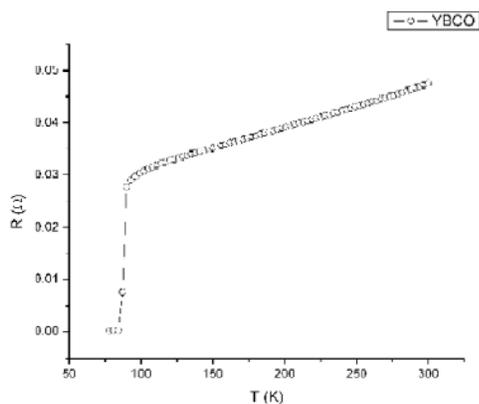

Fig. 9. Resistance as a function of temperature for the sample can be converted to resistivity using the measured sample dimensions.

Figure 9 shows resistance as a function of temperature for the sample can be converted into resistivity using the sample dimensions. Resistance is founf out following the method described in resistance measurement topic.
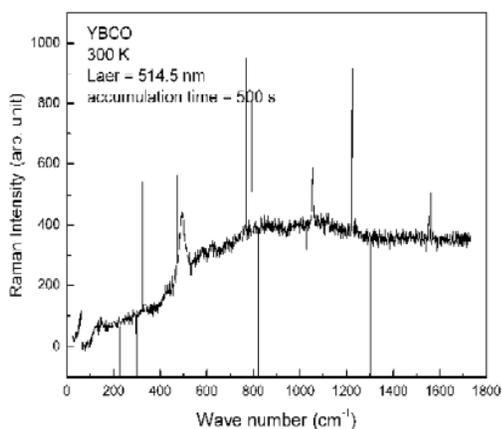

Fig.10. Raw Raman data with sharp singlepoint spikes from cosmic

ray hits. Measurments conducted in Prof. T. Zhou's lab.
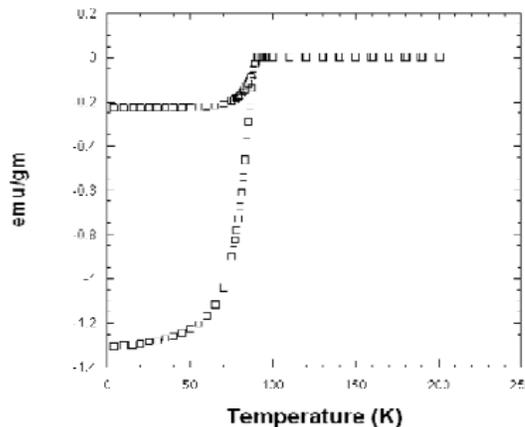
Fig.11. Magnetization measurements in a 100 G field to determine the superconducting fraction of the sample. The zero field cooled portion of the plot can be used to extract the superconducting fraction. Measurements were conducted at Prof. M. Greenblatt's Lab. at Rutgers University.

Figure 10 and 11 shows different other readings taken on the sample materials.

ACKNOWLEDGMENT

REFERENCES

[1] Roland Hott2, Reinhold Kleiner3, Thomas Wolf2 and Gertrud Zwicknagl4. 2005. "Superconducting Materials — A Topical Overview", Publisher: Springer Berlin Heidelberg.

[2] High-Temperature Superconductivity in Cuprates. The Nonlinear Mechanism and Tunneling Measurements Series: Fundamental Theories of Physics, Vol. 125, Mourachkine, A. 2002, 340 p., Hardcover

[3] Engel, S., Thersleff, T., Hühne, R., Schultz, L., Holzapfel, B. Enhanced flux pinning in YBCO layers by the formation of nano-sized $BaHfO_3$ precipitates using the chemical deposition method.

[4] Knoth, K., Engel, S., Apetrii, C., Falter, M., Schlobach, B., Hühne, R., Oswald, S., Schultz, L., Holzapfel, B. Chemical solution deposition of $YBa_2Cu_3O_{7-x}$ coated conductors.

[5] Falter, M., Demmler, K., Häßler, W., Schlobach, B., Holzapfel, B., Schultz, L. Chemical solution deposition (CSD) of $YBa_2Cu_3O_{7-x}$ films and oxide buffer layers by dip coating.

[6] http://hyperphysics.phy-astr.gsu.edu/hbase/Solids/scond.html

Recitation of Load Balancing Algorithms In Grid Computing Environment Using Policies And Strategies - An Approach

M.Kamarunisha, S.Ranichandra, T.K.P.Rajagopal

Abstract -**Grid computing is a term referring to the combination of computer resources from multiple administrative domains to reach common goal. What distinguishes grid computing from conventional high performance computing systems such as cluster computing is that grids tend to be more loosely coupled, heterogeneous, and geographically dispersed. Grid computing is the next generation IT infrastructure that promises to transform the way organizations and individuals compute, communicate and collaborate. The goal of Grid computing is to create the illusion of a simple but large and powerful self-managing virtual computer out of a large collection of connected heterogeneous systems sharing various combinations of resources. Grid Resource Management is defined as the process of identifying requirements, matching resources to applications, allocating those resources, and scheduling and monitoring Grid resources over time in order to run Grid applications as efficiently as possible. Focus of this paper is on analyzing Load balancing requirements in a Grid environment and proposing a centralized and sender initiated load balancing algorithm. A load balancing algorithm has been implemented and tested in a simulated Grid environment.**

I. INTRODUCTION

The rapid development in computing resources has enhanced the performance of computers and reduced their costs. This availability of low cost powerful computers coupled with the popularity of the Internet and high-speed networks has led the computing environment to be mapped from distributed to Grid environments [1]. In fact, recent researches on computing architectures are allowed the emergence of a new computing paradigm known as Grid computing. Grid is a type of distributed system which supports the sharing and coordinated use of geographically distributed and multiowner resources, independently from their physical type and location, in dynamic virtual organizations that share the same goal of solving large-scale applications.

In Grid computing, individual users can access computers and data, transparently, without having to consider location, operating system, account administration,

and other details. In Grid computing, the details are abstracted, and the resources are virtualized. Grid Computing has emerged as a new and important field and can be visualized as an enhanced form of Distributed Computing [2]. Sharing in a Grid is not just a simple sharing of files but of hardware, software, data, and other resources [2]. Thus a complex yet secure sharing is at the heart of the Grid.

II. WHY GRID TECHNOLOGIES?

Computers have been proven to be very efficient to solve complex scientific problems. They are used to model and simulate problems of a wide range of domains, for instance medicine, engineering, security control and many more. Although their computational capacity has shown greater capabilities than the human brain to solve such problems, computers are still used less than they could be. One of the most important reasons to this lack of use of computational power is that, despite the relatively powerful computing environment one can have, it is not adapted to such complicated computational purposes. The following are given the reasons for why we need grid computing.

III. LOAD BALANCING IN GRID ENVIRONMENT

A key characteristic of Grids is that resources (e.g., CPU cycles and network capacities) are shared among numerous applications, and therefore, the amount of resources available to any given application highly fluctuates over time. In this scenario load balancing plays key role. Load balancing is a technique to enhance resources, utilizing parallelism, exploiting throughput improvisation, and to cut response time through an appropriate distribution of the application. To minimize the decision time is one of the objectives for load balancing which has yet not been achieved. As illustrated in Figure1 load balancing feature can prove invaluable for handling occasional peak loads of activity in parts of a larger organization.
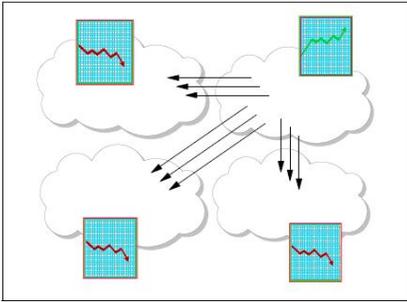
Figure 1: Job Migration

A. Load Balancing Algorithms

Algorithms can be classified into two categories: static or dynamic.

### (i) Static Load Balancing Algorithm
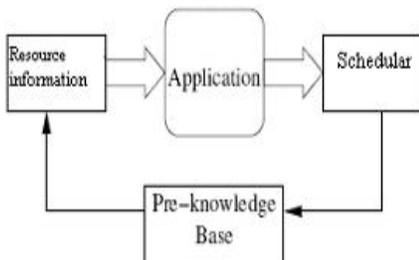
Figure.2: Static Load Balancing

Static load balancing algorithms allocate the tasks of a parallel program to workstations based on either the load at the time nodes are allocated to some task, or based on an average load of our workstation cluster. The decisions related to load balance are made at compile time when resource requirements are estimated.

(ii) Dynamic Load Balancing Algorithm

Dynamic load balancing algorithms make changes to the distribution of work among workstations at run-time; they use current or recent load information when making distribution decisions. Multicomputers with dynamic load balancing allocate/reallocate resources at runtime based on no a priori task information, which may determine when and whose tasks can be migrated. As a result, dynamic load balancing algorithms can provide a significant improvement in Performance over static algorithms.
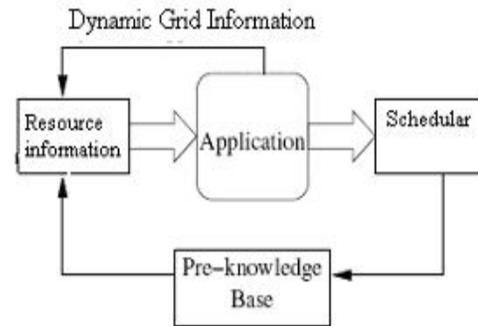

Figure.3: Static Load Balancing

IV. LOAD BALANCING STRATEGIES

There are three major parameters which usually define the strategy a specific load balancing algorithm will employ. These three parameters answer three important questions:
• who makes the load balancing decision
• what information is used to make the load balancing decision, and
• Where the load balancing decision is made.

A. Sender-Initiated vs. Receiver-Initiated Strategies

The question of who makes the load balancing decision is answered based on whether a sender-initiated or receiver-initiated policy is employed. In sender-initiated policies, congested nodes attempt to move work to lightly-loaded nodes. In receiver-initiated policies, lightly-loaded nodes look for heavily-loaded nodes from which work may be received. The sender-initiated policy performing better than the receiver-initiated policy at low to moderate system loads. Reasons are that at these loads, the probability of finding a lightly-loaded node is higher than that of finding a heavily-loaded node. Similarly, at high system loads, the receiver initiated policy performs better since it is much easier to find a heavily-loaded node.

B. Global vs. Local Strategies

Global or local policies answer the question of what information will be used to make a load balancing decision in global policies, the load balancer uses the performance profiles of all available workstations. In local policies workstations are partitioned into different groups. The benefit in a local scheme is that performance profile information is only exchanged within the group. The choice of a global or local policy depends on the behavior an application will exhibit. For global schemes, balanced load convergence is faster compared to a local scheme since all workstations are considered at the same time.

C.Centralized vs. De-centralized Strategies

A load balancer is categorized as either centralized or distributed, both of which define where load balancing decisions are made [44-46]. In a centralized scheme, the load

balancer is located on one master workstation node and all decisions are made there. Basic features of centralized approach are:
• a master node holds the collection of tasks to be performed
• tasks are sent to the execution node
• when a execution process completes one task, it requests another task from the master node

V. LOAD BALANCING POLICIES

Load balancing algorithms can be defined by their implementation of the following policies:
• Information policy: specifies what workload information to be collected, when it is to be collected and from where.

. Triggering policy: determines the appropriate period to start a load balancing operation.
• Resource type policy: classifies a resource as server or receiver of tasks according to its availability status.
• Location policy: uses the results of the resource type policy to find a suitable partner for a server or receiver.
• Selection policy: defines the tasks that should be migrated from overloaded resources (source) to most idle resources (receiver).

The main objective of load balancing methods is to speed up the execution of applications on resources whose workload varies at run time in unpredictable way. Hence, it is significant to define metrics to measure the resource workload. The success of a load balancing algorithm depends from stability of the number of messages (small overhead), support environment, low cost update of the workload, and short mean response time which is a significant measurement for a user. It is also essential to measure the communication cost induced by a load balancing operation.

VI. PROBLEM FORMULATION

In grid environments, the shared resources are dynamic in nature, which in turn affects application performance. Workload and resource management are two essential functions provided at the service level of the Grid software infrastructure. To improve the global throughput of these environments, effective and efficient load balancing algorithms are fundamentally important. The focus of our study is to consider factors which can be used as characteristics for decision making to initiate Load Balancing. Load Balancing is one of the most important factors which can affect the performance of the grid application. This work analyzes the existing Load Balancing modules and tries to find out performance bottlenecks in it. All Load Balancing algorithms implement five policies [3].

The efficient implementation of these policies decides overall performance of Load Balancing algorithm.

The main objective of this paper is to propose an efficient Load Balancing Algorithm for Grid environment. Main difference between existing Load Balancing algorithm and proposed Load Balancing is in implementation of three policies: Information Policy, Triggering Policy and Selection Policy. For implementation of Information Policy all existing Load Balancing algorithm use periodic approach, which is time consuming.

The proposed approach uses activity based approach for implementing Information policy. For Triggering Load Balancing proposed algorithm uses two parameters which decide Load Index. On the basis of Load Index Load Balancer decide to activate Load Balancing process. For implementation of Selection Policy Proposed algorithm uses Job length as a parameter, which can be used more reliably to make decision about selection of job for migration from heavily loaded node to lightly loaded node. Following table discusses the main differences between the proposed algorithm and Condor Load Balancing algorithm.

	Information Policy	Triggering Policy	Selection Policy
Condor Load Balancer (existing)	Load Balancing information is collected using periodic approach	Load Balancer is triggered based on Queue Length	Task is selected for migration using Job Length as criteria.
Proposed Load Balancer	Load Balancing information is collected using Activity based approach	Load Balancer is triggered based on Queue Length and current CPU Load	Task is selected for migration based upon CPU consumption of tasks

Table 1: Comparison between Condor LB Module and Proposed LB Module

VII. PROPOSED LOAD BALANCING ALGORITHM

Load balancing is defined as the allocation of the work of a single application to processors at run-time so that the execution time of the application is minimized. This chapter is going to discuss the design of proposed Load Balancing algorithm.

A. Background

While many different load balancing algorithms have been proposed, there are four basic steps that nearly all algorithms have in common:

1. Monitoring workstation performance (load monitoring)
2. Exchanging this information between workstations (synchronization)
3. Calculating new distributions and making the work movement decision (rebalancing criteria)
4. Actual data movement (job migration)

Efficient Load Balancing algorithm makes Grid Middleware efficient and which will ultimately leads to fast execution of application in Grid environment. In this work, an attempt has been made to formulate a decentralized, sender-initiated load balancing algorithm for Grid environments which is based on different parameters. One of the important characteristics of this algorithm is to estimate system parameters such as queue length and CPU utilization of each participating nodes and to perform job migration if required.

B. Design of Load Balancing Algorithm
Load balancing should take place when the load situation has changed. There are some particular activities which change the load configuration in Grid environment. The activities can be categorized as following:
• Arrival of any new job and queuing of that job to any particular node.
• Completion of execution of any job.
• Arrival of any new resource
• Withdrawal of any existing resource.
Whenever any of these four activities happens activity is communicated to master node hen load information is collected and load balancing condition is checked. If load balancing condition is fulfilled then actual load balancing activity is performed. Following is the proposed algorithm for Load Balancing:

Loop
wait for load change
// depends on happening of any of four defined activities
if (activity_happens ())
If (LoadBalancing_start ())
while HeavilyLoaded_list is not empty
Determine tasks which can be migratable using criteria of CPU consumed by each job which has least CPU consumption selected for being migrated.
Selected job = j;
If LightlyLoaded_list is empty
PendingJob_list = PendingJob_list + j;
Else
Migrate (LightlyLoaded_list [first], HeavilyLoaded_list[n], j);
// update the database

End while
End Loop

Following are some functions used in the above algorithm:
Activity_happens (): this function return Boolean value. If any of above defined activity occurs it returns true otherwise it returns false.
LoadBalancing_start (): this function also return Boolean value. If on the basis of given parameters (CPU utilization and queue length) load balancing will be required it will return true else it will return false. This function also updates two lists:
HeavilyLoaded_list and LightlyLoaded_list: Threshold heavy load and threshold light load is defined initially which depends on the traffic of application on the Grid.

Function: LoadBalancing_start
Return Type: Boolean
Start:
If (Standard Deviation of Load of nodes < SD_Threshold)
If (Load of any node is greater then average Load value of nodes)
HeavilyLoaded_list= HeavilyLoaded_list + l (new selected node);
End if
Else (Load of any node is greater then threshold heavy load value)
HeavilyLoaded_list= HeavilyLoaded_list + l (new selected node);
Else if (Load of any node is less then threshold light load value)
End
Outline of Load Balancing_start Function

Here actual load distribution is performed at a centralized controller or manager node. The central controller polls each workstation and collects state information consisting of a node's current load as well as the number of jobs in the node's queue. The polling is done on basis of occurrence of some defined activity. It is not done periodically. Periodic checking approach is used in Condor. In case of periodic approach Load Balancer collects load sample periodically which is not required and infect creates an overhead also.
In the proposed algorithm information is collected only if there is a change in configuration of rid. This information is used to perform load balancing. Above is the flow diagram of algorithm. First of all it initializes different parameters. Whenever any of four activities which are required to start information policy of load balancing occurs, it starts collecting load balancing information. Once information has been gathered then it is decided that load

balancing is required or not. For this purpose application uses CPU utilization and queue length parameters.

With help of these parameters we decide which resource is heavily loaded and which resource is lightly loaded. After selection of resource the application selects job out of n-jobs running on that resource. This selection is based upon on CPU consumption of different jobs. Least CPU consumed job will be selected for migration. When job is selected, application checks for available lightly loaded resource. If lightly loaded resource is available then migrate selected job from heavily loaded resource to lightly loaded resource. If no lightly loaded resource is available then add selected job to pending job list. This job will be executed later when some lightly loaded resource will be available. Finally all the value will be updated in database.
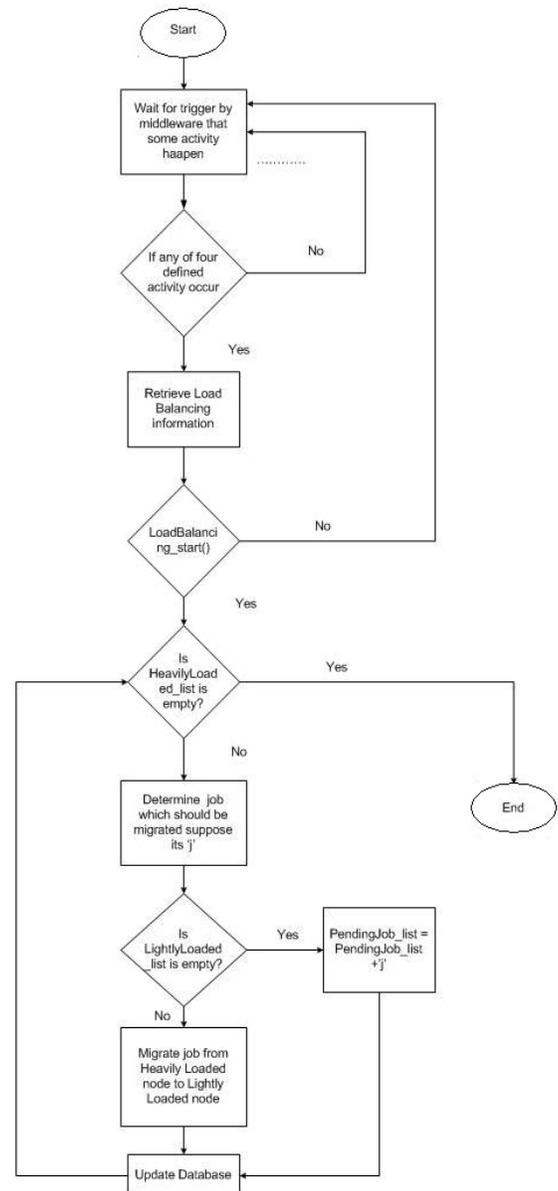

Figure 4: Flow Chart of Algorithm

VIII. IMPLEMENTATION DETAILS

    A Load Balancing Module has been developed which executes in simulated grid environment. This application has been developed using J2EE and MySQL database server.

Screen Shot.1: Image of Load Balancing (1) Page

Above is the image of Load Balancing (1) page. This window appears after Load Balancing has been performed. In normal scenario if sufficient lightly loaded resources are available then after load balancing no heavily loaded resource will be available. Job from all heavily loaded resource will be migrated to lightly loaded resource. This page also gives information about which Job ID is migrated from which resource to which resource.

Figure 5: Overall System Architecture

Above is the overall architecture of the application developed. Information about all resources is stored in resource database. Resources are generated by GridSim. Resource discovery process use resource database to discover all possible match to the resource query. Next process is resource selection and allocation. This process is also done by GridSim. Once resource allocation is done then Load Balancing process come in existence. Execution of Load Balancing depends on condition specified.

IX. EXPERIMENTAL RESULTS

Screen Shot.2: Image of Load Balancing (2) Page

Above is the image of Load Balancing (2) page. This image shows after Load Balancing has been performed but job is not migrated. This is one particular case when heavily loaded resource has been finalize and job which should be migrated has been finalize but there is no corresponding lightly loaded resource is available. In this case job is put in to Pending Job list. When ever any lightly

loaded resource will be available this job will be migrated to the lightly loaded resource.

X. CONCLUSION

Grid application performance remains a challenge in dynamic grid environment. Resources can be submitted to Grid and can be withdrawn from Grid at any moment. This characteristic of Grid makes Load Balancing one of the critical features of Grid infrastructure. Here we have focused on Load Balancing and tried to present the impacts of Load Balancing on grid application performance and finally proposed a efficient Load Balancing algorithm for Grid environment.

In this work we analyzed existing Load Balancing algorithm and proposed an enhanced algorithm which more efficiently implements three out of five policies implemented in existing Load Balancing algorithm. These three policies are: Information Policy, Triggering Policy and Selection Policy. Proposed algorithm is executed in simulated Grid environment.

REFERENCES

[1] Krishnaram Kenthapadi, Stanford University , kngk@cs.stanford.edu and Gurmeet Singh Mankuy , Google Inc., manku@google.com,Decentralized Algorithms using both Local and Random Probes for P2P Load Balancing.

[2] Ian Foster , Carl Kesselman Steven Tuecke , The Anatomy of the Grid Enabling Scalable Virtual Organizations , Intl J. Supercomputer Applications, 2001.

[3] Francois Grey, Matti Heikkurinen, Rosy Mondardini, Robindra Prabhu, "Brief History of Grid", http://Gridcafe.web.cern.ch/Gridcafe/Gridhistory/history.html.

[4]Marcin Bienkowski, Miroslaw Korzeniowski, Friedhelm Meyer aud der Heide,Dynamic Load Balancing in Distributed Hash Tables.

[5]Giuseppe Di Fatta and Michael R. Berthold, Department of Computer and Information Science, University of Konstanz, Decentralized Load Balancing for Highly Irregular Search Problems.

[6]Anthony Sulistio, Chee Shin Yeo, and Rajkumar Buyya, Visual Modeler for Grid Modeling and Simulation (GridSim) Toolkit.

Assessment and Comparison of Soil carbon pool under Silvo-pastoral Agroforestry system in the North Wales, UK

Kasahun Kitila H, Douglas Godbold, Hussein Omed

Abstract - As result of increased emission of green house gases, especially increased emission of Co2, Climate change is the main global challenges that many countries are facing. Increasing carbon sequestration through a forestation, reforestation and appropriate land use practices are considered as means to sink the atmospheric Co2 in terrestrial ecosystem. Agroforestry is recognized as a strategy for soil carbon sequestration (SCS) under the afforestation/reforestation activities in different parts of the world.However, little information is available on soil carbon dynamics under agroforestry systems. This study was aimed to determine the soil organic carbon pool under silvo-pastoral agroforestry system. The study was conducted at Henfeas research center in the north Wales, UK where Sycamore (Acer pseudoplatanus L.) and Red alder (Alnus rubra) were planted in 1992 in integration with the grasslands. The soil samples were collected to the depth of 30cm at different depth intervals (0-10, 10-20 and 20-30cm) under five treatments: under and outside the canopy of both Sycamore (Acer pseudoplatanus L.) and Red alder (Alnus rubra) and under the control grassland. The concentration of soil organic carbon (SOC %) under each treatment were analyzed using LOI (loss on ignition method) where soil samples were burned at 450 oc. The regression formula ($Y = 0.458X - 0.4$ Where, $Y = SOC$ (%) and $X = SOM$ (%)) developed by Ball, 1964, was used to convert soil organic matter to SOC. It was identified that SOC concentration were significantly different at ($P < 0.05$) between the treatments and along the soil profile.

Keywords- SOC, climate change, land use, Soil, Co2 emissions, decomposition, microbial activity

─ ─ ─ ─ ─ ─ ─ ─ ─ ◆ ─ ─ ─ ─ ─ ─ ─ ─ ─

1. Introduction

Climate change is the biggest global challenges affecting environmental, economical and social welfare throughout the world (IPCC, 2007). The latest assessment report of the Intergovernmental Panel on Climate Change (IPCC, 2007) projected that global average temperatures in 2100 will be between 1.8 and 4.0 °C higher than the 1980–2000 average. Precipitation will also expect to be increased by 2.29% at the end of the century. On the other hand, sea levels are expected to increase to 0.59 meters by 2100 based on observed rates of ice flow from Greenland and Antarctica.

Increased emissions of green house gas (GHG) to the atmosphere are the main causes for the Climate change to be happened (IPCC, 2007). According to this report, emissions of GHG due to anthropogenic activities like from different

agricultural land to perennial vegetation, applying different, soil management practices like mulch farming, reduced tillage, integrated nutrient management (INM), and integrated land management like introduction of agroforestry system Freeman et al. ,(2004). They found that agroforestry practices are the most important land use system in enhancing both above and below ground carbon storage.

1.2 Soil carbon pool potential in UK

factories, and land use change contributes significant amount for the increased GHG specially the concentration of CO_2 in the atmosphere. Studies by (Paustian et al., 2000) also suggested that, Climate change is usually attributed to the extreme use of fossil fuel by industrialized countries and the conversion of forests to agricultural and urbanization in the developing or poor countries. Soil is the most important feature where large amount of Soil organic carbon (SOC), which is originated from plants and animal tissue that exist in different stages of decomposition, can be stored (Lal et al., 2001). It was estimated that soils contain more than three times the atmospheric pool and more than four times the biotic pool of carbon

There are different land use management practices suggested to boost the capacity of soil to store carbon. This include conversion of marginal

According to the study by Bradley et al.,(2005) , Soussana et a.,(2004) and UK country soil survey report in 2007, most of terrestrial carbon pool in UK is found below ground up to 30cm depth . The total soil carbon pool was estimated by integrating soil series databases and land cover in the country. According to these reports, the total amount of carbon in the soils of UK was estimated to be 9.8 ± 2.4 billion tones where 6.9 billion tones in Scotland and 2.8 billion tones in England and Wales.Intergovernmental Panel on Climate Change (IPCC, 2007) also suggested that, the UK's Climate Change act committed to cut 80% of greenhouse gases

(GHGs) by 2050 so as to achieve the commitments to stabilize the atmospheric carbon dioxide to below 400PPm and the global average temperature rise to 2°C. In the some report, the UK government also has planned for a low carbon transition in agriculture to voluntarily reduce 6-11% of GHG. Increasing the level of soil organic carbon levels as climate change mitigation strategy was largely being ignored by climate policymakers and analysts in the UK, partly due to the inadequacies of the current agricultural GHG accounting systems (Smith P.,2008). However, recently encouraging agreements have been made at the European level to reduce agricultural greenhouse gas emissions by at least 20% by 2020, primarily by storing carbon in the soil (EU Agriculture Commissioner :Mariann Fischer Boel, September, 2009). Currently, the UK Government published strategy to safeguard soil carbon pool acknowledged that preventing emissions from soil and exploring how to increase existing stores of soil carbon can make an important contribution to meeting the Government's emission reduction targets and carbon budgets.

1.3 Rationale of the study

The environmental, economical and social potential advantage of agroforestry was highly pronounced now days in both temperate and tropical climatic zones. The study by Sinclair F. et al, (2000) and Young, 1997) suggested that agroforestry system provide very diversified benefit like to enhance biodiversity in certain landscapes, to reduce soil erosion and nitrate leaching in intensively managed agricultural landscapes and it enable farmers to cultivate poor soils as arable land by enhancing soil organic matter, improving water holding capacity, and increasing nutrient inputs through nitrogen fixation. This study also indicated that, when agroforestry is immediately established after slash and burn agriculture, 35% of the original forest carbon stock can be restored in improving soil quality and quantity through organic inputs from crop residues and tree litter resulting in the maintenance or increase of soil organic matter (SOM).
International Panel on Climate Change (IPCC 2007) estimated that the current total area under agroforestry is 400 million ha that have a potential to store 0.72 Mg ha^{-1} year^{-1} of carbon (Watson et.al. ,2000). However, the impact of any agroforestry system on soil carbon sequestration mainly affected by the amount and quality of biomass input available by tree, grasses and soil properties that significantly alters the rate of turnover of organic matters Nair *et al*., (2009).

In silivo-pastoral agroforestry practices, there is close interaction between trees and pasture that attribute the system to be the largest potential to sequester carbon in both above and below ground to offsets greenhouse gas emissions associated with deforestation and shifting cultivation (Richard T. Conant e*t al*., 2001). Silvo-pastoral agroforestry is widely practiced in temperate region including in Uk where timber and sheep production are integrated (Buck *et al.,* 1999). The incorporation of trees on farms affects carbon stocks differently compared to a single cropped area since trees are important in key processes such as nutrient cycling (Watson *et al.,* 2000).

Now days, research in temperate and tropical agroforestry systems has focused on role of agroforestry in soil and water conservation, crop and pasture productivity, and nutrient cycling. However, little information is available on role of agroforestry systems in soil carbon dynamics and its potential to store SOC Oelbermann et al., (2007). On the other hand the losses of soil organic carbon (SOC) from forest and grassland due to different anthropogenic activities were studied and documented but there is knowledge gap on whether or not SOC under forest and grass vegetation differs (Barker et al., 1996). This project is required to provide understandings on contribution of silvo-pastoral agroforestry practices in increasing SOC.

2. Research Objectives

1) To quantify and compare the amount of carbon stored under different tree species and grassland in silvo-pastoral agroforestry system

2) To determine the total carbon pool and changes with the depth of soil profiles under the silvo-pastoral agroforestry system.

3. Materials and Methods

3.1 Study area:

The Henfaes experimental site is one of a national network of six sites established across the country investigating the potential of silvo-pastoral agroforestry in UK farms (Sibbald and Sinclair, 1990). It was established in 1992 on 14 ha of agricultural land at the University of Wales, Bangor farm (Henfaes), which is located in Abergwyngregyn, Gwynedd, 12 km east of the city of Bangor. The climate is hyper oceanic, with an annual rainfall of about 1000 mm. The soil is a fine loamy brown earth which was classified as a Dystric Cambisol in the FAO systems of classfication. Topography consists of a shallow slope on a deltaic fan of approximately 1-2^0 and the aspect is north-westerly, at an altitude of 4-14 m above sea level.

According to the study report by Teklehaimanot and Sinclair, (1998), the depth of the water table ranges between 1 - 6 m. The entire site was sown to a mixture of perennial ryegrass (*Lolium perenne* L.) and white clover (*Trifolium repens* L.) in April 1992 at a seed rate of 12.5 kg ha[-1] of *L. perenne* var. Talbot, 12.5 kg ha[-1] of *L. perenne* var. Condessa, 2 kg ha[-1] of *T. repens* var. Gwenda and 2 kg ha[-1] of *T. repens* var. S184.

This experimental site has a common set of core treatments described by Sibbald and Sinclair (1990). These comprise sycamore (*Acer pseudoplatanus* L.) planted at 100 and 400 stems ha[-1] into grazed pasture and at 2500 stems ha[-1] without grazing as a farm woodland control, and pasture without trees as an agricultural control. There are also additional treatments at the Henfaes site: red alder (*Alnus rubra* Bong.) planted at 400 stems ha[-1] into grazed pasture and at 2500 stems ha[-1] without grazing as a farm woodland control, and sycamore (*Acer pseudoplatanus L.*) planted at an overall density of 400 stems ha[-1] integrated with grazing pasture.

3.2 Experimental design

This experiment was conducted in agroforestry research experiments where grassland integrated with both sycamores (Acer pseudoplatanus L.) and red alder (*Alnus rubra)*. These trees were planted 18years ago (in 1992) with the spacing of 400stems per hectare. Both sycamore (*Acer pseudoplatanus L.)* and red alder *(Alnus rubra)* were planted in separated plot integrated with grazing pasture in strip arrangements.

The experiment had five treatments with seven replications from which soil samples were collected. The treatments were:

1. Under the canopy of Sycamore *(Acer pseudoplatanus L.)*
2. Outside the canopy of Sycamore (Acer pseudoplatanus L.))
3. Under the canopy of Red alder (*Alnus rubra)*
4. Outside the canopy of Red alder *(Alnus rubra)*
5. Control (under grassland without trees)

3.3 Soil sampling

Soil samples were collected from both under and outside the canopy of sycamore *(Acer pseudoplatanus L.)* and red alder *(Alnus rubra)* and from the control (grassland with no trees). For this experiment, seven stems or trees were randomly selected for each species. The sample plots under and outside of the canopy of both trees were arranged in a perpendicular to each other. Soil samples were collected from under the canopy of both trees and

from the mid points of the grass strip that is parallel to the selected stem. For soil sampling under the control (grassland with no trees), the transect line was drown starting from the center boundary line. The samples were taken from seven different points along the transect line.

For soil sampling under the control, the transect line was drown starting from the center boundary line. The samples were taken from seven different points along the transect line.

Soil samples were taken to the depth of 30 cm at different depth interval (0-10, 10-20, 20-30cm). This depth was to which SOC is most likely affected due to land use change and this sampling technique was also used in many similar studies to assess the soil carbon pool under different land use system.

3.4 Soil organic carbon analysis using loss on ignition method (LOI)

The soil samples were oven dried at 105 °c for 24 hours to remove the moisture. The dried samples were grinded and sieved to 2 mm size to remove large particles (generally those particles greater than 2-mm in diameter) to make the samples homogenous for further analysis.

The loss-on-ignition (LOI) method was used for the determination of soil organic matter content. About 20gm of oven dried soil sample was added to a ceramic crucible (or similar vessel). The samples were then heated to 450°C overnight (16hrs) to remove all soil carbon (Ball, 1964).
%SOM (Soil organic matter) = (Weight of oven dried-weight after burning) /weight of oven dried X 100. Finally, the loss-on-ignition (LOI) method determines only the organic matter content in the soil. For the sample burned at the temperature of 450°c, there was the regression formula or correction factor developed by (Ball, 1964) to convert soil organic matter (%) to SOC (%). The result was calculated by using this regression formula: Y= 0.458X-0.4, Where, Y= SOC (%) and X= SOM (%) or LOI (%).

4. Data analysis

The data collected during the experiment were analyzed using SPSS16.0 statistical software. Depending on the characteristics of the variables assessed and the distribution of the data, one way and two-way analysis of variance (ANOVA) was used for the analysis to test differences in soil organic carbon (SOC) among the treatments and soil profile. The differences in Soil chemical and physical properties across the treatments and soil profile or depth were also tested at statistically different parameters ($p<0.05$). Post-hoc tests (Tukey HSD) were used to further compare the treatment means.

Correlation analyses were also carried out to detect functional relationships among key soil variables (soil pH, bulk density and moisture content) and their interaction with the change in SOC.

5. RESULT

5.1 Soil organic carbon (SOC) content under Silvo-pastoral agroforestry system

The SOC content was analyzed for each treatment. The mean and standard error of mean for SOC (%) and other soil properties under each treatment were analyzed and summarized in table 1. Fisher's least significant difference (LSD) was used to test the significance difference of means that were considered significantly different at α=0.05 probability level table 2. Based on this analysis, the Mean difference in SOC was highly significant at (P<0.05) between control grassland (CGL) and the rest of the treatments (outside and under the canopy of both sycamore *(Acer pseudoplatanus L.)* and red alder *(Alnus rubra)* .However, SOC content was not significantly different at (P>0.05) between outside and under the canopy of both red alder *(Alnus rubra)* and sycamore *(Acer pseudoplatanus L.)*.

Table 1: Mean ± standard error of mean of SOC, pH, bulk density and moisture content under each treatments

Treatments	Differ			
	Soil properties			
	SOC (%)	pH	Bulk density (g/cm³)	Moisture content (%)
Under control grassland	0.89 ±0.61	6.74± 0.40	2.43±0.15	10.76±2.48
Outside the canopy of red alder	4.30±1.74	5.03±0.05	1.90±0.26	12.73±1.88
Outside the canopy of sycamore	4.51±1.90	5.24±0.10	1.46±0.05	16.73±2.21
Under the canopy of red alder	6.39±2.88	4.98±0.08	1.56±0.31	18.0±2.78
Under the canopy of sycamore	5.81±2.59	4.93±0.07	1.50±0.26 .	18.60±2.97
Ground mean	4.30±2.77	5.38±0.20	1.77±0.42	15.38±3.83

Table 2: Multiple Comparisons of the mean difference of SOC content between the treatments using Fisher's least significant difference (LSD)

Treatments		Mean Difference [a]	Sig. level
Control grassland	Outside canopy of red alder	-3.80600[a]	.045
	Outside canopy of sycamore	-4.01967[a]	.041
	Under canopy of red alder	-5.89733[a]	.006
	Under canopy of sycamore	-5.31733[a]	.011
Outside canopy of red alder *(Alnus rubra)*	Control grassland	3.80600[a]	.045
	Outside canopy of sycamore	-.21367	.903
	Under canopy of red alder	-2.09133	.251
	Under canopy of sycamore	-1.51133	.399
Outside canopy of sycamore *(Acer pseudoplatanus L.)*	Control grassland	4.01967[a]	.041
	Outside canopy of red alder	.21367	.903
	Under canopy of red alder	-1.87767	.300
	Under canopy of sycamore	-1.29767	.467
Under canopy of red alder *(Alnus rubra)*	Control grassland	5.89733[a]	.006
	Outside canopy of red alder	2.09133	.251
	Outside canopy of sycamore	1.87767	.300
	Under canopy of sycamore	.58000	.742
Under canopy of sycamore *(Acer pseudoplatanus L.)*	Control grassland	5.31733[a]	.011
	Outside canopy of red alder	1.51133	.399
	Outside canopy of sycamore	1.29767	.467
	Under canopy of red alder	-.58000	.742

The Mean difference is significant at $P< 0.05$.

5.2 Soil organic carbon analysis at varies depth intervals

The SOC content of soil was significantly different with the depth at ($P<0.05$) (See table 3 below). The Mean difference was highly significant between the depth intervals (0-10cm) and (20-30cm). However, the mean difference in SOC (%) was not significantly different at ($P>0.05$) between 0-10 and 10-20cm depth intervals.

The trend of SOC (%) content along the soil profile was analyzed for each treatment figure 1. Accordingly, Mean of SOC under the canopy of red alder *(Alnus rubra)* is higher than the rest of the treatments while SOC under control grassland is much lower than other treatments.

Table 3: Multiple Comparisons of the mean difference of SOC content between the depth intervals using Fisher's least significant difference (LSD)

Depth (cm)		Mean Difference	Sig. level
0-10	10-20	2.19360	.177
	20-30	3.86340[a]	.027

10-20	0-10	-2.19360	.177
	20-30	1.66980	.296
20-30	0-10	-3.86340[a]	.027
	10-20	-1.66980	.296

a.   The Mean difference is significant at P< 0.05 level
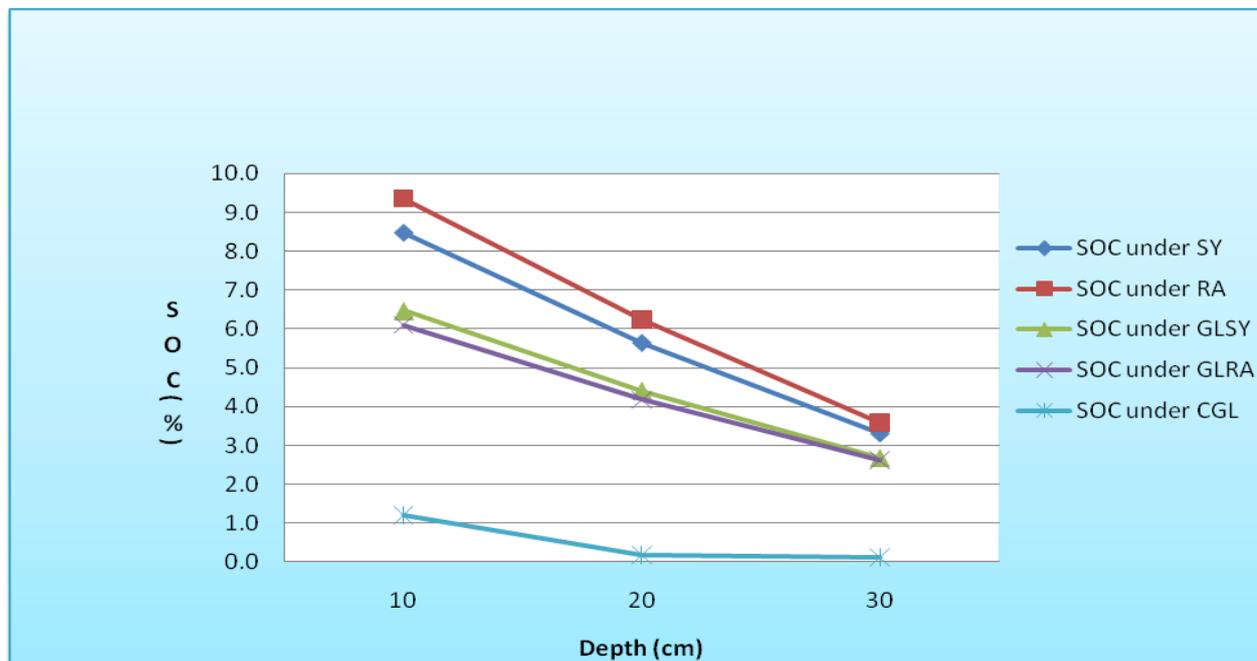
Figure 1: Variation of soil carbon pool at different depth interval under and outside the canopy of both sycamore *(Acer pseudoplatanus L.)* and red alder *(Alnus rubra)* and under the control grassland

Table 4: Mean ± standard error of SOC, pH, bulk density and moisture content in varies depth intervals

Soil depth (cm)	Soil properties			
	SOC (%)	PH	Bulk density(g/cm³)	Soil moisture content (%)
0-10	6.32±3.16	5.19±0.22	1.56±0.44	17.96±3.99
10-20	4.12±2.37	5.30±0.34	1.79±0.40	15.02±2.92
20-30	2.46±1.37	5.65±0.48	1.95±0.41	13.16±3.50
Ground mean	**4.30±2.77**	**5.38±0.20**	**1.77±0.42**	**15.38±3.83**

6.   Discussion
6.1 Soil organic carbon (SOC) under silvo-pastoral agro forestry system

According to this study result, although carbon inputs were higher under the canopy of the trees i.e under the canopy of both sycamore *(Acer pseudoplatanus L.)* and

red alder *(Alnus rubra)*, soil organic carbon content was not significantly different among the agroforestry components or treatments except with the SOC under the control grassland table 2. SOC was significantly lower under the control grassland (on grassland without trees) as compared with the SOC concentration under and outside the canopy of both sycamores *(Acer pseudoplatanus L.)* and red alder *(Alnus rubra)* (under the silvo-pastoral agroforestry system where trees and grasses are integrated). Therefore, this experiment clearly showed that agroforestry land use contributes high significance to store SOC that could be one of the strategies in mitigating climate change or offsetting CO_2 emissions.

This experiment is highly complemented with the previous studies by (Nair and Kumar, 2009). These studies have suggested that agroforestry practices like silvo-pastoral systems are important for conserving and sequestering soil carbon due to the greater interactions between trees and pasture through facilitating carbon input or exchange between the systems. They also indicated that agroforestry land use systems have the potential to offset immediate greenhouse gas emissions associated with deforestation and shifting cultivation.

Some other authors tried to compare the total SOC under agro forestry system and grassland. They have showed that agroforestry have higher potential to sequester carbon than pastures and field crops (Kirby and Potvin, 2007, Haile et al., 2008). The amount of SOC stored under agro forestry system was also compared among other land uses by Watson et al., (2000). This study has suggested that, trees in agroforestry systems can store high SOC pool compared to mono cropped areas. The study also indicated that, agroforestry systems contain 50 to 75 Mg of carbon per hectare compared to row crops that contain less than 10 Mg of carbon per hectare.

Reviewing SOC content in agroforestry system in comparison with other land-use systems, Nair et al. (2009) tried to rank the land-use systems in terms of their SOC content in the order: forests > agroforests > tree plantations > arable crops including grasslands. In addition, IPCC report in 2000 also predicted the Carbon sequestration potential of different land use and management options up to 2040's .According to these report, agroforestry was the promising land use system to sequester high carbon to mitigate the climate change as compared with other land use practices.

Soil organic carbon concentration was significantly varied along the soil profile. It was higher in the upper 0–10cm of the soil layer as compared with the SOC content (10-20cm and 20-30cm) showing the decreasing

trend with the depth under each treatments (See the SOC trend with depth in figure 1.

Studies by (Makumba et al., 2007) suggested that SOC content of the soil varies with the soil profile due to the fact that accumulation of organic matter from the litter fall, dead wood and branches is usually higher to the surface of the soil profile. Soil organic matter accumulation is generally higher to the first 0-20cm depth (Makumba *et al.*, 2007). Makumba and his colloquies also suggested that even though most of the tree roots occur to add substantial amounts of carbon from root exudates and fine-root turnover in the deeper soil layers, SOC content is low.

In this experiment, it was identified that the mean concentration of SOC (%) was higher under the canopy of red alder *(Alnus rubra)* where mean SOC to 30cm depth was 6.4% as compared with under the canopy of sycamore *(Acer pseudoplatanus L.)* (5.8%), outside the canopy of sycamore *(Acer pseudoplatanus L.)* (4.51%), outside the canopy of red alder *(Alnus rubra)* 4.3%, and under the control grassland (2.06%). The reason why SOC was higher under the canopy of red alder *(Alnus rubra)* is not clearly known but it could be due to the fact that red alder *(Alnus rubra)* can actively fix atmospheric nitrogen that is important to increase the biomass production and consequently used as an input for the increment of SOC in the soil (Teklehaimanot and Martin, 1998). This study showed that, red alder *(Alnus rubra)* was introduced at Henfaes to investigate the use of biological nitrogen fixation as an alternative to chemical fertilizer.

The study by (Jackson *et al.,* 2000) also suggested that, in silvo-pastoral agro forestry systems, where trees are allowed to grow in integration with grasses, there are high probability of altering the above- and belowground total productivity and changes in the quantity and quality of litter inputs. Such changes in vegetation component, litter, and soil characteristics modify the carbon dynamics and storage in the ecosystem.

The data analysis result in table 2 clearly showed that, trees positively affect the carbon pool potential under the grassland (outside the canopy of both red alder *(Alnus rubra)* and sycamore *(Acer pseudoplatanus L.)*. The mean SOC content outside the canopy of both red alder *(Alnus rubra)* and sycamore (*Acer pseudoplatanus L.)* were 4.30% and 4.51% respectively. These values are almost two times higher than the SOC content of the soil under the control pasture (2.06%) that was not integrated with any tree species. Therefore, from this result one can conclude that tree have appositive impact in increasing

SOC under the grassland in silvo-pastoral agro forestry system. Studies also showed that the presence of trees have both positive impacts on the productivity of grazed pasture through shelter effects and improving soil fertility or increasing soil organic matter (SOM) (Sibbald et al. 1991) and negative impacts, through competition for light, water and nutrients (Sinclair et al. 2000).

CONCLUSION AND RECOMMENDATIONS

This study examined the contribution of silvo-pastoral agroforestry system to store soil organic carbon. It was identified that the integration of trees with grass in silvo-pastoral agroforestry system is important to sequester more soil organic carbon as compared with mono cropping system. On the other hand, soil organic carbon pools under the agroforestry components were compared. The result showed that red alder (*Alnus rubra*) could store more SOC than under sycamore (*Acer pseudoplatanus L.*) and grassland due to the nitrogen fixing potential of the red alder (*Alnus rubra*) that can facilitate the biomass production and consequently increase SOC contents of the soil. Based on the result of this experiment, SOC under the control grassland was by half lower than the SOC under the trees and grassland integrated with trees (sycamore and red alder).Therefore, this study clearly showed that, it is strongly recommended to incorporate silvo-pastoral agroforestry system as the min strategy to increase soil organic carbon in terrestrial ecosystem.

The study also identified that most of SOC at Henfeas agroforestry experimental site was found to the depth of 10cm showing the decreasing trend with the increased depth. Therefore, any soil disturbance to this depth can affect the SOC stored under this land use practice.

In order for countries to control against increasing atmospheric CO_2, increasing carbon sequestration below ground in the form of SOC is crucial. However, there were different natural and anthropogenic factors that affect the carbon storage in the soil. This study identified some of these factors such as climate change, soil characteristics and disturbances or land use change. Land use changes are the main causes for the depletion of SOC and increased concentration of GHG in the atmosphere. In order to improve the SOC pool under different land uses, land management practices such as conversion of marginal agricultural land to perennial vegetation, introduction of agroforestry system and applying different soil management practices are important.

Acknowledgements

References

1   Barker J.R., Baumgardner, G.A.,Turner, D.P. and Lee, J.J., (1996). Carbon dynamics of the conservation and wetland reserve programs. Journal of Soil and Water Conservation 51, pp. 340–346

2   Bradley, R. Milne, J. Bell, A. Lilly, C. Jordan and A. Higgins (2005). A soil carbon and land use database for the United Kingdom, Soil Use Manage 21 (2005), pp. 363–369

3   Buck L.E., J.P. Lassoie and E.C.M. Fernandes (1999). Agroforestry in Sustainable Agricultural Systems, CRC Press, Boca Raton, FL, USA pp. 416.

4   Freeman, N. Fenner, N.J. Ostle, H. Kang, D.J. Dowrick and B. Reynolds et al., Export of dissolved organic carbon from peatlands under elevated carbon dioxide levels, Nature 430 (2004), pp. 195–198

5   Haile SG, Nair PKR, and Nair VD (2008). Carbon storage of different soil-size fractions in Florida silvopastoral systems. J Environ 37:1789–1797

6   Intergovernmental Panel on Climate Change (IPCC) Working Group III, (2007)

7   Jackson, N.A., Wallace, J.S., Ong, C.K., (2000). Tree pruning as a means of controlling water use in an agroforestry system in Kenya. For. Ecol. Manage. 126, 133–137.

8   Kirby, K. R., Potvin, C. (2007). Variation in carbon storage among tree species: Implications for the management of a small-scale carbon sink project. Ecological Management. 246, 208–215 geo-

engineering

9 Makumba, W. Akinnifesi, F. K., Janssen, B., Oenema, O. (2007). Long-term impact of a gliricidia-maize intercropping system on carbon sequestration in southern Malawi. Agric. Ecosyst. Environ. 118, 237–243.

10 Nair, P.K.R.,and Kumar, (2009). Agroforestry as a strategy for carbon sequestration. J. Plant Nutri. Soil Sci. 172, 10–23.

11 Oelbermann, M. and Voroney, R. (2007): Carbon and nitrogen in a temperate agroforestry system: Using stable isotopes as a tool to understand soil dynamics. Ecol. Eng. 29, 342–349.

12 Paustian, K., Six, J., Elliott, E.T., Hunt, H.W., (2000) . Management options for reducing CO2 emissions from agricultural soils

13 Richard T. Conant, Keith Paustian, And Edward T. Elliott (2001). Grassland management and conversion into grassland: effects on soil carbon

14 Sinclair F., Eason B. and Hooker J. (2000). Understanding and management of interactions. Agroforestry in the UK. Forestry Commission Bulletin 122.Forestry Commission, Edinburgh, pp. 17–22

15 Sibbald A.R. and Sinclair F.L. (1991). A review of agroforestry research in progress in the UK. Agrofor Abst 3:149

16 Smith P. Greenhouse gas mitigation in agriculture. Philosophical Transactions of the Royal Society of London Series B Biological Sciences (2008) 363, 789

17 Soussana, P. Loiseau, N. Vuichard, E. Ceschia, J. Balesdent and T. Chevallier (2004) Carbon cycling and sequestration opportunities in temperate grasslands, Ecol. Eng. 16, 122–125, UK

18 Teklehaimanot Z.and Martin (1998). Diurnal and seasonal patterns of nitrogenous activity of red alder (Alnus rubra) in comparison with white clover in silvopastoral agroforestry systems. Biol Fert Soils 28: 267–270

19 Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J., Dokken, D.J. , 2000 land-use change and forestry, pp 135-140, Cambridge   University