



International Journal of Advanced Computer Science and Applications

Special Issue



International Conference & Workshop On Advance Computing 2013

ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



www.ijacsa.thesai.org



W H E R E W I S D O M S H A R E S

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS



THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org



Associate Editors

Dr. Zuqing Zhu

Service Provider Technology Group of Cisco Systems, San Jose

Domain of Research: Research and development of wideband access routers for hybrid fibre-coaxial (HFC) cable networks and passive optical networks (PON)

Dr. Ka Lok Man

Department of Computer Science and Software Engineering at the Xi'an Jiaotong-Liverpool University, China

Domain of Research: Design, analysis and tools for integrated circuits and systems; formal methods; process algebras; real-time, hybrid systems and physical cyber systems; communication and wireless sensor networks.

Dr. Sasan Adibi

Technical Staff Member of Advanced Research, Research In Motion (RIM), Canada

Domain of Research: Security of wireless systems, Quality of Service (QoS), Ad-Hoc Networks, e-Health and m-Health (Mobile Health)

Dr. Sikha Bagui

Associate Professor in the Department of Computer Science at the University of West Florida,

Domain of Research: Database and Data Mining.

Dr. T. V. Prasad

Dean, Lingaya's University, India

Domain of Research: Bioinformatics, Natural Language Processing, Image Processing, Expert Systems, Robotics

Dr. Bremananth R

Research Fellow, Nanyang Technological University, Singapore

Domain of Research: Acoustic Holography, Pattern Recognition, Computer Vision, Image Processing, Biometrics, Multimedia and Soft Computing

CONTENTS

Paper 1: Tuning of PID Controllers using Advanced Genetic Algorithm

Authors: Ms. Reshmi P. Pillai, Sharad P. Jadhav, Dr. Mukesh D. Patil

PAGE 1 – 6

Paper 2: Removal of False Negatives in Moving Object Detection Using RGB color space

Authors: Shailaja Surkutlawar, Prof. Ramesh Kulkarni

PAGE 7 – 12

Paper 3: Satellite Image Retrieval using Modified Block Truncation Coding and Kekre Transform Patterns

Authors: A R Sawant, Dr. Vinayak A Bharadi, Dr. H B Kekre

PAGE 13 – 20

Paper 4: Performance Evaluation and Statistical Analysis of MANET routing Protocols for RPGM and MG

Authors: Prajakta M. Dhamanskar, Dr. Nupur Giri

PAGE 21 – 26

Paper 5: SWiFiNet : A Task Distributed System Architecture for WSN

Authors: A.W. Rohankar, Mrinal K. Naskar, Amitava Mukherjee

PAGE 27 – 33

Paper 6: GaAs, InP, InGaAs, GaInP, p+-i-n+ Multiplication measurements for Modeling of Semiconductor as photo detectors

Authors: Sanjay.C.Patil, B.K.Mishra

PAGE 34 – 42

Paper 7: Performance of Turbo Product Code in Wimax

Authors: Trushita Chaware, Nileema Pathak

PAGE 43 – 47

Paper 8: e-Restaurant: Online Restaurant Management System for Android

Authors: Dr. Vinayak Ashok Bharadi, Vivek Ranjan, Nikesh Masiwal, Nikita Verma

PAGE 48 – 54

Tuning of PID Controllers using Advanced Genetic Algorithm

Ms. Reshmi P. Pillai

Department of Electronics Engg.
RamraoAdik Institute of Technology
Nerul, Navi Mumbai, India

Sharad P. Jadhav

Department of Electronics Engg.
RamraoAdik Institute of Technology
Nerul, Navi Mumbai, India

Dr. Mukesh D. Patil

Department of Electronics Engg.
RamraoAdik Institute of Technology
Nerul, Navi Mumbai, India

Abstract—Proportional-Integral-Derivative (PID) controllers have been widely used in process industry for decades from small industry to high technology industry. But they still remain poorly tuned by use of conventional tuning methods. Conventional technique like Zeigler-Nichols method does not give an optimized value for PID controller parameters. In this paper, we optimize the PID controller parameter using Genetic Algorithm (GA), which is a stochastic global search method that replicates the process of evolution.

Using genetic algorithm the tuning of the controller will result in the optimum controller being evaluated for the system every time. The GA is basically based on an iterative process of selection, recombination, mutation and evaluation. The performance of Advanced Genetic Algorithm (AGA) is compared with Guo Tao's Algorithm (GTA) and Elite Multi-Parent Crossover Evolutionary Optimization Algorithm (EMPCOA). AGA has a different replacement strategy as compared to EMPCOA which helps to maintain the population diversity and thus reducing the computational time which is proved by the results presented here.

The effectiveness of the AGA is also verified for a system with an unstable plant. The PID controller is also tuned with different error criteria viz. Integral Time Absolute Error (ITAE), Integral Square Error (ISE) and Integral Absolute Error (IAE).

Keywords-PID tuning; Genetic Algorithm; Multi-parent crossover; Elite crossover; Discrete recombination.

I. INTRODUCTION

The PID controller was patented in 1939 by Albert Callender and Allan Stevenson of Imperial Chemical Limited of Northwich, England. The PID controller is widely used in most industrial processes despite continuous advances in control theory. The main reason is due to their simplicity of operation, ease of design, inexpensive maintenance, low cost, and effectiveness for most linear systems. Recently, motivated by the rapidly developed advanced microelectronics and digital processors, conventional PID controllers have gone through a technological evolution, from pneumatic controllers via analog electronics to microprocessors via digital circuits [1]. Most conventional PID tuning methods require considerable technical experience to apply tuning formulas to determine the PID controller parameters. The conventional tuning methods require the process model to be reduced if it is too complicated originally [2]. In practical applications, most of the industrial processes exist to be non-linear, variability of parameters and uncertainty of models are very high, thus using

conventional PID tuning methods the precise control of the process cannot be achieved. Due to this, PID controllers are rarely tuned optimally and thus require improved tuning technology. The above problems can be well addressed by the application of non-conventional methods for tuning of the PID controller. Most practical PID remains poorly tuned leading to deteriorated process performance [3]. The conventional tuning methods require considerable technical experience and are time consuming and do not work well for non-linear, higher order and time-delayed systems and the ones that do not have a precise mathematical model [1]. Non-conventional methods are especially useful for solving problems of computationally complicated and mathematically untraceable. Hence the need arises for an optimization algorithm like Genetic Algorithm (GA).

Genetic Algorithm is a stochastic search and optimization method that mimics the process of natural evolution [4]. John H. Holland formally introduced GA in his book, 'Adaptation in Natural and Artificial Systems' in the 1975 at the University of Michigan, Ann Arbor, United States. GA is one of the Evolutionary Algorithms (EA) methodologies. The key aspect distinguishing an evolutionary search algorithm from traditional algorithms is that it is population-based. Through the adaptation of successive generations of a large number of individuals, EA performs an efficient directed search. Evolutionary search is generally better than random search as EA inspired by the evolution process in nature and try to solve problems by evolving sets of search points. GA imitates natural evolution with survival of the fittest approach. It performs on coding of parameters hence does not depends on the continuity of parameter nor the existence of the derivatives of the functions, thus allowing it to handle multi parameters or multi-model type of optimization problems. GA can also work for non-deterministic systems or the systems that can be only partially modeled. GA uses random choice and probabilistic decision to guide the search, where the population improves toward near optimal points from generation to generation [5]. The main advantage of the GA formulation is that fairly accurate results may be obtained using a very simple algorithm. The GA is basically based on an iterative process of selection, recombination, mutation and evaluation. GA has parallel search techniques, which emulate natural genetic operations. Due to its high potential for optimization, GA has received great attention in control systems such as the search of optimal PID controller parameters.

The paper is organized as follows: Section II gives basic idea of a PID controller. Section III introduces Genetic Algorithm. The algorithm of implementation of Advanced GA is given in Section. The design steps are discussed in Section V. The results are presented in Section VI. Final conclusions are drawn in Section VII.

II. PID CONTROLLER

A PID controller aims at minimizing the error between a measured process variable of the controlled system and a reference, by calculating the error and generating a correction signal to the system from the error. The block diagram of a conventional PID controller is shown in the Fig. 1, where $r(t)$ is the reference value, $y(t)$ is the output of the controlled system, $e(t)$ is the error between $r(t)$ and $y(t)$, whereas $u(t)$ is the output control signal of the PID controller. A conventional PID controller consists of three components: the proportional part, the integral part and the derivative part as shown in Fig. 1. The proportional term produces an output value that is proportional to the current error value. The contribution from the integral term is proportional to both the magnitude of the error and the duration of the error. Derivative control is used to reduce the magnitude of the overshoot produced by the integral component and improve the combined controller-process stability. The output control signal of a PID controller is described as follows,

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de(t)}{dt}, \quad (1)$$

where, $u(t)$ is the output control signal, $e(t)$ is an error signal, and K_p , K_i , and K_d refers to the proportional gain, the integral gain and the derivative gain, respectively.

K_p , K_d and K_i should satisfy the following equations,

$$K_i = K_p T_i, \quad (2)$$

$$K_d = K_p T_d, \quad (3)$$

where T_i and T_d refers to the integration time and derivative time, respectively.

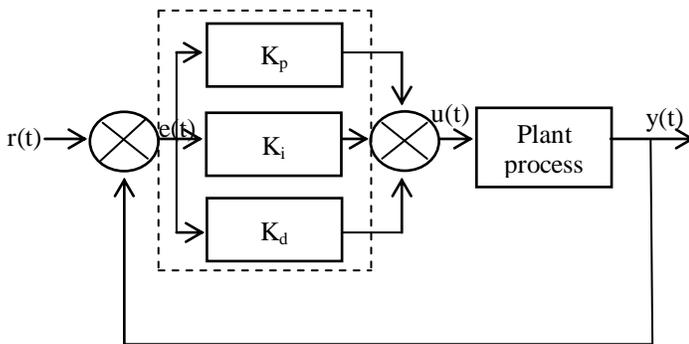


Fig. 1. PID Controller Block Diagram.

The individual effects of these terms on the closed-loop performance are summarized in Table I. Note that this table serves as a first guide for stable open-loop plants only. For optimum performance K_p , K_i and K_d are mutually dependent in tuning.

TABLE I. Effects Of Independent P, I And D Tuning [6]

Closed loop response	Rise Time	Over-shoot	Settling Time	Steady State Error	Stability
Increase K_p	Decrease	Increase	Small Increase	Decrease	Degrade
Increase K_i	Small Decrease	Increase	Increase	Large Decrease	Degrade
Increase K_d	Small Decrease	Decrease	Decrease	Minor change	Improve

The quality of PID tuning rules is of considerable practical importance because a small percentage improvement in the operation of a plant can translate into large economic savings or other benefits.

III. GENETIC ALGORITHM

GA is a probabilistic optimization algorithm with a high probability of finding a good solution in a given search space. Genetic Algorithm can handle multiple variables and only requires the ability to develop a mathematical model to configure a set of inputs (the variables) in order for the model to produce an optimal output. After initialization of population, each string (individual) in the population is evaluated to determine the performance of the string. Then, the higher-ranking strings are mate. The process of crossover is performed by combining strings containing partial solutions. The algorithm favors fittest strings as parents, thus better strings will have more number of offspring. The GA exploits the regions of the solution space, because successive generations of reproduction and crossover produce increasing numbers of strings in those regions. In this paper the offspring replaces the weakest string, thus maintaining the population size same [7]. Lastly, mutations modify a small fraction of the strings. Mutation alone does not generally advance the search for a solution, but it does provide insurance against the development of a uniform population incapable of further evolution [8].

Guo Tao's Algorithm (GTA) is a linear non-convex multi-parent crossover operator (GTX) which is used in optimization of nonlinear continuous functions [9]. The multi-parent crossover utilizes more number of candidate solutions and the replacement strategy implemented is supposed to be minimizing selection pressure. But major limitation of GTA is that it may ignore better solutions in population. To make use of better solutions in population elite preservation strategy is introduced by Xiaoyi Che, Youxin Luo and Zhaoguo Chen implemented in the elite multi-parent crossover evolutionary optimization algorithm (EMPCOA) [10]. The selection scheme and replacement strategy implemented in EMPCOA gives global optima with increase in an execution time. This is mainly due to the decrease in population diversity and therefore requires more number of iterations to converge.

Aimed at these shortages of GTA and EMPCOA we are motivated to implement the Multi-parent Crossover Algorithm with Discrete Recombination [7] with better parts of both algorithms like, fixed population size of GTA and elite preservation strategy of EMPCOA with multi-parent crossover. Here, we aim to reduce the number of iterations and

execution time with improvement in transient (performance) response.

General steps involved in GA are.

- Representation
- Objective function
- Population initialization
- Parent selection mechanism
- Variation operator, crossover (recombination)
- Variation operator, mutation and
- Termination condition.

A GA is typically initialized with a randomly generated population consisting of candidate individuals. Each individual in the population is usually represented by a real-valued number or a binary string. Such strings are called as chromosomes. A set of chromosome or individual is a whole population. Performance of each individual is measured and assessed by the objective function. The objective function assigns each individual a corresponding number called its fitness value. If the termination criteria are not met with the current population then new individuals are created with genetic operators. A survival of the fittest strategy is applied on individuals. The fittest parents are found out by reproduction or selection operator. New individuals are generated by performing operations such as crossover and mutation on the individuals whose fitness has just been measured. The fitness of the offspring is then computed. The offspring is inserted into the population replacing the parents or low-fitness individuals producing a new generation. This cycle is performed until the termination criterion is reached. Such a single population GA is powerful and performs well on a wide variety of problem. Every iteration of GA loop is referred to as a generation. When termination criteria gets satisfied GA stops. The work flow of GA is as shown in Fig 2.

IV. PID TUNING USING GA METHOD

The main features of the algorithm implemented are,

- It uses elite preservation strategy.
- It makes use of multi-parent crossover to create new offspring.
- The performance of the proposed algorithm is better than the existing algorithms like GTA and EMPCO in terms of number of iterations and computational time.
- It gives better transient response as compared to the existing algorithms like GTA and EMPCOA.

The design steps of the algorithm are

- Step 1** Produce initial population $P_0 = X_1; X_2...X_N$ randomly at searching space S , N is number of individuals in populations and $t = 0$.
- Step 2** Arrange the individuals in population P from good to bad according to the fitness of parent. Then still record as $P_t = X_1, X_2...X_N$ after the arrangement, X_1 is the best individual X_N is the worst one.

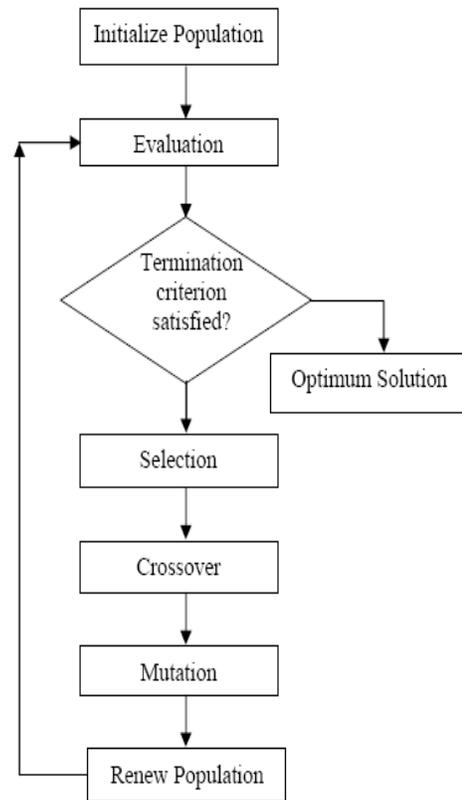


Fig. 2. Workflow of Genetic Algorithm [11].

Step 3 Termination criteria is when the fitness difference between X_{worst} and X_{best} become less than or equal to fitness limit (ϵ) (i.e. the fitness of the worst individual is almost same as the best one), then go to step 7.

Step 4 Choose K ($K \leq m$) best individuals X_1, X_2, \dots, X_K from population P_t , and then, choose $(m-K)$ individuals $X_{k+1}, X_{k+2}, \dots, X_m$ from the rest $(N-K)$ individuals randomly. A subspace V is formed from these m ($m \leq N$) individuals. Then perform multi-parent crossover as given in (4).

$$V = x_c \in S, \quad x_c = \sum_{i=1}^m a_i x_i, \quad (4)$$

$$\sum_{i=1}^m a_i = 1, \quad -0.5 \leq a_i \leq 1.5. \quad (5)$$

Step 5 Compare x_c with X_{worst} , x_c replaces X_{worst} if better (x_c, X_{worst}) condition is true else discard x_c .

Step 6 Go to Step 2

Step 7 Output is the best solution and end.

Here, we implement the algorithm for optimizing PID parameters. The following section describes how gains K_p , K_i and K_d are represented in form of chromosome or individual. The implementation of PID parameters optimization

procedure using GA starts with the chromosome representation as shown in Fig. 3.

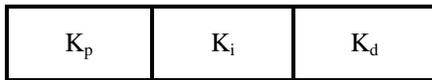


Fig. 3. Chromosome representation

The chromosome is formed by three values that correspond to the three gains to be adjusted in order to achieve a satisfactory behavior. The gains K_p , K_i and K_d are strings of chromosome as shown in Fig. 3. A set of chromosomes or individual forms a generation.

An objective function could be created to find a PID controller that gives the smallest overshoot, fastest rise time or quickest settling time. There are several variables used as the standard to measure system performance. In general, unit step input is used to test the systems, and the output signals is characterized by some standard performance measures like settling time, percentage overshoot, rise time and peak time. All these measures are defined in the time domain response. Each chromosome in the population is passed into the objective function one at a time. The chromosome is then evaluated and assigned a number to represent its fitness, which is its fitness value. The GA uses the chromosomes fitness value to renew population consisting of the fittest members. When the chromosome enters the evaluation function, it is split up into its three terms. The newly formed PID controller is placed in a unity feedback loop with the system transfer function. This will result in reducing the compilation time of the program. The system transfer function is defined in another file and imported as a global variable. The controlled system is then given a step input and the error can be assessed using error performance criterion integral absolute error (IAE).

V. DESIGN PROCEDURE

- The initial population of 10 chromosomes is generated between 0 to 20 using random number generation.
- The objective function is used as error performance criterion. The error criteria used is Integral of absolute error (IAE). The magnitude of this error will be used to assess the fitness of each chromosome.
- The termination criterion for the algorithm is based on fitness limit. The algorithm terminates when the difference between the fitness of best solution X_{best} and worst solution X_{worst} is less than or equal to fitness limit $\epsilon = 1$.
- A subspace is formed using elite preservation strategy and then multi-parent crossover takes place using 5 parents. The 2 best chromosomes are selected and the 3 random parents are selected from the remaining 8 chromosomes.
- In the proposed algorithm the offspring solution (x_c) is compared and replaced with X_{worst} .

VI. RESULT

A. Example 1

We implement all three algorithms on a system with transfer function given as follows.

$$G(s) = \frac{1}{0.004s^3 + 0.08s^2 + 0.5s + 1} \quad (6)$$

The controller parameter and the computational time for all three algorithms are given in Table II. It can be seen that AGA requires least amount of time to obtain the optimum controller gains.

TABLE II. Comparison Of Controller Gains & Computational Time For Example 1

	GTA	EMPCOA	AGA
K_D	16.1815	2.5092	0.948
K_P	14.9747	15.0449	6.847
K_I	17.3709	16.5410	14.72
No. of iterations	15	11	8
Computational time (s)	3.85	6.83	2.06

The closed loop step response of the system using different algorithms is as shown in Fig. 4. It is observed that the transient response of the system using AGA is best because of fastest settling time and least overshoot. The same is summarized in Table III.

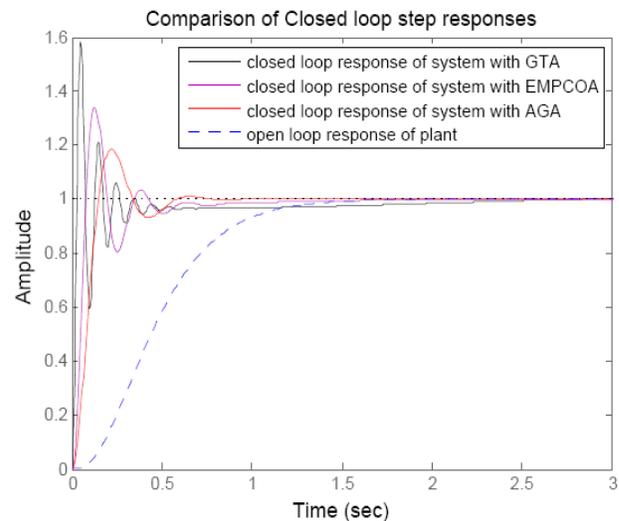


Fig. 4. Comparison of closed loop step responses for GTA, EMPCOA and AGA for the illustrative example 1.

TABLE III. Comparison Of Step Response For Example 1

	GTA	EMPCOA	AGA	Open loop plant
Settling time (s)	1.78	0.812	0.538	1.29
Rise time (s)	0.0186	0.0534	0.101	0.723
Overshoot (%)	59	34	18	0

B. Example 2

We implement all three algorithms on a system with unstable plant whose transfer function is given as follows.

$$G(s) = \frac{6}{s^2 - 0.5s + 6} \quad (7)$$

The controller parameter and the computational time for all three algorithms for example 2 are given in Table IV. It can be observed the using AGA the optimum controller gains are obtained in minimum number of iterations although the amount of time required is slightly greater than that of EMPCOA.

The closed loop step response of the system using different algorithms is as shown in Fig. 5. It is observed from the transient response of the system that AGA has minimum overshoot although the settling time required is more than that required by EMPCOA. The same is summarized in Table V.

TABLE IV. Comparison Of Controller Gains & Computational Time For Example 2.

	GTA	EMPCOA	AGA
K_d	8.3112	7.8441	17.8181
K_p	5.3240	16.1898	12.1858
K_i	11.2521	12.1148	15.9448
No. of iterations	6	11	4
Computational time (s)	2.46	6.49	3.3

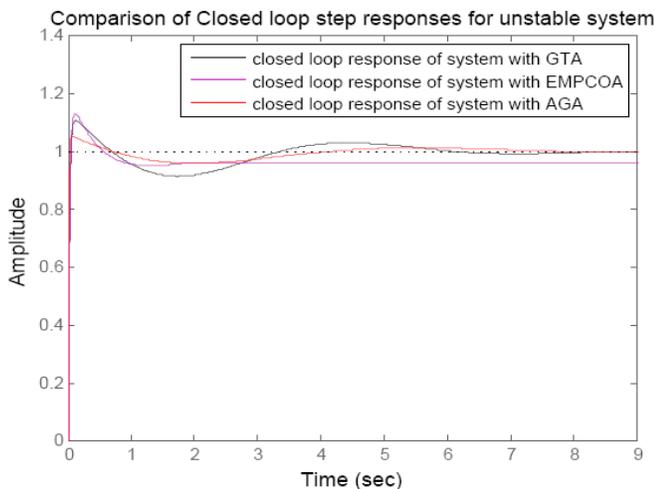


Fig. 5. Comparison of closed loop step responses for GTA, EMPCOA and AGA for the illustrative example 2.

TABLE V. Comparison Of Step Response For Example 2

	GTA	EMPCOA	AGA
Settling time (s)	5.15	2.27	3.31
Rise time (s)	0.0366	0.0366	0.0366
Overshoot (%)	10.6	12.9	5.05

C. Comparison with different error criteria

The parameters of PID controller are obtained with different error criteria for Example 1. The open loop transfer function is given by the (6). The performances of the designed controllers are compared. The controller gains (K_p , K_d and K_i) are obtained using Advanced Genetic Algorithm. The AGA is implemented using different error criteria, viz. ITAE, ISE and IAE as objective function for the system.

The controller gains so obtained with different error criteria are as shown in Table VI. AGA using ISE requires minimum number of iterations to compute the parameter value, but IAE gives best performance with minimum computational time while ITAE requires highest computational time.

A comparison of closed loop step response of the system using AGA with all three error criteria as objective function is as given in Fig.6. The same is summarized in Table VII, which shows that the system with AGA using IAE gives the best performance with fastest settling time and minimum overshoot

TABLE VI. Comparison Of Controller Gains & Computational Time For Example 1 Using Different Error Criteria.

	ITAE	ISE	IAE
K_D	6.3486	9.2521	0.948
K_P	16.2908	7.3565	6.847
K_I	15.7815	13.5914	14.72
No. of iterations	42	3	8
Computational time (s)	100.42	3.55	2.06

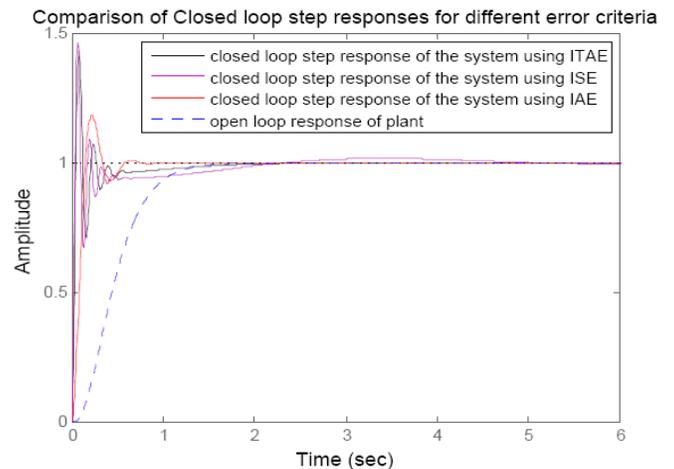


Fig. 6. Comparison of closed loop step responses for AGA using different error criteria for the illustrative example 1.

TABLE VII. Comparison Of Step Response For Example 1 Using Different Error Criteria.

	ITAE	ISE	IAE	Open loop plant
Settling time (s)	1.18	1.8	0.538	1.29

Rise time (s)	0.077	0.0258	0.101	0.723
Overshoot (%)	43.1	46.4	18	0

VII. CONCLUSION

We have implemented Advanced Genetic Algorithm (AGA) along with the existing genetic algorithms - Guo Tao's Algorithm (GTA) and Elite Multi-Parent Crossover Optimization Algorithm (EMPCOA). From the results presented it is observed that AGA requires lesser computational time as compared to GTA and EMPCOA. This is mainly because of the different replacement strategy in AGA which improves the population diversity and hence lesser number of iterations is required. Also from the graph it can be seen that the transient response of the system using AGA is better than that of GTA and EMPCOA. The closed loop step response of the system has a better settling time and maximum overshoot.

To prove the effectiveness of the algorithm, it is tested on an unstable system. From the results presented it is observed that it requires lesser number of iterations and a better transient response due to fast settling time and lesser overshoot.

Further, the PID controller is tuned with different error criteria. The performances of the controllers tuned using different error criteria are compared.

REFERENCES

[1] K. S. Tang, Kim Fung Man, Guanrong Chen and Sam Kwong, "An optimal fuzzy PID controller", IEEE Transactions On Industrial Electronics, vol 48, no 4, pp. 757-765, 2001

[2] W.K. Ho, O.P.Gan, E.B. Tay and E.L. Ang, "Performance and gain and phase margins of well-known PID tuning formulas", IEEE Transactions on Control Systems Technology, vol. 4, no. 4, pp. 473-477, 1996.

[3] Aidan O'Dwyer, "PI and PID controller tuning rules: an overview and personal perspective," Proceedings of the IET Irish Signals and Systems Conference, 2006c

[4] J.H. Holland, "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence", Ann Arbor, MI: University of Michigan Press 1975, Second edition, Cambridge, MA: The MIT Press, 1992.

[5] Teo Lian Seng, Marzuki Bin Khalid and Rubiyah Yusof, "Tuning of a Neuro-Fuzzy," IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, 1999

[6] Kiam Heong Ang, Gregory Chong and Yun Li, "PID control system analysis, design and technology", IEEE Transactions on Control Systems Technology, vol 13, no 4, pp. 555-576, 2005.

[7] Dazhi Jiang, Yulin Du, Jiali Lin, Zhijian Wu, "Multi-parent crossover algorithm with discrete recombination," IEEE, 2010

[8] R. Michalski, J. Carbonell and T. Mitchell, "machine learning: an artificial intelligence approach", Los Altos, Morgan Kaufmann, 1986.

[9] Guo Tao and Kang Li-Shan, "A new evolutionary algorithm for function optimization", Wuhan University Journal of Natural Sciences, vol.4, no.4, pp.409-414, 1999.

[10] Xiaoyi Che, Youxin Luo and Zhaoguo Chen, "Optimization for PID control parameters on hydraulic servo control system based on the elite multi-parent crossover evolutionary algorithm", International Conference on Measuring Technology and Mechatronics Automation, IEEE Computer Society, pp. 845-848, 2010.

[11] B. Nagaraj, P. Vijayakumar, "A comparative study of PID controller tuning using GA, EP, PSO and ACO," Journal of Automation, Mobile Robotics and Intelligent systems, vol 5, no.2, February 2011

Removal of False Negatives in Moving Object Detection Using RGB color space

Shailaja Surkutlawar
EXTC

Vivekananda education society of information technology
Mumbai, India

Prof. Ramesh Kulkarni
EXTC

Vivekananda education society of information technology
Mumbai, India

Abstract— Moving objects detection is a fundamental step in many vision based applications. Background subtraction is the typical method. Many background models have been introduced to deal with different problems. The method based on mixture of Gaussians is a good balance between accuracy and complexity, and is used frequently by many researchers. But it still cannot provide satisfied results in some cases. Video-surveillance and traffic analysis systems can be heavily improved using vision-based techniques to extract, manage and track objects in the scene. However, problems arise due to shadows. In particular, moving shadows can affect the correct localization, measurements and detection of moving objects. This work aims to present a technique for shadow detection and suppression used in a system using RGB color space for moving visual object detection and tracking. Experimental results show that the proposed approach can significantly enhance shadow suppression results and moving objects are detected.

Keywords—Moving objection detection; background subtraction; mixture of Gaussians MoG; Shadow detection; RGB color space.

I. INTRODUCTION

Moving Objects segmentation is a fundamental and critical task in many vision based applications, such as automated visual surveillance, human-machine interface, and very low-bandwidth telecommunications. A common approach is to perform background subtraction, which identifies moving objects from the difference between the current frame and a reference frame (which often called “background model”). The background model must be representation of the scene with no moving objects and must be kept regularly updated because for some cases, the background is changing with time.

To detect a moving object, a surveillance system usually utilizes background subtraction. The key of background subtraction is the background model. In the moving object detection process, one of the main challenges is to differentiate moving objects from their cast shadows. Moving cast shadows are usually misclassified as part of the moving object making the following stages, such as object classification or tracking, to perform inaccurate. The Gaussian mixture model (GMM) [1] represented the statistics of one pixel over time can cope

with multi-modal background distributions. However, a common problem for this approach is to find the right balance between the speed at which the model adapts to changing background, and the stability. The shadow points and the object points share two important visual features: motion model and detectability. Since the most common techniques for foreground object detection in dynamic scene are inter-frame difference or background suppression, all the moving points of both objects and shadows are detected at the same time.

Moreover, shadow points are usually adjacent to object points and with the more commonly used segmentation techniques shadows and objects are merged in a single blob. These aspects cause two important drawbacks called as false negatives: The former is that the object shape is falsified by shadows and all the measured geometrical properties are affected by an error (that varies during the day and when the luminance changes). This affects both the classification and the assessment of moving object position (normally given by the shape centroid), as, for instance, in traffic control systems that must evaluate the trajectories of vehicles and people on a road. The second problem is that the shadows of two or more objects can create a false adjacency between different objects. In this paper an approach of shadow detection and suppression based on RGB color space is defined.

The paper is organized as follows: Section II deals with the proposed method in which background subtraction using Gaussian Mixture Model to classify the pixels as background or foreground by thresholding the difference between the background image and the current image is explained, Section III deals with Post processing techniques for suppressing shadow using RGB color space, Section IV discusses the experimental results of shadow suppression techniques. Finally, the conclusion is given in Section V.

II. PROPOSED METHOD

In the proposed method shown in Fig. 1 the first step is to extract the foreground moving object from the background using mixture of Gaussian is used and then shadow suppression method using RGB color space is described on which area confidence is applied to extract the blobs to be detected.

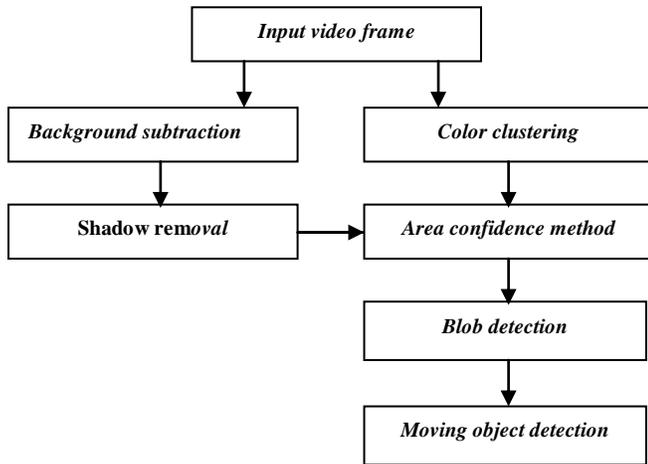


Fig.1. Block diagram of the proposed method

A. Background subtraction

In the model of Mixture of Gauss [1] [4] [5], the background is not a single frame without any moving objects. Gaussian Mixture Model (GMM) is thought to be one of the best background modeling methods and works well when gradual changes appear in the scene [2] [3]. The GMM method models the intensity of each pixel with a mixture of k Gaussian distributions. The probability that a certain pixel has a value X_t at time can be written as

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where k is the number of distributions (currently, 3 to 5 is used), $\omega_{i,t}$ is the weight of the k^{th} Gaussian in the mixture at time t and $\eta(X_t, \mu_{i,t}, \Sigma_{i,t})$ the Gaussian probability density function.

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^3 / 2^3 |\Sigma_{i,t}|^{3/2}} e^{\left\{ \frac{-1(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})}{2} \right\}} \quad (2)$$

Where, $\mu_{i,t}$ is the mean value and $\Sigma_{i,t}$ is the covariance of the k^{th} Gaussian at time t . For computational reasons, the covariance matrix is assumed to be of the form

$$\Sigma_{k,t} = \sigma^2 \cdot I \quad (3)$$

Where σ is the standard deviation.

This assumes that the red, green, and blue pixel values are independent and have the same variance, allowing us to avoid a costly matrix inversion at the expense of some accuracy.

Thus, the distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. A new pixel value will, in general, be represented by one of the major components of the mixture model and used to update the model.

However, it fails when there are sharp changes, such as sudden illumination changes or sudden partial changes in the background. To tackle this problem, some improvement has

been made in recent researches. In [6], every frame is processed on pixel level, region level and frame level with color and gradient information to overcome the problem caused by sudden illumination changes based on GMM. In [7], a hierarchical GMM using state models without temporal correlation on different scales is proposed to handle sharp changes. Zivkovic presented an improved GMM algorithm automatically fully adapting to the scene, by choosing the number of components for each pixel in an online procedure [8] [9], which leads to big improvement in reduced processing time and slight improvement in segmentation result.

If the pixel process could be considered a stationary process, a standard method for maximizing the likelihood of the observed data is expectation maximization. Unfortunately, each pixel process varies over time as the state of the world changes, therefore an approximate method which essentially treats each new observation as a sample set of size 1 and uses standard learning rules to integrate the new data.

If lighting changes occurred in a static scene, it would be necessary for the Gaussian to track those changes. If a static object was added to the scene and was not incorporated in to the background until it had been there longer than the previous object, the corresponding pixels could be considered foreground for arbitrarily long periods. This would lead to accumulated errors in the foreground estimation, resulting in poor tracking behavior. These factors suggest that more recent observations may be more important in determining the Gaussian parameter estimates. Since there is a mixture model for every pixel in the image, implementing an exact Expectation maximization algorithm on a window of recent data would be costly. Instead, we implement an on-line K-means approximation. Every new pixel value X_t is checked against the existing k Gaussian distributions, until a match is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution. GMM algorithm can be summarized as:

- Initialize each pixel of the scene with k Gaussian distributions
- Every new pixel value X_t , is checked against the existing Gaussian distributions until a match is found.
- A match is defined as a pixel value within 2.5 standard deviations of a distribution.
- If none of k -distributions match current pixel value, least probable distribution is go out.
- A new distribution with current value as mean value, an initially high variance, and low prior weight, is entered.
- Prior weights of k distributions at time adjusted as

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \quad (4)$$

- μ_t and σ for unmatched distributions remain the same.
- Parameters of distribution matching new observation are updated as :

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho(X_t) \quad (5)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \quad (6)$$

$$\rho = \alpha\eta(X_t | \mu_{t-1}, \sigma_{t-1}) \quad (7)$$

- Gaussians are ordered by the value of ω/σ
- 1st B distributions are chosen as background model, where

$$B = \text{arg}g_b^{\min}(\sum_{k=1}^b \omega_k > T) \quad (8)$$

III. MOVING SHADOW SUPPRESSION USING RGB COLOR SPACE

Shadows are due to the occlusion of light source by an object in the scene. In particular, that part of the object not illuminated is called self-shadow, while the area projected on the scene by the object is called cast shadow [10]. This last one is more properly called moving cast shadow if the object is moving.

A. Normalized RGB color space

The Normalized RGB space aims to separate the chromatic components from the brightness component. The red, green and blue channel can be transformed to their normalized counterpart by using the formulae

$$l = R + G + B, r = R/l, g = G/l, b = B/l \quad (9)$$

When $l \neq 0$ and $r = g = b = 0$ otherwise.

One of these normalized channels is redundant, since by definition r , g , and b sum up to 1. Therefore, the Normalized RGB space is sufficiently represented by two chromatic components r and g and a brightness component l . Normalized RGB suffers from a problem inherent to the normalization namely noise sensor or compression noise at low intensities results in unstable chromatic components.

Under the consideration of saving computational cost, RGB space based method proposed by Horprasert in [4] is adopted. The basic idea in [4] is that shadow has similar chromaticity but lower brightness. For a given observed pixel value I_i , a brightness distortion, α_i , and a color distortion CD_i , is calculated by,

$$\alpha_i = \text{arg}g_{\alpha_i} \min(I_i - \alpha_i E_i)^2 \quad (10)$$

$$CD_i = \|I_i - \alpha_i E_i\| \quad (11)$$

adopted. The basic idea in [4] is that shadow has similar chromaticity but lower brightness. For a given observed pixel

Where E is the expected chromaticity line, α_i equals 1 if the brightness of the given pixel in the current frame is the same as in the background image. α_i , is less than 1 if it is darker and greater than 1 if it becomes brighter than the expected brightness. Then, the criteria for shadow pixels simply becomes,

$$\begin{cases} \tau_a < \alpha_i < 1 \\ CD_i < \tau_{CD} \end{cases} \quad (12)$$

In [11], τ_a and τ_{CD} are predefined thresholds $\tau_a = 0.7$ and $\tau_{CD} = 5$, in our experiments.

IV. EXPERIMENTAL RESULTS

Original frames for the experiments shown in Fig. a, Fig. b, Fig. c, Fig. d. On this original frames Background subtraction using GMM is applied, as a result background pixels, foreground pixels and some shadow pixels (falsely segmented as foreground pixels) shown as black and white respectively in Fig. e, Fig. f, Fig. g, Fig. h. Post processing techniques for shadow suppression using RGB color space applied on Fig. e, Fig. f, Fig. g, Fig. h and results shown in Fig. i, Fig. j, Fig. k, Fig. l. Results show that shadows and noise pixels eliminated in most of the images but suppression and blob size can be enhanced using HSV color space.



a) Original Frame 1



b) Original Frame 2



c) Original Frame 3



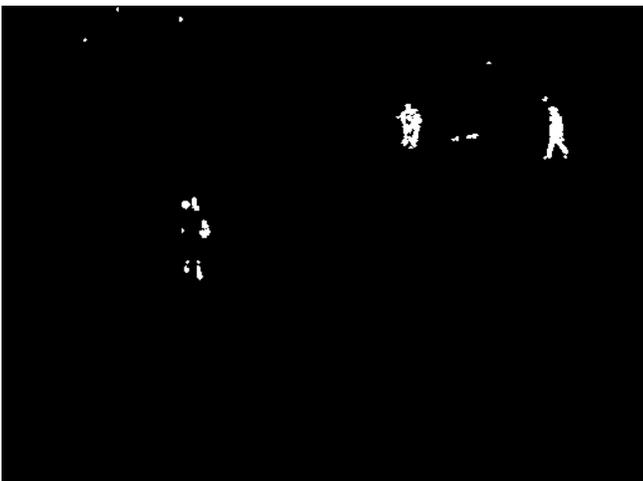
d) Original Frame 4



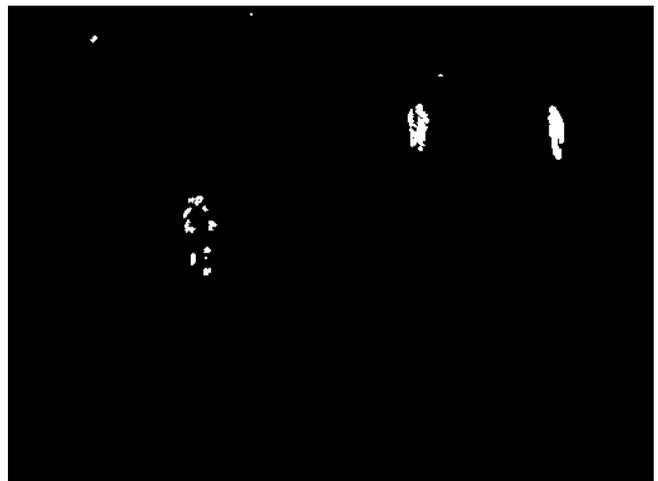
e) Background Subtraction of Frame 1



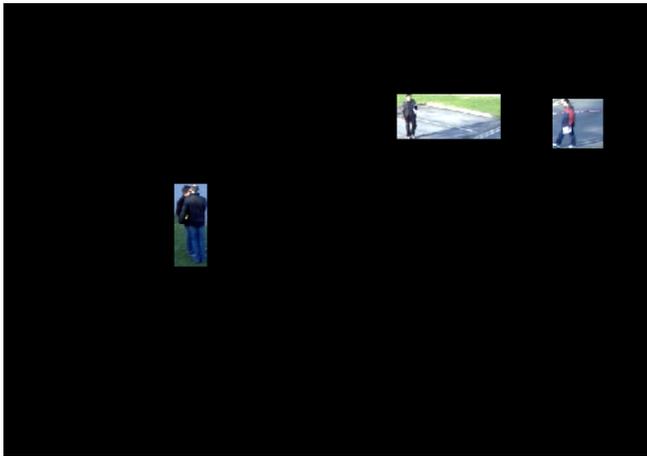
f) Background Subtraction of Frame 2



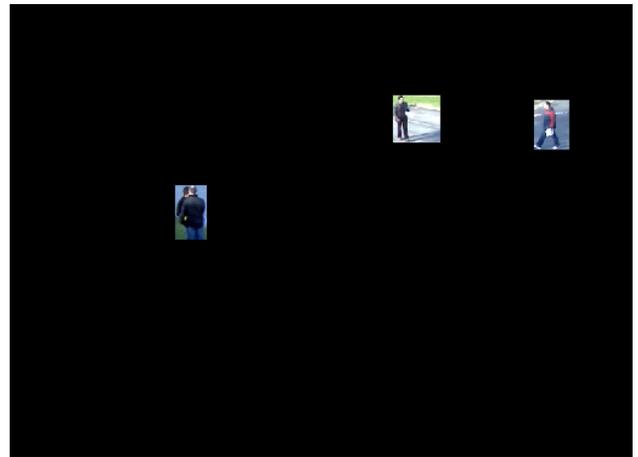
g) Background Subtraction of Frame 3



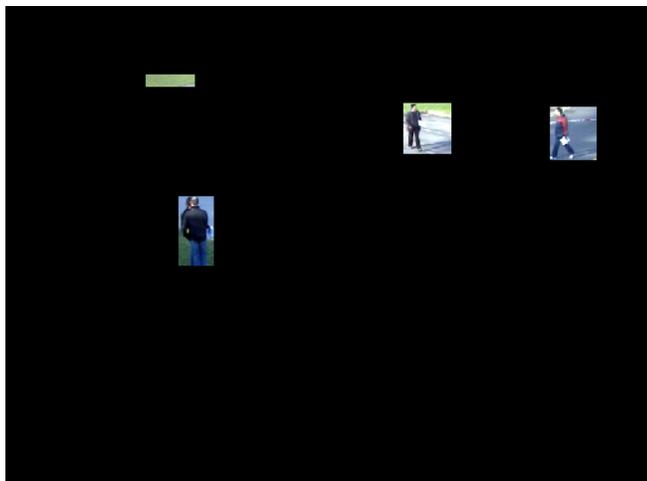
h) Background Subtraction of Frame 4



i) RGB Result Of Frame 1



j) RGB Result Of Frame 2



k) RGB Result Of Frame 3



l) RGB Result Of Frame 4

VI. CONCLUSION

Moving objects detection and segmentation is a fundamental step in many applications based on vision. Mixture of Gaussians is the frequently used method to subtracting moving objects from background. But its results are not good enough in some cases. In this paper, a post-processing method is proposed to solve this problem. The results with more complete boundaries provided by the color clustering is used to verify the outputs of mixture of Gaussians, and thus two possible false segmentations can be corrected effectively. Moving shadow suppression using RGB and HSV colour spaces and small region median filter are also adopted. This paper compare shadow suppression results using RGB and HSV colour space and found that results of HSV are good over RGB colour space.

REFERENCES

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Ft. Collins, 1999: pp.246-252.
- [2] C. Stauffer, and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. on PAMI, Vol. 22, No. 8, 2000 pp. 747-757.
- [3] P. Kumar, and A. Mittal, "Study of robust and intelligent surveillance in visible and multimodal framework," informatica, Vol. 31, 2007 pp. 447-461.
- [4] Horprasert T, Harwood D, Davis L S. "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," in Proceedings of IEEE ICCV' 99 Frame-Rate Workshop, 1999, pp.1-19.
- [5] KaewTraKulPong P., Bowden R, "An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection," in proceedings of 2nd European Workshop on Advanced Video Based Surveillance Systems, Sept 2001, Pages:1-5.

- [6] O.Javed, K.Shafique, and M .Shah., “A hierarchical approach to robust background subtraction using color and gradient information,” In Workshop on Motion and Video Computing, Dec. 2002, pp. 22–27.
- [7] Y.D.Sun, and B. Z. Yuan, “Hierarchical GMM to handle sharp changes in moving object detection,” in Electronics Letters, Vol. 40, No. 13, 2004, pp. 801-802.
- [8] Z.Zivkovic, “Improved adaptive gaussian mixture model for background subtraction,” in proceedings ICPR, 2004.
- [9] Z. Zivkovic and F.vander Heijden, “ Efficient adaptive density estimation per image pixel for the task of background subtraction,” Pattern Recognition Letters, Vol. 27,No. 7, 2006, pp. 773-780.
- [10] C. Jiang and M.O. Ward, “Shadow identification,” in Proceedings of IEEE Int’l Conference on Computer Vision and Pattern Recognition, , 1992 ,pp. 606–612.
- [11] I. Haritaoglu, D. Harwood, and L.S. Davis, “W4: Real-time surveillance of people and their activities,”in proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. 2000, pp. 809–830.
- [12] J. Stauder, R. Mech, and J. Ostermann, “Detection of moving cast shadows for object segmentation,”in proceedings of IEEE Transactions on Multimedia, vol. 1, no. 1, Mar. 1999, pp. 65–76.
- [13] M. Kilger, “A shadow handler in a video-based real-time traffic monitoring system,” in Proceedings of IEEE Workshop on Applications of Computer Vision, 1992, pp. 11–18.
- [14] A. Elgammal, D. Harwood, and L.S. Davis, “Non-parametric model for background subtraction,” in Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.
- [15] T. Horprasert, D. Harwood, and L.S. Davis, “A Statistical approach for real-time robust background subtraction and shadow detection,” in Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.
- [16] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Statistical and knowledge-based moving object detection in traffic scene,” in Proceedings of IEEE Int’l Conference on Intelligent Transportation Systems, Oct. 2000, pp. 27–32.
- [17] I. Mikic, P. Cosman, G. Kogut, and M.M. Trivedi, “Moving shadow and object detection in traffic scenes,” in Proceedings of Int’l Conference on Pattern Recognition, Sept. 2000.
- [18] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, “A color segmentation scheme for object-based video coding,” in Proceedings of the IEEE Symposium on Advances in Digital Filtering and Signal Processing, 1998, pp. 25–29.

Satellite Image Retrieval using Modified Block Truncation Coding and Kekre Transform Patterns

A R Sawant
M. E. EXTC
TCET, Kandivali
Mumbai-400 101

Dr. Vinayak A Bharadi,
Bijith Markarkandy
IT Department, TCET, Kandivali,
Mumbai-400 101

Dr. H B Kekre
Computer Department
MPSTME, NMIMS
University, Mumbai-28

Abstract--Satellite images and Content Based Image Retrieval are interesting topics of research since years. Specifically, CBIR is on developing technologies for bridging the semantic gap that currently prevents wide-deployment of image content-based search engines. Image search engines currently in use are mostly rely on human generated data, such as text. Annotation of an image is totally depend on the person's perception who is going to store it into database. It is time-consuming as well as error prone. Therefore search engine using text input results in various non-relevant images. To overcome drawbacks of text based image retrieval, Content based image retrieval is introduced where retrieval of images is totally depend on the features of images. Mostly, content-based methods are based on low-level descriptions, while high-level or semantic descriptions are beyond current capabilities. In this paper, we will try to implement the technique to fill this gap. This technique can eventually be extended to allow for content-based similarity type of search, to find out different query blocks from satellite image. We will fire an object as a query and most similar blocks will get retrieved. This will help out to trace particular object from large satellite image.

Keywords--Image retrieval; CBIR; MBTC; Kekre's pattern; Satellite image retrieval.

I. INTRODUCTION

Retrieval is the wide topic of research from the decades, because it is a challenge to reduce the semantic gap Basically retrieval of data means to get desired data from the database. It may image, text , audio or video as per requirement of user. The basic types of retrievals are mentioned in figure(1).

From ages images have been the mode of communication for human being. Today we are able to generate, store, transmit and share enormous amount of data because of the exhaustive growth of Information and Communication Technology. Much of this information is multimedia in nature, which consists of digital images, video, audio, graphics, and

text data [1], [2]. But all that information is only useful if one can access it efficiently. This does not only mean fast access from a storage management point of view but also means that one should be able to find the desired information without scanning all information manually.

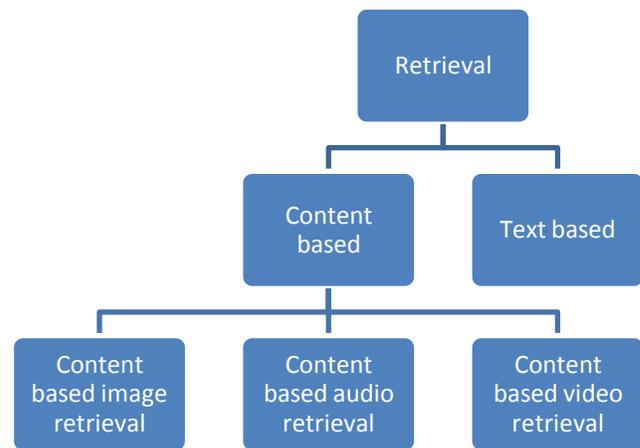


Fig.1. Types of retrieval

Previous method used for image retrieval is Text based image retrieval. The advantage of textual indexing of image is that it can provide user with key word searching, catalogue browsing and even with query interface. But the major drawback of text based image retrieval are, annotation depends on the person who adds it. [4] Due to all these drawbacks, Content Based Image Retrieval is introduced.

Here we will be dealing with three features of images shape, texture and color.

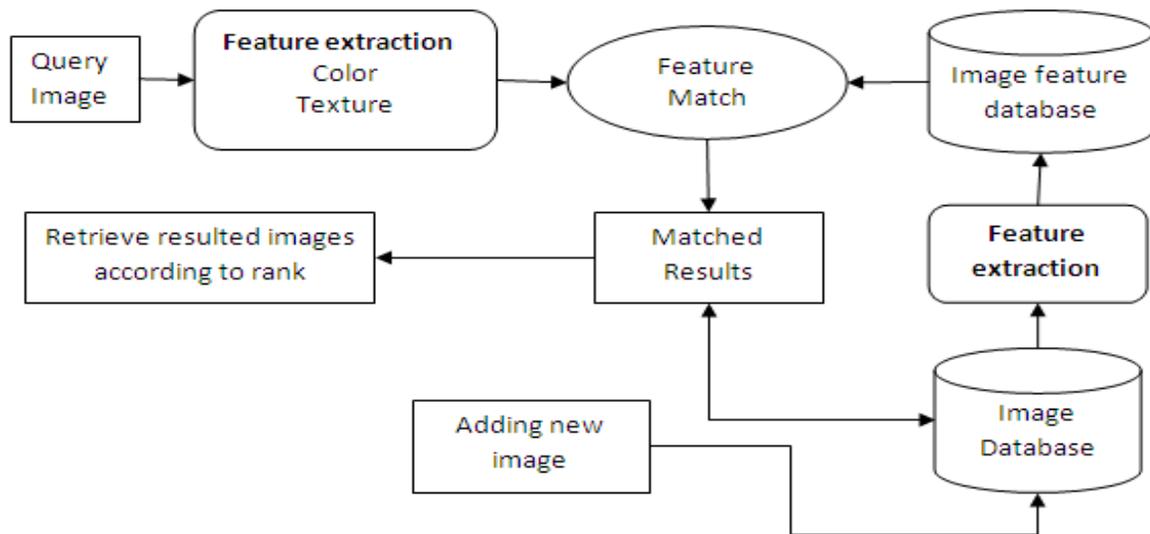


Fig.2. .Basic block diagram for planned work

II. CBIR EXISTING TECHNIQUES

Kekre, Bharadi et al. have introduced Content Based Image Retrieval using Fusion of Gabor Magnitude and Modified Block Truncation Coding [4]. Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. And modified block truncation is used to retrieve color feature from image. They proved that the proposed system is giving higher Precision and Recall as compared to only Gabor and Only MBTC based CBIR. Gabor feature gives good response to texture of the image and Modified BTC give good response to color content of image.

J. Zhang and W. Zou [5] have presented a novel technique that employs both the color and edge direction features for Content-Based Image Retrieval (CBIR). In this method, a given image is first divided into sub-block which has the same size and then the color and edge direction features of each sub block can be extracted. Next, it constructs a codebook of color feature using clustering algorithm and then each sub-block is mapped to the codebook. The color feature is used to retrieve images, and the edge direction feature is the weight of the similarity measure for the color feature.

Rose and Shah carried on a research project to improve the accuracy of CBIR Using Gradient Projections [6]; the image's structural properties were examined to distinguish one image from another. By examining the specific gray level of an image, a gradient can be computed at each pixel. Pixels with a magnitude larger than the thresholds are assigned a value of 1. These binary digits are added across the horizontal, vertical, and diagonal directions to compute three projections. These vectors are then compared with the vectors of the image to be matched using the Euclidean Distance Formula. These numbers are then stored in a bookmark so that the image needs only be examined once. A program has been developed for

Matlab that performs this method of projecting gradients. Three databases were amassed for the testing of the proposed system's accuracy: 82 digital camera pictures, 1,000 photographic images, and a set of object orientated photos. The program was tested with 100% accuracy with all submitted images to the database, and was able to distinguish between pictures that fooled previous CBIR engines. The weakness of this project was its color-blindness.

A CBIR method based on color-spatial feature has been proposed by Lei, Fuzong & Zhang[7] . They proposed a fast algorithm which could include several spatial features of color in an image for retrieval because except for the color histogram information, the position information of each color plays an important role too. These features are area and position, which mean the zero-order and the first-order moments, respectively. By computing the moments of each color region the similarity of two images according to the weight of each factor can be computed. In fact, these features are a kind of representation for image in the scale of low resolution, and the sample image given by a user is usually a draft drawn by hand. Moreover, when a user judges the similarity between two pictures, he will firstly judge them in coarse scale. In this sense, this method is close to the vision model of our eyes. Because the features are simple and can be calculated in fast speed, better result can be made easily through training.

Dr H B Kekre ,S D Thepade et al. introduced Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC [9] and tested on generic image database with 1000 images spread across 11 categories. The average precision and recall of all queries are computed and considered for performance analysis. Gradient operators used for shape extraction were Robert, Prewitt, Sobel and Canny which are known as 'Mask-Shape-BTC' CBIR techniques. The problem with these Mask-Shape-CBIR methods is the need of resizing the database images to match it with the size of query. This drawback is removed using proposed Mask-Shape-BTC-CBIR methods. In

proposed image retrieval techniques the feature vectors are formed by applying the block truncation coding (BTC) on the shape image obtained using slope magnitude applied on gradient of the image in both horizontal and vertical direction.

Y.N.Mamatha and A.G. Ananth worked on Content Based Image Retrieval of Satellite Imageries Using Soft Query Based Colour Composite Techniques[14]. They realized that using colours as a content, content based image processing have been carried out for a sample of high resolution urban image and low resolution rural image scenes obtained from satellites.

A research paper on Content-based satellite cloud image retrieval by Deepak Upreti [15] has been developed using gray level, texture and shape as retrieval features from the satellite image repository. The system allows the user to search for an image on the basis of any of the three features alone or in combination by assigning weights to the features. The histogram approach is used to extract the gray level feature, texture feature is extracted using gray level co-occurrence matrix method and the shape feature is extracted using the morphological operations. The images and the extracted feature vectors are stored in the Oracle 10g database. Euclidean distance metric is used to compute the similarity between the images. The system is robust as it provides search based on the multiple features. The performance of the system was evaluated by analyzing the retrieval results using precision. The proposed method of image retrieval performs well for the query image. The precision in retrieving the images is high in the proposed method as compare to the other methods of image retrieval.

III. PROPOSED TECHNIQUE

As we have discussed in previous section there are many techniques which are introduced to reduce semantic gap and to retrieve similar images. But still scope of improvement is possible. Non-relevant images get retrieved along with relevant images. It is difficult to implement technique which will provide zero non-relevant image retrieval. But we can try to reduce this possibility and may design a technique with maximum relevant images retrieval. To achieve this, here we are going to merge to techniques which are mentioned below. The overview of proposed technique is shown in figure(2).

a. MODIFIED BLOCK TRUNCATION CODING

Block truncation coding (BTC) is a relatively simple image coding technique developed in the early years of digital imaging. This method first divides the image into small non-overlapping image blocks. The small blocks are coded one at a time. For each block, the original pixels within the block are coded using a binary bitmap the same size as the original block and two mean pixel values[4]. The method first computes the mean pixel value of the whole block and then each pixel in that block is compared to the block mean. If a pixel is greater than or equal to the block mean, the corresponding pixel position of the bitmap will have a value of 1, otherwise it will have a value of 0. The simplest extension was to view a color image as consisting of three independent grey scale images and apply BTC to each color plane independently. Most color images are recorded in RGB space, which is perhaps the most well-known color space [11].

In modified BTC to create a binary bitmap in the RGB space, an inter-band average image (IBAI) is first created and a single scalar value is found as the threshold value.

First we will resize the image in 256*256pixels. Now let $X=\{r(i,j),g(i,j),b(i,j) \ i =1,2,\dots,m. \ j=1,2,\dots,n\}$ be an $m \times n$ color block in RGB space where $m=n=256$.

$$i_b = \frac{1}{3} r(i, j) + g(i, j) + b(i, j) \dots(1)$$

The threshold is computed as the mean of each color and using this threshold binary of bitmap can be calculated[12] as shown in figure(3).

Fig.3. Bitmap of image

After the creation of the bitmap, two representative (mean) colors are then computed. The two mean colors are $MC1 = \{Cr1, Cg1, Cb1\}$ and $MC2 = \{Cr2, Cg2, Cb2\}$. Where $MC1$ represents uppermean components and $MC2$ represents lowermean components[12].

b. PATTERN GENERATION FOR TEXTURE

Patterns are generated for extracting texture feature of image. Each pattern defines different formats of textue. The idea is to map the image equivalent to patterns and then the different texture pattern ‘16-pattern’ generated using Kekre’s transform matrix.

Number of patterns can be generated using transform matrices namely 4 pattern, 16 pattern, 64 pattern. $N \times N$ matrix can be used to generate N^2 patterns. For example, if we want to generate 16 pattern then 4×4 matrix need to be used. Element wise multiplication of each row of the transform matrix is taken with all possible rows of the same matrix.

The 16 Kekre texture patterns [13] is generated using Kekre transform matrices of size 4×4 . Figure 4 gives 2×2 kekre matrix and generation of four Kekre texture patterns. 2×2 Kekre transform matrix is shown in figure, each row of this matrix is considered one at a time and is multiplied with all rows of the same matrix to generate Kekre texture patterns as shown in figure 5. The 4×4 Kekre transform matrix is given in figure 6 and visualization of 16 Kekre transform patterns generated using it is shown in figure 8, where black and grey colour scaled between 1 to 256 in the pattern and 0 is represented by white colour.

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Fig.4. 2x2 Kekre Matrix

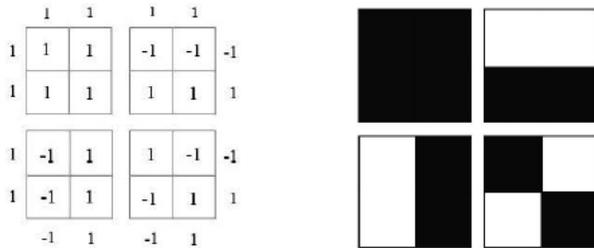


Fig.5. 4-Kekre Pattern Generation

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Fig.6. 4x4 Kekre Matrix

c. FUSION OF MBTC & PATTERNS FOR SATELLITE IMAGE RETRIEVAL

Satellite image consist of different classes are stored in hard disk. These images are divided into 256x256 pixel size blocks passed through feature vector extraction module and resultant feature vector is again stored as a MAT file [4].

A. Feature Extraction Module

- i. Any query image or block of satellite image will be first divided into 4x4 pixel block and MBTC technique is applied on each block. MBTC basically separates each block into three components i.e. red, green and blue. For each component Uppermean and Lowermean is calculated. It results into six components separated from one image i.e. uppermean red (Cr1), uppermean green (Cg1), uppermean blue (Cb1) and lowermean red (Cr2), lowermean (Cg2), lowermean blue (Cb2) [4],[5].
- ii. We calculated 16 grey patterns using 4x4 Kekre's transform matrix i.e. P1, P2,P3,P4,P5,P6 and so on
- iii. Then we will check for the grey levels existing in Kekre's pattern and quantize six components calculated by step (i) into these grey levels. Here we will get quantized grey components.
- iv. Now to form a feature vector, each pattern calculated at step (ii) will be superimposed on Cr1, Cg1, Cb1, Cr2, Cg2,

Cb2. E.g. first we will superimpose P1 with Cr1 (as shown in figure 4.14) and check for how many times P1 exist in Cr1 (or count of P1 present in Cr1). This count will be stored as the first element of one matrix.

- v. Step (iv) is repeated to find P2 to P16 count in Cg1, Cb1, Cr2, Cg2 and Cb2 components of the image.
- vi. Step (iv) & (v) will be resulted into 12x8 matrix which is declared and stored as a 'feature vector'.

When Query object image is fired, it is forwarded to feature vector extraction module and feature vector also calculated for it. Then satellite image (figure-4.15) resized in multiple of 256 and divided into blocks of 256x256 pixel size and each block is passed through feature extraction module. These extracted features for each blocks are stored in different matrices.

Feature vector of query object image is compared with feature vector of satellite image blocks stored in database goes through similarity measurement. Here we compared feature vectors using absolute distance. According to distance measured between query object and blocks of satellite image blocks, similar blocks are retrieved.

IV. RESULT AND DISCUSSION

For each image red, green and blue components are separated from uppermean and lowermean as shown in figure 7 and quantized in pattern values as shown in figure 8,

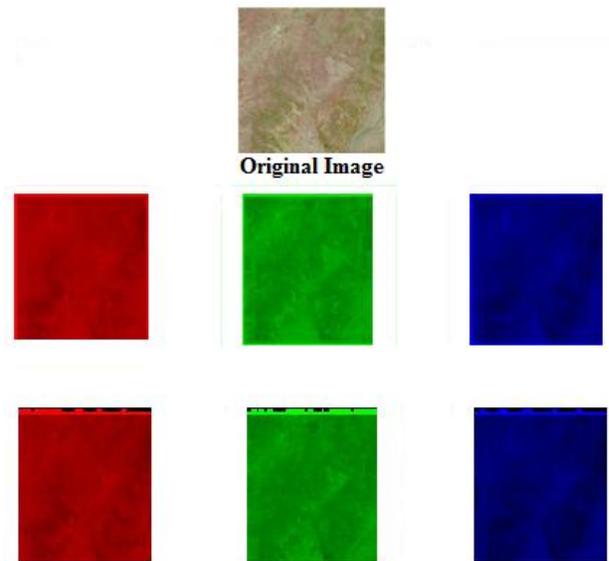


Fig.7. Upperman and lowermean for R,G,B components

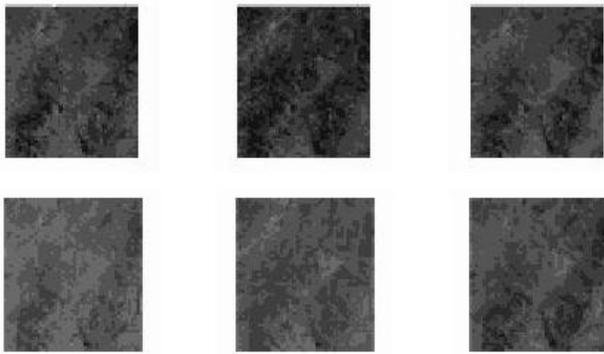


Fig.8. Quantized upperman and lowermean matrix

Using 4x4 Kekre matrix 16-pattern transform generated as shown in figure 9.

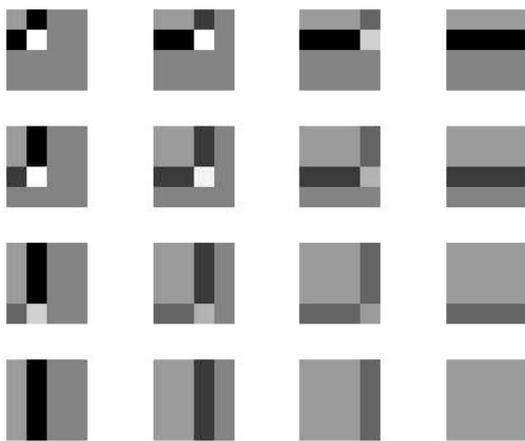


Fig.9. 16-Kekre Pattern Generation

Find co-occurrence of each pattern shown in figure 9 is checked with components in figure 8. This co-occurrence matrix is nothing but feature vector as show in figure 10,

1888	1651	1262	999	740	514	337	234	1191	1063	957	879
1651	1651	1651	1651	514	514	514	514	1063	1063	1063	1063
1262	1651	1888	2118	337	514	740	944	957	1063	1191	1291
999	1651	2118	2717	234	514	944	1759	879	1063	1291	1573
2062	1831	1270	980	946	704	467	319	1363	1255	1132	1053
1831	1831	1831	1831	704	704	704	704	1255	1255	1255	1255
1270	1831	2062	2277	467	704	946	1160	1132	1255	1363	1451
980	1831	2277	2827	319	704	1160	1998	1053	1255	1451	1787

Fig.10. (8x12) feature vector

A satellite image is divided into blocks of 256x256 pixel size as shown in figure 11.

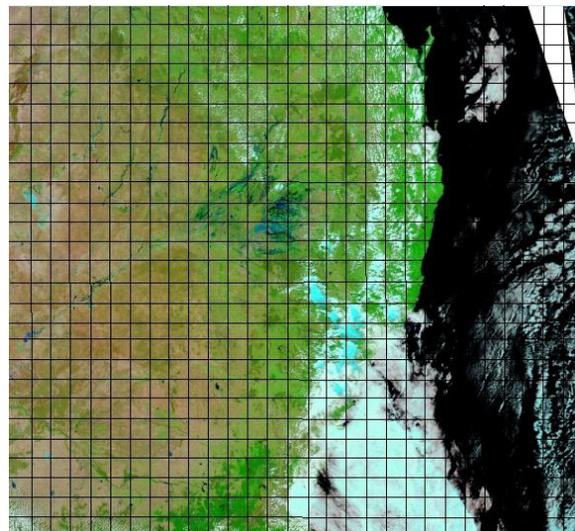


Fig.11. Original Satellite image of size [6800x6400] resized into multiple of 256 and divided into blocks of (256x256) pixel size

Now we will apply fusion based Content based image retrieval technique on each block and save feature vector of each block on hard disk. Each block is of (256x256) pixel size and after applying discussed technique on each block we will get feature vector of (8x12) size for each block. For this image we got 675 blocks of size (256x256) pixel on which we had applied CBIR technique and stored in a MAT file.

Then we had calculated FV for query object which is sample picture (size 256x256) of sand and greenery portion shown in figure 7.

Now to test the retrieval technique,we had segregated five query object block from original satellite image as shown in figure 12.

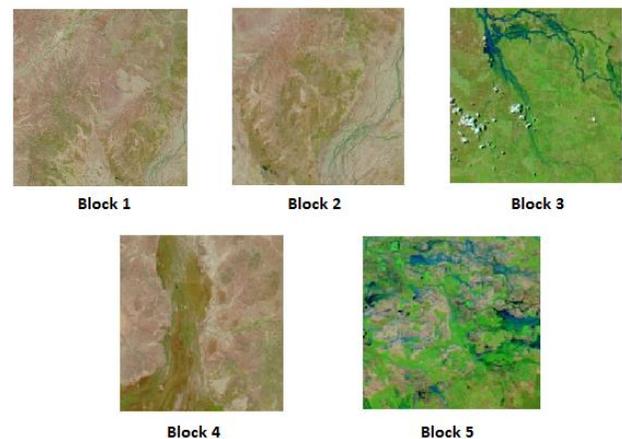


Fig.12. Segregated blocks from original image

Now we will fire each block as a query object to find out similar objects from the satellite image shown in figure 11. Retrieval results are shown below from figure 13 to figure 17. Relevant blocks to the query object are highlighted in results below.

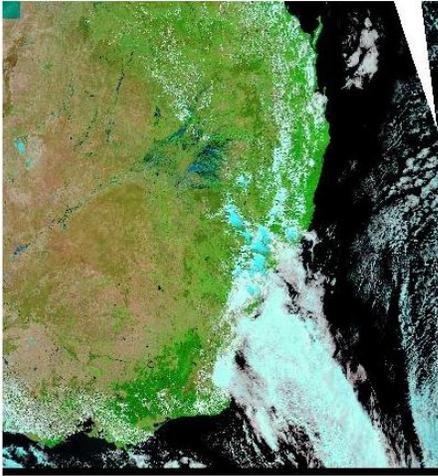


Fig.13. Retrieval output for block 1 from image 1

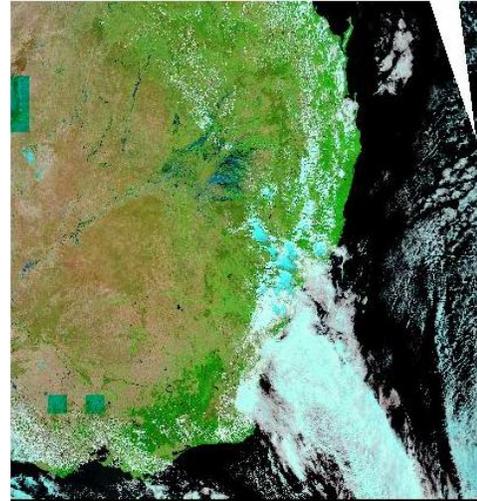


Fig.16. Retrieval output for block 4 from image 1

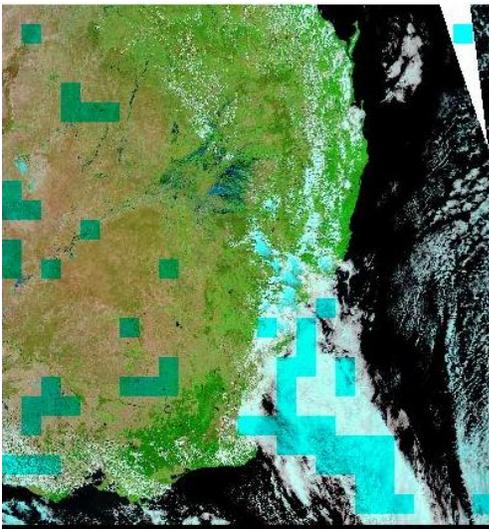


Fig.14. Retrieval output for block 2 from image 1

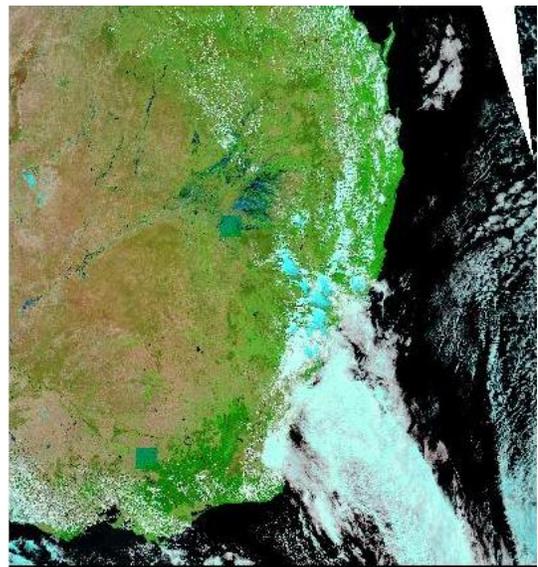


Fig.17. Retrieval output for block 5 from image 1

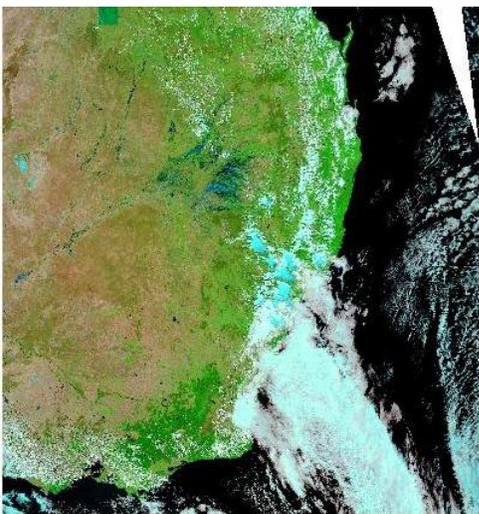


Fig.15. Retrieval output for block 3 from image 1

TABLE I. TEST RESULTS OF SATELLITE IMAGE RETRIEVAL

Satellite Images	Block No.	Total (256x256) Blocks in Images	Retrieved Relevant Blocks	Query Block Retrieved or not	Probability of block Retrieval	Efficiency of Retrieval (%)
Image 1	1	675	1	Yes	1	100
	2	675	103	Yes	0.01	1
	3	675	1	Yes	1	100
	4	675	5	Yes	0.20	20
	5	675	2	Yes	0.50	50
Image 2	1	64	1	Yes	1.00	100
	2	64	1	Yes	1.00	100
	3	64	2	Yes	0.50	50
	4	64	2	Yes	0.50	50
	5	64	1	Yes	1.00	100
Image 3	1	864	13	Yes	0.08	8
	2	864	2	Yes	0.50	50
	3	864	138	Yes	0.01	1
	4	864	1	Yes	1.00	100
	5	864	3	Yes	0.33	33
Image 4	1	168	6	Yes	0.17	17
	2	168	2	Yes	0.50	50
	3	168	20	Yes	0.05	5
	4	168	1	Yes	1.00	100
	5	168	1	Yes	1.00	100
Image 5	1	360	41	Yes	0.02	2
	2	360	1	Yes	1.00	100
	3	360	2	Yes	0.50	50
	4	360	1	Yes	1.00	100
	5	360	6	Yes	0.17	17

Similar test has been performed on other four satellite images. Five block queries are segregated from other satellite images and retrieved from each image. Retrieval results for each image is as shown in table I. From the results we can conclude that for all images we had retrieved relevant blocks using the implemented retrieval technique. But efficiency for each test is different. Analysis of efficiency of retrieval we achieved is as shown in table 2.

Table II - Efficiency of retrieval

Min. efficiency	Max. efficiency	Avg. efficiency
1%	100%	52.16%

Analysis shows that average efficiency for implement technique is greater than 50%.

From table 1, we can observe that there are 2 test which are completely failed where we got efficiency of 1%.

Therefore, $(2/25)*100= 8\%$ of the tests are unsuccessful. Six tests are below 33% efficiency are partially successful i.e. $(6/25)*100= 24\%$ of the total test are giving average retrieval results. And 15 out of 25 tests had achieved the efficiency above 50% i.e. $(15/25)*100= 60\%$ tests of the total tests are able to retrieve highly localized area.

V. CONCLUSION

In this paper we have proposed a satellite Image Retrieval technique using Color & texture features. This technique is

based on MBTC & Kekre Transform Patterns. This method is used for specific region block extraction from the high resolution Satellite image. Max retrieval accuracy was 100%, out 25 tests 60% tests were giving accuracy above 50% and high localization of the query block. This technique is very much suitable for searching a particular region from the satellite image.

REFERENCES

- [1] W.M. Smeulders, "Content-Based Image Retrieval at the End of the Early Years", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 12, DECEMBER 2000
- [2] G. Rafiee, S.S. Dlay, and W.L. Woo, "A Review of Content-Based Image Retrieval", IEEE, 2010. Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
- [3] Y N Mamatha and A.G Ananth and S O Neil, "Content Based Image Retrieval of Satellite Imageries Using Soft Query Based Color Composite Techniques", IEEE Trans on Acoustic speech signal processing, Vol 1, No.3, pp. 1278- 1288, 1986
- [4] Kekre H.B., Bharadi, V.A., Thepade S.D., Mishra B.K., Ghosalkar, S.E., Sawant S.M., "Content Based Image Retrieval Using Fusion of Gabor Magnitude and Modified Block Truncation Coding", IEEE computer society, 2010 IEEE
- [5] J. Zhang and W. Zou, "Content-Based Image Retrieval Using Color and Edge Direction Features", 2010 IEEE
- [6] Jed Rose and Mubarak Shah, "Content-Based Image Retrieval Using Gradient Projections", 1998 IEEE
- [7] Zhang Lei, Lin Fuzong, Zhang Bo, "A CBIR method based on color-spatial feature", 1999 IEEE
- [8] Zhao Hai-ying, Xu Zheng-guang, Penghong, "A Texture Feature Extraction Based On Two Fractal Dimensions for Content_based Image Retrieval", 2008 IEEE
- [9] Dr.H.B.Kekre, S.D. Thepade et al., "Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC", International Journal of Computer Applications (0975 – 8887) Volume 6– No.8, September 2010
- [10] Dr.H.B.Kekre, S. D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST - GVIP Journal, ISSN: 1687-398X, Volume 10, Issue 1, February 2010
- [11] H B Kekre and V A Bharadi, "Modified BTC & Walsh coefficients Based Features for Content Based Image Retrieval" NCICT, India.
- [12] A R Sawant, Dr. V A Bharadi, Dr. H B Kekre
- [13] B. Markarkandy "Color & Texture Based Image Retrieval using Fusion of Modified Block Truncation Coding (MBTC) and Kekre Transform Patterns", IJACSA Special Issue on Selected Papers from International Conference & Workshop On Emerging Trends In Technology 2012
- [14] Dr.H.B.Kekre, S.D. Thepade et al., "Performance Comparison of Gradient Mask Texture Based Image Retrieval Techniques using Walsh, Haar and Kekre Transforms with Image Maps", International Conference on Technology Systems and Management (ICTSM) 2011
- [15] Y N Mamatha and A.G Ananth and S O Neil, "Content Based Image Retrieval of Satellite Imageries Using Soft Query Based Color Composite Techniques", IEEE Trans on Acoustic speech signal processing, Vol 1, No.3, pp. 1278- 1288, 1986
- [15] Deepak Upreti, Dr. Sameer Saran, Dr. Nicholas Hamm, "CONTENT-BASED SATELLITE CLOUD IMAGE RETRIEVAL", Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente

Performance Evaluation and Statistical Analysis of MANET routing Protocols for RPGM and MG

Prajakta M. Dhamanskar
Fr. C.R.C.E., Mumbai
Lecturer, IT Dept .India

Dr. Nupur Giri
V.E.S.I.T., Chembur
Professor, Computer Engg. Dept.
HMCC, Mumbai, India.

Abstract—Mobile Ad Hoc Network is a collection of mobile nodes forming temporary network. In MANET routing protocols are classified as Proactive, Reactive and Hybrid. The work presented here evaluates performance of three Reactive routing protocols such as AODV, DSR and TORA under six performance metrics such as packet delivery ratio, routing overhead, packet loss, normalized routing load, throughput and end to end delay. The nodes follow two realistic mobility models such as Reference Point Group Mobility model (RPGM) and Manhattan Grid (MG) model. This work also presents the statistical analysis of AODV, DSR and TORA in Manhattan Grid (MG) model and Reference Point Group Mobility (RPGM) model in low load with low speed, average load with average speed, high load with high speed and very high load with high speed. Contribution in this work is beneficial in deciding which protocol to choose for better QoS.

Keywords— MANET; AODV; DSR; TORA; RPGM; MG.

I. INTRODUCTION

In MANET as the nodes are moving, the topology of the network changes dynamically. Also when packets are forwarded from source to destination, before the packets reach to the destination many routes break and many new routes are constructed dynamically. So an efficient routing algorithm is required to be used [4]. In MANET the routing algorithms are classified as On Demand i.e. Reactive, Table Driven i.e. Proactive and combination of both as Hybrid. AODV, DSR and TORA are Reactive, DSDV and OLSR are Proactive and ZRP is Hybrid routing algorithms. This work evaluates performance of only Reactive routing algorithms for nodes following Reference Point Group Mobility Model (RPGM) and Manhattan Grid Model (MG).

Mobility models are categorized as Entity Mobility e.g. Manhattan Grid and Group Mobility e.g. RPGM and define the pattern in which the nodes are moving. In Manhattan Grid the area is divided into rows and columns. Nodes can move only horizontally along the rows and vertically along the columns. Nodes can choose a random destination and move towards this destination with a predefined speed range upon reaching to the destination pause for some time and again repeat the same process. Whereas in RPGM nodes form groups and move in a coordinated manner. The logical center of the group is the group leader. Group leader determines the group member's speed and direction [8].

II. SIMULATION MODEL

Bonnmotion which is a mobility generator tool is used to generate the scenarios for RPGM and MG [9]. Network traffic is generated by network traffic generating tool supported by ns2 which is in \$NS2_HOME/indep-utils/cmu-scen-gen/cbrgen.tcl [11]. The simulation is carried out by increasing the number of nodes in the network as well as by increasing the speed of mobile nodes. And effect of these two factors i.e. network scalability and speed of mobile nodes on above mentioned six performance parameters is observed. The different simulation parameters are explained in Table 1 below.

III. PERFORMANCE PARAMETERS

Following performance parameters were used for evaluation of Reactive routing protocols.

- 1) Packet Delivery ratio i.e. PDR is defined as total packets received by constant bit sources (CBR) divided by total number of packets sent by CBR sink at destination.
- 2) Normalized Routing Load is defined as total routing control packets transmitted divided by total received data packets.
- 3) The packet End-to-End Delay is the average time that packets take to traverse the network. This is the time from the transmission of the packet by the sender up to their reception at the destination's application layer.
- 4) Throughput is defined as total delivered data packets divided by simulation time. Throughput is the successful message delivery over the communication channel.
- 5) Packet Loss is defined as total number of dropped packets divided by total number of data packets transmitted by sources.
- 6) Routing overhead is defined as total routing control packets generated during the simulation time.

TABLE I. SIMULATION PARAMETERS

Parameter	Number of nodes			
	25	50	75	100
Simulation area	500*500	700*700	1000*1000	1200*1200
Traffic Nodes	15	30	50	75
Nodes per group	5	10	15	20
Simulation time	300 sec			
Speed	2 m/s, 30 m/s, 60 m/s			
Pause	10 sec			
Traffic rate	2.5 Mbps			
Traffic type	Constant Bit Rate (CBR)			
Mobility models	RPGM, MG			

IV. SIMULATION RESULTS AND DISCUSSION

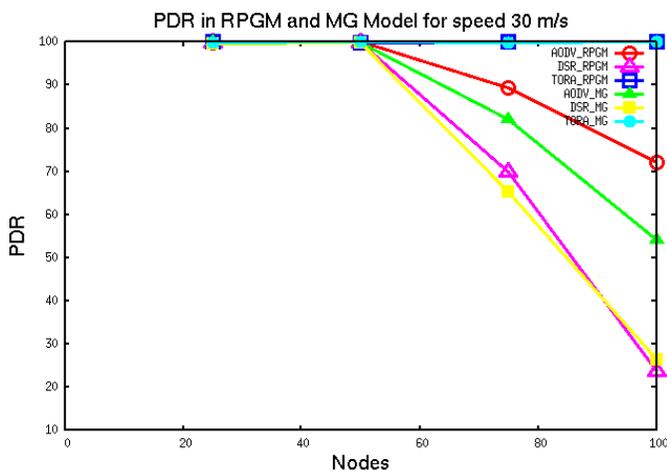


Fig. 1. PDR Vs Nodes

From Fig. 1, it is observed that PDR in TORA is highest among AODV, DSR and TORA because TORA protocol finds multiple paths from source to destination. So even if the network topology changes frequently due to mobility of nodes, TORA does not react at all. And therefore PDR in TORA is highest as compared to AODV and DSR.

Fig. 2 shows that NRL of DSR is the least because of route cache which is available at each DSR node. In case of route failure the DSR node refers to this cache for selecting new route and the probability of route discovery is reduced so routing overhead is reduced.

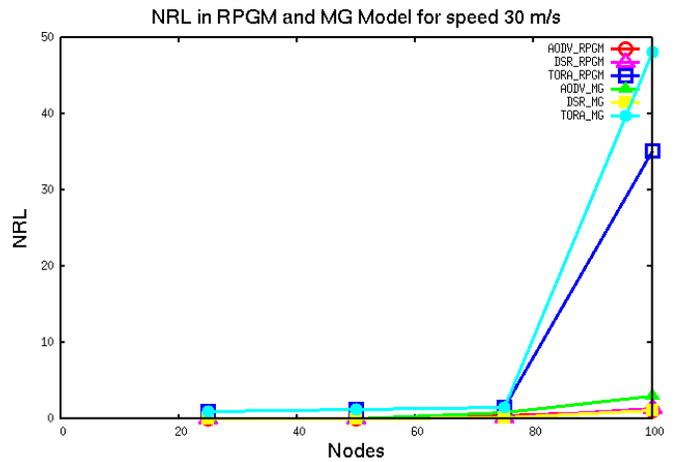


Fig. 2. NRL Vs Nodes

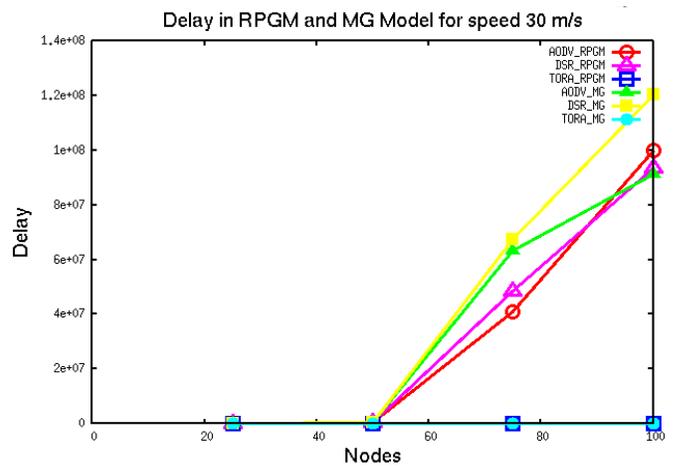


Fig. 3. Delay Vs Nodes

From Fig. 3 it is observed that delay of DSR protocol is greater in Manhattan Grid model because there is restriction on node movement as nodes can move only in four directions like left, right, top and down. And speed of node is restricted by preceding node in the same route.

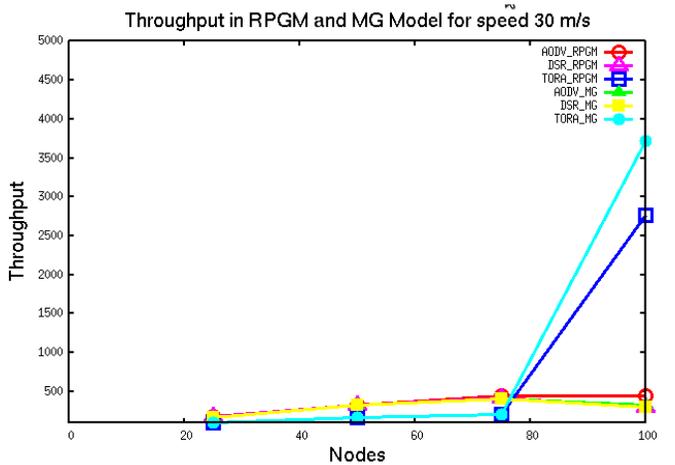


Fig. 4. Throughput Vs Nodes

Fig. 4 shows that throughput of TORA protocol is higher than AODV and DSR. Figure 4 shows that up to 75 nodes throughput of TORA protocol for RPGM and MG both is approximately same but after 75 nodes as the nodes increases throughput in Manhattan Grid is higher than RPGM because of increased number of nodes, the group size increases, congestion in the network increases and the probability of successful message delivery decreases so it is less in RPGM.

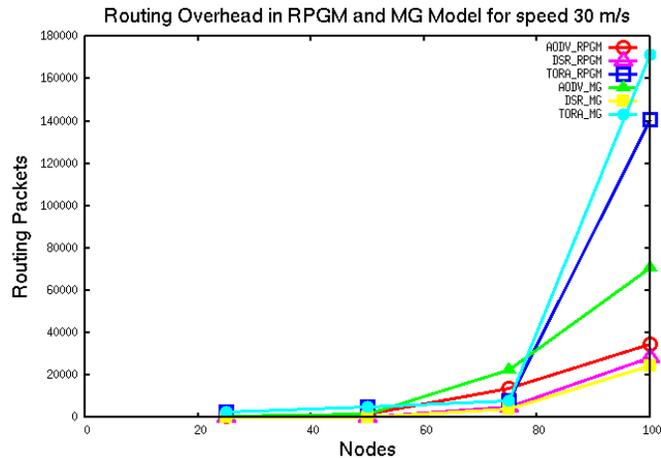


Fig. 5. Routing Overhead Vs Nodes

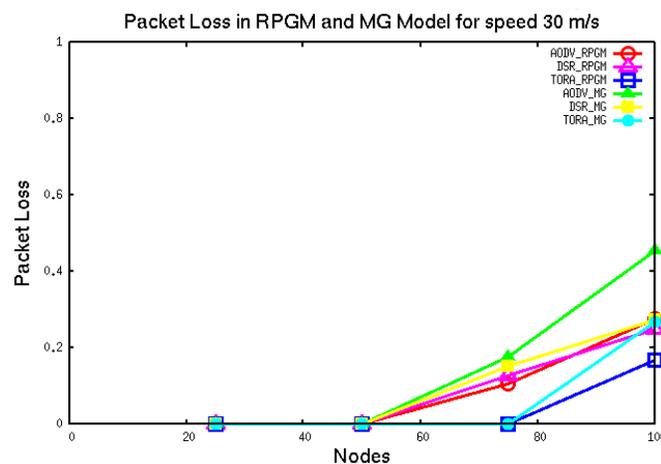


Fig. 6. Packet Loss Vs Nodes

Fig. 5 shows that routing overhead is least in DSR protocol because each node has its own route cache which it refers in case of route failure because of mobility of nodes. So less routing packets are generated in DSR.

Fig. 6 shows that as the number of nodes increases and goes beyond 50, packet loss increases because congestion increases and also because of node mobility probability of routes breaking frequently increases so packet loss increases. Packet loss of AODV is higher than DSR and TORA because there is no route cache as DSR also there are no multiple routes as TORA. Also figure shows that in Manhattan Grid model as there is restriction on node movement packet loss is higher. And as RPGM is group mobility model and group leader determines the group motion behavior and each member of the

group is uniformly distributed in the neighborhood of the group leader and each node deviates its speed and direction from that of the group leader. So Packet loss is less in RPGM as compared to MG.

V. PERFORMANCE ANALYSIS

The following KIVIAT diagrams in figure 7, 8 and 9 present the overall performance of AODV, DSR and TORA for 100 nodes and speed of mobile nodes as 30 m/s, 2 m/s and 60 m/s respectively. These diagrams help in quick identification of performance evaluation under the six performance metrics such as PDR, NRL, Delay, Throughput, Routing Overhead and Packet Loss. Each axis represents one parameter as shown in the figure.

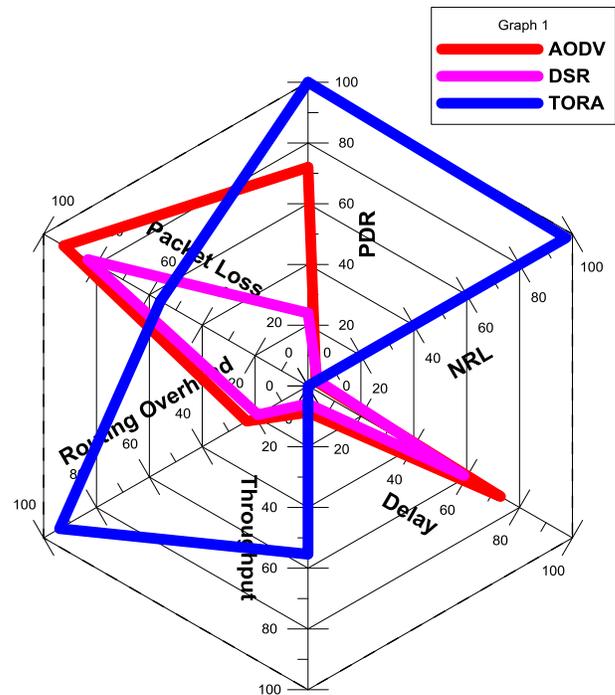


Fig. 7. KIVIAT diagram for routing protocol comparison with speed of mobile nodes 30 m/s

The above fig. 7 shows that for speed 30m/s TORA protocol performs well for PDR, Throughput, Packet Loss and Delay and performs badly for NRL and Routing Overhead. AODV performs average for PDR, Throughput and Routing Overhead but performs badly for Packet Loss and Delay. Whereas DSR performs average for NRL, Delay and Packet Loss and performs badly for throughput and PDR.

For speed of mobile nodes 2 m/s TORA protocol performs well for PDR, Packet Loss, Delay, and Routing Overhead. But it performs badly for NRL, Throughput. So selecting TORA is good choice. DSR performs good for Throughput, performs average for Delay, Routing Overhead, Packet Loss, NRL and PDR. So selecting DSR is average choice. AODV performs badly for Delay, Routing Overhead, and Packet Loss and performs average for PDR, NRL and Throughput. So selecting AODV is bad choice.

AODV performs badly for Packet Loss and performs average for Delay, Routing Overhead, PDR, NRL, and Throughput. So selecting AODV is average choice.

VI. COMPARATIVE ANALYSIS

Comparison of AODV, DSR and TORA for speed of mobile nodes 30 m/s can be presented in the following table. Value '1' represents the good choice, '2' represents average choice and '3' represents bad choice.

TABLE II. COMPARISON OF MANET ROUTING PROTOCOLS FOR SPEED OF MOBILE NODES 30 M/S

	AODV	DSR	TORA
PDR	2	3	1
NRL	1	2	3
Delay	3	2	1
Throughput	2	3	1
Routing Overhead	2	1	3
Packet Loss	3	2	1

VII. STATISTICAL ANALYSIS

The number of nodes and speed of mobile nodes in low load with low speed, average load with average speed, high load with high speed and very high load are 25 and 2 m/s, 50 and 30 m/s, 75 and 60 m/s and 100 and 60 m/s respectively.

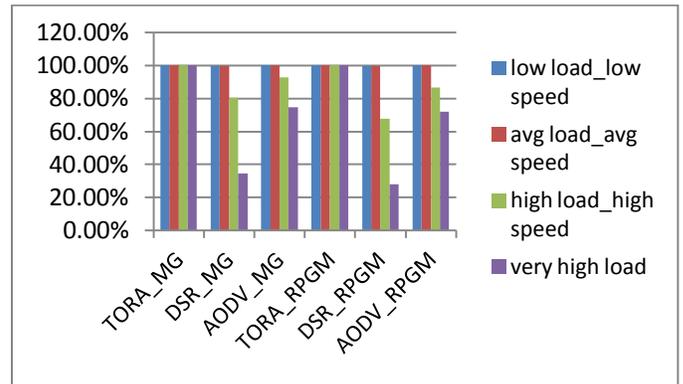


Fig. 10. PDR in low, average, high and very high load and speed.

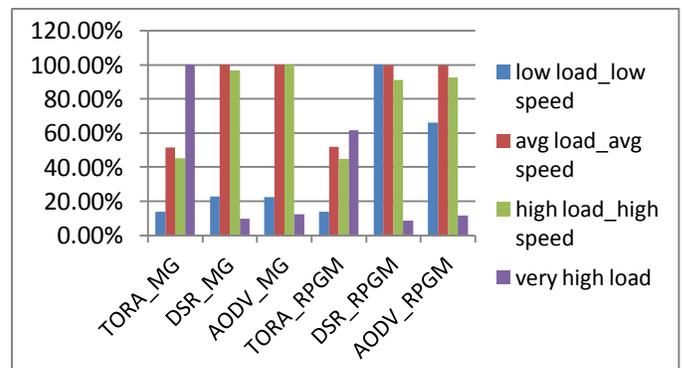


Fig. 11. Throughput in low, average, high and very high load and speed.

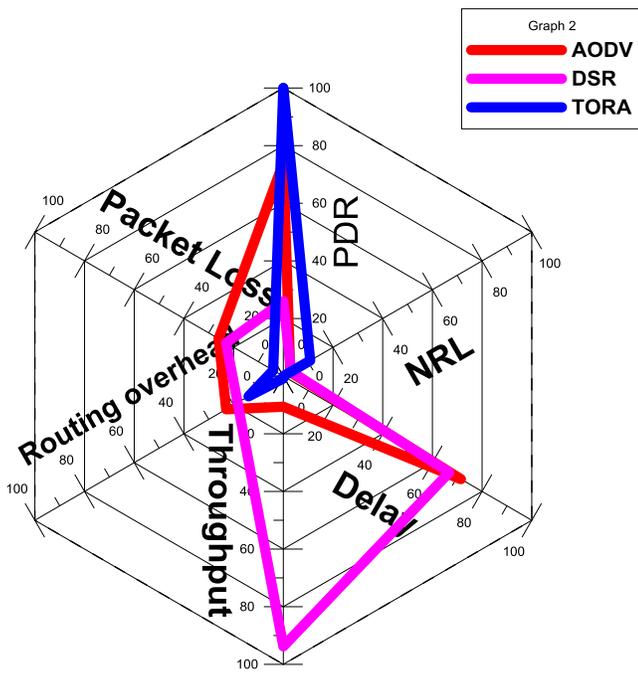


Fig. 8. KIVIAT diagram for routing protocol comparison with speed of mobile nodes 2 m/s

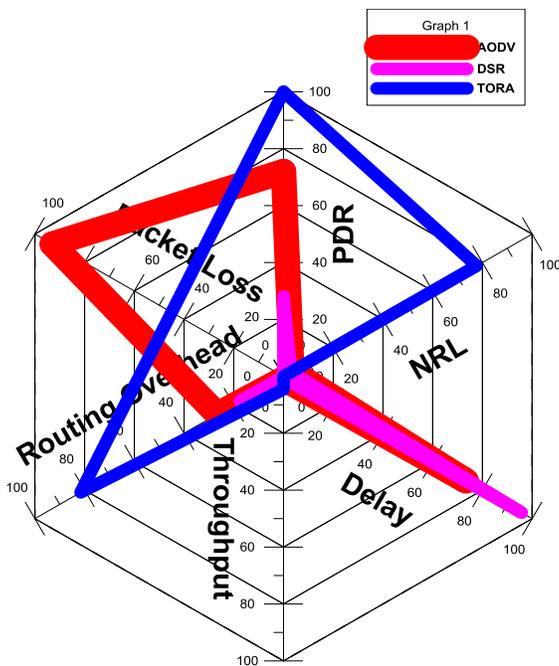


Fig. 9. KIVIAT diagram for routing protocol comparison with speed of mobile nodes 60 m/s

For speed of mobile nodes 60 m/s TORA performs well for PDR, Delay, and Throughput and performs badly for Routing overhead and NRL and performs average for Packet Loss. So selecting TORA is good choice. DSR performs good for Packet Loss, NRL and Routing Overhead, performs badly for PDR, Delay and Throughput. So selecting DSR is bad choice.

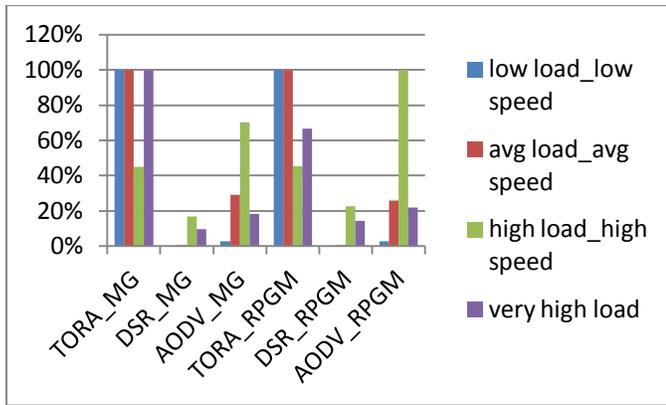


Fig. 12. Routing Overhead in low, average, high and very high load and speed.

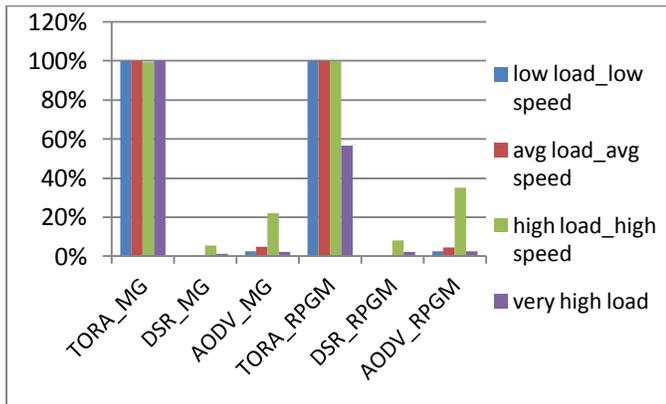


Fig. 13. NRL in low, average, high and very high load and speed.

PDR is observed to be maximum in low load and low speed and average load and average speed and least in very high load from fig. 10. PDR of DSR in Manhattan Grid and RPGM model is dropped by 20 % and 33 % respectively in high load and high speed and it is dropped by nearly 66 % and 73 % respectively in very high load because as the speed of mobile nodes increases the probability of link breaking increases, congestion also increases with increased load.

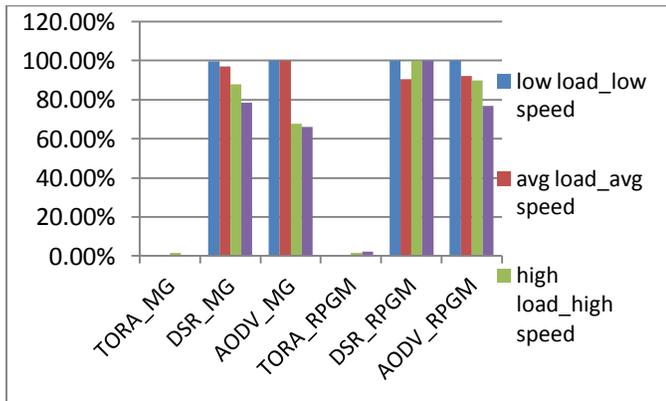


Fig. 14. Delay in low, average, high and very high load and speed.

From the fig. 11 it is observed that in low load and low speed DSR in RPGM model has maximum throughput and it is

dropped by 9 % and 92 % in high load with high speed and very high load respectively. AODV has maximum throughput in average load with average speed and high load with high speed and it is dropped by 78 % and 88 % in low load with low speed and very high load with high speed respectively because the packets delivered in low load are always lesser than that in average load and high load but further in very high load as the network becomes very congested and at high speed because of frequent topological changes throughput of the protocol is dropped.

Routing overhead of TORA is observed to be maximum in low load with low speed, average load with average speed and very high load from fig. 12. It is observed to be least of DSR in low load with low speed and average load with average speed and it is increased by 9 % and by 16 % in very high load and high load with high speed respectively because with increasing speed and increasing load the probability of topology changes increases with frequently route breakings and the routing control packets required for route discovery and route creations also increases resulting in increased routing overhead.

From fig. 13, NRL of TORA is observed to be highest in both MG and RPGM models and it is observed to be least in DSR. NRL of DSR is increased by 5 % in high load with high speed and NRL of AODV is increased by 2 % and 20 % in average load with average speed and high load with high speed because the NRL increases with increasing speed and load as routing packets are increased and it increases up to a threshold value in high load but after that it is again decreased in very high load.

From fig. 14 it is observed that Delay is least in TORA protocol in both RPGM and MG models and it is observed to be maximum in AODV in low load with low speed and average load with average speed. Delay of AODV in MG model is increased by 9 % and 19 % in high load with high speed and low load with low speed respectively because the nodes are moving with speed of 2 m/s in low load with low speed. Also in high load, high speed there is more congestion, more queuing delays and time consumed in route discovery and route creation so delay of DSR increases with increased load.

VIII. CONCLUSION

TORA outperforms AODV and DSR in PDR, Delay, Throughput and Packet Loss; DSR outperforms the other protocols in Routing Overhead and AODV outperforms the rest protocols in NRL. AODV has the worst performance in Delay and Packet loss and DSR has worst performance in PDR. The overall performance of these protocols is better when the nodes follow Reference Point Group Mobility model than Manhattan Grid model.

Our simulations do provide a link between the theoretical concepts associated with ad hoc routing protocols and the expected performance in practical implementations.

From this work, the conclusion is that among the protocols considered, there is no single one with an overall superior performance. One protocol may be superior in terms of routing overhead while others may be superior in terms of packet delivery ratio, packet end-to-end delay or throughput etc. The choice of a particular routing protocol will depend on the

intended use of the network. The work in this paper is beneficial in selecting the protocol for better QoS.

REFERENCES

- [1] Gupta A. K., Sadawarti H., Verma A. K. , Performance analysis of AODV, DSR & TORA Routing Protocols. IACSIT, 2010
- [2] Tyagi S.S., Chauhan R.K., Performance Analysis of Proactive and Reactive Routing Protocols for Ad hoc Networks. 2010 International Journal of Computer Applications
- [3] Kuppusamy P., Thirunavukkarasu K., Kalavathi B., A study and Comparison of OLSR, AODV, and TORA Routing Protocols in Ad Hoc Networks. 2011 IEEE.
- [4] Maan F., Mazhar N.: MANET Routing Protocols Vs Mobility Models, A Performance Evaluation. 2011 IEEE.
- [5] Liu Tie-yuan, CHANG Liang, Gu Tian-long., Analyzing the Impact of Entity Mobility Models on the Performance of Routing Protocols in the MANET. 2009
- [6] Shrestha A., Tekiner F., On MANET Routing Protocols for Mobility and Scalability
- [7] Aschenbruck N., Ernst R., Gerhards-Padilla E., Schwamborn M., BonnMotion -Mobility Scenario Generation and Analysis Tool
- [8] <http://www-public.it-sudparis.eu/~gauthier/MobilityModel/mobilitymodel.htm>
- [9] http://net.cs.unibonn.de/fileadmin/ag/martini/projekte/BonnMotion/src/BonnMotion_Docu.pdf
- [10] T. Valentina, S. Mirjana, B. R. Slavica., MANET Routing Protocols vs. Mobility Models: Performance Analysis and Comparison
- [11] P. Dhamanskar, N. Giri, Performance Evaluation of On Demand MANET routing protocols for different mobility models. IJETAE, 2012

SWiFiNet : A Task Distributed System Architecture for WSN

A.W. Rohankar

Dept. of IT
Sinhgad College of Engineering,
Pune University, India.

Mrinal K. Naskar

Dept. of ETCE,
Jadavpur University
Kolkata 700 032, India

Amitava Mukherjee

IBM India Pvt. Ltd.
Salt Lake,
Kolkata 700 091, India

Abstract—WSN is a technology, straddling many application areas of this millennium. The research and practices in WSN have not mended their ways due to many unjust requirement specifications for WSN architecture. The paradox is WSN characteristics and traditional approaches for network operation are diagonally opposite. Power optimization, resource constraints and small form factor are the main characteristics of the WSN node. Taking into consideration these main characteristics, the ‘SWiFiNet’: a task distributed reusable architecture for WSN has been proposed in this paper. The focus of this architecture developed is to identify and reuse system components while preserving the sensor node characteristics. The complex network functionality is pushed onto overlay second tier devices, leaving sensor free for application development. This work demonstrates the implementation of ‘SWiFiNet’ on hardware platform and network simulator using complete portability of reusable system components. Simulation and hardware results have been presented which illustrate that ‘SWiFiNet’ performs better and are application independent generic framework for WSN application development.

Keywords—Wsn; Reusable; Reconfigurable; Network Architecture

I. INTRODUCTION

WSN technology is holding many promises for the future short-range wireless communication enabled applications. The applications will be two fold, firstly replacement of existing wired systems and secondly many newer systems, which could not be realized due to wired technology constraints. The resulting systems would be agile in development and deployment yet economically viable. The unjust requirement specifications coupled with WSN characteristics have failed to address issues of integrated reusable generic WSN application framework that can be used across the WSN application domains. Research and development in wireless communication are an ongoing process due to dynamic and unpredictable behavior of wireless radios but most of the research work is directed towards specific isolated ideas and areas like scattered, random topology, Ad-hoc, nondeterministic network [1, 2]. Present Wireless communication research and development still follows the communication research and development still follows the footprints of existing communication standard and protocols coupled with inflated specifications of WSN. It is influenced by the traditional network characteristic like mesh topology, self organizing network, large scalable network etc. Integrated network architecture for the application domains at large are

not thought of. Typical approach has been to develop powerful smart wireless interfaces, which supports the important features/requirements for a particular class of applications like military, environment sensing or more focused applications like fuel-level control in automobiles. The result is a plethora of wireless interfaces appropriate for a certain class of applications; but almost no interoperability between them [3]. This approach has resulted in horizontal system models leaving little space for reusability and slimmness, which are most desirable requirements of WSN. The survey of WSN applications and their distribution shows that, despite many proposals, no common benchmarks or detailed, large-scaled experiments have been published. The research seems to focus either on node implementations or theoretical work on distinct aspects, such as routing algorithms, without a realistic relation to physical platforms [4].

The diverse application areas have inflated requirement specifications for WSNs. There is a need to synthesize the WSN application requirements, WSN characteristics and design the system accordingly. We believe that rational specifications could lead to a reusable WSN framework that would be useful for many of the WSN applications. We take bottom to top approach for the system design. The end node hardware and software architecture is designed meeting requirements of WSN characteristic, and then the rest of the system is built as a support system. This work presents ‘SWiFiNet’ (Swift Wi-Fi Network); a reusable task distributed WSN architecture. The architecture is based on practical assumptions drawn from the WSN application requirements. SWiFiNet is an application independent task distributed architectural framework that is usable in a large number of various WSN applications.

Section 2 talks about the characteristics of a reusable and reconfigurable WSN in terms of node specifications, network type and size, network life time and energy optimization, network dynamics, self organization and scalability, data transmission and communication reliability and lastly overlay backbone support. Section 3 describes the earlier studies that have motivated Design of ‘SWiFiNet’. Section 4 presents the features of ‘SWiFiNet’, designed as per the requirements explained in earlier sections. Section 5 presents implementation on hardware as well on simulated platform. Section 6 evaluates the SWiFiNet with performance parameters and section 7 concludes the study.

II. CHARACTERISTICS OF REUSABLE, RECONFIGURABLE WSN

Hierarchical network, network dynamics, data transfer models, scalability and energy optimization using scheduling are desirable features of the reusable WSN architecture.

A. Node Centric System Characteristics

WSN is defined as congregation of small tiny wireless sensor nodes. The sensor nodes are lightweight, power aware, small size, low cost and are in large numbers to be deployed in the field. The node description translated into hardware and software specifications of the end node imply the use of 8 bit/16 bit processors having less computing power, limited memory size, less peripheral resources, small size and low cost. The power source is often small battery, prompts for aggressive energy optimization at all levels.

B. Hierarchical Network, Network Size

Hierarchical networks are suitable to deploy and manage the network. The WSNs are deployed with specific tasks and aims, in most of the cases it is data gathering from the deployment field. Real field deployments of WSN demonstrate that the network size is limited often to 10s to 100s of nodes satisfying large share of WSN application domains [1,5]. The network design specifications like network throughput, reliability and lifetime would demand deterministic behavior of network devices in terms of energy consumption, reliability metrics.

C. Network Lifetime and Energy Optimization

WSN lifetime can be described as time until the last useful message is delivered from the network to the system. This implies maximum lifetime for the end node. Major energy consumption is in radio communication. The radio communication can be reduced by having an optimum sleep - wake up schedule satisfying requirements of the application

D. Network Dynamics, Self-Organization, Scalability

Advantage of wireless network is easy deployment and immediate startup of the network. The network topology can change over time or due to application demand. Scalability is often an issue in the networks. The network must support a sufficient number of nodes and their addition and deletion in the field. A resource constrained sensor node cannot handle these issues like traditional networks with resourceful and constraint less end device. Most of this functionality can be pushed onto backbone network devices.

E. Data Transmission Models and Communication Reliability

There are three types of data transfer which take place in WSN. Periodic data transfer, event driven data transfer and queried data transfer. The network layer should support all these transport models. The communication reliability is a measure of success of receiving the message at the destination.

F. Overlay backbone support

The hierarchical architecture has second tier devices having more resources than end node. We further expand this concept to overlay infrastructure that will act as message catching layer, which will be well equipped for handling the complexity of network and application functionality in

distributed form. While overlay infrastructures are free from the constraints imposed by WSN characteristics, it will provide increased network and computational capacity enabling large-scale deployments

III. BACKGROUND

'SWiFiNet': a task distributed reusable architecture for WSN has been inspired by earlier researchers who worked on the concepts based on the parameters/keywords discussed above. Hierarchical task distributed WSN architectures for example have been proposed earlier also in the WSN research domain.

Theme of such proposals is providing an intermediate overlay layer in terms of system components of high resources to facilitate the system interface to the high end networks. Many of them have proposed 802.11 based dynamic or fixed backbone devices so that rest of the system can make use of existing system software and hardware to interface with internet cloud. Some of the researchers have proposed centralized network management. An infrastructure based deployment (i.e. Fixed wireless deployment) [3], Tenet architecture based on three important tenets [6], three layered hierarchical architecture presented in iCASS [7] are few examples. Leel et al. have presented ART wise gateway architecture [8]. The envisioned architecture is two-tiered, IEEE 802.15.4/ZigBee for Tier 1 and IEEE 802.11 for Tier-2.

According to Linfeng (2010) there are two challenges in WSN architecture design: the first is the message interactions among different modules; the second is how to reuse the communication protocols. The authors have presented Environment-Adaptive Architecture Model for Wireless Sensor Networks (EAWNA) The tier 2 is capable of the overlay WLAN must be able to deal with a large number of nodes so that its increase does not affect the behavior of the two-tiered architecture [9]. The ANGEL architecture is presented for health care application, where the authors claim that it can be reused for any other application as well [10]. In Reconfigurable wireless networks H. Ramamurthy has presented a reusable solution for industrial wireless sensor networks [11].

Triantafyllidis has proposed an open and reconfigurable Wireless Sensor Network (WSN) for pervasive health monitoring, with particular emphasis on its easy extension with additional sensors and functionality by incorporating embedded intelligence mechanisms [12]. ZUMA [13], is a centralized future smart-home platform that interconnects all kinds devices in the home environment, Mote herding uses a mix of many 8-bit sensor nodes (motes) and fewer but more powerful 32-bit sensor nodes (micro-servers) [14], EMMON a system architecture for large-scale, dense, real-time embedded monitoring[15] are few more works have some resemblance. EMMON provides hierarchical communication architecture together with integrated middleware, command, control software and follows the lead line of having overlay dynamic/fixed second tier resourceful devices.

IV. DISTRIBUTED SYSTEM ARCHITECTURE: 'SWIFINET'

'SWiFinet' is designed as distributed task architecture for the WSN where the emphasis is on task distribution and reusability of the system components. A task-layered architecture is shown in figure 1. The network is envisaged as hierarchical architecture, master sink node, router node and sensor node. The routers are used for extending the physical distance connectivity and network management. The base layer is 802.15.4 MAC/PHY layer. Built on top of this layer is a common DLL task layer. This layer provides basic functionality for transmitting and receiving data and control packets. The DLL layer provides functionality for network communication handling and is transparent to the user. The network task stack grows as per the device hierarchy.

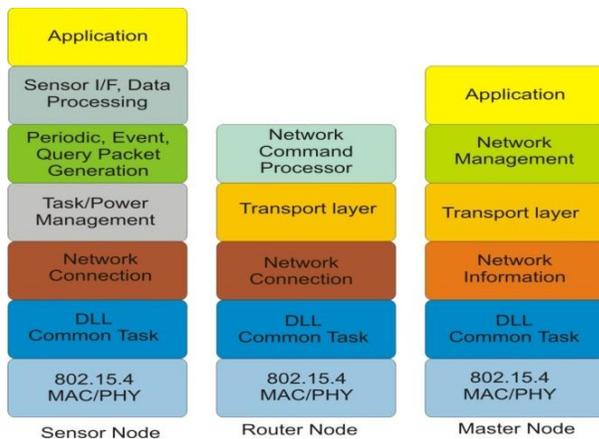


Fig.1. Distributed Task architecture for WSN

A. Sensor Node

The sensor node architecture carries the only network task restricted to connecting to the available neighboring overlay router nodes or to the master. The transport mode that is with 'ack' or without 'ack' and message priority is embedded in communication packet. Once this packet is handed over to the parent device, it is the responsibility of the parent device to transfer the message to the destination device.

The network layer maintains a routing table containing two parent device entries. The first entry is a primary parent address and the second entry is a secondary parent address. The network connection layer invokes connection functionality when there is no entry in the routing table and a data packet request is generated from upper layer. A hello packet is broadcasted in this case. The neighboring parent device issues join request. This join request is accepted, if the received packet meets LQI threshold criteria. If there are multiple join requests received by the sensor node then minimum hop criteria is applied. Task/Power management layer provides support for wakeup-sleep schedule for the node. This schedule is application dependant and can be configured runtime by the master sink node. The tasks are managed by task scheduler if available or can be a super loop structure. Sensor interface and data processing layer contain drivers for I2C, SPI, ADC, DAC devices and data processing routines. The application layer manages the application deployed with associated configuration parameters set at lower layers.

B. Router Node

The router node is used if a sensor node is not in the range of the master sink node. The router node can also be used if there is more number of devices and clustering is required in case of network management. The router joins to the network in a similar way as explained for the sensor node except here the router starts connection operation immediately instead waiting for the data transfer request to arrive from the upper layer. The additional task of generating a join request is added in this layer. The join request is generated if the device receives hello packet from the sensor node or neighbor router node. The router can respond to the hello packet only if he has joined to the network and secondly the requesting device meets the threshold LQI criteria. The router maintains a similar table like sensor node having parent device information leading to the master sink node. Another table maintains device addresses for the child routers down the network line. The router does not maintain any database for end node devices. Once the routing node receives the data packet, it is the routing nodes responsibility to transfer to the next destination node and so on. The command processor layer provides functionality for data aggregation, data fusion depending on the configuration parameters. A master node can set the configuration parameters at run time.

C. Master Node

The master node maintains complete topology information of the network. The network information layer is responsible for maintaining routing tables for each node in the network. Mainly this is required for sending command packets and querying the end node. The routers and end nodes will send their routing tables every time these tables are updated. The network and application parameters can be tuned by configuring routers and sensor nodes configuration tables.

D. SWiFiNet: Points of Comparison

SWiFiNet even though has similarities with the architectures explained in section 3.0 also has major differences in many ways. The second tier in the SWiFinet architecture uses the same hardware as sensor nodes except sensor part and is based on 802.15.4 compliant trans-receivers unlike many of the above architecture having higher bandwidth radios like 802.11 likes at the second tier. Sleep-wake schedule in SWiFinet is application dependant and is dictated by an end node. This solves two major issues, firstly the sensor node life is deterministic and application dependant and secondly no time synchronization overhead incurs saving program complexity and power consumption in communication. The overlay devices are always active when they need to hear the end node communication. The configuration parameters at each task layer provide facility to tune the network and application specific parameter. In SWiFiNet network operation is completely decoupled from the application functionality. SWiFiNet aims to provide an integrated reusable and reconfigurable architectural framework for a variety of applications.

V. IMPLEMENTATION

'SWiFiNet' is implemented on hardware platform as well in simulator ns-2 platform. This gave an advantage to test all

the logic in a simulation environment. Particularly difficult and complex situations that are difficult to realize during hardware testing were simulated and problems were resolved on simulation platform. The system software code was reused with hardware and in creating agent 'SWiFi' in ns-2 simulator.

A. Hardware Implementation

SWiFiNet architecture is realized using a single chip CPU plus trans-receiver solution. The implementation is not processor specific and can be ported to any other hardware platform. The Single chip processor is 32-bit RISC processor operating at 16 MHz, with 2.4 GHz. 802.15.4 compliant trans-receiver. Receiver sensitivity is -97dB, consumes 17.5 mA in active mode. Transmission power is +0.5 to +2.5 dB with programmable stages. The processor has SPI, I2C, UART bus, along with timers and IO ports. Figure 2 Shows sensor node hardware. The same hardware can be used for router and master node realization.

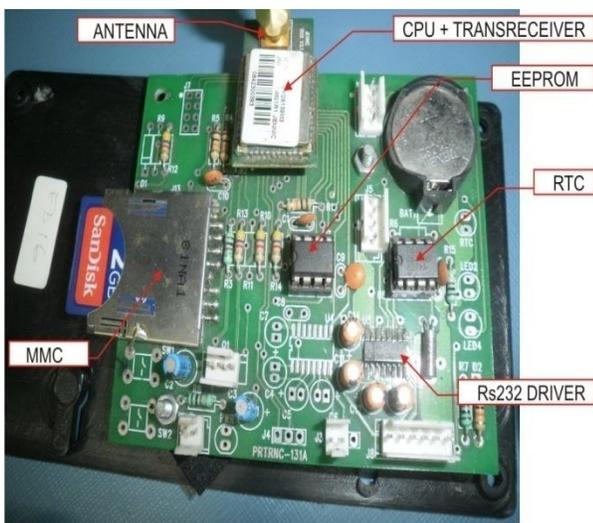


Fig.2. Hardware platform used for WSN nodes

B. ns-2 Simulation of SWiFinet

Simulation has its own advantage of checking all the operational logic and allows testing limiting conditions which otherwise could not be tested in real field situations. The ns-2 network simulator provides basic radio models and base functionality for packet communication. It allows building proprietary protocol model by creating network agent. This agent can be invoked in the TCL interface script to verify the behavior of the protocol in different scenarios.

C. Agent SWiFi

Agent 'SWiFi' software architecture is shown in figure 3. The portability of the software was truly tested while porting the software blocks in ns-2 environment. There was 100% success in porting software architecture between ns-2 environment and hardware platform. Network common tasks are used as basic interface between device specific tasks.

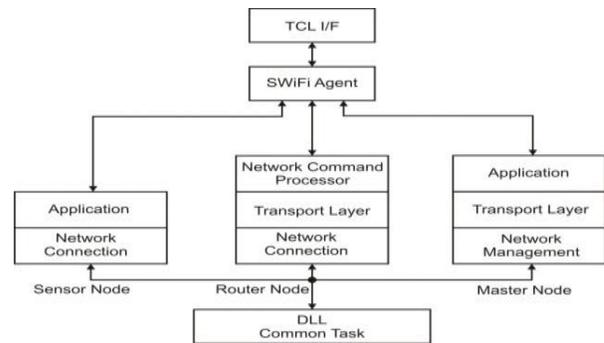


Fig.3. Block schematic of network component simulation in ns-2

C. Architecture and building interface

The devices were identified by the accessible variable passed from the TCL script while invoking the 'SWiFi' agent. Value 1, 2, 3 were attached to master, router and sensor node sink node functionality respectively. An application layer at end node was used for generating packets at a desired sampling rate, packet size and data mode transfer. An application layer at the master sink node was tested by developing PING application.

VI. RESULTS AND DISCUSSIONS

SWiFiNet shares a common characteristic with Dynamic Source Routing (DSR) protocol as they both initiate routing activity on an on demand basis. The reactive nature of this protocol is a significant departure from proactive traditional network protocol. The motivation behind this is the reduction of the routing load. Since the routing load is an important parameter in WSN in perspective of power consumption SWiFiNet adopted reactive approach to set up the network. Although DSR is meant for mobile ad-hoc network, it will be interesting to compare them in static topology. SWiFiNet also supports mobility of the devices within a network. Simulation in ns-2 is carried with propagation model: Two Ray Ground model, net-interface: Phy/WirelessPhyExt and mac layer MAC/802_11Ext. Field size is set to 1000 m x 1000 m and bandwidth is set at 1 Mbps. The results are averaged over three sets of experimental values. Hardware results are presented in comparison with simulation. Standalone hardware results are presented with a variation in packet interval.

I. Performance Parameters

The normalized routing load is number of routing packets transmitted per data packets sent to the destination and also each forwarded transmission is calculated as one transmission. Average route acquisition time is the time taken by the network devices to join the network and become functional. An average transmission delay is an end-to-end delay for the message to reach destination averaged over total message transmitted.

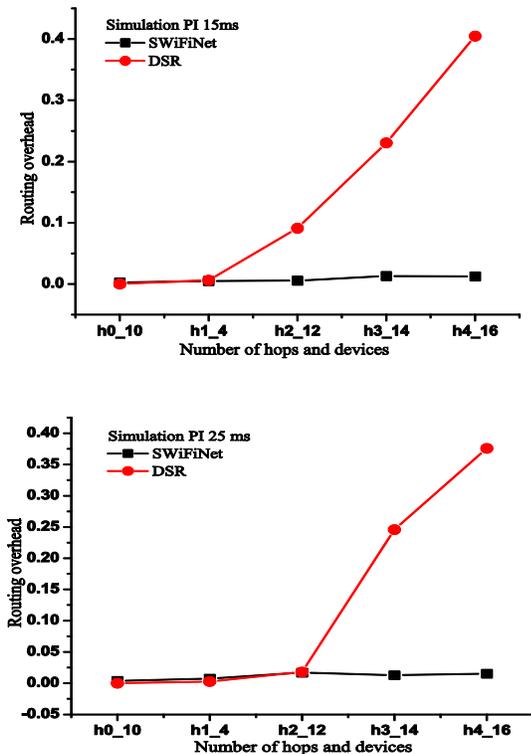
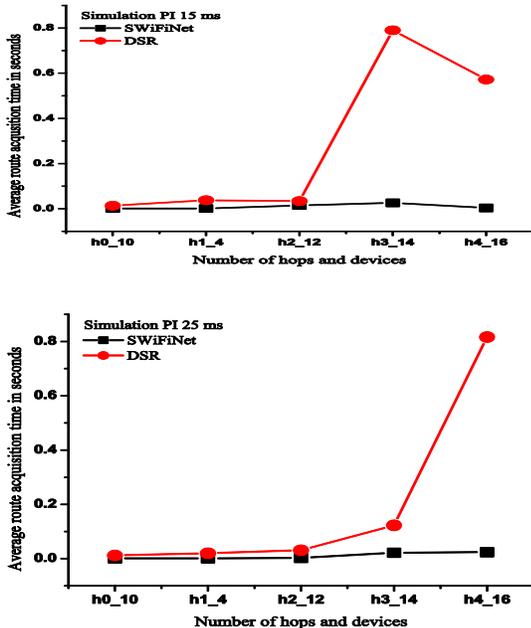
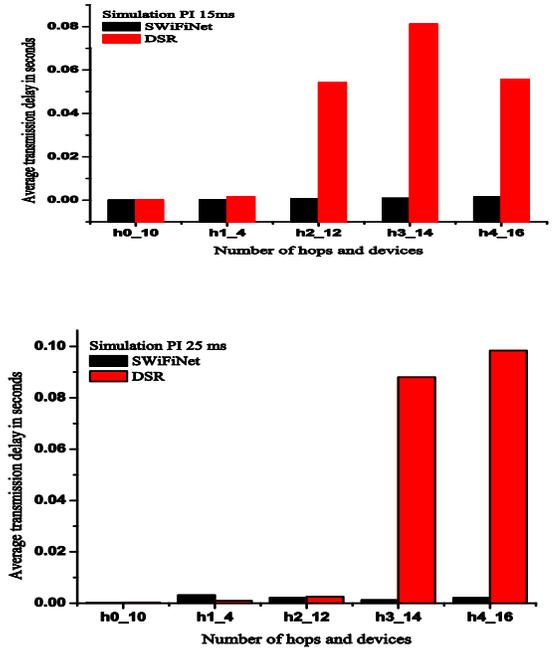


Fig.4. (a). Normalized routing load



(b). Average route acquisition time



(c). Average transmission delay

Figure 4 (a) shows normalized routing overhead. The routing overhead increases sharply with the increase in number of hops and devices with DSR. The routing overhead more or less remains constant with SWiFiNet. This is because SWiFiNet is based on deterministic communication protocol. The joining process is simple and involves minimum message communication. Secondly, the parent devices respond to neighboring device only when it is part of the network. This deterministic behavior is supplemented with results of route acquisition time shown in figure 4 (b). The SWiFiNet route acquisition time shows marginal variations with an increase in hops and devices as compared with DSR. DSR shows varying and steep increase in time to increase in the hops and number of devices. Average transmission delay is plotted in Figure 4 (c). The SWiFiNet shows almost constant delay value since routing paths are fixed. Large variations in average transmission delay are observed with DSR due to non-deterministic routes formed by the DSR protocol as expected.

II. Hardware Results

Hardware experimentation is carried out within an open air field in line of sight devices. A set of 50 ms and 250 ms packet interval are compared considering the 802.15.4 radio 250 kbps bandwidth. Figure 5 (a) shows normalized routing overhead for 25 ms packet interval simulation results compared with 250 ms packet interval for hardware experiment result set. The plot shows an elevated overhead graph for hardware results incurring more command packets per data packet as compared with simulation results. This elevation is attributed to difference in simulation radio model and field radio.

The field radio requires more command packet transmission retries owing to field environment. Figure 5 (b) and figure 5 (c) shows average route acquisition time and average transmission delay respectively. The route acquisition time increases with decrease in packet interval rate. This increase in time is due to the mobility mechanism introduced into the protocol.

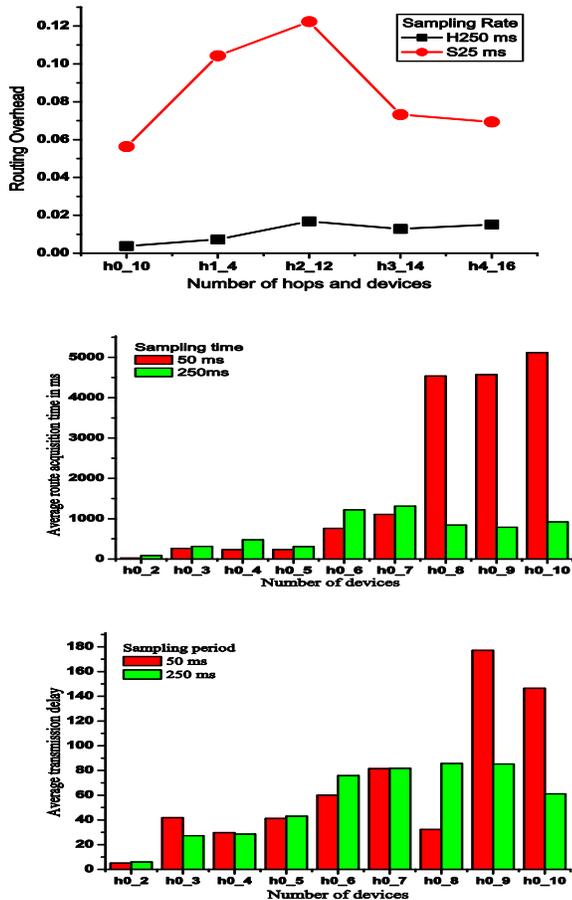


Fig.5. (a). Normalized Routing Load Figure 5 (b). Average route acquisition time Figure 5 (c). Average transmission delay

If a threshold repeats failure count crosses the set threshold value, the transmitting device assumes that the path is broken and starts a fresh joining process. The average transmission delay time increases as the communication congestion increases in the network due to higher packet interval rate and the number of devices.

VII. CONCLUSION

SWiFiNet is task distributed generic reusable architecture for WSN. The sensor node has been spared from wireless network management complexity, keeping network component required for joining the network. The network complexity of reliable data transfer, network management and other network related operation has been pushed on to overlay second tier devices having more power resources. This will make the node; small size, low cost and power optimized as

envisioned by the WSN characteristics. The results show that SWiFiNet performs better on many network parameter accounts. The deterministic behavior of the network operating at end node will allow the application designer to calculate network lifetime to accurate degree. The reusable modules will pave the way for building and deploying the application quicker. The design is scalable and self-organizing after deployment. The configuration parameter at various levels allows the designer to tune the network as per the application requirement.

VIII. ACKNOWLEDGMENTS

The authors wish to acknowledge the support given by M/S Partonics & Sinhgad Technical Education Society.

REFERENCES

- [1] K. C. Bhaskaran Raman, "Censor Networks: A Critique of "Sensor Networks" from a Systems Perspective," in *proceeding of ACM SIGCOMM Computer Communication Review* vol. 38, pp. 75-78, July 2008
- [2] P. D. David Culler, Cheng Tien Ee, Rodrigo Fonseca, Jonathan Hui, Philip Levis, Joseph Polastre, Scott Shenker, Ion Stoica, Gillman Tolle, and Jerry Zhao, "Towards a Sensor Network Architecture: Lowering the Waistline," pp. 1-5.
- [3] B. S. P. Harish Ramamurthy, Rajit Gadh, "Reconfigurable Wireless Interface for Networking Sensors (ReWINS)," in *proceeding of the 9th IFIP International Conference on Personal Wireless Communications (PWC 2004)*, 2004.
- [4] M. H. a. a. Mauri Kuorilehto, Timo D. H'am" al "ainen, "A Survey of Application Distribution in Wireless Sensor Networks," *EURASIP Journal on Wireless Communications and Networking* vol. 5, pp. 774-788, 2005.
- [5] M. K. N. A. W. Rohankar, Amitava Mukherjee "A STEP TOWARDS REUSABLE WSN ARCHITECTURE," *International Journal of Research and Reviews in Wireless Sensor Networks (IJRRWSN)* vol. 2, 2012.
- [6] E. K. D. E. R. Govindan, F. Bian, K. Chintalapudi, O. Gnawali, S. Rangwala, R. Gummadi, and T. Stathopoulos, "Tenet: An Architecture for Tiered Embedded Networks," in *Proceedings of Tech. Rep.*, 2005, November 10
- [7] M. Cinque, Di Martino, C. And Testa, A, "iCAAS: Interoperable and Configurable Architecture for Accessing Sensor Networks," in *Proceedings of the 3rd International Workshop on Adaptive and Dependable Mobile Ubiquitous Systems - ADAMUS 2009*, pp. 19-24, 2009, July 13-17.
- [8] J. C. Leal, A. Alves, M. Koubaa, A., "On a IEEE 802.15.4/ZigBee to IEEE 802.11 gateway for the ART-WiSe architecture," in *Emerging Technologies and Factory Automation, 2007. ETFA. IEEE Conference on*, 2007, pp. 1388-1391.
- [9] L. Linfeng, "Research on Environment-Adaptive Architecture Model of Wireless Sensor Networks," in *Networks Security Wireless Communications and Trusted Computing (NSWCTC), 2010 Second International Conference on*, 2010, pp. 130-133.
- [10] A. H. Willig, J. H. Karowski, N. Baldus, H. Huebner, A., "The ANGEL WSN Architecture," in *Electronics, Circuits and Systems, 2007. ICECS 2007. 14th IEEE International Conference on*, 2007, pp. 633-636.
- [11] B. S. P. H. Ramamurthy, and R. Gadh, "Reconfigurable Wireless Interface for Networking Sensors (ReWINS)," in *Proceedings of IFIP TC6, 9th International Conference, PWC 2004*, 2004, September
- [12] A. K. Triantafyllidis, V. Chouvarda, I. Maglaveras, N., "An open and reconfigurable Wireless Sensor Network for pervasive health monitoring," in *Pervasive Computing Technologies for Healthcare, 2008. Pervasive Health 2008. Second International Conference on*, 2008, pp. 112-115.
- [13] M. N. K. V. G. Soini, J. Rabaey, J. M. Sydanheimo, L. T., "Beyond Sensor Networks: ZUMA Middleware," in *Wireless Communications*

- and Networking Conference, 2007.WCNC 2007. IEEE, 2007, pp. 4318-4323.*
- [14] L. G. T. Stathopoulos, J. Heidemann, and D. Estrin, "Mote herding for tiered wireless sensor networks," in *proceeding of Technical Report 58, CENS*, 2005, Dec. 7.
- [15] M. e. B. Stefano Tennina, Pedro Braga, Ricardo Gomes, Mario Alves, Farrukh Mirza, Vincenzo Ciriello, Gabriella Carrozza, Pedro Oliveira, Vinny Cahill, "EMMON: A WSN System Architecture for Large Scale and Dense Real-Time Embedded Monitoring," presented at the Ninth IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, 2011.

GaAs, InP, InGaAs, GaInP, p+-i-n+ Multiplication measurements for Modeling of Semiconductor as photo detectors

Sanjay.C.Patil

(Research Scholar at NMIMS MUMBAI)
Parshvanath College of Engineering, THANE (W),
Mumbai, 400601 India

B.K.Mishra

Thakur College of Engineering and Technology,
Kandivali (E) Mumbai, 400101 India

Abstract—Optoelectronic is one of the thrust areas for the recent research activity. One of the key components of the optoelectronic family is photo detector to be widely used in broadband communication, optical computing, optical transformer, optical control etc. Present paper includes the investigation, carried on the basis of the. Multiplication measurements on GaAs, InP, InGaAs, GaInP, p+-i-n+s with – region thicknesses, with investigation of applicability of the local ionization theory. A local ionization coefficient to be increasingly unrepresentative of the position dependent values in the device as is reduced below 1 μm .

Keywords— Photo detectors; Impact ionization.

I. INTRODUCTION

With the advent of broadband communication, INTERNET communication, optical computing, the research in the optical devices has emerged as one of the thrust area for the active and consistent research. One of such primary devices is the photo detector. Therefore, recent years have seen tremendous research activities in the area of photo detector mathematical modeling and simulation. It includes the domain like optical control of r.f./microwave devices, broadband communication like optical communication, as an optoelectronic component to support green technology, etc.

Basic objective of the active research includes the reduction in the device dimension, reduction in power utilization, better performance against noise, ease of integration to make complete system on chip, better reliability against all possible adverse condition and so on. At present the commercially available photo detectors for the various includes PIN diode, APD, photo transistor and the recent device like PIN FET. A photo detector operates by converting light signals that hit the junction to a voltage or current. The junction uses an illumination window with an anti-reflect coating to absorb the light photons. The result of the absorption of photons is the creation of electron-hole pairs in the depletion region. Impact Ionization is the process in a material by which one energetic charge carrier can lose energy by the creation of other charge carriers. For example, in semiconductors, an electron (or hole) with enough kinetic energy can knock a bound electron out of its bound state (in the valence band) and promote it to a state in the conduction band, creating an electron-hole pair. Three-five materials refer to compound semiconductors made from one element from

Group III on the periodic chart (arsenic in the case of GaAs) and one from Group V (gallium in the case of GaAs). Other three-five (or III-V in Roman numerals) semiconductors include indium phosphide and gallium nitride.

The band gap energy of semiconductors tends to decrease with increasing temperature. When temperature increases, the amplitude of atomic vibrations increase, leading to larger inters atomic spacing. In a regular semiconductor crystal, the band gap is fixed owing to continuous energy states. In a quantum dot crystal, the band gap is size dependent and can be altered to produce a range of energies between the valence band and conduction band. It is also known as quantum confinement effect.

Band gaps also depend on pressure. Band gaps can be either direct or indirect, depending on the electronic band structure.

Gallium arsenide (GaAs)

A compound of the elements gallium and arsenic it is an III/V semiconductor, and is used in the manufacture of devices such as microwave frequency integrated circuits, monolithic microwave integrated circuits, infrared light-emitting diodes, laser diodes, solar cells, and optical windows.

Indium phosphide (InP)

A binary semiconductor composed of indium and phosphorus. It has a face-centered cubic crystal structure, identical to that of GaAs and most of the III-V semiconductors. InP is used in high-power and high frequency electronics because of its superior electron velocity with respect to the more common semiconductors silicon and gallium arsenide. It also has a direct bandgap, making it useful for optoelectronics devices like laser diodes.

Indium gallium arsenide (InGaAs)

Indium gallium arsenide (InGaAs) is a semiconductor composed of indium, gallium and arsenic. It is used in high-power and high-frequency electronics because of its superior electron velocity with respect to the more common semiconductors silicon and gallium arsenide. InGaAs band gap also makes it the detector material of choice in optical fiber communication at 1300 and 1550 nm. Gallium indium arsenide (GaInAs) is an alternative name for InGaAs.

Gallium indium Phosphide (GaInP)

GaInP, a wide band gap semiconductor lattice matched to GaAs, is of interest for a variety of device applications such as heterojunction bipolar transistors (HBTs) And heterojunction field-effect transistors (HFETs) The advantages of GaInP over GaAlAs for GaAs-based HBT applications include its large energy band gap (1.9 eV), lower conduction band offset, reduced deep-level concentration, and easier selective etching. GaInP-emitter HBTs with high current gain and good microwave performance have been reported

II. THEORY

To design structures for high-voltage or high-power applications, an accurate knowledge of impact ionization coefficients in GaInP is necessary for calculating breakdown characteristics of junctions. In the work reported here, photocurrent multiplication was used to measure the electron and hole ionization coefficients in ~100 GaInP by illuminating $p^+ - n^- - n^+$ diode structures from either side with above band gap radiation.6 The results show that GaInP has significantly lower values of α and β than those of GaAs or InP, a promising indication for high-voltage applications.

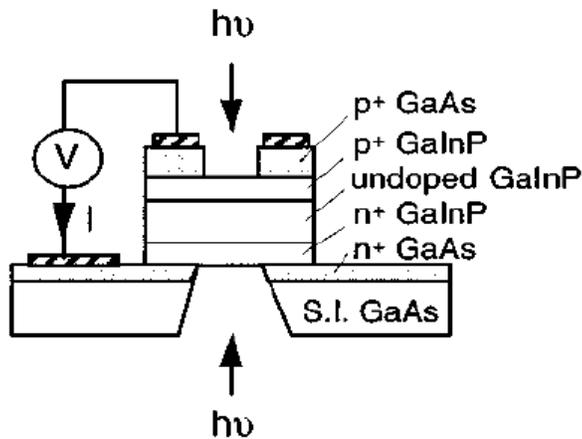


Fig.1. Basic structure of photo detector

Accurate determination of the electron and hole ionization coefficients, α and β respectively, is important, since these are used to determine avalanche multiplication characteristics and breakdown. Conventionally α and β are assumed to depend only on the local electric field, and the mean multiplication due to an electron-hole pair generated at position is given by

$$M(x_0) = \frac{\exp \left[- \int_{x_0}^W (\alpha - \beta) dx \right]}{1 - \int_0^W \alpha \exp \left[- \int_0^x (\alpha - \beta) dx \right] dx}$$

As described by Stillman and Wolfe [7] where W is the total depletion width. The electric field exists between $x=0$ and $x=W$ causing electrons to move from left to right. This expression is also traditionally used to derive the values of, α

and β from photo multiplication measurements performed with carrier injection from the depletion region edges. [7]-[8]

For electron multiplication, $M_e, x_0 = 0$ and for pure hole multiplication $M_h, x_0 = W$. However, carriers entering the high field region with energy much less than the ionization threshold must traverse a dead space, distance d_e , for electrons or d_h for holes, before they acquire sufficient energy to impact ionize. With estimation this dead space distance is given by equating it to the ballistic distance a carrier requires to reach the ionization threshold energy,

$$E_{th} \text{ i.e. } d = E_{th} / qF .$$

For a device where electrons are injected at $x = 0$, these corrections usually disallow electron ionization in a dead-space region from $0 < x < d_e$ and hole ionization in the region from $W - d_h < x < W$. Okuto and Crowell [9] presented an approximate expression relating multiplication to the ionization coefficients while accounting for the reduced multiplication by these regions. Bulman et al[10]. simplified their expression to interpret the measured multiplication results from p-n-n junctions by assuming no electron-initiated ionization occurs within a distance d_e from their injection point.

Physics of Impact Ionization

Avalanche photo detectors (APDs) use the impact ionization principle to detect and amplify very low light signals and they are still used as the most sensitive detector for most systems. It is found that to have a high-gain, low-noise APD, the ionization rate for electrons and holes needs to be very different. Efforts to find the best material to perform impact ionization reveal that silicon has the largest ionization rate difference between electrons and holes among all semiconductors. Due to its cut-off wavelength around 1 μm , however, silicon has rarely been employed for today's optical communication systems that use 1.55 μm wavelengths light. Suitable materials using combinations of binary and ternary/quaternary III-V semiconductors such as InGaAs on InP substrates have provided solutions for longer wavelength detection.

Impact Ionization Gain Mechanism

The impact ionization gain mechanism can be demonstrated by using a thermally generated electron, or an electron created by an absorbed photon, travelling inside a semiconductor where a depletion region is formed. Figure 6 illustrates the impact ionization process in a reverse biased PIN. As shown in the figure, electrons can gain sufficient kinetic energy while travelling in a high electric field. If the electric field is high enough, this high-energy electron may initiate the electron-electron scattering so that an electron in the valence band can be excited to the conduction band. As a result of this, another electron-hole pair is produced by promoting an electron from the valence band into the conduction band. Due to the strong electric field, the subsequent electron and hole will continue to collide with the

lattice and create more electron-hole pairs. Therefore, numerous carriers are generated and the result is a multiplied current output. This phenomenon is sometimes referred to as the avalanche breakdown

Ionization Threshold Energy

The minimum energy required to excite an electron, however, is more than just freeing an electron from the valence band. In order to decide the ionization threshold energy, various methods including from the valence to the conduction band is equal to the band gap energy of the semiconductor. Impact ionization parabolic, non parabolic, realistic, and non local pseudo potential band structure have been utilized. It is found that the threshold energy differs for most semiconductors. It is a function of the band structure, effective mass ratio between electron and hole, density of state, phonon interaction, and spin-orbit splitting energy.

The best way to estimate the ionization threshold energy involves a simple two parabolic band model. In this model, we consider one conduction band with effective mass m_e and one valence band with mass m_h . As shown in Fig. 6, prior to the collision, the electron travelling from the left-hand side has a kinetic energy of $\frac{1}{2} m_e v_i^2$ and a momentum of $m_e v_i$, where v_i is the initial velocity of the electron. After collision, three carriers exist: a new electron-hole pair plus the original electron. Electrons continue travelling to the right and the hole to the left.

Assuming that the collision is elastic, the conservation of energy and momentum must be satisfied these two assumptions are summarized as

$$\frac{1}{2} m_e v_i^2 = E_g + \left(\frac{1}{2} m_e v_e^2 \times 2 + \frac{1}{2} m_h v_h^2 \right) \quad (1)$$

And

$$m_e v_i = m_e v_e \times m_h v_h \quad (2)$$

Where E_g is the energy band gap of the semiconductor, v_e is the electron velocity, and v_h is the hole velocity after the collision. Note that when $m_e = m_h$ and $v_e = v_h = v_i$, (1) and (2) can be derive that the required initial electron energy for ionization process as

$$E_i = \frac{1}{2} m_e v_i^2 = 1.5 E_g \quad (3)$$

This is the well-known 3/2-band-gap rule for ionization threshold energy in semiconductor [1].

Ionization Coefficients and Gain Equations

To derive the gain equations for impact ionization process, we need to first define the impact ionization coefficient. For most semiconductor materials, the impact ionization process is

asymmetric to some degree for electrons and holes. For example, the ionization rate for electrons is about 5 times greater than holes in Germanium. For silicon, however, electrons can ionize 50 times to 1000 times more readily than holes depending on the electric field amplitude. The probability for initiating impact ionization is quantified as the impact ionization coefficient. It is defined as the reciprocal of the average distance travelled by an electron or hole to produce an electron-hole pair. Therefore, its unit is cm^{-1} . For electrons, the coefficient is denoted as α . For holes, it is denoted as β . Materials with very different value for α and β , such as in silicon can create less impact ionization noise because only one type of carrier is dominant during the ionization process. First, the length of the semiconductor is long ($L > 1 \mu m$) such that non-local theories [4] are not considered. Secondly, the current density is low to avoid the space charge effect which may screen the electric field and lower the gain. A differential equation can be used to describe the multiplication process in terms of ionization rates α and β . If β is much smaller than α , (4) can be rewritten as

$$M_n = \frac{1}{1 + \frac{\alpha}{\alpha - \beta} [\exp(-x(\alpha - \beta))]_0^L} = \frac{(\alpha - \beta) \exp(L(\alpha - \beta))}{\alpha - \beta \exp(L(\alpha - \beta))} \quad (4)$$

$$M_p = \frac{1}{1 + \frac{\beta}{\beta - \alpha} [\exp(-x(\beta - \alpha))]_0^L} = \frac{(\beta - \alpha) \exp(L(\beta - \alpha))}{\beta - \alpha \exp(L(\beta - \alpha))} \quad (5)$$

$$M_n \approx \frac{\exp(L\alpha)}{1 - \frac{\beta}{\alpha} \exp(L\alpha)} \quad (6)$$

A positive feedback factor (β/α) in (6) shows the effect of ionization coefficients on multiplication gain. As β gets bigger, but still much smaller than α , M_n can reaches breakdown in a shorter distance.

If we consider an extreme case, $\beta = 0$, (2.6) can be further simplified to

$$M_n = \exp\left(\int_0^L \alpha dx\right) = \exp(\alpha L) \quad (7)$$

From (7) we observe that when $\beta = 0$, there is no avalanche breakdown because M_n just continues to increase exponentially with αL . A plot is given in Fig.2 using (6) to demonstrate the gain versus αL for various value of (α/β)

Impact Ionization Coefficient Measurement

We know that α and β are strong function of electric field in the multiplication region. Electric field, however, also depends on the bias voltage, doping profile, and the device geometry. In this section, we will explain how to measure the ionization coefficients accurately for different device structures including p-i-n diodes.

In measurement, a very high energy light is used to illuminate the diode. Due to the high energy of the photons, electron-hole pairs can be created very close to the surface to obtain the pure electron or hole injection condition. Electron ionization is measured by shining light from the p+ side. Photon-excited electrons are injected into the intrinsic layer while holes are swept to the left. Hole ionization is measured by illuminating light from the n+ side, injecting holes into the intrinsic layer.

The electron gain M_n can be calculated by dividing the total current measured for a given electric field by the photocurrent created by the light. The photocurrent can be precisely measured when the diode is operated without avalanche gain. The hole gain M_p can be obtained through the same process

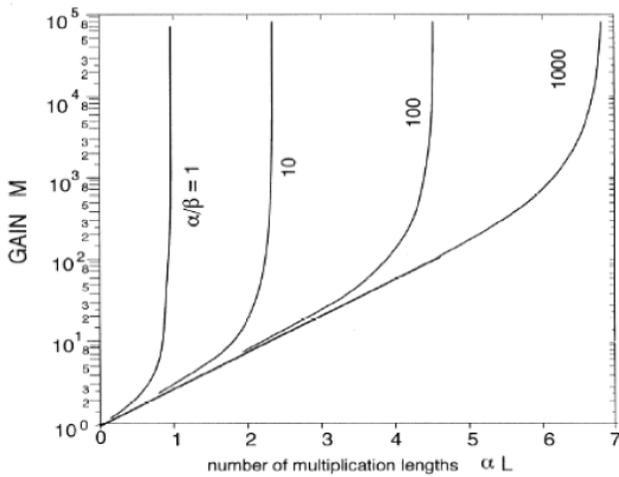


Fig.2. Multiplication gain M versus αL for pure electron injection. Various α/β value are used to demonstrate its effect on avalanche breakdown curve

Once M_n and M_p are known, we can use equation (4) and (5) to derive formulas related α and β to M_n and M_p . They are calculated to be

$$\alpha = \frac{1}{L} \left(\frac{M_n - 1}{M_n - M_p} \right) \ln \left(\frac{M_n}{M_p} \right) \quad (8)$$

And

$$\beta = \frac{1}{L} \left(\frac{M_p - 1}{M_p - M_n} \right) \ln \left(\frac{M_p}{M_n} \right) \quad (9)$$

Impact Ionization Response

The carrier build-up time in the multiplication process depends on the contribution of carrier feedback. In the ideal condition, where only one type of carrier is capable of initiating impact ionization process (electrons for $\beta = 0$), the output current pulse increases with the transit time for the

initially injected electron. The output current decreases to zero as all the ionized holes arrive at the negatively biased electrode. Thus, the current pulse lasts about twice as long as the transit time. Since the pulse width is independent of the multiplication gain, there is no gain-bandwidth product limitation when β or $\alpha = 0$.

It is found that the avalanche multiplication process does not affect the device bandwidth as long as the dc multiplication gain M is less than α/β . On the other hand, if $M > \alpha/\beta$, the multiplication gain becomes a function of frequency and is expressed as

$$M(\omega) = M_0 / (\omega^2 M_0^2 \tau_1^2)^{1/2} \quad (10)$$

Where τ_1 is an effective transit time and is approximately $\tau_1 = N(\beta/\alpha)\tau$. N is a number varying slowly from 1/3 to 2 as β/α varying from 1 to 10^{-3} , and τ is the transit time equal to L/v_s where L is the length of the avalanche gain region and v_s is the saturation velocity

The gain-bandwidth product for $M > \alpha/\beta$ can be obtained using (10) for high frequencies and is expressed as

$$M(\omega) \times \omega = \frac{M_0 \omega}{(\omega^2 M_0^2 \tau_1^2)^{1/2}} = \frac{1}{\tau_1} = \frac{1}{N(\beta/\alpha)\tau} \quad (11)$$

Equation (11) indicates the basic requirements for an impact-ionization based device to obtain a high gain-bandwidth product. These requirements include a small β/α value and short intrinsic time. Therefore, a correct choice of material, multiplication layer thickness, and carrier transport velocity are essential.

Using (10), the calculated frequency response for a P-I-N structure with $1\mu m$ thick gain region when operating at $M=50$ is given in Figure 3.

Using Matlab to simulate the both the measure of α and β since Bulmanet al [10]. found that $\beta > \alpha$, these two sets of parameters for the coefficients since they enable the data to be more accurately quantified over the wide field range the coefficients are parameterized in The impact ionization coefficients can be fitted into an exponential form:

$$\alpha(E) = A_e \exp(-B_e / E)$$

$$\beta(E) = A_h \exp(-B_h / E)$$

$$\alpha_{eH} = A_{eH} \exp[-(B_{eH} / \xi)^{C_{eH}}]$$

$$\beta_h = A \exp[-(B / \xi)^C]$$

For the local calculation the values of α and β were taken from [10] for electric fields, F . This investigation involved measuring both the multiplication and excess noise

characteristics of several samples with overlapping field regions and represents the most extensive and rigorous to date

Impact Ionisation coefficient and Multiplication of electrons & holes for GaAs

Impact Ionisation Coefficients for GaAs

$$\alpha(E) = 1.89 \times 10^7 [(\exp(-5.75 \times 10^7) / E)^{1.82}] \text{cm}^{-1}$$

$$\beta(E) = 2.21 \times 10^7 [(\exp(-6.57 \times 10^7) / E)^{1.75}] \text{cm}^{-1}$$

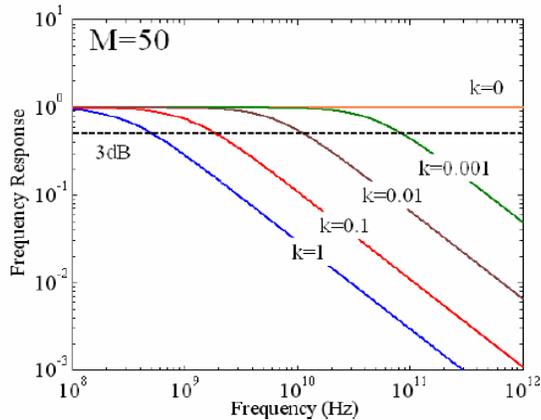


Fig.3. Frequency Response of a P-I-N photodiode with $1 \mu\text{m}$ thick multiplication layer. Notice that parameter k is define as β/α

Multiplication of electrons and holes for GaAs

Simulation of calculate Electron and hole multiplication factors for a range of ideal p-i-n s with from 1 um down to 0.48um for GaAs with temperatures of 20K to 500K, Shown in figure 5 increasing temperature causes the multiplication of electrons and holes to shift to higher voltage therefore higher electric fields are required to offset the increase in carrier cooling by phonon scattering and maintain multiplication

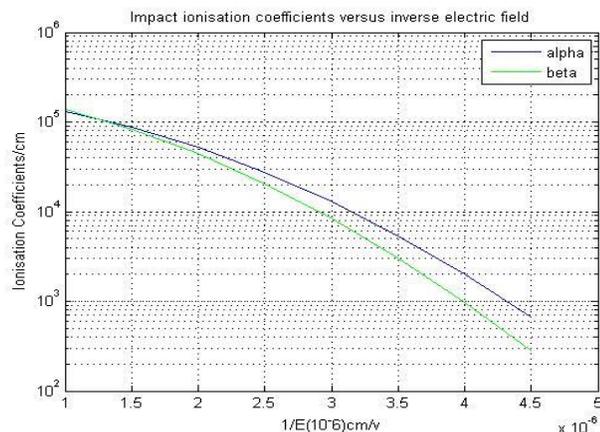


Fig.4. Impact Ionisation Coefficients for GaAs versus inverses electric field

Impact Ionisation coefficient and Multiplication of electrons & holes for InGaAs

$$\alpha_{eH} = A_{eH} \exp[-(B_{eh} / \xi)^{C_{eH}}]$$

$$\beta_h = A \exp[-(B / \xi)^C]$$

Impact Ionisation Coefficients for InGaAs (narrowband gap semiconductor) great importance in the case of an ionization coefficient which increases with temperature that can result in an unstable positive power dissipation feedback which the voltage breakdown can be obtained by extrapolating the multiplication curveAs shown in fig 6

For alpha

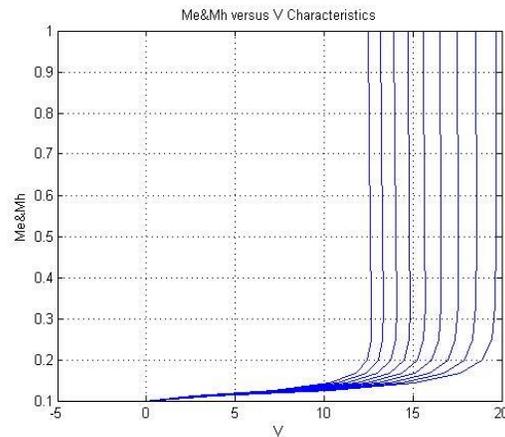


Fig.5. simulated of multiplication $w = 0.48 \mu\text{m}$ device as a function of temperature from 20K to 500K

$$A_{eH} = 7.2597 \times 10^4 - (24.204T) + (0.3259T^2) \text{cm}^{-1}$$

$$B_{eh} = 5.9988 \times 10^5 + (3.4763 \times 10^2 T) + (2.4768T^2) \text{V} / \text{cm}$$

$$C_{eH} = 1.783 + (7.2548 \times 10^{-4} T)$$

For beta

$$A = 6.1026 \times 10^5 + (9.6637 \times 10^2 T) + (1.1384T^2) \text{cm}^{-1}$$

$$B = 1.3394 \times 10^6 + (1.0699 \times 10^3 T) + (20.4507T^2) \text{V} / \text{cm}$$

$$C = 1.0910 - (2.3505 \times 10^{-4} T)$$

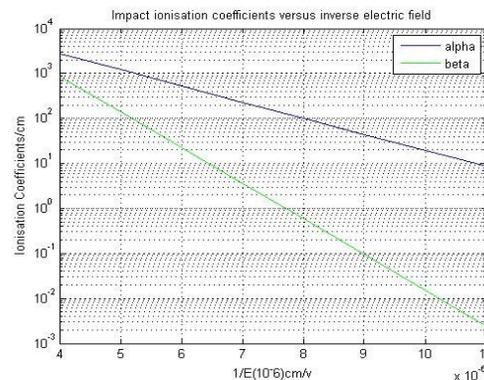


Fig.6. Impact Ionisation Coefficients for InGaAs versus inverses electric field

Multiplication of electrons and holes for InGaAs

Simulation of Electron multiplication factors that depends on temperature with the thickness of 1.3um and 1.9 um for InGaAs p-i-ns from 20K to 400K and hole multiplication with thickness of 3.0um as n-i-p from 20K-300K.

The results from figure 7 show a very limited increase in photocurrent initially and then the sudden and clear onset of the avalanche multiplication process. T

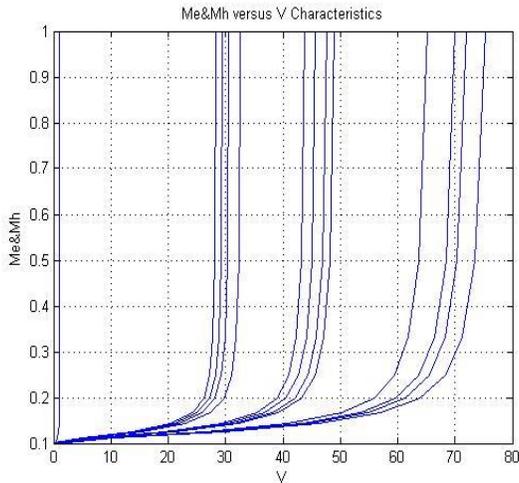


Fig.7. Simulated of multiplication characteristics of 1.3 and 1.9 um thick InGaAs p-i-n diode from 20–400 K

Impact Ionisation coefficient and Multiplication of electrons & holes for InP

The electron and hole ionization coefficients can be extracted from the measured multiplication results if both electrons initiated and hole initiated multiplication results are available for the same structure. When an electron (or hole) initiates the multiplication process, an electric current is induced by the moving electrons and holes within the multiplication region.

InP has high electron peak velocities resulting from large inter valley separation and good breakdown properties owed to a relatively low electron impact ionization coefficient.

Compared to the data of GaAs and InP illustrated in Fig.6 and Fig. 8, GaInP has the lowest values of α and β , signifying a higher breakdown voltage. Moreover, the slopes of the curves of α and β vs $1/E$ for GaInP shown in Fig. 7 are steeper than for the others.

Impact Ionisation Coefficients for InP

$$\alpha(E) = 2.93 \times 10^6 [\exp(-(2.64 \times 10^6) / E)] \text{cm}^{-1}$$

$$\beta(E) = 1.62 \times 10^6 [\exp(-(2.11 \times 10^6) / E)] \text{cm}^{-1}$$

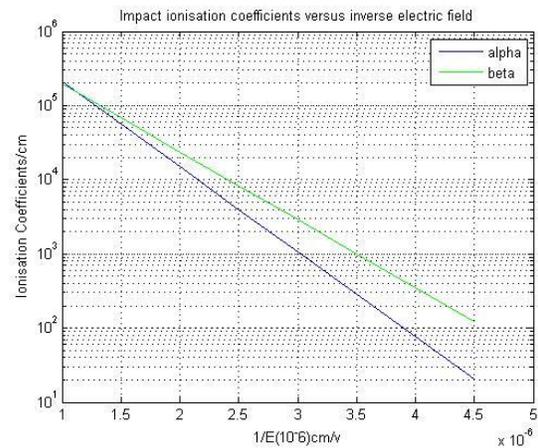


Fig.8. Impact Ionisation Coefficients for InP versus inverses electric field

Multiplication of electrons and holes for InP

The results from figure 9 show that the temperature dependence of multiplication is not uniform, since the multiplication curves with thickness of the materials from 0.24 to 2.40um respectively. Simulation of Electron multiplication factors for a range of ideal p-i-ns with from 0.24um down to 2.40um for InP with temperatures of 20K to 500K.

As a test of the parameterized coefficients, test shows for all the materials, the predicted multiplication for an ideal diode, compared with the measured characteristic. It can be seen that the agreement is good, both at low and high multiplications using matlab for these simulations for wideband gap semiconductor and narrowband gap semiconductor.

The structures for high-voltage or high-power applications, an accurate knowledge of impact ionization coefficients in GaInP is necessary for calculating breakdown characteristics of junctions.

In the work reported here, photocurrent multiplication was used to measure the electron and hole ionization coefficients from either side with above bandgap radiation.

The results show that GaInP has significantly lower values of α and β than those of GaAs or InP, a promising indication for high-voltage applications.

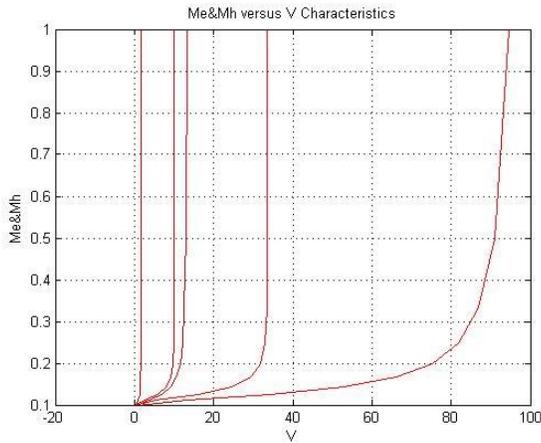


Fig.9. Simulated of multiplication characteristics of 0.24 and 2.40 um thick for InP p-i-n diode from 20-400 K

Impact Ionisation coefficient and Multiplication of electrons & holes for GaInP

Impact Ionisation Coefficients for GaInP(wideband gap semiconductor)

$$\alpha(E) = 3.85 \times 10^6 [\exp(-3.17 \times 10^6 / E)] \text{cm}^{-1}$$

$$\beta(E) = 1.71 \times 10^6 [\exp(-3.19 \times 10^6 / E)] \text{cm}^{-1}$$

The electron and hole multiplication factors, Mn and Mp, are defined as the total output photocurrent divided by the electron or hole current injected at the contacts. Results for Mn and Mp from respective measurements are shown in Fig 10.

In these results, slight corrections were made to the injected currents due to slightly voltage-dependent intrinsic reverse currents. Because the widths of the depletion regions depend on bias, the number of thermally generated minority carriers that diffuse into the depleted n^- region also varies. This leads to a slight increase of the injected current as the reverse bias is increased

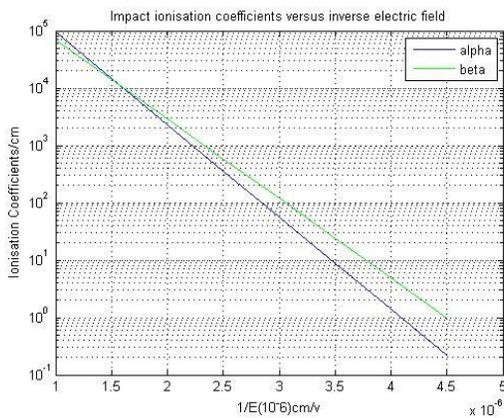


Fig.10. Impact Ionisation Coefficients for GaInP versus inverse electric field

Multiplication of electrons and holes for GaInP

Simulation of Electron multiplication factors for a range of ideal p-i-n s with from 0.24um down to 2.40um for GaInP with temperatures of 300K

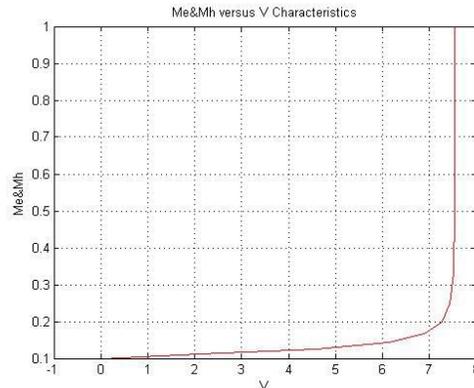


Fig.11. Simulated of multiplication characteristics of 0.24 um thick for GaInP p-i-n diode from 300 K

III. RESULT AND DISCUSSION

Voltage breakdown due to avalanche multiplication, also defined as the reverse bias voltage where multiplication rate goes to infinity naturally of great practical interest.

The voltage VB is defined as the reverse-bias voltage across the multiplication region at which the mean gain becomes infinite.

In doing so rather than testing with hands on materials in fab or clean room physically testing the materials, with this simulation we can create a model and simulate any materials first with respect to their voltage breakdown versus thickness (width) without wasting unnecessary time and cost for testing of materials

Case 1

GaAs wide band gap, highly resistive which makes it a very good electrical substrate therefore makes it a very good material for ideal material for microwave and millimeter wave integrated circuits. It can also be operated at higher power levels because they have higher breakdown voltage.

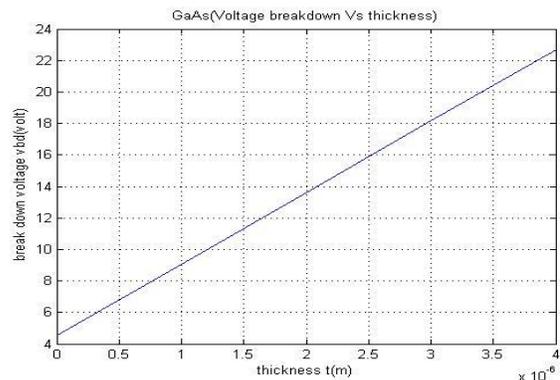


Fig.12. Simulated of voltage breakdown versus thickness for GaAs

Case 2

GaInP wideband gap semiconductor exhibits high breakdown voltage characteristics indicating that it is a good choice of material for high power applications. This suggests that the ideal electric field distribution assumption is valid and that edge effects are not important in this material system for mesa geometry structures. Note that since these punches through devices have a low carrier concentration in the n^- region and high breakdown voltage, the effects of dead space could be ignored. α And β data were used to calculate the expected breakdown voltages for p-i-n diode with various thicknesses

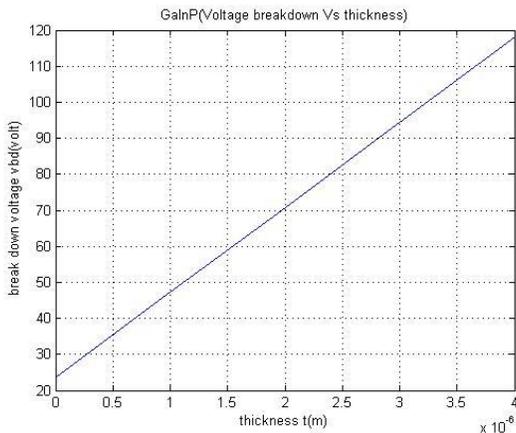


Fig.12. Simulated of voltage breakdown versus thickness for GaInP

GaInP wideband gap semiconductor exhibits high breakdown voltage characteristics indicating that it is a good choice of material for high power applications. This suggests that the ideal electric field distribution assumption is valid and that edge effects are not important in this material system for mesa geometry structures. Note that since these punches through devices have a low carrier concentration in the n^- region and high breakdown voltage, the effects of dead space could be ignored. α And β data were used to calculate the expected breakdown voltages for p-i-n diode with various thicknesses

Case 3

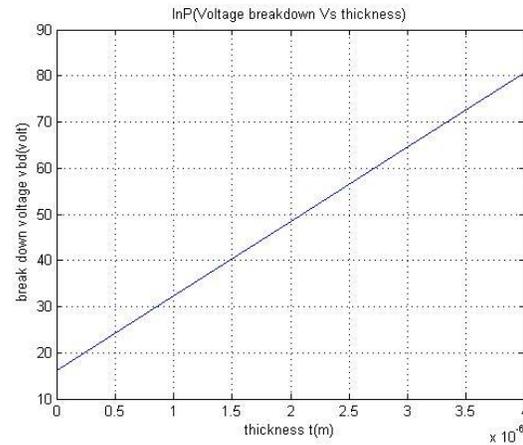


Fig.13. Simulated of voltage breakdown versus thickness for InP

InP wideband gap have an advantage compared to GaAs for many applications when used in high-field regions of the device profile, can significantly improve device performance. These applications include high performance power amplifier for cellular phones, ultra-efficient ultra-linear power amplifiers ideally suited for digital communication systems and satellite networks ICs, and highly integrated mixed signal and high-speed fiber-optic circuits. Wide bandgap semiconductors are associated with a high breakdown voltage. This is due to a larger electric field required to generate carriers through impact mechanism.

Case 4

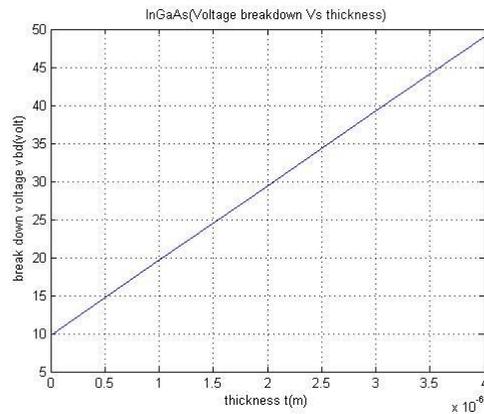


Fig.14. Simulated of voltage breakdown versus thickness for InGaAs

InGaAs with narrow band gap is also a popular material in infrared detectors and some short wave infrared cameras. It also has lower multiplication noise than germanium when used as the active multiplication layer of an avalanche photodiode. Also has a low breakdown voltage

IV. CONCLUSION

For this research work materials are simulated base on p-i-n diodes therefore very little studies on Avalanche Photodiode (APD) and Schottky Barrier diodes.

The carriers are absorbed in a π^- region. The absorption leads to the generation of electron-hole pairs in this region.

The electric field in the π -region is high enough to separate the carriers. The electric field across the π -region is not high enough for the charge carriers to gain enough energy for multiplication to take place.

REFERENCES

- [1] [1] B.K.Mishra, computer Aided modeling of solid state photodetectors Ph.D thesis Birla Institute of Technology, Mesra, Ranchi 1995
- [2] T. P. Pearsall, "Impact ionization rates for electrons and holes in Ga_{0.47}In_{0.53}As," *Appl. Phys. Lett.* 36, 218-220 (1980)
- [3] S. Wang, R. Sidhu, X. G. Zheng, X. Li, X. Sun, A. L. Holmes, Jr., and J. C. Campbell, *IEEE Photonics Technol. Lett.* 13, 1346 (2001).
- [4] B.K.Mishra, Lochan Jolly, S.C.Patil, "In_{1-x}Ga_xAs a next generation material for photodetectors", JSAM.2011
- [5] R.Poerschke, Springer-Verlag, Madelung, O.(ed.), *Semiconductor: group IV elements and III-V compound. Series "Data in science and technology"*, Berlin, 164 (1991).
- [6] J. R. Chelikowsky, and M. L. Cohen, *Phys. Rev. B* 14, 556 (1976).
- [7] R. J. McIntyre, "A New Look at Impact Ionization – Part I: A Theory of Gain, Noise, Breakdown Probability, and Frequency Response," *IEEE Trans. Electron Devices*, vol. 46, 1623–1631 (1999).
- [8] H. Ando and H. Kanbe, "Ionization coefficient measurement in GaAs by using multiplication noise characteristics," *Solid-State Electron.*, vol. 24 pp. 629–634, 1981.
- [9] O.Konstantinov, Q.Wahab, N. Nordell, and U. Lindefelt, "Ionization rates and critical fields in 4H silicon carbide," *Appl. Phys. Lett.*, vol. 71, July 1997
- [10] Y.Okuto and C. R. Crowell, "Ionization coefficients in semiconductors," *Phys. Rev. B.*, vol. 10, pp. 4284–4296, Nov. 1973.
- [11] G. E. Bulman, V. M. Robbins, and G. E. Stillman, "The determination of impact ionization coefficients in (100) gallium arsenide using avalanche noise and photocurrent multiplication measurements," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 2454–2466, Nov. 1985.
- [12] M. M. Hayat, B. E. A. Saleh, and M. C. Teich, "Effect of dead space on gain and noise of double-carrier-multiplication avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. 39, pp. 546–552, Mar. 1992.
- [13] M. M. Hayat, W. L. Sargeant, and B. E. A. Saleh, "Effect of dead space on gain and noise in Si and GaAs avalanche photodiodes," *IEEE J. Quantum Electron.*, vol. 28, pp. 1360–1365, May 1992.
- [14] A. Di Carlo and P. Lugli, "Dead-space effects under near breakdown conditions in AlGaAs/GaAs HBT's," *IEEE Electron Device Lett.*, vol. 14, pp. 103–105, Mar. 1993.
- [15] S. P. Wilson, S. Brand, and R. A. Abram, "Avalanche multiplication properties of GaAs calculated from spatially transient ionization coefficients," *Solid-State Electron.*, vol. 38, pp. 2095–2100, Nov. 1995.
- [16] A. Di Carlo and P. Lugli, "Dead-space effects under near breakdown conditions in AlGaAs/GaAs HBT's," *IEEE Electron Device Lett.*, vol. 14, pp. 103–105, Mar. 1993.
- [17] S. P. Wilson, S. Brand, and R. A. Abram, "Avalanche multiplication properties of GaAs calculated from spatially transient ionization coefficients," *Solid-State Electron.*, vol. 38, pp. 2095–2100, Nov. 1995.
- [18] *IEEE TRANSACTIONS ON ELECTRON DEVICES*, VOL. 50, NO. 10, OCTOBER 2003 2027
- [19] *IEEE TRANSACTIONS ON ELECTRON DEVICES*, VOL. 50, NO. 10, OCTOBER 2003 2021 Temperature Dependence of Breakdown and Avalanche Multiplication in In_{0.53}Ga_{0.47}As Diodes and Heterojunction Bipolar Transistors M. Yee, W. K. Ng, J.P. R. David, Senior Member, IEEE, P. A. Houston, C. H. Tan, and A. Krysa
- [20] Shanghai Institute of Technical Physics Theory Study of SAGCM InP L. Lin, W. J. Wang, N. Li, X. S. Chen and W. Lu Shanghai Institute of Technical Physics, Chinese Academy of Sciences National Lab for Infrared Physics,
- [21] Impact ionization coefficients in (100) GaInP S.-L. Fu, T. P. Chin, M. C. Ho, C. W. Tu, and P. M. Asbeck Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093

Performance of Turbo Product Code in Wimax

Trushita Chaware

Department of Information Technology
Thakur College of Engineering and Technology
Kandivali(E), Mumbai, India

Nileema Pathak

Computer Engineering Department
Atharva College of Engineering
Malad(W), Mumbai, India

Abstract— IEEE 802.16 is a standard for Broadband Wireless Access (BWA) air interface. 802.16e supports mobile broadband wireless access, which is an additional feature over its predecessors, which support fixed wireless access. Binary Convolutional Turbo Coding (CTC) is used as mandatory Forward Error Correction method in 802.16e. In this paper the performance of a simple and efficient optional coding scheme namely Turbo Product Code (TPC) is proposed for 802.16e system and is compared with CTC.

KEYWORDS— CTC; eBCH; AWGN; Code Rate; OFDM

I. INTRODUCTION

IEEE 802.16 standard system or commonly called Worldwide Interoperability for Microwave access (WiMAX), provides specifications for both fixed Line of Sight (LOS) communication in the range of 10-66 GHz (802.16c), and fixed, portable, Non-LOS communication in the range of 2-11GHz (802.16a, 802.16d). The IEEE 802.16e provides mobility and also enhanced performance. The architecture is based on scalable sub channel bandwidth using variable size FFT according to channel bandwidth. Forward Error correction is done by mandatory Convolution Codes. While doing channel coding, the limit on data rate is given by Shannon's limit is as $R < W \log_2 (1 + S/N)$ bits/sec, which sets a limit on bandwidth and signal to noise ratio. Efficient communication systems are systems that permit a high rate of information to be communicated with the lowest possible power and least BER.

IEEE 802.16 wireless broadband standard is one such promising future wireless system, primarily because it offers the potential for high spectral efficiency, flexible spectrum options (2-6 GHz), scalable carrier bandwidth options (1.25 MHz to 20 MHz), multiple duplexing options (time and frequency division duplex), various sub channelization options, and also mobility.

II. 802.16 SYSTEM

IEEE 802.16 standard for Broadband Wireless Access and its associated industry consortium, WiMAX forum promises to offer high data rates over large areas to a large number of users where broadband is unavailable.

Taking the advantage of OFDM technique the physical layer of Wimax system is able to provide robust broadband service. The basic principle of OFDM is to divide a high-rate data stream into N lower rate streams and to transmit them at the same time over a number of subcarriers. In OFDM high bit rate data is divided into N low bit rate parallel data streams

and then transmitted simultaneously with deferent frequencies. OFDM systems are implemented using a combination of fast Fourier Transform (FFT) and inverse fast Fourier Transform (IFFT) blocks. The effect of ISI on an OFDM signal can be further improved by the addition of a guard period to the start of each symbol. This guard period is a cyclic copy that extends the length of the symbol waveform [1]. The block diagram of a general Wimax system is shown in figure 1.

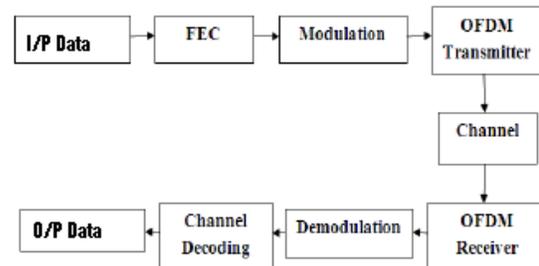


Fig.1. WiMAX System

The performance of this system can be further enhanced by using forward error correcting techniques. Commonly used error correcting method for a wireless medium for 802.16 standard is Convolutional Turbo Codes (CTC). Turbo Product Code (TPC) can be used instead of CTC and the performance of the wireless communication system can be analysed. TPC with eBCH as a constituent code provides further benefits, because eBCH code can be decoded easily using the syndrome method and it can be used for multiple random error correction. General points of comparison for CTC and TPC are that CTC perform best for low code rate applications while TPC perform best for high code rate applications. CTC will have difficulty achieving high data rates, whereas TPC can operate at high data rates. CTC exhibit error floor at BER below 10^{-5} , while TPC error floor is less pronounced at lower BER values [2].

III. CONVOLUTIONAL TURBO CODE

In the Wireless MAN system, the Convolutional Codes(CC) is the only mandatory coding scheme, all the others like CTC and TPC are optional. Convolutional codes map information to code bits sequentially by convolving a sequence of information bits with "generator" sequences. A convolutional encoder encodes K information bits to $N > K$ code bits at one time step. Maximum Likelihood decoding can be done using Viterbi algorithm, other decoding algorithms

such as SOVA (Soft Output Viterbi Algorithm) and BCJR (Bahl, Cocke, Jelinek and Raviv) algorithm, can be used [3].

The CTC is a parallel concatenated convolutional code or turbo code. An overview of the CTC encoder is depicted in figure. 2. It consists of a CTC encoder with a natural rate of 1/3 followed by an additional interleaver and a final puncturing to obtain the desired rate. The CTC encoder consists of two identical constituent encoders separated by a CTC internal interleaver. The constituent encoder has a natural rate of 2/4 and its minimal realization is depicted in the lower part of figure. 2. It consists of $m = 3$ memory elements and consumes two input bits per time instance and produces four output bits. Furthermore, the encoder is recursive and systematic. The CTC encoding procedure is described below. The two information bits A and B are fed directly to the output and in a first encoding step additionally into the constituent encoder, producing the parity bits Y1 and W1. Afterwards, in a second encoding step, the interleaved information symbols (A and B) are again fed into the constituent encoder, now producing the parity bits Y2 and W2. This means the info/code tuple of the CTC encoder is AB/ABY1W1Y2W2 and its natural rate is therefore 1/3. CTC is decoded using the well known iterative decoding process using Log Maximum A Posteriori (MAP) algorithm, where the two decoders exchange information based on log likelihood ratio of the information bits. The complexity of this decoding process increases exponentially as they get closer to optimality.

IV. TURBO PRODUCT CODE

A two-dimensional product code is built from two component codes with parameters C 1 (n_1, k_1, d_1) and C2 (n_2, k_2, d_2), where n_i, k_i, d_i stands for code word length, number of information bits, and minimum hamming distance respectively [3]. The product code $P = c_1 \times c_2$ is obtained by placing ($k_1 \times k_2$) information bits in an array of k_1 rows and k_2 columns. The parameters of product code P are $n = n_1 \times n_2, k = k_1 \times k_2, d = d_1 \times d_2$ and code rate is $R = R_1 \times R_2$, where R_i is the code rate of C_i .

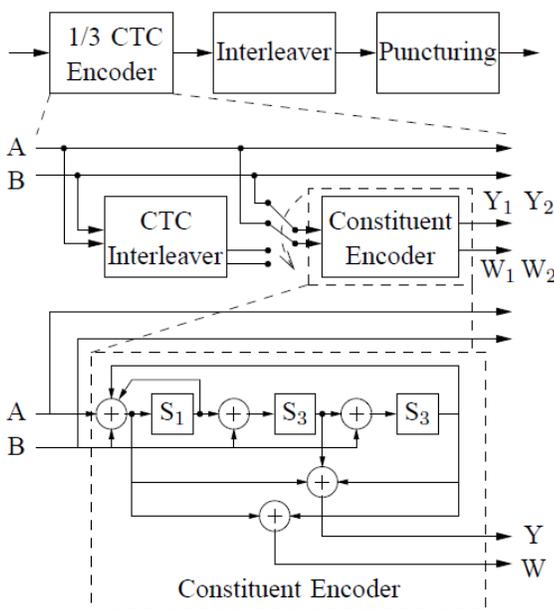


Fig.2. Convolutional Turbo Code Encoder

Thus very long block codes can be built with large minimum Hamming distance. Figure. 2 shows the procedure for construction of a 2D product code using two block codes C1 and C2. All the rows of matrix P are the code words of C1 and all the columns of matrix P are code words of C2 [4].

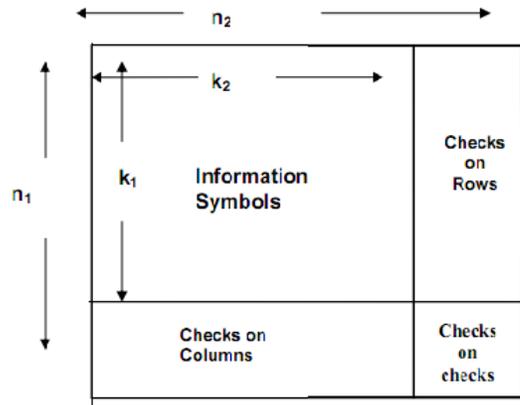


Fig.3. An example of a 2D product code constructed using two component codes

Other benefits in using TPC are as follows

- Longer battery life - Less transmit power required for the battery operated device using TPC.
- IP free encoder - No IP license required for encoding.
- Availability of IC's – Standard IC's are available providing the required functionality thus, the TPC implementation price increase is minimal.
- Lowest cost solution - A very low cost encoder, smaller batteries, smaller packages, it all adds up to reduced cost and improved performance [5].

A. DECODING OF TPC

a) Soft Decoding of Linear Block Codes

In hard decision decoding, received signal is sampled and the resulting voltages are compared with a single threshold. If a voltage is greater than the threshold it is considered to be definitely a 'one', regardless of how close it is to the threshold. If it is less, it is definitely 'zero'. In soft decision decoding we get not only the 1 or 0 decision but also an indication of how certain we are that the decision is correct. Few bits may be used as 'confidence' bits to indicate the certainty of soft decision. Turbo Product codes can be decoded by sequentially decoding the rows and columns of product code P, in order to reduce decoding complexity. However, to achieve optimum performance, one must use Maximum Likelihood Decoding (soft decoding) of the component codes. Thus, we need soft-input/ soft-output decoders to maintain optimum performance when decoding the rows and columns of product code P.

b) Chase Algorithm

The Trellis based Maximum A posteriori Probability (MAP) decoding (soft decoding), used for CTC decoding,

provides a good BER performance but it is very complex and computationally difficult.

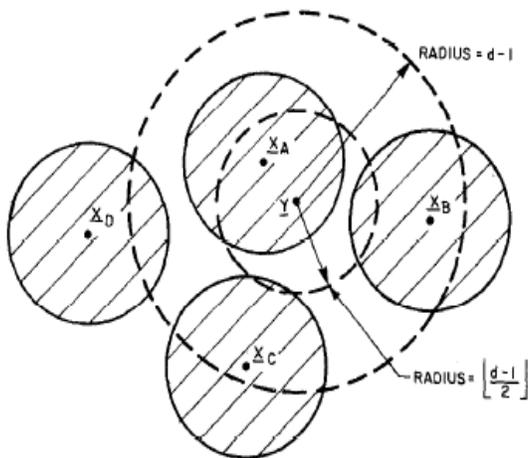


Fig.4. Geometric sketch for decoding with channel measurement information

Instead of using trellis based MAP algorithm, the Chase algorithm is repeatedly applied along rows and columns of TPC in order to obtain extrinsic information for each bit position. Chase algorithm is used to obtain soft output. Chase initially used a method of finding the Euclidean distance between code words (2^k) for filtering them for decoding. Each codeword is surrounded by a sphere of radius $(d - 1)/2$. Thus, an unique codeword, or equivalently a unique error pattern, is obtained by a binary decoder if the received sequence is within one of these spheres. In our case there is a unique error pattern $Z = Y \text{ XOR } X_A$ within the sphere of radius $(d - 1)/2$ which surrounds Y as shown in figure 4 [6].

c) Chase – Pyndiah Algorithm

This algorithm generates Test Patterns using least reliable bits p. These least reliable bits are found using a reliability sequence obtained by soft decoding of received signal. Test patterns for p=2 are shown in figure 5.

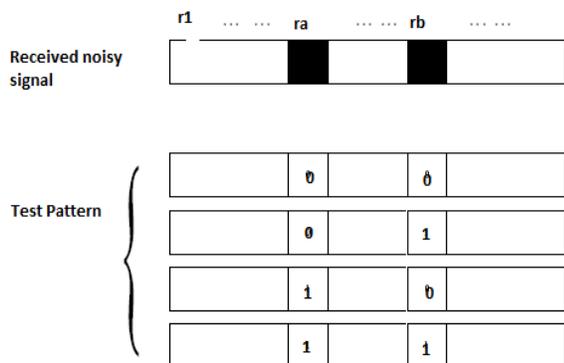


Fig.5. Test Pattern generated by Chase Algorithm, p=2

Steps for Chase Pyndiah Algorithm

1. Generate a reliability sequence $rabs = (|r1|, |r2|, \dots, |rn+1|)$ and a binary received sequence $y = (y1, \dots, yl, \dots, yn)$. Determine the p least reliable bit positions of sequence y using $rabs$.
2. Form 2^p test pattern sequences $t^j, j = 1, \dots, 2^p$ which consist of all combinations of binary sequences containing the p least reliable bits of y .
3. Determine the 2^p perturbed sequences $zj = y \oplus t^j, j = 1, 2, \dots, 2^p$.
4. Decode perturbed sequences zj and obtain valid codeword set $cj, where j = 1, 2, \dots, 2^p$
5. Calculate analog weight of valid codeword set.
6. Estimate the maximum likelihood (ML) codeword d from valid codeword set
7. Compute extrinsic information for received signal and is used to arrive at a solution i.e. the decoded message [4].

The decoding procedure described below is generalized by cascading elementary decoders illustrated in figure 6. Let us consider the decoding of the rows and columns of a product code P described above and transmitted on a Gaussian channel. On receiving matrix [R] corresponding to a transmitted codeword [E], the first decoder performs the soft decoding of the rows (or columns) of P using as input matrix [R]. Soft-input decoding is performed using the Chase algorithm (as given in above Section) and the soft output is computed [7].

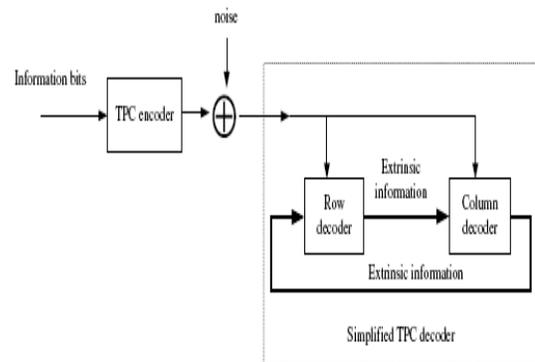


Fig.6. Iterative Decoding of TPC

V. APPLICATIONS OF TPC

The iterative decoding of product codes is also known as Block Turbo Code (BTC), because the concept is quite similar to turbo codes based on iterative decoding of concatenated recursive convolutional codes. TPC (eBCH (64, 51, 6) as constituent code with code rate of 0.635 and 6 iterations in an AWGN channel provides a BER of 10^{-6} at an $Eb/N0$ of 2db [4].

TPC are the most efficient known codes for high code rate applications. For code rates greater than 0.95, digital transmission systems can transmit data on a Gaussian channel

at more than 98% of channel capacity, $R/C > 0.98$, by guarantee of a minimum distance of 16 or more. While the minimum distance of a CTC can be relatively small. Because of high d_{min} , typically 16 or higher, there is no error floor for TPC.

Another attractive application for TPC concerns high data rate systems. Indeed, the decoding speed of a TPC can be increased by using several elementary decoders for the parallel decoding of the rows (or columns) of a product code since they are independent [10]. These features, of TPC can be effectively used in a high data rate application like the Wimax 802.16 systems.

VI. SIMULATION RESULT FOR TPC IN 802.16

In this paper, results of using TPC as the Forward Error Correction (FEC) method in 802.16e systems, is studied. TPC block is generated using eBCH (64, 39). Chase Pyndiah algorithm as explained above is applied for 10 iterations for decoding the TPC at the receiver.

TPC is implemented in a 128 point FFT OFDM system. Modulation method used is QPSK. The results, which are plotted using Monte Carlo simulation method for iteration 1, iteration 4, iteration 6, iteration 8 and iteration 10, are shown in Fig. 7.

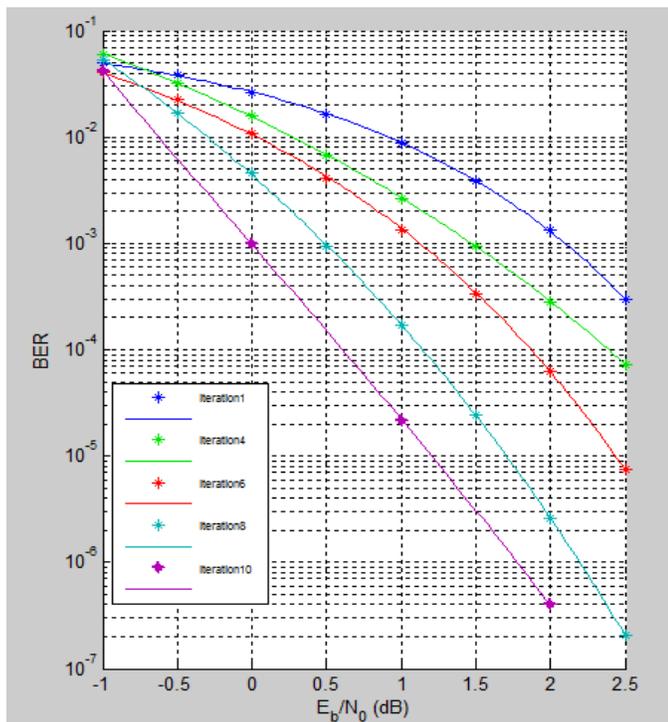


Fig.7. BER verses Eb/N0 for TPC in 802.16 systems

VII. COMPARISON OF PERFORMANCES

A. COMPARISON OF CTC AND TPC IN 802.16 SYSTEM

CTC can be used as a FEC method in 802.16e systems. The BER performance of the IEEE 802.16e with CTC decoder using rate 1/3 QPSK code versus the number of iterations is shown in Fig. 8. It is observed that the CTC exhibits a BER greater than 10^{-6} , but at E_b/N_0 ratio of 2.5db [7].

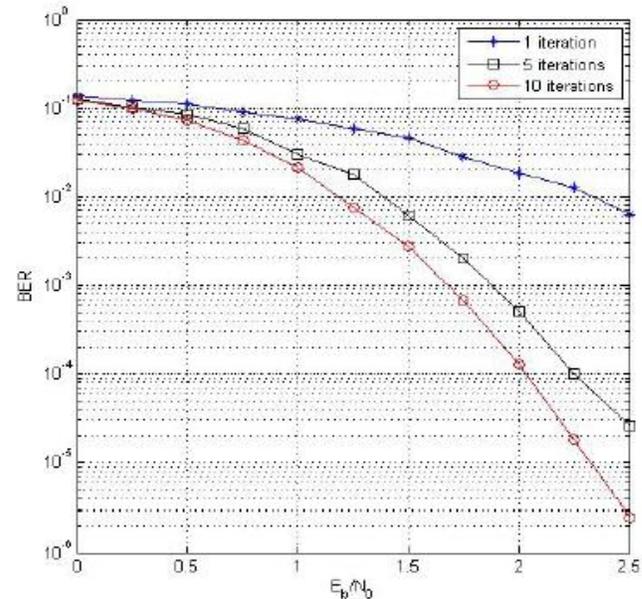


Fig.8. BER verses SNR for CTC in 802.16 systems

Fig. 7, which is plotted for TPC in 802.16e and Fig. 8, which is plotted for CTC in 802.16e, are compared. It is observed that TPC provides the bit error rate less than 10^{-6} at E_b/N_0 of 2.5db, which is better than CTC in 802.16e. From Fig.7 and 8 the comparative performance of CTC and TPC, also shows that there is a gain of around 1 db in using TPC.

B. COMPARISON OF HYBRID DECODER AND TPC

Hybrid decoder is a concept, where soft and hard decoding techniques are combined. Initial m iterations are run by using soft decoding and $n-m$ iterations are run by hard decoding, where n is the total number of iterations. Value of n depends on the application and performance [11].

Fig. 9 below shows the BER Vs. E_b/N_0 performance for the standard SISO and Hybrid decoder. Although the number of calculations are reduced, the performance is not very good. Comparing the Fig. 7 and Fig. 9, it is observed that TPC with SISO gives the better performance as compared to hybrid decoder.

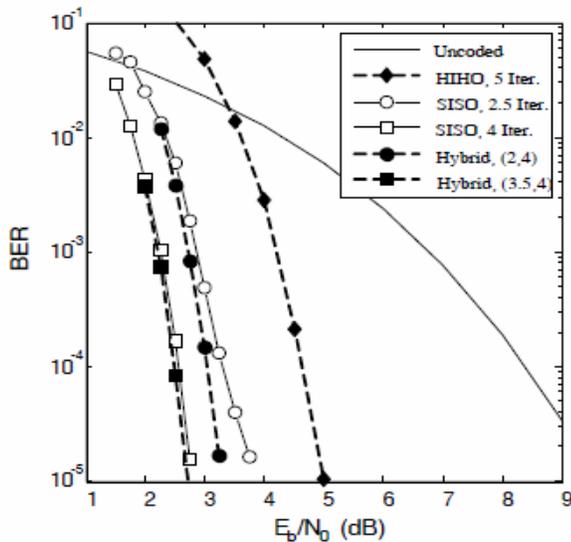


Fig.9. BER versus E_b/N_0 for standard SISO and hybrid decoder for eBCH (32,21,6)

C. COMPLEXITY COMPARISON

Other advantage of using TPC over CTC in 802.16e systems is that, there is no need of using the complex interleaver as in CTC. As illustrated in Fig.2 CTC encoder are constructed using interleaver optimization techniques [2]. The need of using complex interleaver and doing its optimization using different techniques for obtaining better BER can be avoided by using TPC. This ultimately reduces the complexity of the decoding technique, which leads to performance improvement.

The table below shows the comparison between CTC and TPC based on the graph which is shown in fig. 7 and fig. 8. It is observed that the performance of TPC is much better as compared to CTC for a given E_b/N_0 range. This can be further improved by increasing the range for E_b/N_0 depending on the application.

TABLE I. E_b/N_0 verses BER for TPC and CTC

E_b/N_0 in db	BER for TPC at 10 th iteration	BER for CTC at 10 th iteration
0	1×10^{-3}	1×10^{-1}
0.5	1×10^{-4}	8×10^{-2}
1	1×10^{-5}	2×10^{-2}
1.5	3×10^{-6}	3×10^{-3}
2	5×10^{-7}	1×10^{-4}
2.5		3×10^{-6}

VIII. CONCLUSIONS

In this paper we had implemented 802.16e system using TPC as forward error correction technique. The results for TPC shows that this can be used for high code rate and data

rate applications. TPC with SISO decoding also gives better performance as compared to hybrid decoder. Also there is no need of using complex interleaver or its optimization as in CTC. TPC have satisfied the performance/complexity tradeoff, to meet almost any requirement whether it is a single carrier or multiple carrier OFDM system.

IX. FUTURE SCOPE

Future scope of this project is improving the performance of TPC by reducing the number of test patterns used to decode the constituent codes of the block of TPC. This is achieved by using syndrome analysis technique. TPC can achieve a throughput of Giga bits per second by parallel decoding of rows and columns. The hardware can be designed to support conflict free interleaver memory access model to support parallel decoding [10]. Combining encryption and encoding is the new field of interest for researchers. TPC can be combined with chaotic encryption for different type of data, and performance can be analyzed.

ACKNOWLEDGMENT

This work was supported by the coordinators and faculty members of Thakur College of Engineering and Technology. Without the support of whom this project would not be a success. We would like to extend our gratitude and sincere thanks to them for their exemplary guidance and encouragement.

REFERENCES

- [1] Paul Guanming Lin "OFDM SIMULATION in MATLAB " thesis submitted for project in California Polytechnic State University
- [2] <http://www.aha.com>. "Improving Bandwidth Utilization with Turbo Product Codes", pp 32-33
- [3] B. Baumgartner, M. Reinhardt, G. Richter, M. Bossert "Performance of Forward Error Correction for IEEE 802.16e" University of Ulm Germany Dept. of Telecommunications and Applied Information Theory
- [4] R. Pyndiah, "Near-Optimum Decoding of Product Codes: Block Turbo Codes" IEEE Trans. Comm., vol. 46, pp.1003-1010
- [5] Brian A. Banister, Ph.D "Using Turbo Product Codes in Client Station Uplink for Reduced Power Consumption" Comtech AHA Corporation., pp 1-3
- [6] D. Chase "A class of algorithms for decoding block codes with channel measurement information" IEEE Transactions on Information Theory, IT-18 No.1, 1972, pp 172
- [7] Yejun He, Francis C.M. Lau *, Chi K. Tse, "Study of bifurcation behavior of two-dimensional turbo product code decoders", pp 503
- [8] Mohamed Amr Mokhtar "Study the Performance of Mobile WiMAX Convolutional Turbo Code" Proceedings of the 13th WSEAS International Conference on COMMUNICATIONS, pp 170
- [9] Bernard Sklar, Digital Communications Fundamentals and Applications, Pearson, 2ND Edition, 2010, pp 483-489
- [10] Richard E. Blahut "Theory and Practice of Error Control Codes" Addison-Wesley, 1983.
- [11] Li Zhou, Hengzhu Liu and Botao Zhang "Flexible and high efficiency turbo product code design" National University of Defense Technology China.
- [12] A1-Dweik, S. Le Goff, B. Sharif, IEEE, "A Hybrid Decoder for Block Turbo Codes", IEEE Transactions on communications, Vol. 57, No. 5, MAY 2009, 1229 -1232.

e-Restaurant: Online Restaurant Management System for Android

Dr. Vinayak Ashok Bharadi

Associate Professor, IT Department, Thakur College of
Engineering & Technology
Kandivali(E) , Mumbai, India

Vivek Ranjan, Nimesh Masiwal, Nikita Verma

UG Student, IT Department, Thakur College of
Engineering & Technology Kandivali(E) , Mumbai, India

Abstract—The simplicity and ease of access of a menu are the main things that facilitate ordering food in a restaurant. A Tablet menu completely revolutionizes the patron's dining experience. Existing programs provide an app that restaurants can use to feed their menus into iOS & Android based tablets and make it easier for the diners to flip, swipe & tap through the menu. We here aim to provide the restaurants with a tablet menu that would recommend dishes based on a recommendation algorithm which has not been implemented elsewhere. In addition to this we run the app on an Android based tablet & not on an iOS based tablet which is more expensive alternative. We use a cloud-based server for storing the database which makes it inexpensive & secure.

Keywords—*Recommendation; Tablet; menu; Intelligent; Android application; restaurant*

I. INTRODUCTION

Over the years, technology has tremendously revolutionized the restaurant industry. But much of the innovation has been with point-of-sale (POS) operations. Yet other areas of a restaurant are ripe for innovation, such as the menu.[4] Traditional restaurant service requires waiters to interact with customers directly before processing their orders. However, a high-quality recommendation service system would actively identify customers and their favorite meals and expenditure records.[1]

There is a famous saying that “People eat with their eyes”.[6] The e-Menu provides additional information about menu items and drinks than a traditional paper menu. With interactive pictures it gives additional information about the food item. Tablets are said to eliminate order-taking errors from the waiters. In the kitchen, there is less confusion as everything is now written clearly. Developers of similar applications maintain that customers who seat at tables outfitted with tablets spend about 10% more than those at other tables (“people buy more when they can do so instantly, without waiting for service”).[9] With the visuals, you know exactly what you're going to get in your plate The service goes quicker. Tablets are said to allow cutting the labor expenses. Customers feel more involved in the process. Restaurants can build their e-reputation and customer community in live.

The restaurant menu, as we know it, has evolved from its humble beginnings on carte chalkboards and imageless print to today's detailed, colorful displays. With the emergence of digital tablets and user-friendly touch screen technology

menus can move to a whole new surface. With this electronic menu, orders can be taken correctly the first time. There is no need to run back and forth to a distant terminal, because the terminal is always with the server. Every order is associated with an individual seat at the table, and orders are built one customer at a time, just like on paper, but with greater accuracy. Items can also easily be shared by the whole table, moved or modified, and noted and the cost can be calculated in real time.[10]

The Recommendation algorithm suggests dishes to the patrons based on previous orders. It makes it easier for the customer to build his/her order and also view the most popular dishes. Moreover, various dimension filters can be used according to individual preferences e.g. Price, taste, quantity, etc.

In a study earlier [1], a preliminary experiment was conducted in a restaurant, and a questionnaire survey was administered to fifteen waiters and forty-five customers. The survey result was encouraging. In addition, extensive interviews with restaurant owners were conducted and the results indicated that the proposed system is useful in reducing running cost, enhancing service quality as well as customer relationship.[9]

II. EXISTING SYSTEMS

A. Conventional Systems

Restaurant services such as making reservations, processing orders, and delivering meals generally require waiters to input customer information and then transmit the orders to kitchen for meal preparation. When the customer pays the bill, the amount due is calculated by the cashier. [1]

Although this procedure is simple, it may significantly increase the workload of waiters and even cause errors in meal ordering or in prioritizing customers, especially when the number of customers suddenly increases during busy hours, which can seriously degrade the overall service quality. [1]

B. Electronic POS Terminals

A very commonly implemented system, currently being used by numerous restaurants and chains all over the world, is the electronic point-of-sale terminal system.

Here, the servers/waiters generally take the order from the customer and head onto a terminal, where they can feed the order into a computer. The order can then be transmitted to the

kitchen automatically via the terminal through a network, or it may even be delivered manually by the server to the kitchen.

Although a huge improvement over the pen and paper still prevalent over the world, this does not have much value addition for the customer and mostly only benefits the establishment and the administration of the establishment.

C. Tablet Based Menus

With the current onslaught and popularity of touch based devices, especially the Apple iPad, it did not take long for the tablet based menus to make an entrance in to the market.

In their current state, these menus are just a glorified version of their paper based counterparts.

The workflow, even with these tablet menus, remains the same as the one mentioned in section 2 (A) or a combination of the previous two. There is very little value addition to the customer and the establishment or its staff.

The available technology is definitely not being utilized to its maximum potential.

III. PROPOSED SYSTEM

A. System Architecture

The following figure (Fig 1), demonstrates the basic architecture of our system.

Understanding the other intricacies and details of the system will become a lot easier if one goes through this figure, before diving into the rest of it.

The system will consist of the following main components: The backend, which is made up of the web server and the database, and the frontends that include both the patron frontend (delivered as a native mobile application) and the administration or the kitchen frontend (delivered as a web application).

This system is based on the very popular Model-View-Controller (MVC) architecture. MVC is most commonly used in websites, very popular and tried and tested.

None of the frontends directly “talk” to the database. They instead rely on RESTful web services that can be used to perform CRUD operations on the database.

B. System Overview

The most important components of this system are the database and the patron frontends or the tablet applications.

Providing value to both the business and its patron is an important objective, but we believe that one follows the other.

Following that belief, the customer is given a whole lot of importance.

The following figure (Fig 2) demonstrates the flow of our tablet application

The patron will be presented with a tablet, running the Android OS. This tablet will be synchronized with the database running on our centralized cloud powered servers.

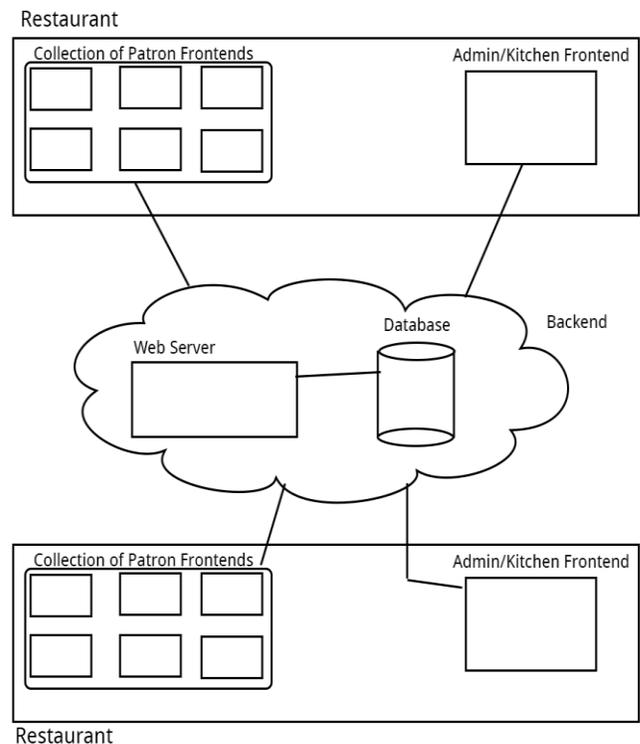


Fig.1. System architecture & the main components

The menu data, upon synchronization, is stored on locally on the tablets so that the user, i.e. patron, need not wait for the menu to be downloaded from the servers. This will allow faster access to the menu.

The user can then browse the menu however they want to, sorting the items on various dimensions like price, popularity, ratings, etc. The user can also click through to view more information about any item like nutritional information, ingredients, trivia and any other content that the restaurant administration may feel like including.

The user can also view personalized recommendations for items that they may like. This is one of the most important aspects of our system that not only enhances the customer experience but can also help increase revenue for the business.

While browsing through the menu, the customer may add items to his/her order. This process is commonly known as “building the order”. After the order is built and read, the user may go ahead and place the order. The staff will automatically and almost instantly be notified about the new order so that they can act on it.

If the establishment allows, the user may even track the status of their order so that the customer may know when to expect their food and drinks to land up on their table.

C. Recommendation Overview

The recommendation algorithm is an innovative feature that we aim to include in our menu. When most tablet menus provide the customers with only a simple menu, this system

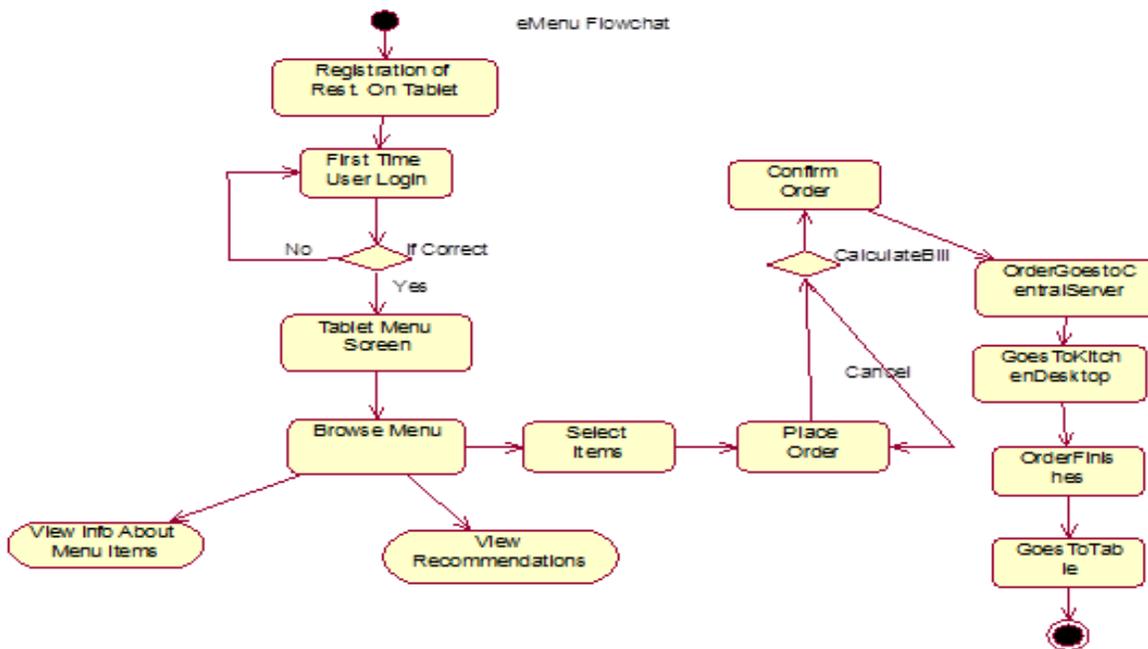


Fig.2. Flow chart of the tablet application

will provide recommendations which will make it easier to build an order considering what other customers have ordered

previously or the similarities between various dishes. Recommendation systems using sets were considered [2]. We finally decided to use the below methodology, which has been discussed in an earlier study [3].

D. Methodology For Recommendation System

The algorithm mainly has 5 parts:

1. Users - a certain number of people are made to rate individual food items.
2. Entities - the food items.
3. Value Dimensions - the categories that are formed to rate the food items e.g. Price, quality, meat content, etc.
4. Belief System - is personal to each user & allows to tell the system what ideal value they want each value dimension to have.

Fig 1: Flow chart of the tablet application

Each User rates a food item on a scale of 1 to 5 with respect to two things:

1. User’s ideal value dimension
2. The weight or the importance of that value dimension

With the food items set in place, our next task was to analyze the various attributes that were associated with each food item. We applied a food item click counter to the entire data set, which produced a list of the most viewed food item.

After listing out the Top N clicked food-items by the customers, we apply normalization to the retrieved list of food-items, to filter out redundant clicks.

We further investigate the levels of commonality that existed between various pairs of food items. Jaccard’s coefficient was used to calculate the degree of similarity.

The Jaccard index, also known as the Jaccard similarity coefficient (originally coined coefficient de communauté by Paul Jaccard), is a statistic used for comparing the similarity and diversity of sample sets.

The Jaccard coefficient measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:[12]

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Collaborative filtering-based recommender systems rely on information derived from social activities of the users, such as opinions or ratings, to form predictions or produce recommendation lists. Existing collaborative filtering techniques involve generating a user-item matrix, from which recommendation results could be derived Content-based filtering focuses on the selection of relevant items from a Content-based filtering focuses on the selection of relevant items from a large data set, things that a particular user has a high probability of liking. This involves training the data set with machine learning techniques. Clustering involves sectioning the data set into particular sets, each of which correspond to certain preference criteria. Also, typical recommendation systems output their results either as predictions, a numerical ranking value corresponding to a particular item or recommendations, a list of relevant items .The conventional approaches to computing similarity involve

the use of two popular techniques: Pearson correlation similarity given by Equation 1

$$\text{Pearson Correlation Similarity} = \frac{\sum_{j=1}^l (r_{ij} - \bar{r}_i)(r_{kj} - \bar{r}_k)}{\sqrt{\sum_{j=1}^l (r_{ij} - \bar{r}_i)^2} \sqrt{\sum_{j=1}^l (r_{kj} - \bar{r}_k)^2}} \quad (1)$$

and cosine similarity, given by Equation 2.

$$\text{Cosine Similarity} = \sum_j \frac{r_{ij}}{\sqrt{\sum_j r_{ij}^2}} \frac{r_{kj}}{\sqrt{\sum_j r_{kj}^2}} \quad (2)$$

We used Jaccard's coefficient in our models,

to compute similarity. by generating a list of top-N food-items clicked, then applying a re-ranking algorithm to this list. In this manner, the recommendation list becomes more refined by using the information derivable from the original data set.[11]

E. Comparison of Existing vs. Proposed systems

TABLE I. COMPARISON OF SYSTEMS

	<i>Existing System</i>	<i>Proposed System</i>
Operating Systems	iOS	Android
Communication Channel	Communication between the customer & the waiter	Communication between customer & the terminal
Customer identification	RFID used to identify customers Additional hardware required	User accounts maintained in the database server
Server location	Localized server	Centralized server
Recommendation System	Recommendations not implemented	Recommendations implemented
Exclusivity	Exclusive to every establishment	Can be extended to be used by multiple establishments

F. Technologies

The technologies being used to build the system are cutting and largely open source. Open source technologies help in keeping the costs in check, thus enabling the various establishments to use this setup without any cause of concerns with regards to the costs involved.

Amazon Web Services has a very effective and inexpensive service known as Elastic Cloud Compute which allows one to setup servers on the fly with the specifications one requires. We will be using the same (or similar) service to keep the costs down and maintain scalability.

The server uses a software known nginx. nginx [engine x] is an HTTP and reverse proxy server, as well as a mail proxy server, written by Igor Sysoev. For a long time, it has been running on many heavily loaded Russian sites including Yandex, Mail.Ru, VKontakte, and Rambler. According to Netcraft nginx served or proxied 12.81% busiest sites in February 2013.

The database uses the ever popular MySQL as its DBMS. The web services mentioned earlier are powered by a web framework called Symfony that uses PHP and an ORM called Doctrine. OAuth and SSL can and will be used to implement security.

The tablets supplied to the patrons will be running Android OS.

G. Features Overview

In this section we won't go into the detailed features of the system, but instead take a bird's eye look at the same.

1. Intuitive, Beautiful & User friendly

The end users, i.e. the restaurant customers, will have maximum interaction with this system. This interaction will mostly occur through the tablet application.

Unlike most applications that have a targeted user base, our application will be used by all & sundry. It could be used by an 8 year old or an 80 year old person, from varying cultures and background. This ensures the need for a design that is intuitive, user friendly and beautiful thus making it a pleasure to use.

2. Scalable

The system has the potential of being used by millions of users at any given time, with possible peaks at various rush hours at the establishments this may be implemented at. Thus, the system will have to be scalable and should be able to accommodate as many users and as much data as required.

3. Secure

Since private information about so many businesses and individuals will be stored in our database. We will need to ensure that the whole system - the web service, the app, the database as well as the server be secure from external as well as internal threats.

H. Assumptions

The success of the proposed system is based on the following assumptions that:

- End users should be able to use tablets.
- Network connection will always remain stable.
- Security of server is maintained.
- Tablets should work fault free.

I. Risks

The above mentioned assumptions, give rise to a bunch of risks. Those risks are mainly the following:

- Intrusion of confidential data

- Server goes down
- Tablets suffer a defect
- End users are not able to use the tablet devices

J. Advantages

Most, if not all, of the current tablet based menu systems use Apple's iPad device. This is mostly due to the fact that the current generation of iPads were the first tablet devices to feature high resolution displays and the perceived brand recognition and status of the brand that is Apple.

It is a well-known fact that Apple devices sell at a premium. So, if an establishment were to implement this system, they would unfortunately have to pass on the extra costs to their customers that they incurred in the procurement of these devices.

High resolution Android tablets are now available at extremely economical rates with varying feature sets. Since all our application needs is a Wi-Fi connection and good display, the costs of the tablet devices can be kept to a minimum.

The centralized servers can allow the sharing of customer data between restaurants, if allowed to do so explicitly by both the customer and the business, then this can help in providing better recommendations and user experience to the patrons. This is really not possible in the stand alone systems that currently exist.

A high-quality service system should be customer-centric, i.e., it should immediately recognize the identities, favourite meals and expenditure records of customers so as to provide customer-centric services. Therefore, using advanced technologies to improve service quality has attracted much attention in recent years.

In recent years, various product recommendation systems have been developed to enhance customer satisfaction and perceived value. Defined as a system which recommends an appropriate product or service after learning the customers' preferences and desires, recommendation systems are powerful tools that allow companies to present personalized offers to their customers. Extracting users' preferences through their buying behaviours and histories of purchased products is the most important element of such a system.

The mobile device-based service unit enables instant transmission of customer orders via Internet to the kitchen for meal preparation. In addition, the expenditure information can be sent to the cashier for bill pre-processing. The restaurant managers can access the database to evaluate the business status anytime and make appropriate redeployments for food materials. Notably, all ordering and expenditure information is digitized for database storage, which allows restaurant owners to consider discounts or promotion to customers based on expenditure statistics. Customers can thus appreciate high-quality service, which in turn highly promotes enterprise image and increases business revenue for the restaurant.

For our backend server, where the admin and the kitchen frontends reside, uses the Amazon Web Service Elastic Cloud

Compute (AWS EC2) servers. The advantages of using these EC2 servers are manifold and a few are listed below for reference.

- You can easily scale vertically (upgrade to a larger 2 or 4 CPU instance) or horizontally (add instances).
- You get a lot of RAM. A small instance has 1.7Gb. A large instance has 7.5Gb.
- You can more easily and cheaply leverage S3 for backup, storage, and serving of large files.
- They have excellent bandwidth.
- Less worry about hardware failure. Failures do occur, although it should occur less often than a dedicated server, and recovery is much easier.
- No CPU throttling or other usage limitations. At a shared host, it is common practice to kill long running scripts that are using significant CPU.
- Dedicated IP address. It's yours and yours alone, as long as you keep your instance running.

IV. RESULT

The following screenshots show our tablet application (earlier referred to as the patron frontend) running on a virtual android machine on a standard PC (Intel i5, Windows 7).

For our test environment we used a distribution of Android known as AndroidVM. This distribution is available in different versions for phones and tablets. We are, as visible, running the tablet version of the same. The Android version being used is 4.1.1, although the application is theoretically compatible with versions 2.2 onwards. This compatibility has not been tested

A. Testing

We performed the following Test

Test-id	Test Case	Expected Result	Actual Result	Pass/Fail
01	Does the total cost change when items are added/removed from the cart?	Yes	Yes	Pass
02	Does the total cost change when the quantity is reduced/increased?	Yes	Yes	Pass
03	Do all the images load on clicking?	Yes	Yes	Pass
04	Is it easy to navigate through all the categories e.g. Desserts/Veg/Non Veg/Cocktails,etc.	Yes	Yes	Pass

05	Can an order be cancelled within a selected time frame(eg. Time frame= 10 mins)	Yes	Yes	Pass
06	Is it possible to order more food items after specific time intervals?	Yes	Yes	Pass
07	If discount offers on specific card payments can be availed?	Yes	Yes	Pass
08	Does order get placed & customer gets billed if there is a loss in connectivity	No	No	Pass

All the above Test cases are executed in Aakash Tablet with configuration : 4.0.1 ICS , 1GB Ram, 1Gz Processor

B. Test Results Screenshots

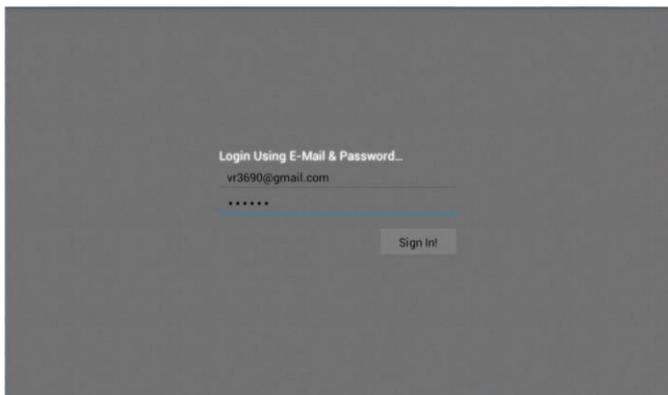


Fig.3. Login screen of Android app

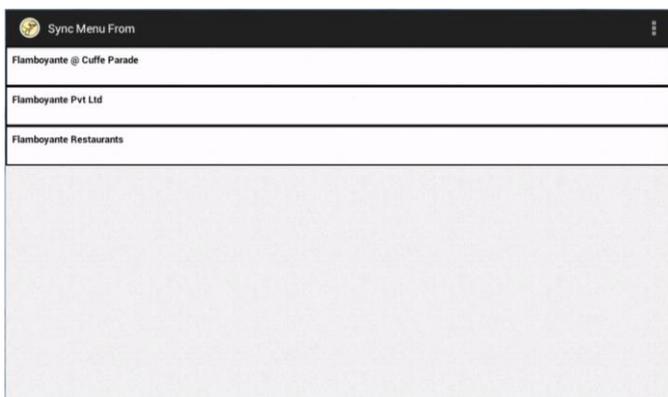


Fig.4. Choose establishment

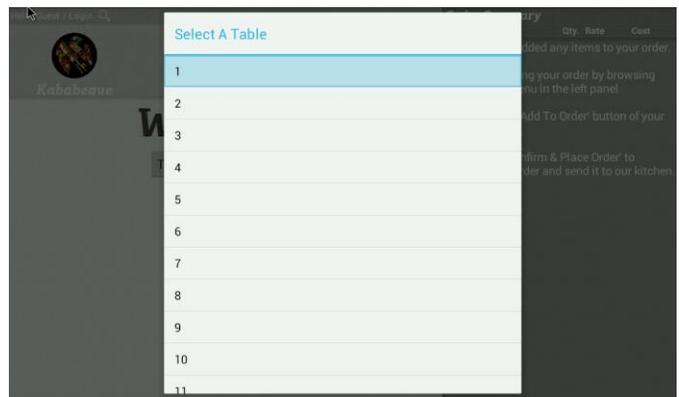


Fig.5. App Welcome Screen



Fig.6. Browsing Menu



Fig.7. Add items to order

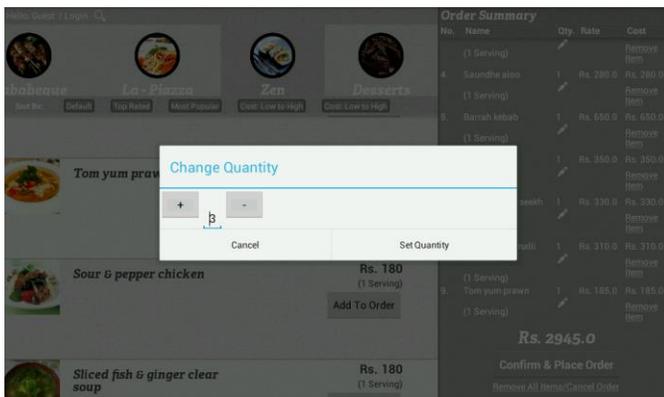


Fig.8. Edit quantity

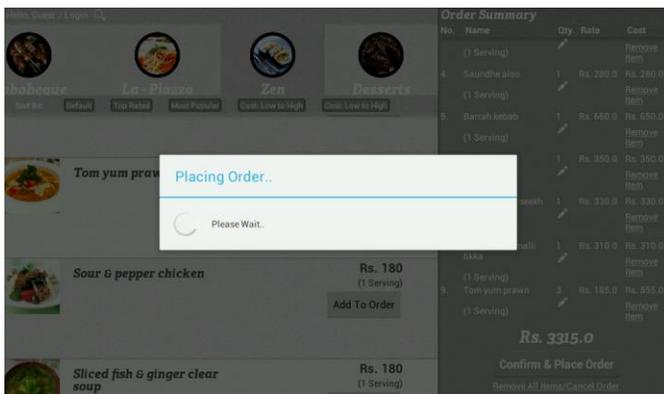
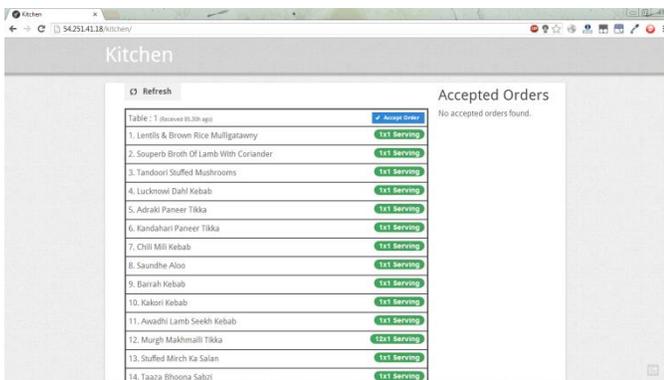


Fig.9. Place Order



Order received at kitchen

Fig 10 depicts the kitchen interface running in the latest version of the Chrome browser. The main aim of the kitchen frontend is to display the latest orders, place by the customers using the patron frontend, in a comprehensive manner. An AJAX query is fired at a pre-defined regular interval to check if any new orders have been placed and the same is displayed.

ACKNOWLEDGMENT

First & foremost we would like to thank our mentor & guide Dr. Vinayak Ashok Bharadi, for his constant support & guidance. His active cooperation & involvement have helped us through the various stages of project development.

We would also like to express our gratitude to Dr. Kamal Shah, H.O.D., Information Technology for her thoughtful recommendations & suggestions.

Last but not the least we would like to thank Dr. B.K. Mishra for extending his support.

References

- [1] Tan-Hsu Tan, Ching-Su Chang, Yung-Fu Chen, Yung-Fa Huang, Tsung-Yu Liu, "Developing an Intelligent e-Restaurant With a Menu Recommender for Customer-Centric Service", Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions.
- [2] Tomoko Kashima, Shimpei Matsumoto, and Hiroaki Ishii, "Recommendation Method with Rough Sets in Restaurant Point of Sales System", PIMECS 2010 Vol III
- [3] Ali Akhtarzada, Cristian S. Calude and John Hosking, "A Multi-Criteria Metric Algorithm for Recommender Systems", CDMTCS-400
- [4] <http://www.forbes.com/sites/tomtaulli/2012/01/13/will-tablets-kill-the-traditional-menu/>
- [5] http://www.emenuworld.com/emenu/features_22
- [6] <http://www.deplod.com/tablet/ipad/>
- [7] <http://aptito.com/Why-Aptito>
- [8] <https://www.menupad.com/features>
- [9] <http://www.waitersrace.com/tablets-and-e-menu-debates-around-the-restaurants-tables>
- [10] <http://www.touchbistro.com/features>.
- [11] <http://www.acsu.buffalo.edu/~suchismi/iRec.pdf>
- [12] http://en.wikipedia.org/wiki/Jaccard_index