1-9-2012

Proceedings of International Conference on Information and Communication Technology

Prof.Srikanta Patnaik Mentor
IRNet India, patnaik_srikanta@yahoo.co.in

Proceedings of International Conference on

Information and Communication Technology

(IJCICT-2012)
9th January, 2012
BHUBANESWAR, India

Interscience Research Network (IRNet)
Bhubaneswar, India

Editorial

If we make a review of the 21st century generation, their planning and activities, their interest and involvement driven by some electronic device coupled with an advanced technical functioning. There is a spectacular focus on Information science and Communication technology as it drives the present socio-technical system of the global society. ICT appears as an effective tool for empowering all the civic and anti civic systems of the world. ICT is emerging as an investment area in the millennium development goals of various organizations like UNO, WTO, IBRD and other international apex bodies. It has become an integrated discipline in all the academic spheres as it increases global adaptability.

Information Economy report produced by the UNCTAD establishes the fact that developing nations are making a global pace by gaining mileage from ICT. Let me include some of the contents from report "In rich countries, broadband subscribers increased by almost 15% in the last half of 2005, reaching 158 million. Business broadband connectivity grew most significantly -- in the European Union, for example, from 53% of enterprises in 2004 to 63% in 2005. Broadband enables companies to engage in more sophisticated e-business processes and to deliver a greater range of products and services through the Internet, thus maximizing the benefits of information and communication technology (ICT). The use of broadband directly increases competitiveness and productivity, the report says -- and that, in turn, has an impact on macroeconomic growth. It is estimated that broadband can contribute hundreds of billions of dollars a year to the Gross Domestic Products (GDPs) of developed countries over the next few years.

The mushrooming growth of the IT industry in the 21st century determines the pace of research and innovation across the globe. In a similar fashion Computer Science has acquired a path breaking trend by making a swift in a number of cross functional disciplines like Bio-Science, Health Science, Performance Engineering, Applied Behavioral Science, and Intelligence. It seems like the quest of Homo Sapience Community to integrate this world with a vision of Exchange of Knowledge and Culture is coming at the end. Apparently the quotation "Shrunken Earth, Shrinking Humanity" holds true as the connectivity and the flux of information remains on a simple command over an internet protocol address. Still there remains a substantial relativity in both the disciplines which underscores further extension of existing literature to augment the socio-economic relevancy of these two fields of study. The IT tycoon Microsoft addressing at the annual Worldwide Partner Conference in Los Angeles introduced Cloud ERP (Enterprise Resource Planning,) and updated CRM (Customer Relationship Management) software which emphasizes the ongoing research on capacity building of the Internal Business Process. It is worth mentioning here that Hewlett-Packard has been with flying colors with 4G touch pad removing comfort ability barriers with 2G and 3G. If we progress, the discussion will never limit because advancement is seamlessly flowing at the most efficient and state-of-the art universities and research labs like Laboratory for Advanced Systems Research, University of California. Unquestionably apex bodies like UNO, WTO and IBRD include these two disciplines in their millennium development agenda, realizing the aftermath of the various application projects like VSAT, POLNET, EDUSAT and many more. 'IT'

has magnified the influence of knowledge management and congruently responding to social and industrial revolution.

The conference is designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these two integrated disciplines. Even a fraction of active participation deeply influences the magnanimity of this international event. I must acknowledge your response to this conference. I ought to convey that this conference is only a little step towards knowledge, network and relationship.

I congratulate the participants for getting selected at this conference. I extend heart full thanks to members of faculty from different institutions, research scholars, delegates, IRNet Family members, members of the technical and organizing committee. Above all I note the salutation towards the almighty.

Editor-in-Chief

Prof. (Dr.) Srikanta Patnaik
President, IRNet India and Chairman IIMT
Intersceince Campus, Bhubaneswar
Email: patnaik_srikanta@yahoo.co.in

Design of A Probe-Fed Circularly Polarized Microstrip Patch Antenna For WLAN Application

Chitta Ranjan Das

Electronics and Communication Engineering Department
C. V. Ramana College of Engineering (B.P.U.T), Bhubaneswar, Odisha, India

Abstract **-** Microstrip Antenna is used in wide range of application but narrow bandwidth often limits there more widespread use. The thesis covers two aspects of Microstrip antenna designs the first is the analysis and design of single element rectangular Microstrip antenna which operates at the central frequency of 5 GHz and the second aspect is the design of circularly polarized Microstrip patch antenna. The antenna is single feed circular polarized Microstrip Antenna .This miniaturized Microstrip antenna has wide bandwidth in the frequency band of WLAN and exhibits circularly polarized far field with very good axial ratio bandwidth. The simulated result using IE3D software is verified by measurement. For rectangular Microstrip antenna design fabricate on Glass Epoxy substrate based, Microstrip board with dielectric constant 4.36 and the substrate height is 1.57 mm and loss tangent is 0.001. The properties of antenna such as bandwidth, S-Parameter has been investigated and compared between different optimization scheme and theoretical results.

I. INTRODUCTION

Microstrip antenna is defined as: "An antenna which consists of thin metallic conductor bonded to thin grounded dielectric substrates" [1].. The most impotent high data rate wireless broadband networking systems for future wireless broadband networking systems for future wireless communications are High Performance Local area Network type1(HIPERLAN/1) and High Performance Local area Network type 2(HIPERLAN/2) which use the frequency bands 5.150GHz-5.350GHz and 5.470Ghz-5.275GHz respectively with Omni-directional antennas. But for short distance indoor LAN communication, directive antenna can be used HIPERLAN/2 has a very high transmission rate up to 54Mbits/s [2,3].Single layer Microstrip antenna have narrow bandwidth, but using multi-layered configurations like proximity coupled Microstrip antennas or aperture coupled Microstrip antennas or aperture coupled Microstrip antennas ,higher bandwidth can be achieved[2] .Circularly Polarized Microstrip patch antenna has very small size, wide bandwidth, moderate gain and very good axial ratio bandwidth required for communication using HIPERLAN/ 2.HIPERLAN/2 is principally used for indoor wireless local area network and for indoor signal propagation due to multiple reflection from walls and other human made configuration, signal change its direction and hence signal from other directs except null direction will be received by receiving antenna.[11]

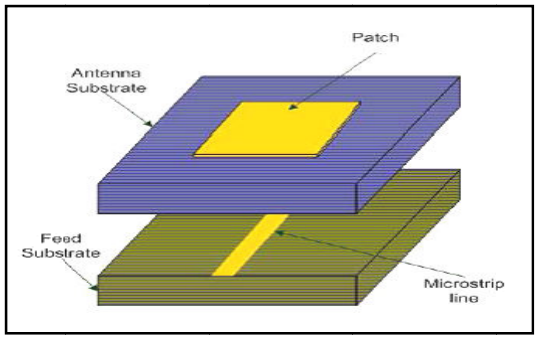


Fig.1 : Proximity–coupled Feed

II. DESIGN SPECIFICATIONS

IE3D is an integral full-wave electromagnetic simulation and optimization packages for analysis and design of 3D and planner microwave circuits MMIC, RFIC,RFID, antennas ,digital circuits and high speed printed circuit broads(PCB).since its formal introduce in 1993 IEEE international Microwave Symposium(IEEE IMS 1993),IE3D has been adopted as industrial standard in planner and 3D electromagnetic simulation. [13,14].The essential parameters for the design of a circularly polarized Microstrip Patch Antenna are: Frequency of operation (*fo*): The resonant frequency of the antenna must be selected appropriately. The high data rate wireless broadband networking systems for future wireless communications are High Performance Local Area Network type 1 (HIPERLAN/1)and High Performance Local Area Network type 2 (HIPERLAN/2)which h use the frequency bands 5.150

GHz–5.350 GHz and 5.470 GHz–5.725 GHz respectively with Omni-directional antennas. Hence the antenna designed must be able to operate in this frequency range. The resonant frequency selected for my design is 5.0 GHz. Dielectric constant of the substrate (ε_0): The dielectric material selected form design is fabricated on Glass Epoxy which has a dielectric constant of 4.36. Height of dielectric substrate (h): the height of the dielectric substrate is selected as 1.57 mm.

III. EXPERIMENAL RESULT USING MATLAB

From Fig.2, conclude that resonance frequency is at 5 GHz and give return loss of nearly -32db.

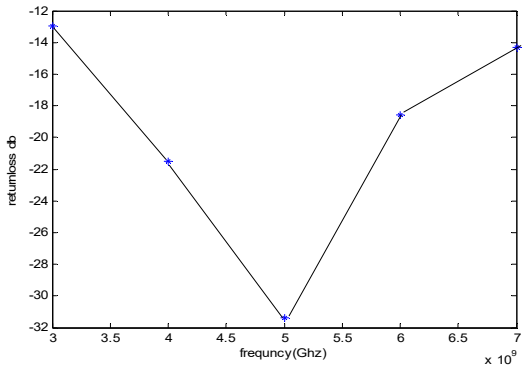


Fig. 2 : Return loss vs. Impedance match as a function of frequency plot (MATLAB)

From Fig.3 frequency vs. V.S.W.R plot, VSWR is 1.054db at frequency range 5 GHz. Thus V.S.W.R plot for antenna 1.05:1.this considers a good value of level of mismatched is not very high. High level of mismatch means port not properly matched.

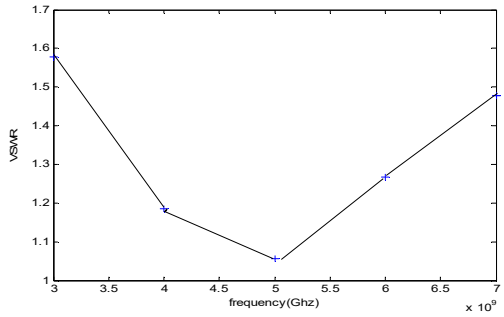


Figure 3 shows the impedance match as a function of frequency vs. VSWR

IV. SIMULATED RESULT USING IE3D

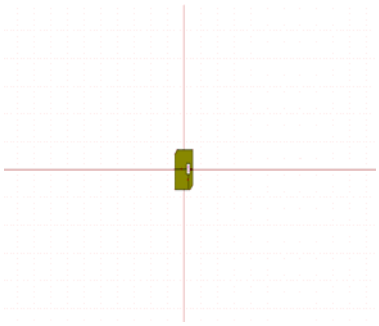


Fig. 4 : Structural view of patch after circularly polarized

From Fig.5 conclude that resonance frequency is not exactly at 5 GHz, It is resonating at about 4.89 GHz and give return loss of nearly -20db.Return loss of antenna which should be -10db for good performance.-10db return loss means 90% of power is radiated.

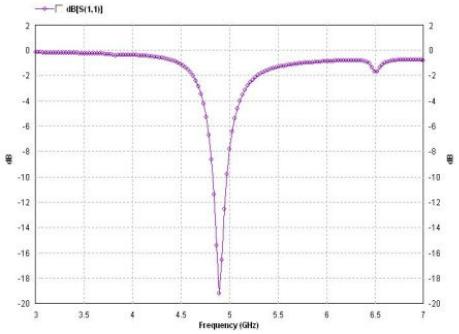


Fig 5 Frequency vs. return loss

Normalized scaling allows Smith chart to be used for problem involving any characteristics impedance or system impedance, although by far most commonly used is 50 ohms. From fig.6 Smith Chart shows that resonance frequency is not exactly at 5 GHz, It is resonating at about 4.89 GHz.

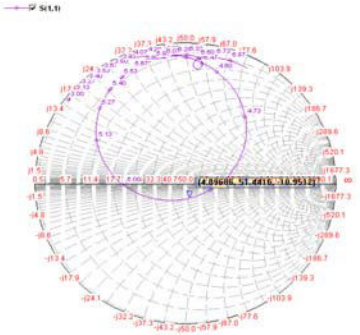


Fig.6 : Smith chart of circularly polarized Microstrip antenna

From Fig. 7 Frequency vs. V.S.W.R plot, VSWR is 1.054db at frequency range 5 GHz. Thus V.S.W.R plot for antenna 1.05:1.this considers a good value of level of mismatched is not very high. High level of mismatch means port not properly matched.

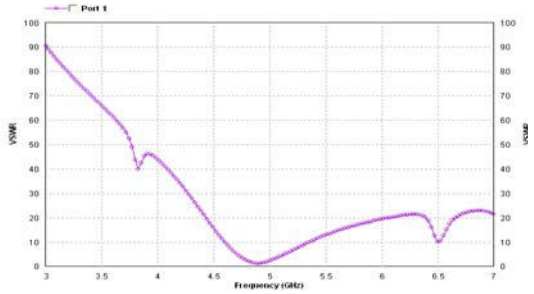


Fig.7 V.S.W.R vs. Frequency plot

Fig. 8 shows Z-parameter of real and imaginary part of antenna at frequency range 5GHz

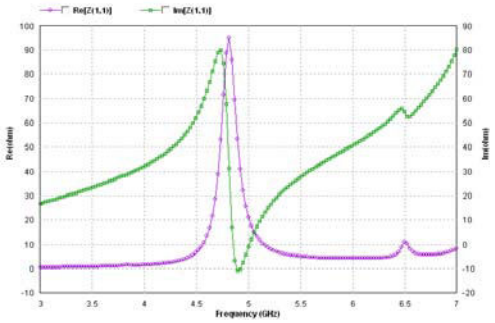


Fig.8 : Z-parameter of circularly polarized Microstrip Antenna

Fig.9 shows Elevation pattern gain display at $\varphi\ is\ 0^0 to 90^0$ at, which pattern gain of H-plane or plane of magnetic field, is in range 2.84db to15.24db (positive) and 2.84db to-87.7(negative) x-y plane.

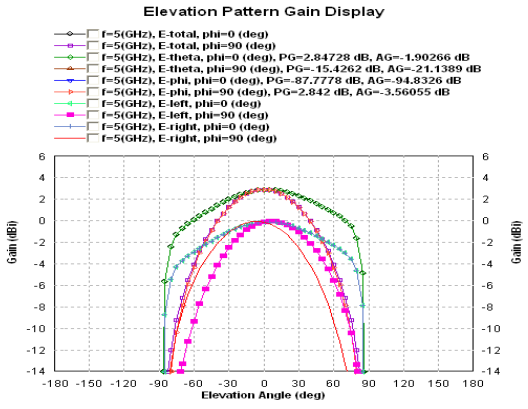


Fig. 9 : Elevation pattern gain display at $\varphi = 0^0 to\ 90^0$

Fig.10 shows axial ratio axial ratio of 0 dB in central frequency (3.8 GHz).Axial ratio is parameter for determine how good circular polarization. At different angles axial ratio is different.

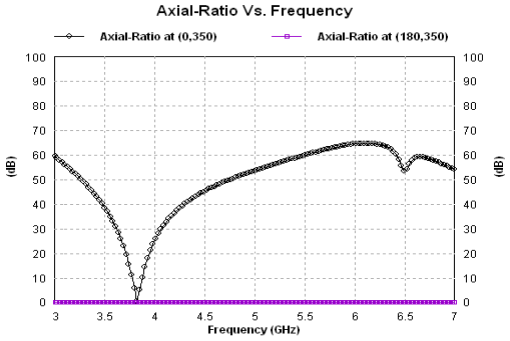


Fig .10 Axial-ratios vs. frequency plot

Fig.11 shows Elevation pattern gain display at $\theta\ is\ 0^0 to 90^0$ at, which pattern gain of E-plane or plane of electric field, is in range 2.84db to82.1db (positive) and 2.84db to -98.7db (negative) x-y plane.

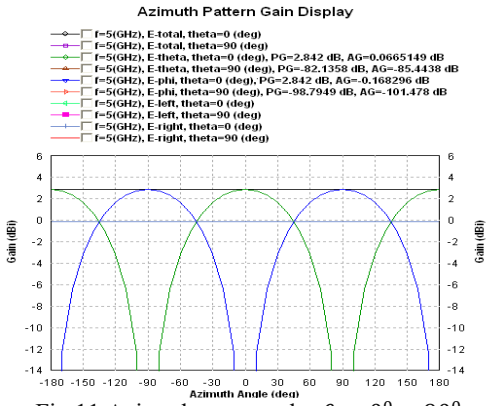


Fig.11 Azimuth pattern plot $\theta = 0^0 to\ 90^0$

Fig.12 shows total gain vs. frequency plot, which shows that at frequency 4.8GHz range maximum gain of 5.31

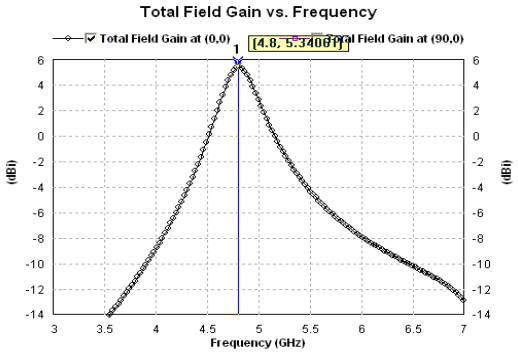


Fig.12 Total gain vs. Frequency plot at 0 to 90^0

Fig.13 shows polar pattern of Azimuth pattern which is a co-polarization pattern of E-plane or electric field (horizontal plane) produce by antenna

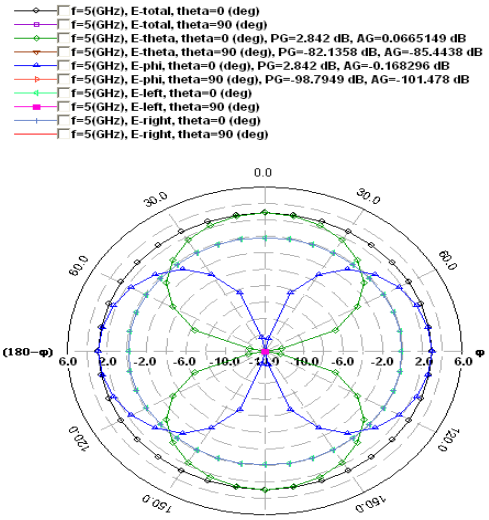


Fig.13 Polar plot of Azimuth pattern gain display at $\theta = 0^0 to 90^0$

Fig.14 shows polar pattern of Elevation pattern which is a cross-polarization pattern of H-plane or magnetic field (vertical plane) produce by antenna.

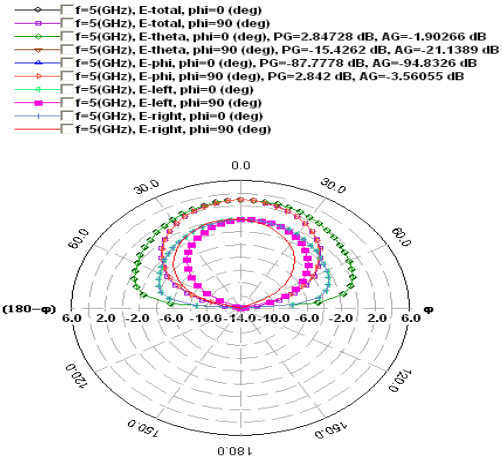
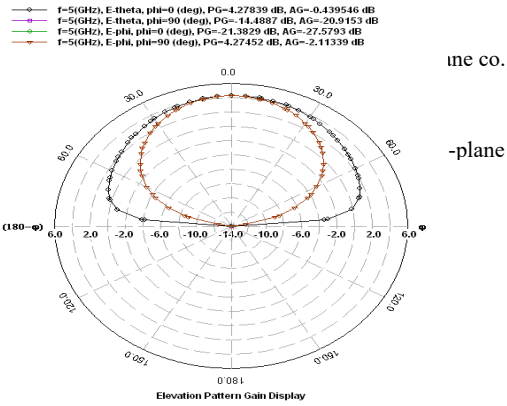


Fig 14 Polar plot of elevation pattern gain display $\varphi = 0^0 to 90^0$

Fig.15 shows 2-Dpattern of H-plane and E-plane co-polarization pattern of antenna .Maximum gain obtain in broadside region.



ıne co.

-plane

Fig.15 : 2-D pattern display of E-plane and H-plane when $\emptyset = 0^0 and \emptyset = 90^0$

Fig.16 shows 3-D pattern of antenna at frequency range 5GHz, here color represent dbi[Gain] of antenna which is 4.28db maximum.

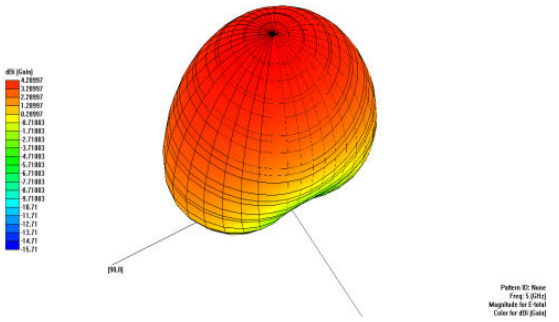


Fig.16 : 3-D (TRUE) Polar Plot E- total of Patch Antenna Element

Fig.17 shows 3-D pattern of axial ratio of antenna at frequency range 5GHz, here color represent Axial Ratio of antenna which is 28db maximum.

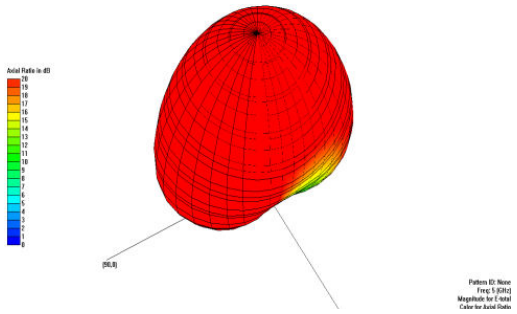


Fig.17 : 3-D (TRUE) polar plot for axial ratio of Antenna Element

3-D current distribution plot gives co-polarization and cross-polarization components. Moreover it gives clear picture of nature of field propagation through patch antenna. Fig.18 clearly shows that antenna is circularly polarized.

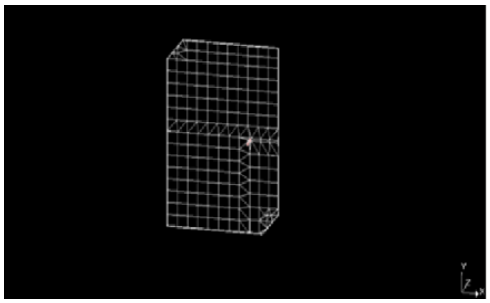


Fig.18 : Current distribution of antenna

Fig.19 shows highest possible degree of accuracy .The term "grid generation" is often used interchangeably. The triangulated zone region shows on grid for current distribution.

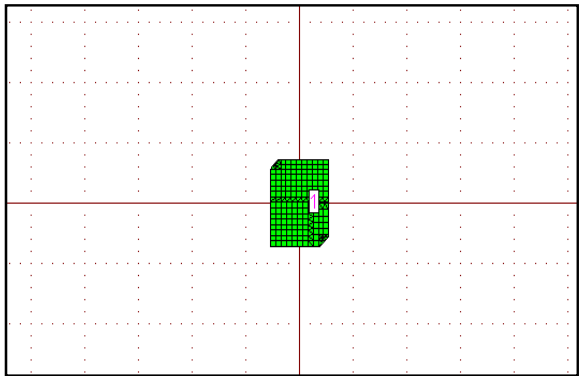


Fig. 19 : Display messing of antenna

V. CONCLUSION:-

The principles of design of a circularly polarized Microstrip Antenna for the application in WLAN are described that operate 5 GHz frequency range. Very important work is done in this thesis is to implement of Circular Polarized Microstrip antenna using IE3D and MATLAB. The simulated and experimental results on proposed circularly polarized proximity coupled Microstrip antenna show that impedance bandwidth and return loss bandwidth of the antenna is very good at 5GHz WLAN band and may be used for HIPERLAN/2. The simulation gave results good enough to satisfy our requirements to fabricate it on hardware which can be used wherever needed.

The theoretical results are compared with the simulated data obtained from IE3D. Results shows that proposed of circularly polarized Microstrip patch antenna has very small size, wide bandwidth, moderate gain and very good axial ratio bandwidth, required for communication using HIPERLAN/2.

REFERENCES:

[1] Christopher J. Booth, editor. The New IEEE Standard Dictionary of Electrical and Electronics Terms. IEEE Press, Piscataway (New Jersey), 5th edition, 1995

[2] Design of a circularly polarized Microstrip Antennas for WLAN, Progress in Electromagnetic Research M, Vol. 79-90,2008

[3] Grieg,D.D.,and Engleman,H.F., "Microstrip- A New Transmission Techniquic for the Kilomegacycle Range, "Proceeding of The IRE,1952, Vol.40,No.10, PP. 1644-1650

[4] Howell,J.Q.,"Microstrip Antenna", IEEE International Symposium Disest On Antenna and Propagation, Williamsburg Virginia,1972 pp.177-180

[5] C. A. Balanis, "Antenna Theory, Analysis and Design", JOHN WILEY & SONS, INC, New York 1997.

[6] D.M.Pozar," Microstrip Antenna" Proc.IEEE.Vol.80, No.1, pp.79-81.Januarry 1992

[7] Ramesh Garg, Prakash Bartia, Inder Bahl, Apisak Ittipiboon, ''Microstrip Antenna Design Handbook'', 2001, pp. 1-68, 253-316 Artech House Inc. Norwood, MA.

[8] Why circular polarization? By FRC Group 1511 South Benjamin Avenue, MasonCity, Iowa50401. www.frccorp.com/.../Why%20 Circular%20 Polorized% 20Antenna.pdf

[9] M. Olyphant, Jr. and T.E Nowicki, ''Microwave substrates support MIC technology'' Microwaves, Part I, Vol. 19, No. 12, pp 74-80, Nov, 1980.

[10] Duffy, S.M.,"An enhanced bandwidth design technique for electromagnetically coupled microstrip fed patch antennas," IEEE Trans. Antenna and Propagat.,Vol.4,236-238,2000

[11] Design comparison between HiperLAN/2 and IEEE802.11a services, By Emil Edbom, Henrik Henriksson (2001).

[12] D. M. Pozar and D. H. Schaubert, Microstrip Antennas: The Analysis and Design of Microstrip By the formula: Antennas and Arrays, IEEE Press, 1995.

[13] Roy, J. S. and M. Thomas, "Compact and broadband Microstrip antennas for next generation high speed wireless communication using HIPERLAN/2," International Journal of Microwave Science and Technology, Vol. 2007, 1–4, 2007.

[14] Zealand manual and Zealand software, INC, E-mail: zeland@zeland.com, Web Site: http://www.zeland.com

◈◈◈

Hardware - Supported Fault - Tolerance For Interconnect Concurrent Transputer Based Multiprocessor Systems

S.S. Nayak[1], R.K. Mishra[2], M.N. Murty[3] & B. Padhy[4]*

[1]Department of Physics, JITM, Centurion University, Paralakhemundi, Odisha, India,
[2]Department of Electronic Science, Berhampur University, Berhampur, Odisha, India
[3]Department of Physics, NIST, Berhampur, Odisha, India,
[4]Department of Electronic Science, Berhampur University, Berhampur, Odisha, India,

Abstract - This paper presents the design of a high performance transputer based fault-tolerant multiprocessor system for critical applications such as aircraft control, nuclear power station control, satellite applications etc. Fault-tolerant building blocks designed with potential for real time processing. The systematic architecture, regularity and recursiveness enables the system to be more fault-tolerant. The design is based on dynamic Triple Modular Redundancy such that the application process can survive upto the concurrent faults and masking the faculty one. The link failure, processor failure and system reliability is discussed in this paper.

Key words *- Fault-tolerance, Multitransputer System, Parallel processing, Reliability, Reconfiguration*

I. INTRODUCTION

Dependability and reliability are important aspects of complex parallel computing system used for various real time application such as aircraft control, nuclear power station control, satellite applications can not afford potential catastrophe in which human lives are at stake due to computer faults. Therefore fault-tolerant building blocks are designed based on transputers (IMS-T800) flexible and cost effect in nature.

This design provides transparent protection from permanent module failures based on multiple modular redundancy. Though, transporter is a link based processor, it is easy to design a parallel machine that encapsulate acceptance testing, fault masking, reconfiguring the network. A dynamic triple modular redundancy scheme employed basing on four transputers. Four transputers form a module and four modules from a nature with fault-tolerant mechanism interconnecting the four bidirectional links of each module. One line of each module dedicated to link adaptor for I/O to peripheral instruments. The reference becomes crazy due to permanent faults, transient faults, software faults, operation errors etc. Hardware faults due to memory crazy and short circuiting etc[1],[2] and errors due to information failure[2]. Faults due to severe environmental conditions includes the transient faults[1]. To challenge the above fault scenario, multilink processor[18] is preferred while designing the network.

Hence, the present transputing systems provides dynamic fault recovery applications in MIDM architecture[3]. The multiple link mechanism is adopted for better group communication[4] in the network.

II. TRANSPUTER OVERVIEW

The IMS T800 transputer is a 32 bit CMOS microcomputer with a 64 bit floating point unit and graphics support It has 4 Kbytes on-chip RAM for high speed processing, a configurable memory interface and four standard INMOS communication links. The instruction set achieves efficient implementation of high level languages and provides direct support for the Occam model of concurrency when using either a single transputer or a network. Procedure calls, process switching and typical interrupt latency are sub-microsecond. For convenience of description, the IMS T800 operation is split into the basic blocks shown in Fig. 1

The processor speed of a device can be pin-selected in stages from 17.5 MHz up to the maximum allowed for the part. A device running at 30 MHz achieves an instruction throughput of 30 MIPS peak and 15 MIPS sustained. The extended temperature version of the device complies with MIL-STD-883C.

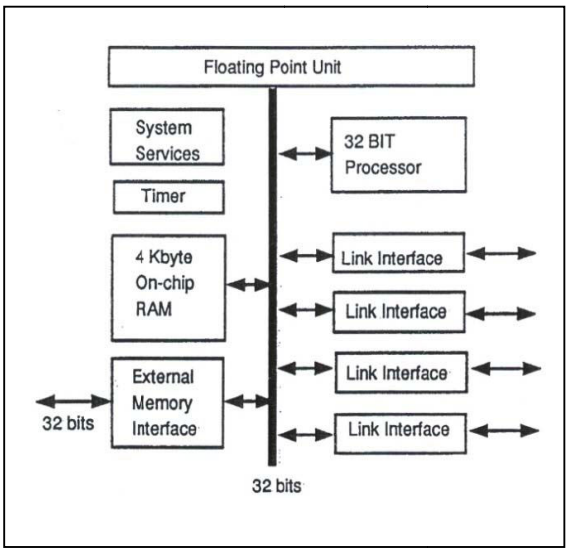


Figure-1

The IMS T800 provides high performance arithmetic and floating point operations. The 64 bit floating point unit provides single and double length operation to the ANSI-IEEE 754-1985 standard for floating point arithmetic. It is able to perform floating point operations concurrently with the processor, sustaining a rate of 2.2 Mflops at a processor speed of 20 MHz and 3,3 Mflops at 30 MHz.

High performance graphics support is provided by microcoded block move instructions which operate at the speed of memory. The two-dimensional block move instructions provide for contiguous block moves as well as block copying of either non-zero bytes of data only or zero bytes only. Block move instructions can be used to provide graphics operations such as text manipulation, windowing, panning, scrolling and screen updating.

Cyclic redundancy checking (CRC) instructions are available for use on arbitrary length serial data streams, to provide error detection where data integrity is critical. Another feature of the IMS T800, useful for pattern recognition, is the facility to count bits set in a word.

The IMS T800 can directly access a linear address space of 4 Gbytes. The 32 bit wide memory interface uses multiplexed data and address lines and provides a data rate of up to 4 bytes every 100 nanoseconds (40 Mbytes/sec) for a 30 MHz device. A configurable memory controller provides all timing control and DRAM refresh signals for a wide variety of mixed memory systems.

System Services include processor reset and bootstrap control, together with facilities for error analysis. Error signals may be daisy-chained in multi-transputer systems.

The standard INMOS communication links allow networks of transputer family products to be constructed by direct point to point connections with no external logic. The IMS T800 links support the standard operating speed of 10 Mbits/sec, but also operate at 5 or 20 Mbits/sec. Each link can transfer data bi-directionally at up to 2.35 Mbytes/sec.

The transputer is designed to implement the Occam language, detailed in the Occam Reference Manual, but also efficiently supports other languages such as C, Pascal and Fortran. Access to the transputer at machine level is seldom required, but if necessary refer to the *Transputer Instruction Set - A Compiler Writers'Guide.*

This data sheet supplies hardware implementation and characterization details for the IMS T800. It is intended to be read in conjunction with the Transputer Architecture chapter, which details the architecture of the transputer and gives an overview of Occam.

III. SYSTEM ARCHITECTURE

The architecture is designed is regular and recursive and the resultant system yields low cost for fault tolerance due to avoidance of roll-back accuracy and small scale of synchronizations.

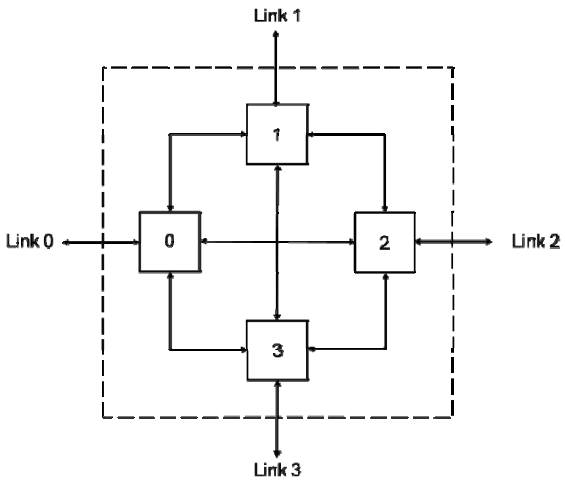


Figure-2 : Single Module

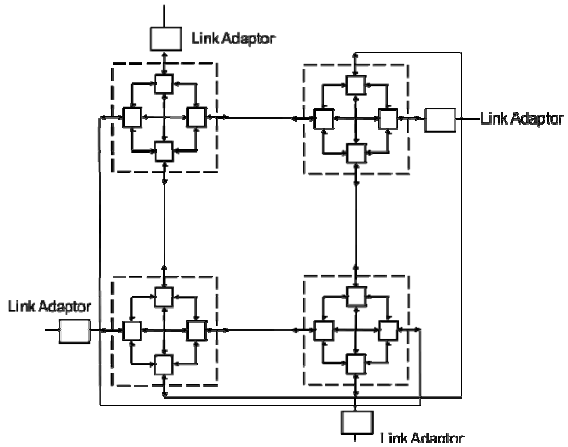


Figure-3 : Hardware Architecture [4 modules]

Each module consisting of four transputer nodes with their own private memory four bidirectional communication links and a copy of operating system. The hardware architecture is shown in figure in the form of grid connection.

a) Operation during multiple link failure

Multiple link failure may result in a network partition. Processor failure is assumed when all the internal links appear to have failed. Consider first the simple one of processors two links filing as in figure. The '0' node scatters the data packet to the available working links, only one node receives it and ACKs are then scattered by the recipient in all the working internal links in response to the packet.

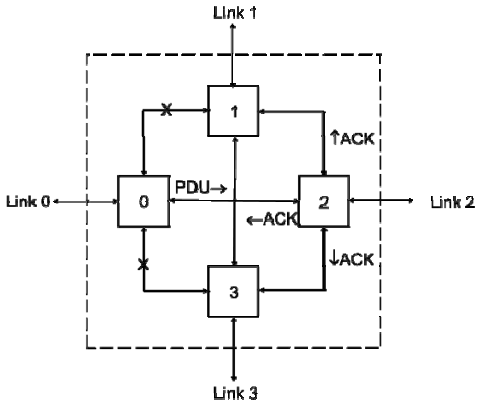


Figure-4 : Node '0' scatters PDU. Node '2' receives PDU

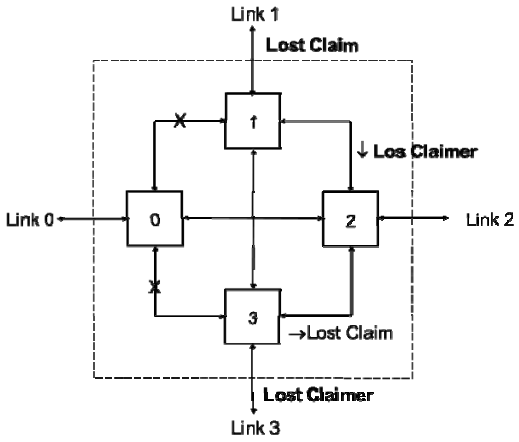


Figure-5 : Node '1' and Node '3' observe and claim it.

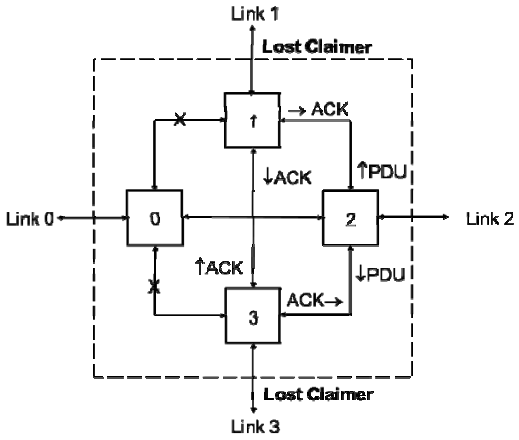


Figure-6 : Lost claimers ACK the PDU received.

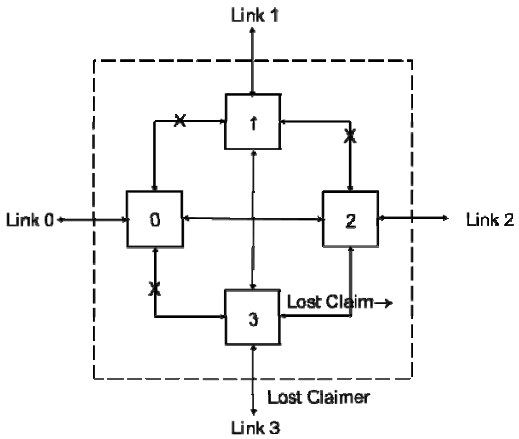


Figure-7 : Direct ACKs are returned to the Originator

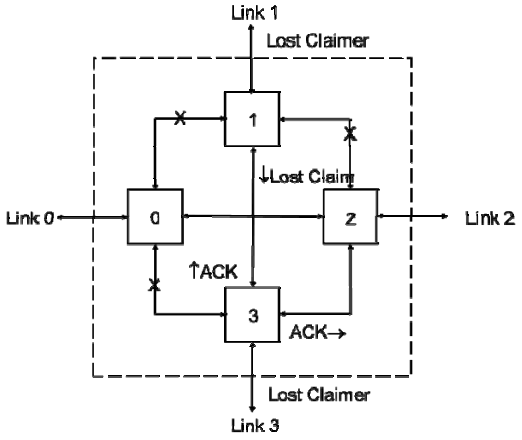


Figure-8 : Node '1' scatter PDU and Node '2' receives and ACKs it.

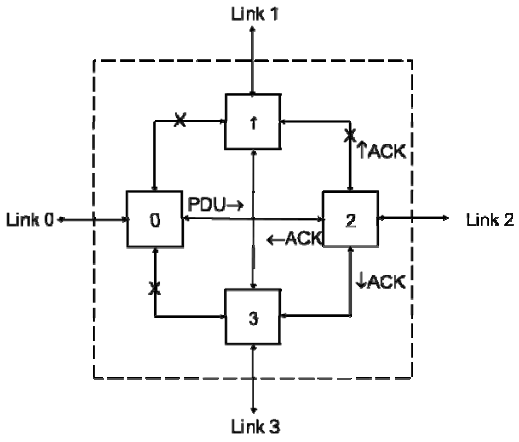
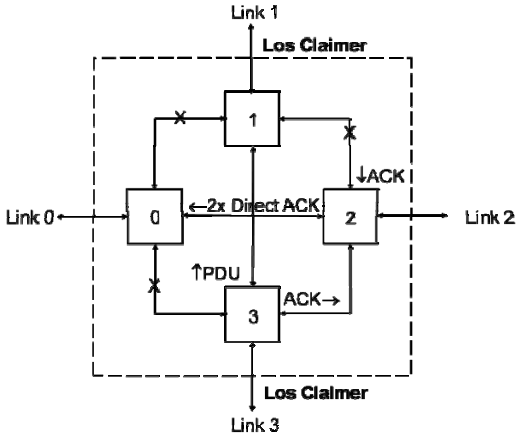
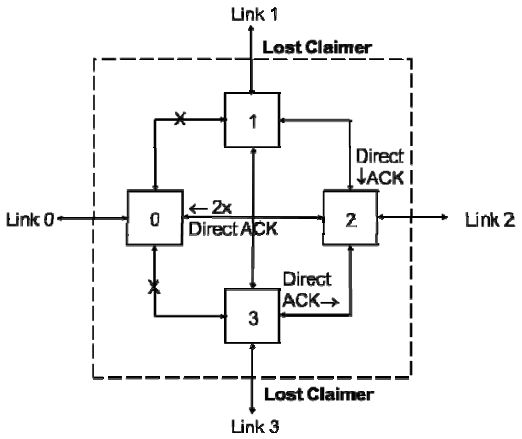


Figure-9 a,b,c : Node 1 receives the PDU and ACKs the reception of the PDU



(a)



(b)

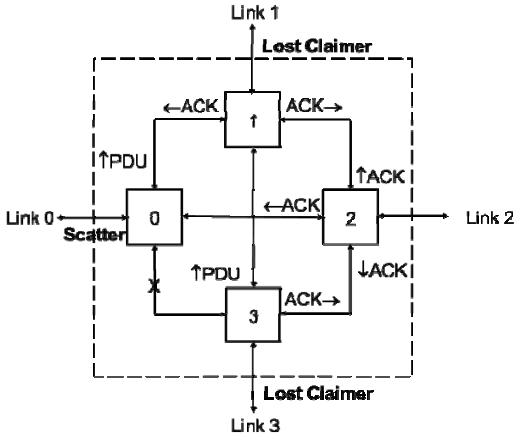
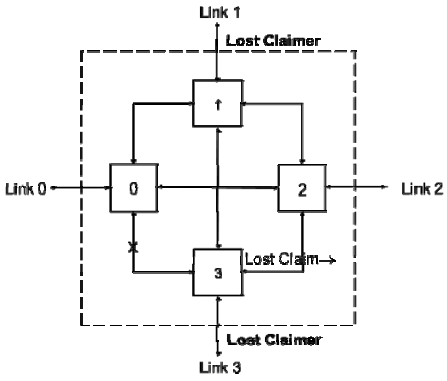


Figure-10 : Node '0' scatters PDU and Node '1' and '2' receives and ACKs it.



(a)

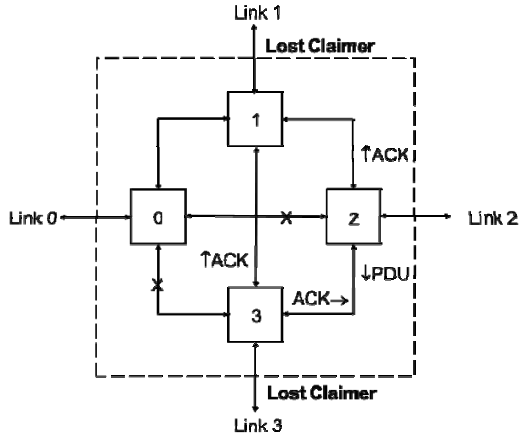


Figure -11a&b : Node '2' and 'e' observe and claim it.

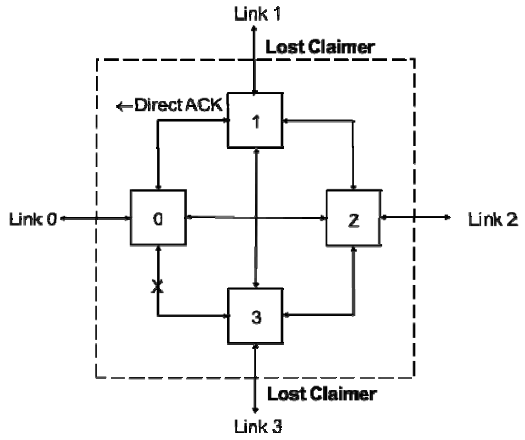


Figure-12 : Direct ACK can reach the originator

The other two links of the recipient are working other two nodes 2 and 1 observe the reception through the ACKs. The recipient sends the data packet to the other two nodes in response to the reception of the lost claim packets. The ACKs are returned to the originator (direct ACK) are also directed to the recipient (Node '0'). When ACKs are scattered form the lost claimer the ACK distributed to the originator is 1st sent to the Node '0' that provided the claimed packet. It is then forwarded by the peer who should have a working link to reach the originator, provided that no further link failure occurred so far.

If the link between the two claimers is working, the observed ACK from either should be received by the other, by now both claimers should have gathered the set of packets i.e. one message packet and two observed ACKs. The previous recipient should now also have gathered the packets in response to the scatter.

Now, there the case of three links fail, and a single link connection in the network. Same procedure adopted as shown in figure.

The direct ACK is returned to a claimer, but this time the link is not working any longer.

The algorithm is designed so that if any other link is still working than ACK is forwarded to that link otherwise the ACK will have to be returned to the link through which it came.

The signal ACK sender now train the only link left to the processor to reach the originator.

b) Operation in the event of processor failure

The detection of processor failure is very important to stop meaningless waiting or sending attempts. To know the status of processor, each time data packet and control packet is exchanged, the update knowledge about link status of all processor attached to the header of the packet. Each node periodically updates the link status vectors as the exchange of packets takes place with other nodes. A faulty processor will be excluded from the network during the protocol operation until it is repaired.

c) Watchdog monitoring

A watchdog timer is associated with each link and gets reset when a correct packet is received from the associated link. When a pending link output queue is empty, the write worker will send a control packet to each link on a periodic basis[18] One a link failure is detected, the associated failure flag is turned on. Next set of timers is employed to detect the node failures. Each application process has an associated timer.[4] If a packet as an observation of new node service arrives at any module, the associated timer for the identified process is started to detect the replica process failure to participate in the network communication.

III. SYSTEM RELIABILITY

To keep the system in reliable operation mode the following parameters are used. Let 'C' be the probability that the fault is detected and 'r' be the probability that the fault is repaired after the fault detected in the system. 't' be the processing time and 'λ' be the fault rate during power on. Taking the modular designs, the fault distribution Rm(t) is expontial and identical for all modules i.e. Rm = e-λt since the modules are symmetric independent of each other except when the repair occurs.

Let the time between the consecutive acceptance tests ≤T. Where T is some constant C and 'r' are constants [5]

To obtain system reliability R….. the system failure probability …..., has three components No. 1 F1 is due to concurrent faults occur during T. No. 2 F2 is the sequential fault occur over time t where t >>T No. 3 F3 results from fault repairs corresponding to the case when faults occur in two modules. One fault is detected when the other is not and the undetected fault in the repairer.

$$F_1 = (1-R_m)^4 + \binom{4}{3}[1-R_m]^3 R_m \quad \text{if } t \le T \quad (1)$$

Where F1 is the probability that all four modules have faults or three out of four modules have faults during 'T'.

Putting $R_m = e^{-\lambda t}$ in Eqn. (1).

$$F_1 = \left(1-e^{-\lambda t}\right)^4 + \binom{4}{3}\left[1-e^{-\lambda t}\right]^3 e^{-\lambda t} \quad t \le T$$

The failure probability F_2 comes from the situation in which the system has suffered two sequential faults. Since all modules are symmetric to each other, so F2 is the sum of all the sequences[6]. The fault detection probability 'c' and the probability successful repair 'r' can change the system failure probability substantially, for four sequences

$$F_2 = 4\left[(1-c)+c(1-r)\right]^2 \int_0^t \frac{d}{dt_1}(1-R_m) \int_{t_1}^t \frac{d}{dt_2}(1-R_m) \int_{t^2}^t \frac{d}{dt_3}(1-R_m) R_m dt_3 dt_2 dt_1, \quad \text{if } t > T \quad \ldots (2)$$

Where (1-c) + c (1-r) is the probability that a fault is not detected.

Substituting $R_m = e^{-\lambda t}$ in (2)

$F_2 = [(1-c) + c(1-r)2]\ [1-6e^{-2\lambda t} + 8\ e^{-3\lambda t} - 3e^{-4\lambda t}]$ if $t > T$

Or $F_2 = [(1-c) + c(1-r]^2\ (1-\ e^{-\lambda t})^3\ (1+3\ e^{-\lambda t})\ t > T$

… (3)

Let F3 is the probability that faults occur on only two modules, one of them is detected and other not and the latter appears to have successfully repaired the former using its runtime context[8].

$$F_3 = 2x\frac{1}{3}\binom{4}{2}(1-c)cr(1-R_m)^2 R_m{}^2 \quad t \le T \quad \ldots (4)$$

The system failure probability is

$$F_c^1 = \begin{cases} F_1 + F_3 & \text{if } t \le T \\ F_2 & \text{if } t > t \end{cases} \quad \ldots (5)$$

Thus the reliability is $R_c^T = 1 - F_c^T$

Eqn (3) reflects the contribution to the reliability from the online forward fault repair. It shows that higher values of the probability C or r, lower than system failure probability [7][9]

V. NETWORK RELIABILITY

The probability that packet exchanges between a pair of nodes can be conducted in the event of link failure in the network is defined as the network reliability Rc[11]. The reliability between two nodes of the network is employed by the very such that a packet sent from one module with two extra ACKs. So, that the receiver to know the packet for proper action, thus the reliability between two modules in the system is increased by the use of redundant ACKs [10].

Let Rl be the reliability of a transputer link Rc be the reliability of the connection between two modules across a link. The probability of a connection failure is

$$F_c = (1-R_l)^2 + 2\ (1-R_l)^2\ R_l[(1-R_l)^2 + R_l\ (1-R_l^2)] \quad \ldots (6)$$

Eqn. (6) can be simplified as

$$F_c = (1-R_l)^3\ (1+2\ (1-R_l)^2)\ R_l) \quad \ldots (7)$$

So, the network reliability Rc is

$$R_c = 1-\ (1-R_l)^3\ (1+2(1-R_l^2)R_l) \quad \ldots \quad (8)$$

There are five possible routes in all possible link failure provided that the system is still connected. The

extra ACKs scattered in response to the data packet received, transform the transmission in the four opportunities for the intended receiver for active actions.

VI. FAULT INJECTION

Several research papers have been published on fault injection into the live systems[16][17]. Several research groups have developed powerful tools to inject faults by software[12],[13]. The major advantage of simulation based fault injection[14] is the observability of all components which have been module. Our network system depicts a simulation based fault injection approach as in Fig. 12. All systems components have to be modelled in the VHDL hardware description language[15] by using standard synthesis tool and gate level descriptions. Our mechanism uses extended cell library to evaluate fault coverage and fault latency automatically.

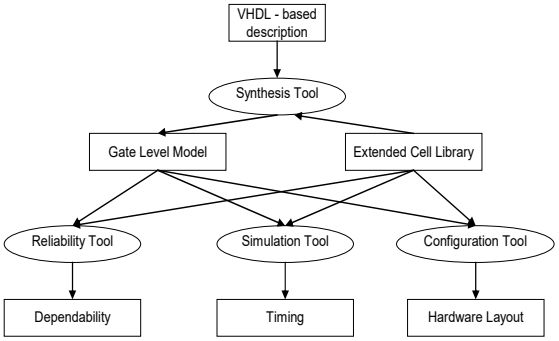


Fig. 13 : Evaluation of Dependability using reliability tool

VII. CONCLUSION

The designers of the parallel system can not just demand that no node or module of the system should fail, in the event of failure, it is required to add redundancy and reconfigurability. This network provides very good group synchronisation improves communication mechanism. The low cost in transferring the ACKs for packet routing reduces the network operative overhead. In our design, the system reliability can be remarkably improved with the mechanism of forward fault repair and redundancy saves the cost of accessing persistent I/O devices. The concurrency control overhead is eliminated due to non-sharing of virtual memory. Definitely this design is highly efficient compared to existing configurations.

REFERENCES

[1] D. Siewiorek, and R. Swarz "The theory and practice of reliable System Design", 1982 by Digital Press.

[2] R. Koo and S. Toueg, "Check pointing and Rollback - Recovery for Distributed Systems", IEEE Trans on Software Engineer, Vol. SE-13, No. 1, January 1987.

[3] R. Beton, J. Kingdon and C. Upstil, "Highly Availability Transputing Systems", Proceedings of the World Transputer User Group Conference. April, 1991.

[4] F. Christian, B. Dancey, and J. Dehn. "Understanding Fault-Tolerant Distributed Systems", Invited paper, 20th Annual International Symposium on Fault-Tolerant Computing, June, 1990.

[5] K. Kim and J. Yoon, "Approaches to Implementation of a Repairable Distributed Recovery Block scheme", Ann Int. Symp. On Fault-Tolerant Computing, 1988.

[6] O. Selin, "Fault-Tolerant Systems in Commercial Applications", Computer, IEEE, August 1994.

[7] J. Ortiz, "Transputer Fault-Tolerant Processor", Proceedings of the Third Conference of the North American Transputer Users Group, 1990.

[8] R. Oates, J. Kerridge, "Adding fault Tolerance to a Transputer-based Parallel Database Machine", Proc. Of the World Transputer User Group Conference, 1991.

[9] R. Strom, D. Bacon, S. Yemini, "Volatile Logging in n-Fault-tolerant Distributed Systems", Annu Int. Symp. On Fault-Tolerant Computing, 1988.

[10] Y. Chen, T. Chen, "Implementing Fault-Tolerance via Modular Redundancy with Comparison", IEEE Trans. On Reliability, Vol. 39, No. 2, June, 1990.

[11] D. Cheriton and W. Zweanepoel, "One-to-Many Interprocess Communication in the V-System",

Report STAN-CS-84-1011, department of Computer Science, Stanford University, August 1984.

[12] J. Barton, E. Czeck, Z. Segall and D. Siewiorek, Fault Injection Experiments using FI-AT, IEEE TOC, Vol. 39, No. 4 [1990], 575-582.

[13] J. Carreira, H. Medeira and J.G. Silva, Software fault injection and monitoring in processor function units, In: Preprints DCCA-5, Dependable computing for critical Applications, Urbana Champaign 1995, 135-149.

[14] E. Jenn, J. Arlat, M. Rimen, J. Ohlsson and J. Karisson, Fault Injection into VHDL Models: The MEFISTO Tool, In: Proc FTCS-24 [1994], 66-75.

[15] P. Ashenden, The VHDL - Cookbook, Technical Report, Univ of Adelaide, South Australia [1990].

[16] R.K. Iyer, Experimental Evaluation, in : Proc. FTCS-25 [1995] 115-132.

[17] J. Karlsson, P. Liden, P. Dahlgren, R. Johansson and U. Gunneflo, Using heavy-ion radiation to validate fault - handling mechanisms, IEEE Micro Vol. 14, No. 1 [1994], 8-23.

[18] Inmos, “The Trnasputer Data Book”, Second Edition, 1989.

◈◈◈

Parallelization of Hierarchical Text Clustering on Multi-core CUDA Architecture

Atul Bagga & Durga Toshniwal

Department of Electronics and Computer Engineering Indian Institute of Technology Roorkee, Roorkee-247667, India

Abstract **-** Text Clustering is the problem of dividing text documents into groups, such that documents in same group are similar to one another and different from documents in other groups. Because of the general tendency of texts forming hierarchies, text clustering is best performed by using a hierarchical clustering method. An important aspect while clustering large text databases is that of high dimensionality of the representation space. Not only does it take lot of space in storing hierarchy trees but also a lot of time is spent in similarity calculations while clustering these documents. In this paper we propose to parallelize a method which uses a tree based summarization technique to store cluster summaries in a tree stored in the memory at all times of processing. The results show that our method shows good accuracy along with a good speed up in calculating clusters.

Keywords*-hierarchical clustering; Text clustering; Tree based compression ; BIRCH clustering;.*

I. INTRODUCTION

Large amounts of data are being generated nowadays in form of text. Internet which is growing very rapidly has data which is mostly in form of text and this text is in the unstructured format. Emails, blogs, web pages, online news, the sources are just too many. The capacity to store data(text) has increased manifolds and cost to do so has decreased. Just to have a better idea of how large the text database has grown Eric Schmidt, Google CEO in 2005 stated that Google which is considered the best search engine ever has indexed about 170 TBs of data and that it is just a small fraction of the total internet data which is expected to be around 5 million TB at that time. And as we know bulk of this internet data is text which is there in an unorganized way, and does not follow any specific structure for representation. Thus it is difficult to fully utilize this text. So to make use of this text we need to keep it in an organized way. Text clustering is the method of organizing text documents into groups of similar documents.

Clustering is an important data mining problem. There have been several studies and researches on the subject of clustering specialized in the field of text.[1,2,3]

A good survey of clustering algorithms exist in [4] which covers research done in the field in the past as well as recent times. After the study of the various techniques that have been proposed to be used in text it is inferred that for the purpose of clustering text we need to adopt a clustering method which fits the needs of the high dimensional, large datasets of text. And as text generally follows a hierarchical classification we need to adopt clustering method which can detect clusters or multi cluster belongings of a document at various levels of hierarchy. [5,6]

Unstructured text unlike other forms of data does not have a standardized representation. We require adopting methods to transform text documents into numerical vectors on which clustering could be performed. Even preprocessing becomes an important part in text mining algorithms because data to be clustered properly first needs a good representation model. Thus in our approach we have designed a preprocessor which takes several small measures to make sure that the vectors we get represent the text in a best possible way.

In this paper we parallelize a hierarchical clustering algorithm for clustering unstructured text documents. The clustering algorithm chosen is fast as well as depends on compact representation such that the scalability of the algorithm remains high and it can be used for very large sized datasets as well. Through parallel processing of computations involved on the high dimensional vectors we decrease the time complexity of the algorithm and achieve high speed-ups on running the code on machine with a multi core CUDA GPU.

The paper is further organized as follows- In section 2 we discuss the multi-core CUDA architecture, in section 3 BIRCH clustering approach is explained. Section 4 explains the proposed parallelization of BIRCH and using it for text clustering. Section 5

contains the details of the dataset used in our experiments and obtained results and the paper is concluded in section 6.

II. MULTICORE CUDA ARCHITECTURE

CUDA (Compute Unified Device Architecture) is a parallel programming model and software environment developed by NVIDIA [7]. CUDA gives the developers access to virtual instruction set and memory elements capable of parallel computation in form of CUDA GPUs.

CUDA GPUs act like CPUs with parallel throughput that is instead of a fast processing of a single thread multiple cores of a GPU process multiple threads slowly but in parallel.[8]

The GPUs initially were specialized processor designed to answer the demands of real-time high resolution 3D graphics. But recently the modern GPUs have evolved into highly parallel multi core systems. These systems allow very efficient manipulation of large blocks of data. Thus the parallel processing GPU design is more effective than general-purpose CPUs and especially for algorithms where processing of large blocks of data is required in a quick time. [9]

The present generation GPUs are so fast that if the size of data is small we o reaching the bottleneck of performance and the power of the GPU is not getting fully utilized.

Thus GPUs can implement many parallel algorithms directly using graphics hardware. Well-suited algorithms that leverage all the underlying computational horsepower often achieve tremendous speedups.

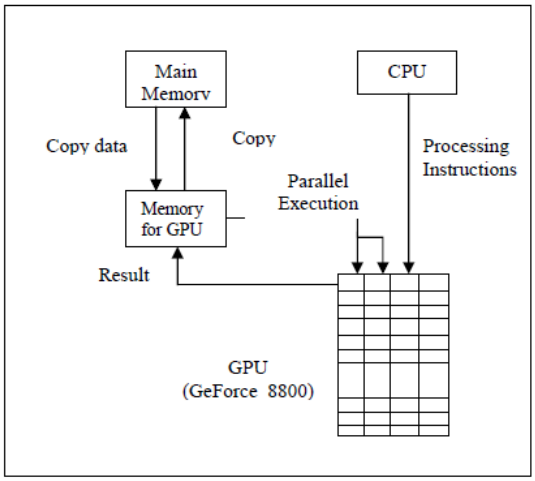


Fig. 1 : Processing flow on CUDA

III. BACKGROUND THEORY

Various approaches could be used for clustering. For the case of text hierarchical methods should be preferred because texts show natural presence of hierarchies in taxonomy. The main drawback of hierarchical clustering algorithms is the high time complexity and consumption of too much space to store the hierarchy tree formed during clustering process.

Here we have explained a hierarchical clustering approach which uses a hybrid of hierarchical clustering and iterative partitioning to form clusters. The use of hybrid approach in this algorithm makes it fast. And the use of compression tree data structure called clustering feature tree to store the tree reduces the memory space consumed in storing the trees.

The algorithm is explained below-

Balanced Iterative Reducing and Clustering Using Hierarchies (BIRCH)

Our technique is based on the popular BIRCH clustering approach [10]. BIRCH is a very effective hierarchical clustering approach especially for large datasets. The complexity of BIRCH algorithm is O(n) which is very fast compared to other hierarchical clustering methods.

BIRCH uses a two way clustering, where in the first part or micro clustering phase it prepares a tree of cluster summaries stored in the main memory. In the second stage or the macro clustering phase it uses an iterative partitioning approach to allocate clusters to each item.

The summarization tree is known as Clustering Feature tree and the summary vectors are called clustering feature vector. A clustering feature vector stores a triplet <n,LS,SS> Where n is the number of items in the cluster, LS is the linear sum of vectors, and SS is the squared sum of vectors.

The summary stored in the clustering feature vector contains enough information to calculate parameters such as centroid, radius and diameter of the cluster, given by eq. 3.1, 3.2 and 3.3 respectively.

$$x_0 = \frac{\sum_{i=0}^{n} x_i}{n} \qquad (3.1)$$

$$R = \sqrt{\frac{\sum_{i=1}^{n} (x_i - x_0)^2}{n}} \qquad (3.2)$$

$$D = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n(n-1)}} \qquad (3.3)$$

BIRCH summary is both compact and accurate method to store the tree. Compact because it stores only the triplet summarizing the cluster information instead of the whole data. And accurate because the summary still contains enough information to calculate centroid, radius and diameter of the cluster precisely.

Once the cluster tree is formed in the main memory, in the second stage items are allocated clusters based on nearest centroid of the clusters in the tree.

The similarity in case of BIRCH is calculated on the basis of Euclidean distance formula. Therefore BIRCH is good only to find spherical clusters.

The similarity calculations in the case of BIRCH applied to text take place on very high dimensional vectors. To speedup the clustering process these calculations could be processed by CUDA GPUs in parallel. The next section explains what functions and calculations can be parallelized in text clustering using BIRCH.

III. IMPLEMENTATION FRAMEWORK

The text documents are initially in the unstructured format. The first and foremost step is to preprocess this text to form numeric vectors. The preprocessing phase consists of the following steps-

Tokenisation: In the step the document is broken into words or tokens.

Stop-Word Removal: Stop words are the words that are irrelevant to the processing to be performed. Words like articles, conjunctions, verbs etc. which have no impact what so ever in calculating similarity between documents are removed.

Stemming: It is the process of converting each word to its root form. Porter's stemmer or Porter's suffix stripping algorithm [11] can be used to perform this task.

Tf-idf: once the document has gone through the first 3 steps we get a bag of words where each word is in the root form. We apply the tf-idf weighing scheme to convert the document from bag of words to numeric vectors where each word in the vocabulary now represents a dimension and each document is represented in this high dimensional space.[12,13]

Special feature implemented in the preprocessor:

The weighing scheme is coded such that the words in the introductory paragraph of the text get more weight as compared to words occurring later in the text. This is done because generally the starting lines of any text contain the crisp of the matter talk about in the text. For example a news paper article will start by saying which place, person, or field it is linked to.

The words which occur rarely like a word occurring only 2-3 times in a text does not represent the document's domain. Such words can be removed to reduce the dimensionality of the vector space. And to limit the dimensionality the number of words taken to represent each document is limited.

Standard implementation of BIRCH works for numeric data. Once we have converted our documents into numeric vectors we can easily use the BIRCH algorithm to find hierarchical clusters in our documents.

When the number of documents in the database is large, the dimensionality increases manifolds and in that case the calculations which take place in creating the clustering feature tree increases. For each entry to be made to the tree it needs to compute the nearest cluster at each level of hierarchy. This distance is computed between the cluster centroid which can be calculated from the clustering feature vector of that cluster. All these calculations involve the high dimensional vector. In our approach we simply used the CUDA GPUs parallel processing to speed-up the part of the algorithm which involves these high dimensional computations. We have used CUDA's thrust libraries for C++, the library provides in built function implementations to carry out primary tasks like sorting, adding, negating etc in parallel using the multiple cores of GPUs.

The functions that can be parallelized are the ones which perform computations on the high dimensional text vectors.

For each entry made to the CF tree it involves multiple distance calculations between clusters. To make these calculations it involves computations between the CF vectors which further consist of LS and SS vectors. Performing all these calculations serially produces a significant overhead of time. Therefore to achieve speed-up all these computations have been implemented using the standard template library (thrust) for CUDA which provides simple built in functions to implement computations in parallel. The library is built to perform the task optimally using as many threads as it can use according to the size of vectors involved in calculations.

IV. EXPERIMENTAL RESULTS

The experiments are performed on the dataset commonly used for text mining. A brief explanation of the dataset is given below.

Datasets Used

20-Newsgroups dataset

Dataset consists of 20000 news articles. These articles fall into 20 categories with 1000 documents falling in each category. On a broad scale these 20 categories can be divided into 6 groups. Thus the dataset is a perfect example of texts falling into hierarchical distribution. [14]

Various versions of the data set are available on the internet hosted by various data repositories. We have used the raw text format version and preprocessed it on our custom designed preprocessing module for text.

20 newsgroups dataset is a perfect example showing that how texts generally fall under multiple categories in multiple levels of hierarchy when it comes classifying them.

The table below shows the various groups into which the documents belong. The bold headings in the table denote the primary level grouping while the sub categories are given in each of the 6 main groups.

Table 1: Hierarchical distribution of documents within the dataset

Computers	**Recreation**	**Science**
Graphics **MS Windows** **IBM Hardware** **MAC** **Hardware** **Windows X**	Autos Motorcycles Sports-Baseball Sports-Hockey	Cryptography Electronics Medicine Space
Miscellaneous	**Politics**	**Religion**
Forsale	Guns Mideast Miscellaneous	Miscellaneous Atheism Christianity

On conducting experiments on the dataset with various number of categories and sub categories we obtained the following results on accuracy of our method.

Table 2 Accuracy of our method at 2 hierarchical levels of clusters with varying no. of documents

No. of categories	**No. of sub categories**	**No. of documents**	**Accuracy**	
			Level 1	**Level 2**
2	2	2000	95	-
4	4	4000	91	-
4	8	8000	86	79
4	12	12000	78	72

All CUDA and sequential experiments were conducted on a PC with 4 GB RAM and an Intel Core 2 Duo E6750 2.66GHz processor running Windows 7 64-bit. An NVIDIA 8800GTX GPU with 768MB of memory was used for all experiments. All of the CUDA algorithms were written using CUDA version 2.3, while the sequential algorithms were written in C++ using Visual Studio 2010. NVIDIA graphs driver version 197.45 was used for the project. Thrust version 1.2 was used for the applicable algorithms.

The experiments were conducted on the dataset dividing it into various groups taking varying number of documents from different categories and the following results were obtained.

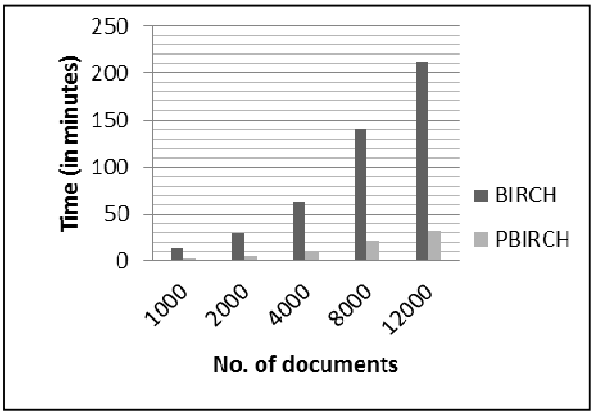
The table 3 shows the variation with no. of documents taken during each run of the experiment. And the relative speedup obtained in each case.

Table 3 Speedup obtained by using the parallel version compared to the serial implementation

No. of Documents	**Vocabulary Size (No. of words)**	**Timing for Creating Clustering Tree (In minutes)**		**Relative Speed-up Obtained**
		BIRCH (text)	**PBIRCH (text)**	
1000	7805	14.3	2.6	5.5
2000	12539	29.5	5.5	5.36
4000	16843	63.1	10.25	6.18
8000	18392	139.8	21.7	6.45
12000	21560	212.5	31.95	6.7

The graph below shows the time taken by the serial processing and parallel processing version of the BIRCH algorithm for text.

We get a 5-7 times speed up on the parallel version of the algorithm, represented as PBIRCH for text clustering. As we increase the number of documents we see that the speed up increases because of multiple parallel computations and compensation of the time taken in tree construction.



VI. CONCLUSION

In our work we have proposed the use of a hierarchical clustering algorithm for unstructured text. Thus the documents are grouped together at various levels of hierarchy.

Due to the use of data summarization tree the method works well even for large datasets. Thus limitation on the size of datasets is increased and algorithm scales well for even large sized datasets.

By use of parallel processing on Nvidia GPUs to compute similarity between document vectors we have reduced the time taken to compute the in memory clustering feature tree.

In the next generation of graphics cards that Nvidia plans to launch the cards will loaded with a feature that GPU cores can access main memory directly in that case the overhead of reading the data from the main memory and writing the results back will be reduced. As this overhead is significantly high for large datasets, the future generation cards will be able to produce even higher speedups.

REFERENCES

[1] Cui, X.; Potok, T.E.; Palathingal, P.; , "Document clustering using particle swarm optimization," Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE, vol. no. , pp. 185- 191, 8-10 June 2005

[2] Liping Jing, Michael K. Ng, Joshua Zhexue Huang, "An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data," IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 8, pp. 1026-1041, June 2007

[3] Gil-Garcia R., Pons-Porrata A., Dynamic hierarchical algorithms for document clustering. Pattern Recognition Letters 31 (6), pp.469-477, 2010.

[4] Premalatha K. and Natarajan A.M, A literature review on document clustering, Information Technology Journal 9, 993-1002, 2010

[5] H. Sun, Z. Liu, and L. Kong, "A Document Clustering Method Based on Hierarchical Algorithm with Model Clustering", in Proc. AINA Workshops, pp.1229-1233, 2008.

[6] Gil-Garcia R., Pons-Porrata A., Dynamic hierarchical algorithms for document clustering. Pattern Recognition Letters 31 (6), pp.469-477, 2010.

[7] D. Luebke and G. Humphreys, "How GPUs Work," Computer Magazine, IEEE Computer Society, vol. 40, no. 2, pp. 96-100, Feb 2007.

[8] NVIDIA Corporation, CUDA Programming Guide, Version 3.0, 2010.

[9] Nvidia Corporation, CUDA Architecture Overview, Version 1.1, 2009

[10] Tian Zhang, Raghu Ramakrishnan, Miron Livny; "BIRCH: an efficient data clustering method for very large databases", In SIGMOD '96: Proceedings of the 1996 ACM SIGMOD international conference on Management of data, pp. 103-114, 1996

[11] Porter, M.F.: An algorithm for suffix stripping. Program, Vol. 14, No. 3, 1980

[12] Na Wang; Pengyuan Wang; Baowei Zhang; , "An improved TF-IDF weights function based on information theory," Computer and Communication Technologies in Agriculture Engineering (CCTAE), 2010 International Conference On , vol.3, no., pp.439-441, 12-13 June 2010.

[13] Lee, D.L.; Huei Chuang; Seamons, K.; , "Document ranking and the vector-space model," Software, IEEE , vol.14, no.2, pp.67-75, Mar/Apr 1997

[14] 20 newsgroups homepage, http:// people.csail.mit.edu/ jrennie/ 0Newsgroups/ , Online Accessed 10th of April 2011.

◈◈◈

Facility Planning & Design And the Use of Software

R. C. Nayak, S. R. Patnaik & H. S. Moharana
Dept. of Mechanical Engineering, Raajdhani Engineering College, Bhubaneswar

Abstract **-** The layout of facilities is an important determinant of operating efficiency and costs. Whenever the flow of materials or people is complex, alternative approaches for facilities design offer feasible means of developing and evaluating alternative arrangements is necessary. Facility planning is very important in a manufacturing process due to their effect in achieving an efficient product flow. It is estimated that between 20%-55% of the Indirect Operating Expenses in manufacturing is related to material handling. This cost can be reduced until 30% through an effective facility planning. Proper analysis of facility layout design causes to improve the performance of production line such as decreasing bottleneck rate, minimizing material handling cost, reducing idle time, raising the efficiency and utilization of labour, equipment and space. A new type of computer-aided engineering environment is envisioned which will improve the productivity of manufacturing/industrial engineers. This environment would be used by engineers to design and implement future manufacturing systems and subsystems. By using software it is possible as overall vision of the proposed environment, identifying technical issues which must be addressed, and describes work on a current prototype computer-aided manufacturing system engineering environment.

I. INTRODUCTION

Facility layout design determines how to arrange, locate, and distribute the equipment and support services in a manufacturing facility to achieve minimization of overall production time, maximization of operational and arrangement flexibility, maximization of turnover of work-in process and factory output in conformance with production schedules. Computerized tools are used on a very limited basis. Given the costs and resources involved in the construction and operation of manufacturing systems, the engineering process must be made more scientific. Powerful new computing environments for engineering manufacturing systems could help achieve that objective.

II. SOFTWARE FOR PROJECT PLANNING

The Project Planning tasks ensure that various elements of the Project are coordinated and therefore guide the project execution.

- Project Planning helps in :

  - Facilitating communication
  - Monitoring/measuring the project progress, and
  - Provides overall documentation of assumptions/ planning decisions.

- The Project Planning Phases can be broadly classified as follows:
  - Development of the Project Plan
  - Execution of the Project Plan
  - Change Control and Corrective Actions

The Project Planning tasks ensure that various elements of the Project are coordinated and therefore guide the project execution.

Tree Grid Gantt Chart is a project management tool for building Gantt charts online on web, Resources for tasks, work and material resource types, task price calculation Resource availability and usage charts, standalone or included in Gantt. Tasks filtering by resources possible, Discrete bars for a real flow - actual completion of a task usable along with a task bar, Flags - any custom icons displayed along with a defined tooltip on certain dates, Tree Grid Gantt Chart provides all basic features for project management like Primavera or Microsoft Project.

Dependencies like descendants (successors) or ancestor (predecessors) or both (mirrored) between tasks.

III. SOFTWARE FOR FACILITY LAYOUT

The goal of the engineering process by using software is to find the best solution to a problem, i.e. a factory or subsystem implementation, given a specific set of requirements and constraints. Engineers must address the entire factory as a system and the interactions of that system with its surrounding environment. Component elements of the factory system include:

- The physical plant or buildings to house facility,
- The production facilities which perform,
- The technologies used in the production facility, i.e., processes, methods, and techniques,
- The work centers/stations, machinery, equipment, tools, and materials which comprise or are used by the production facilities,
- The various support facilities and systems which move and store materials, handle manufacturing by-products and waste, manage information resources, maintain machinery and information systems, and support other needs of personnel,
- The staff organization and mechanisms which are instituted to operate and maintain the manufacturing facility,
- The interface between the factory and its environment, e.g. movements of goods and materials, human access to the facility, links to utilities, and the controls on various forms of environmental impact.

LayOPT™ is an innovative facilities layout analysis and optimization software package which can be used by layout planners in the optimal solution of single and multiple floor facility layout problems.

It is a Windows-based software system with all the amenities of a user-friendly interface, including pull down menus, toolbars, status bars, user-defined window sizes, and an on-line help system. It comes with a User's Guide/Reference and a Training Manual. LayOPT's algorithm is a steepest-descent, two-way exchange optimization routine . In each iteration, the algorithm picks the department pair whose exchange leads to the largest reduction in the objective function. It then automatically exchanges the pair to proceed to the next iteration. The objective function minimized by the LayOPT algorithm is the sum of the parts flows multiplied by the appropriate costs and expected distances between all department pairs with non-zero parts flow between them.

LayOPT is an improvement algorithm that starts with an existing block layout, and given the flow and cost data, attempts to improve it by exchanging the locations of defined departments. While several available improvement algorithms perform basically the same function, many are severely limited by the kinds of exchanges they could perform.

In manufacturing systems, the three main types of layout are product layout, process layout, and group layout, which are further categorized into flow line, cell, and centre arrangements. The distinction among these types of layouts is made based on system characteristics such as production volume and product variety. Product layout, also called flow shop layout is associated with high volume production and low product variety, while process layout (job shop) is associated with low-volume production and high product variety.

Creating a project in LayOPT involves defining the basic input data required to describe an initial layout and execute an optimization run, namely: (a) the building or facility, (b) departments and departmental properties, (c) flow and cost values, and (d) an initial departmental block arrangement.

IV. TOOLS FOR OPTIMIZING LAYOUT DESIGN

The most well known heuristic methods in optimizing layout design are Tabu Search (TS),

Simulated Annealing (SA), and Genetic Algorithms (GA). The popularity of these heuristics has flourished in recent years and several published studies can be found in the literature.

Tabu Search is a mathematical optimization method, belonging to the class of local search techniques. Tabu search enhances the performance of a local search method by using memory structures, once a potential solution has been determined, it is marked as taboo, so that the algorithm does not visit that possibility repeatedly.

A genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems . Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (EA) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover, Simulated annealing (SA) is generic probabilistic meta heuristic for the global optimization problem of applied mathematics, namely locating a good approximation to the global minimum of a given function.

V. FACILITY DESIGN BY SOFTWARE

Design and layout represent the supporting facility component of service package. Factors influencing facility design: Nature and objective of organization; land availability; flexibility; security; aesthetics; community and environment. Design of facility has the greatest important where it directly affects the society.

The virtual buffer zone

The Virtual Facility acts just like a buffer zone, bringing all aspects of data center design, optimization

and management together in a flexible design area that can be tried and tested in as many ways as the planner likes, without risk to the real facility.

- The lifeline of any data center can be divided into two distinct phases:
  1. Design, construction and commissioning.
  2. Day-to-day operational management.

  Simulation in the data canter industry has focused on the design phase and point solutions. Coupled with the Six Sigma DC suite, the Virtual Facility enables to support the operational phase of the lifeline too.

A facilities design criteria are recommended by the management to improve customer service, material handling flows, and space utilization. The use of facilities design software and simulation tools, propose to develop a facilities design study and develop manufacturing simulations to consider what-if scenarios to improve current practices with the facilities redesign study.

There are many aspects of Facility Design including Plant Layout, Cell design, Material Handling, Warehousing and Distribution. Import CAD layouts to add a dynamic simulation of the movement of people and parts, Simulation models can also be used to improve and optimise Manufacturing Cell Design Watch model run with realistic animation and real-time statistics, Analyse the performance of the facility and identify improvements in the 3D results viewer, Plot how queues, lead-times and other KPIs change over time, Demonstrate how the facility runs in 3D with Virtual Reality Fly-through functionality.

ProModel is used by companies who are planning major layout changes to their factory, or warehouse. They may be extending existing facilities or even consolidating two or more operations into one. Inevitably issues of material flows, logistics and location of machines will affect the successful operation of the new facility.

ProModel simulation enables companies to evaluate alternative methods before implementing for real. Today, most companies can't afford the risk of "getting it wrong", let alone the cost. That's why simulation is common sense. It takes the risk out of major decisions and reduces the uncertainty of change. For example: Industrial Engineers use simulation to design new layouts, design cells, or change the production process. Logistics managers use simulation to help in determining the specification of equipment such as conveyors, sortation systems, racking, etc. before committing to purchase.material handling suppliers use ProModel as a sales aid to support proposals to clients. By simulating various options, both supplier and client can evaluate alternatives and both will then have confidence that the agreed solution will work.

VI. CONCLUSION

Analysis of facility design such as layout and material handling system is important in a manufacturing industry. Proper analysis of existing layout design could improve the performance of production line. It could decrease bottleneck rate, minimize material handling cost, reduces idle time, raise the efficiency and utilization of labour, equipment and space.

Facility layout improvements are possible to current systems. These benefits when are possible to be quantified represent important opportunities of improvement in the organization. Simulation tools and lean concepts are beneficial toolsets for assessment of major design criteria in facilities layout. The simulation output results give us a substantial reduction in cycle time. This will radically reduce WIP and with no additional floor-space needs. This will allow a greater inventory control with less investment and cost reductions in material handling with less quality control. With market challenges and new global competition, manufacturing companies must dramatically reduce product delivery time from conception to production in order to gain, or even retain, competitiveness.

REFERENCE

1. Arostegui. M, Kadipasaoglu. S., Khumawala. B. 2006. An empirical comparison of Tabu Search, Simulated Annealing,and Genetic Algorithms for facilities location problems. Springerlink, International journal of Production Economics 103 (2006) 742– 754
2. Askin. R. G. and Standridge. C. R., 1993, Modelling and Analysis of Manufacturing Systems, Wiley, New York.
3. Balakhrishnan, J., Cheng, C.H., and Kam, F.W.,. 2003. FACOPT : A user friendly Facility Layout optimization System. Journal of Computers and Operation Research. Vol. 30, pp. 1625-1641.
4. Balakrishnan. J, Cheng, C.H.; 2007. Multi-period planning and uncertainty issues in cellular manufacturing: A review and future directions, European Journal of Operational Research. Ed.177. page 281–309.
5. Ekren, B.Y., Ornek, A.M., 2008. A Simulation based experimental design to analyze factors affecting production flow time, Simulation Modeling Practice & Theory vol. 16, p. 278- 293,

◈◈◈

Self Diagonistics And Troubleshooting A Modern Day Smart Home Network

Debajyoti Pal

Department of Information, Camellia Institute of Technology, Kolkata, India

Abstract **-** The complexity of home networks has evolved to a greater level of sophistication and complicacy in the recent times comprising of heterogeneous components like at least two computers, web-enabled high-definition television sets, net-enabled blue ray disc players, iPods and many other such devices. Troubleshooting such a sophisticated *smart home network* in case of a malfunction by the novice end users seems to be very demanding. The paper proposes a Smart Home Network Monitoring System that provides a *centralized*, *general-purpose, automatic and convergent logging facility* with the purpose to auto-detect and possibly correct all such failure issues by having a well-defined set of *adaptive and incremental rule engine* that needs to be applied to the entire network in general. Logging of all *events* that happened *before* trouble appeared may give a greater insight and hence help in providing an effective and permanent troubleshooting mechanism. This paper also reports the initial experience of deploying such a facility.

Keywords*-Smart Home Network Monitoring System, General-purpose logging facility, Adaptive and Incremental Rule-Engine, Event, Troubleshooting.*

I. INTRODUCTION

Penetration of cheap broadband service in the past few years has lead to a surge in the home networking environment and subsequently the problems associated with it are becoming well-known. The ultimate goal of providing an integrated multimedia entertainment service has resulted in the emergence of smart home networks consisting of a number of sophisticated yet complex products like laptops, HDTV, Blue-ray disc players, tablets, etc that have ultimately created a plethora of problems for the ultimate home users. In fact the problems that plague the smart home networks though simple are a cause of great confusion and frustration among the end users because of their lack of knowledge and expertise. Misconfigured home networks are a great deal of concern from the security point of view also because they serve as attractive trap-doors for external attackers to exploit. The causes for home network failure can be many and thus tools and logging facilities that enable us to automatically monitor, record, detect and correct such issues will be welcomed.

Continuous monitoring and logging of home network traffic (both inward and outward) can be helpful to provide an insight to the problems that arise in such a network. Specifically what event(s) led to the malfunctioning of the home network might come into limelight by maintaining such a log and can come to be handy in designing an automated, adaptive, and incremental self-diagnostic rule matching system(engine).

Packet-monitoring tools like *tcpdump*, *Wireshark*, *Kismet,* etc helps us to monitor and log all the incoming and outgoing network traffic. All of them however suffer from the same drawback of being tied down to one specific host at a time. Further, in most of the homes presence of a NAT enabled router/gateway for establishing an Internet connection to the ISP server complicates the issue in the sense that it renders the outward traffic monitoring useless. Also due to being tied down to one specific host, multiple tools need to be present one for each host that is a part of the smart home network.

This clearly gives rise to redundant data that serves as a bottleneck for the bandwidth which is shared between the different active devices.

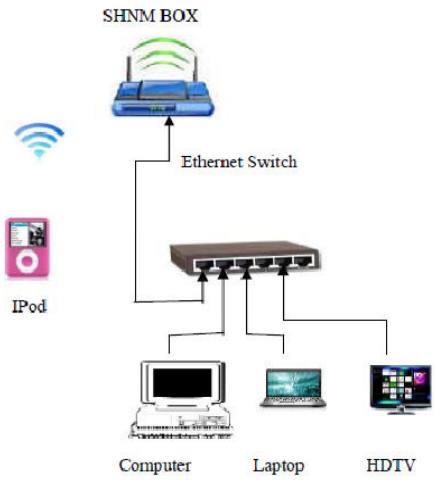
In this paper we propose a centralized, general-purpose, automatic, convergent logging facility that serves as a basis to auto-detect and correct all possible smart home network failures. Since we used Wireshark as the packet monitoring tool, hence a centralized logging facility is required so as to ensure that redundant data flow and hence bandwidth clogging is minimized. Thus the home network implies the presence of client/server architecture which ensures that all the incoming/outgoing traffic is forced to pass through the centralized device that houses the packet monitoring tool. The logging platform is a general purpose one because not only does the packet monitoring tool we deploy operating system neutral but it also supports a wide variety of network protocols from the application,

transport, network and data-link layers and of the TCP/IP stack. The logging facility is automatic because the packet monitoring tool at all times is running in the background and storing the events in a specified location of the storage disk. When a particular limit of the disc usage space is reached the recorded events are transferred to another secondary storage medium. The overall reliability of the system is increased by having the idea of primary/secondary storage in the event of primary storage failure. The centralized architecture that we follow automatically forces the entire system to be a convergent one because traffic from all possible locations are ultimately redirected to a centralized server that we already discussed. The aforesaid facility has got a close resemblance with a typical "Black-box" present in the aircrafts and we refer to the system that houses the monitoring, logging and troubleshooting facility as the Smart Home Network Monitoring System (SHNM).

The outline of the paper is as follows. Section 2 describes the configuration and functionality of our SHNM system. Section 3 deals with the potential applications of such a system. Section 4 deals with the specific requirements and the challenges that can be faced by the system. Ultimately Section 5 gives the detail of our experimental test bed and the scope of future work.

II. SHNM SYSTEM CONFIGURATION AND FUNCTIONALITY

The place of deployment of the SHNM logging facility is of utmost importance. We follow the typical client/server architecture model wherein a single system houses the SHNM facility and all the outgoing or incoming network traffic of any kind must flow through it. Such a scheme has been shown in the Figure 1 below. The configuration has provision both for wired as well as wireless devices. Since the total number of active network points can go well beyond 5 very easily we choose a 10 port Ethernet switch for the wired section which in turn is connected to one of the Ethernet ports of the SHNM system.



Wireless support is also provided by the SHNM system directly in the form of IEEE 802.11a/b/g/n standards. Although home networks with a more complex configuration can do exist, but we assume ours to be a sufficient one for at-least a couple of years to come by. Figure 2 depicts the overall SHNM system functionality.

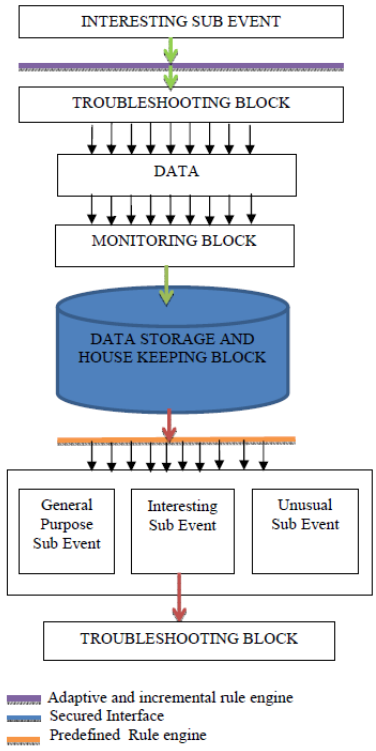


Figure 2. Overall SHNM System Functionality

As shown the SHNM system can be subdivided into 4 blocks namely:

i) Monitoring Block- It actually houses the packet monitoring software like Wireshark which is responsible for capturing all the home network data that are being generated and subsequently transmitted.

ii) Data Storage and Housekeeping Block- This block is responsible for storing all the packets that are being sensed by the Monitoring Block. Each and every packet is opened up and depending upon its contents a pre-defined rule-set is applied and the packets are transferred to a proper Event Generation Block.

iii) Event Generation Block- It actually can be subdivided into the following sub-blocks:

a) General-Purpose Event Sub-Block which contain the logs of all the incoming and outgoing network traffic under normal and healthy network operating conditions (i.e. no network malfunction).

b) Unusual Event Sub-Block which contains the logs of some rare network traffic like a new MAC address appearing for the first time, or modification of the configuration settings of a file that is rarely touched.

c) Interesting Event Sub-Block which contains the logs of certain filtered network traffic that might be attempting to update an operating system, updating some antivirus software or searching for device driver software's for a newly installed piece of hardware (maybe like a graphics card) or any other such related items.

iv) Troubleshooting Action Block- It is this block that has access to the Interesting Event Sub-block and is of prime importance. It houses specialized application program that takes appropriate troubleshooting measures if such a condition is detected. Thus this sub-block should obviously have access to all the sensitive user-data also that might pose to be a security threat or a breach of privacy. Hence the interface that is used by this sub-block to access the user-data should be done through a secure channel as depicted by a dotted line in Figure 2.

The primary application of the SHNM system is to provide support for troubleshooting and diagnosis when some things fail on a smart home network. If the SHNM system is widely adopted then such a service might be provided by a third party provider or by the ISP itself as a value added service on a chargeable basis.

III. APPLICATIONS OF SHNM SYSTEM

In this section we consider some applications of our SHNM System.

i) Automatic Troubleshooting and Future Prediction- This obviously is the prime reason to have our SHNM system in place. Studies by Sheehan *et al* shows that end-users often seek online help to troubleshoot problems of their home network that they are facing. Gathering the knowledge about millions of such end users spread all over the world we could easily produce a list that consists of the most commonly occuring home networking problems. Thus the key to success is to both learn and share any new information with everybody else on the community as and when it appears. So, by collaborating the experiences from different such households the troubleshooting block rule engine can be made to adapt itself to such changes and consequently update its own rule engine. Given a considerable period of time our SHNM System would gradually evolve to an automated Expert System wherein, it might suggest for example, a particular brand of network connected HDTV's creating some sort of a network configuration problem based upon the experience of other households. Thus, given an existing smart home network it can give a suggestion to the users before buying about the best possible alternatives of devices that are available in the market and which are compatible with their own home network thereby ensuring a quality and hassle-free service.

ii) Ensuring Quality of Service(QoS) in terms of Internet Speed- Poor Internet speeds are a common cause of concern in almost every household. It can be due to a improperly configured network or due to policies set forward by the ISP itself. To detect situations wherein a user's ISP is the cause of performance degradation (relative to speed)[8] Mukarram Bin Tariq *et al* have developed the Network Access Neutrality Observatory (NANO), which collects network-flow statistics from different households and attempts to isolate the cause of such performance degradations based upon a statistical model. Thus, this opens up an opportunity to intermix the NANO agent with our SHNM system so as to improve its intelligence to understand the reasons of poor internet speeds if any and hence take appropiate measures.

iii) As a means to improve Network Security- Intrusion Detection Systems, antispywares, antivirus softwares and other network security algorithms depend heavily on their ability to collect different types of relevant data from as many sources as possible to keep themselves updated to the latest available threats. Modern day scenario presents us with a very dangerous situation where the attackers could well be present in a smart home network as ours. The problem is even more

complicated because different home networks may be subscribed to different ISP's and generally they work in isolation to each other. Thus a collaborative SHNM system should be in place wherein the SHNM systems from different home networks interact among themselves, sharing the data they have with the sole aim of detecting any possible new vulnerabilities arising out of such network traffics.

IV. SPECIFIC REQUIREMENTS AND CHALLENGES FACED BY THE SHNM SYSTEM

i) Issue of privacy and its legal implications- It is evident by this time that in order to ensure effective troubleshooting, SHNM systems from various homes should inter co-operate among themselves. But in doing so we risk sensitive user informations and their personal preferences like the type of websites visited, personal credit card informations and so on to be at stake. Obviously, no user would ever want any outsider to have a see into the daily happenings of their household which should be kept as a secret. But in doing so the very basic concept of collaborative information collection mechanism would be violated. To further complicate matters in a country like India the Information Technology Act poses a hefty penalty or imprisonment for upto a few years on the ISP's who violate the privacy of their customers.

Thus the only solution that can be provided is to keep the SHNM System within the premises of a household only and to let the user of such a smart home network make a choice about which information is to be shared and which is to be not. Although it might sound to be a conservative approach, but right at this point of time it is the only best possible alternative available. Signing of a customer agreement form between the service provider and the customer may also be feasible solution.

The concept of automatic operating system software updates will work well with the SHNM system too given their widespread acceptance. In that case the SHNM system which is present in the household would regulary contact a centralized server of the service provider providing the SHNM service and keep the smart home network up to date.

ii) Storage Limitations- The problem of limited storage space is a very important one. The configuration that we used to test the system consisted of a modest 320 GB hard disk drive. Experiments revealed that for a full day of heavy Internet usage (consisting of 3 movie downloads, browsing the Internet and some e-mail exchanges) roughly 3 GB of disk space was utilized for storing the necessary records. This combined with the live streaming features being used on the HDTV's took up another 1 GB. Thus the entire disc space would be consumed in no more than 3 months. Hence periodic removal of the stored data to some offsite network storage device should be done at regular intervals. For example, data transfer from the SHNM system to any offsite network storage device can be scheduled at midnight of Sunday every week.

iii) Reliability Issue- The SHNM system we described should be robust and reliable. Specifically it should be immune for an acceptable period of time to power failures, or certain hardware configuration changes in the machine it is housed in. The design should be such that, in the event of any hardware failure the loss of log data should be minimum.

V. EXPERIMENTAL TEST BED AND SCOPE OF FUTURE WORK

We have implemented an initial prototype of the SHNM system as a tool to understand what exactly goes on in a home network. Our prototype design is based upon an Intel based system running Windows 7 Home Premium Edition as the operating system. Further a software called CCProxy is also installed to handle multiple connections(both wired and wireless) simultaenously. Internet is accessed through high speed EVDO technology being provided by BSNL.

Our SHNM system hardware has an Intel based system consisting of a Core i3-350M , 2.26 GHz processor, 3 GB RAM ,500 GB hard disk drive, 2 ethernet ports and support for Wireless LAN(802.11b/g/n). Default configuration restricts the commencement of an Internet session from inside the house only. The SHNM features are primarily being provided by an open source packet monitoring tool called Wireshark that has been customized as per our requirement.

Monitoring of packets by Wireshark is being done at the application, transport, network and data-link layer levels. A certain region in the hard disk drive has been reserved as the data storage and housekeeping block wherein all the packets that are being captured are stored. Certain rules have been developed that are applied to this section so that the stored packets are seggregated into the General-Purpose subevent, Interesting subevent and Unsusual subevent block.

The algorithm that has been formulated to be the rule engine is fairly simple and has its base on the application layer and data-link layer only of the TCP/IP model.

As the utility of the SHNM System depends primarily on the recorded data, we investigate the reliability of the Wireshark software to capture the packet events. Experiments were carried out on 3

different hosts in the home to simulate conditions of low, medium and heavy loading conditions. All the tests were carried out for a time span of 1 hour and the percantage of packet loss was calculated. Light loading condition consisted of a music video download and general surfing of the websites. Medium loading condition consisted of 2 torrent downloads(total file size >= 1GB) followed by the normal website surfing. Heavy loading condition consisted of 6 torrent downloads (file size >=5 GB), online video streaming using YouTube, video conferencing for 20 minutes using Skype apart from the normal website surfing.

Table 1 below summarizes the results:

Load Type	Low	Medium	Heavy
Time duration	1 hr.	1 hr.	1 hr.
Packets Captured	46,265	1,36,999	2,51,176
% of Packets Lost	0.032	0.101	0.332

A graphical plot of the loading condition(i.e packets captured) on the X- axis v/s the percentage of packets lost on the Y- axis has been generated in figure 3 below from the simulated results. The graph is of linear nature which gives us a clear indication that when using Wireshark as a packet monitoring tool the chances of packet loss increases proportionately as the network traffic increases. In fact towards the heavy loading condition the curve becomes much more steeper indicating that the packet losses are even more in the higher end region

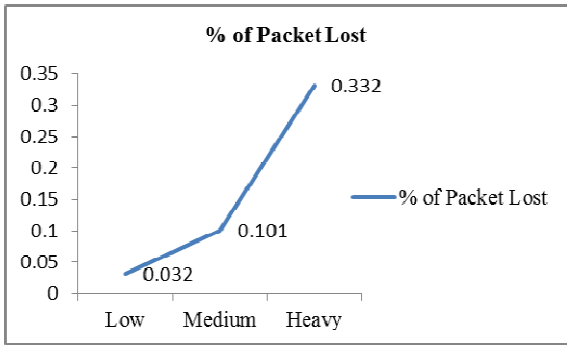


Figure 3

Thus it is evident from the experimental result that for the configuration that we use under heavy loading condition the percentage of packet loss becomes more. Thus, the SHNM System box that we use provides a satisfactory platform for the troubleshooting purpose.

In the near future we expect to improve the capabilities of our prototype so that it can capture all the network events that have been described earlier. We strongly have an intution that the techniques used by any Intrusion Detection System can be extended to our SHNM system also and so we intend to judiciously mix the functionalities of both. We also have a vision to build up an Extensive Data Search Engine that will have intelligence of its own to detect the causes of home network disruption.

REFERENCES

[1] R. Grinter, W.Edwards, M.Newman and N.Ducheneaut, The work to make a Home Network Work in Proceedings of 2005 European Conference on Computer Supported Co-operative Work, Volume 18, page 22, Springer Publicaion.

[2] M.Allman and V. Paxson, Issues and Etiquette Concerning Use of Shared Measurement Data in Proceedings of 2007 ACM Internet Measurement Conference, pages 135-140, San Diego, October 2007.

[3] J.Y.S Marshini Chetty and R.E Grinter, How Smart Homes Learn: The evolution of the networked home and household in Proceedings of 2007 Ubicomp, Innsbruk, Austria, 2007

[4] Erika Sheehan, Marshini Chetty, Rebecca E. Grinter and Warren Keith Edwards, More Than Meets the Eye: Transforming the User Experience og Home Network Management in Proceedings of 2008 ACM Conference on Designing Interactive Systems (DIS 2008), Cape Town, South Africa, February 2008.

[5] S. Kandula, R. Mahajan, P. Verkaik, S. Agarwal, J. Padhye and P. Bahl in Proceedings of 2009 ACM SIGCOMM, Barcelona, Spain, August 2009.

[6] J. Yang. Eden Home Network Management System, Ph.D dissertation, Georgia Tech, 2009.

[7] Erika Sheehan Poole, Marshini Chetty, Tom Morgan, Rebecca E. Grinter and W. Keith Computer Help at Home: Methods and Motivations for informal technical support in Proceedings of 2009 ACM Conference on Human Factors in Computing Systems (CHI 2009) , Boston, MA, April 2009

[8] M. bin Tariq, M. Motiwala, N. Feamster and M. Ammar, Detecting Network Neutrality Violations with casual Inference in Proceedings CoNEXT, December 2009

[9] M. Chetty, R. Banks, R.Harper, T.Reagan, A.Sellen, C.Gkantsidis, T.Karagiannis and P.Key, Who's Hogging the Bandwidth? In Proceedings of 2010 ACM Human Factors in Computing Systems(CHI) Conference, Atlanta, GA, April 2010.

[10] K. L. Calvert, W. K. Edwards, Nick Feamster, R. E. Grinter, Ye Deng, Xuzi Zhou Instrumenting Home Networks in Proceedings of 2010 ACM SIGCOMM Workshop on Home Networks, Home Nets '10, pages 55-60, New York, USA, 2010.

❑❑❑

Real-Time Hand Tracking for Human- Computer Interaction

[1]Ayush Tripathi & [2]S.S Dhotre

[1]Kanishk Puri, Nilesh Srivastava, Prateek Dham
[2]Computer Science Dept., Bharati Vidyapeeth Deemed University College of Engineering Pune, Maharashtra(India)

Abstract - The proposed work is part of a project that aims for the control of a mouse based on hand gesture recognition. This goal implies the restriction of real-time response and unconstrained environments. This is basically a vision based skin-colour segmentation method for moving hand in real time application [3].

This algorithm is based on three main steps: hand segmentation, hand tracking and gesture recognition from hand features. For the hand segmentation step we use the colour cue due to the characteristic colour values of human [1].

I. INTRODUCTION

Nowadays, the majority of the human-computer interaction (HCI) is based on mechanical devices such as keyboards, mouse, joysticks or gamepads. In recent years there has been a growing interest in a class of methods based on computational vision due to its ability to recognise human gestures in a natural way .These methods use as input the images acquired from a camera or from a stereo pair of cameras. The main goal of these algorithms is to measure the hand configuration in each time instant. Our application uses images from a low-cost web camera placed in front of the hand.

The hand must be localized in the image and segmented from the background before recognition. The pixels are selected from the hand. The selected pixels are transformed from the RGB-space to the HSL-space for taking the Chroma information: hue and saturation.

The hands are recognized by the computer using the skin colour as one of the basic features for the hand recognition. The important feature is the accurate segmentation of hands [3].

Figure 1: Configuration for the hand Recognition.

II. HAND SEGMENTATION

The hand must be localized in the image and segmented from the background before recognition. Colour is the selected cue because of its computational simplicity, its invariant properties regarding to the hand shape configurations and due to the human skin-colour characteristic values. Also, the assumption that colour can be used as a cue to detect faces and hands has been proved in several publications. For our application, the hand segmentation has been carried out using a low computational cost method that performs well in real time.

We have encountered two problems in this step that have been solved in a pre-processing phase. The first one is that human skin hue values are very near to red colour, that is, their value is very close to 2π radians, so it is difficult to learn the distribution due to the hue angular nature that can produce samples on both limits. To solve this inconvenience the hue values are rotated π radians. The second problem in using HSL-space is when the saturation values are close to 0, because then the hue is unstable and can cause false detections. This can be avoided discarding saturation values near 0[4].
Another problem to hand segmentation is the number hand movements. Hand is allowed to move around 360 degree which makes it difficult for the camera to capture every single frame. For the capturing the sensitivity of the program has to be increased as per the hand movements.

III. TRACKING

USB cameras are known for the low quality images they produce. This fact can cause errors in the hand segmentation process. In order to make the application robust to these segmentation errors we add a tracking algorithm. This algorithm tries to maintain and propagate the hand state over time [5].

IV. GESTURE RECOGNITION

Our gesture alphabet consists in four hand gestures and four hand directions in order to fulfil the application's requirements. The hand gestures correspond to a fully opened hand (with separated fingers), an opened hand with fingers together, a fist and the last gesture appears when the hand is not visible, in part or completely, in the camera's field of view. These gestures are defined as *Start*, *Move*, *Stop* and the *No-Hand* gesture respectively. Also, when the user is in the *Move* gesture, he can carry out a *Left*, *Right*, *Front* and *Back* movements. For the *Left* and *Right* movements, the user will rotate his wrist to the left or right. For the

Front and *Back* movements, the hand will get closer to or further of the camera. Finally, the valid hand gesture transitions that the user can carry out.

The process of gesture recognition starts when the hand's user is placed in front of the camera field of view and the hand is in the *Start* gesture, that is, the hand fully opened with separated fingers [5].

The image is captured by the web camera and the count of number of fingers is shown by the program code. The count of fingers can help in controlling the mouse movements. For instance count of 2 can guide the mouse to stop. Count of 3 can guide the mouse for movement [3].

For better count the hand has to be kept stationery so that the web can capture the frame for the proper count and movement of the mouse. As around 360 degree hand movement is possible, the hand to be in proper mode under proper light conditions so that count can be maintained properly.

The main method of recognizing the hand is by detecting the skin colour of the hand. Proper colour spacing is used to detect the skin colour. Colour Spacing used are RCB

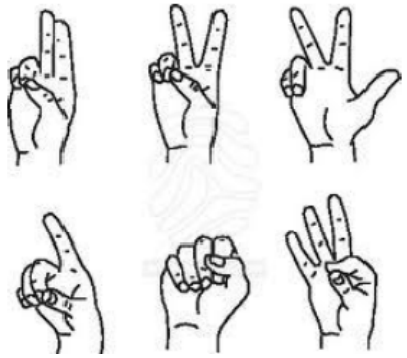


Figure 2: Hand Gestures

V. CONVERSION OF IMAGE

The image captured by the web cam is converted into the grayscale. The conversion of image into the grayscale is done for the reduction in the size of the image. The reduction helps in the better processing of the image [5].

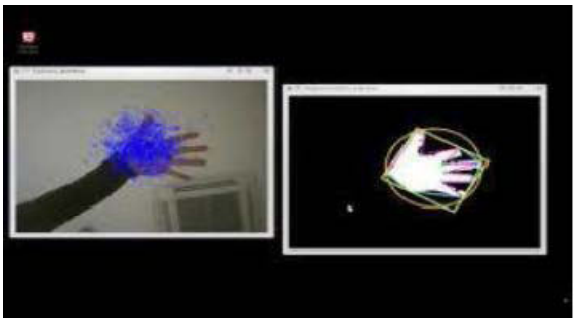


Figure 3: Conversion to Grayscale

VI. MATHEMATICAL MORPHOLOGY

Mathematical morphology is a theory and technique for the analysis and processing of geometrical structures, based on set theory, lattice theory, topology, and random functions.

Mathematical morphology is applied to the hands for the analysis of the hands. Binary morphology is applied to the hands [6].

Binary Morphology: - In binary morphology, an image is viewed as a subset of an Euclidean space R^d or the integer grid Z^d, for some dimension *d*.

Structuring elements: - The basic idea in binary morphology is to probe an image with a simple, pre-defined shape, drawing conclusions on how this shape fits or misses the shapes in the image. This simple "probe" is called structuring element, and is itself a binary image (i.e., a subset of the space or grid).

Basic Operators [6]: -

Erosion

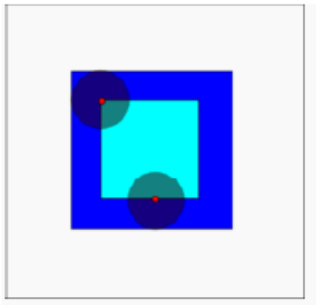


Figure 4: Erosion

The erosion of the dark-blue square by a disk, resulting in the light-blue square.

The erosion of the binary image A by the structuring element B is defined by:

$$A \ominus B = \{z \in E | B_z \subseteq A\},$$

Where B_z is the translation of B by the vector z,

$$B_z = \{b + z | b \in B\} \quad \forall z \in E.$$

When the structuring element B has a center (e.g., B is a disk or a square), and this center is located on the origin of E, then the erosion of A by B can be understood as the locus of points reached by the center of B when B moves inside A. For example, the erosion of a square of side 10, centered at the origin, by a disc of radius 2, also centered at the origin, is a square of side 6 centered at the origin.

The erosion of A by B is also given by the expression:

Dilation

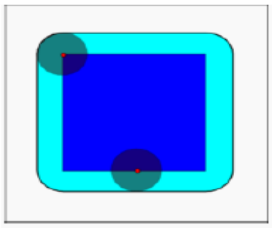


Figure 5: Dilation

The dilation of the dark-blue square by a disk, resulting in the light-blue square with rounded corners.

The dilation of A by the structuring element B is defined by:

$$A \oplus B = \bigcup_{b \in B} A_b$$

The dilation is commutative, also given by:

$$A \oplus B = B \oplus A = \bigcup_{a \in A} B_a$$

If B has a center on the origin, as before, then the dilation of A by B can be understood as the locus of the points covered by B when the center of B moves inside A. In the above example, the dilation of the square of side 10 by the disk of radius 2 is a square of side 14, with rounded corners, centered at the origin. The radius of the rounded corners is 2.

The dilation can also be obtained by: -

$$A \oplus B = \{z \in E | (B^s)_z \cap A \neq \varnothing\}$$

where Bs denotes the symmetric of B, that is,

$$B^s = \{x \in E | -x \in B\}.$$

Opening

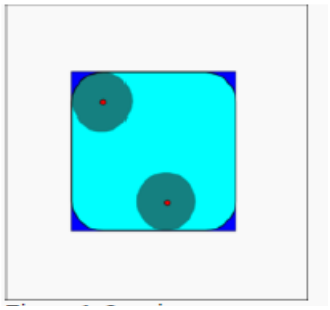


Figure 6: Opening

The opening of the dark-blue square by a disk, resulting in the light-blue square with round corners.

The opening of A by B is obtained by the erosion of A by B, followed by dilation of the resulting image by B:

$$A \circ B = (A \ominus B) \oplus B$$

The opening is also given by:

$$A \circ B = \bigcup_{B_z \subseteq A} B_x$$

which means that it is the locus of translations of the structuring element B inside the image A. In the case

of the square of radius 10, and a disc of radius 2 as the structuring element, the opening is a square of radius 10 with rounded corners, where the corner radius is 2.

Closing

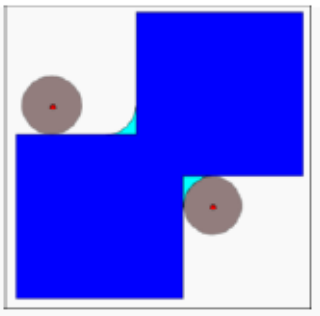


Figure 7: Closing

The closing of the dark-blue shape (union of two squares) by a disk, resulting in the union of the dark-blue shape and the light-blue areas.

The closing of *A* by *B* is obtained by the dilation of *A* by *B*, followed by erosion of the resulting structure by *B*:

$$A \bullet B = (A \oplus B) \ominus B$$

The closing can also be obtained by

$$A \bullet B = (A^c \circ B^s)^c$$

Where *Xc* denotes he complement of *X* relative

to *E* (that is,

$$X^c = \{x \in E | x \notin X\}.$$

The above means that the closing is the complement of the locus of translations of the symmetric of the structuring element outside the image *A*.

VII. FORMATION OF CONVEX HULL

In mathematics, the convex hull or convex envelope for a set of points *X* in a real vector space V is the minimal convex set containing *X*.

In computational geometry, a basic problem is finding the convex hull for a given finite nonempty set of points in the plane. It is common to use the term "convex hull" for the boundary of that set, which is a convex polygon, except in the degenerate case that points are collinear. The convex hull is then typically represented by a sequence of the vertices of the line segments forming the boundary of the polygon, ordered along that boundary [7]

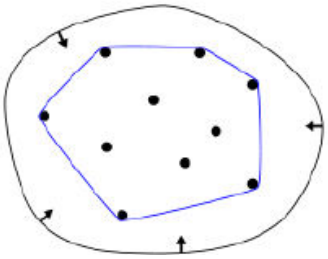


Figure 8: Formation of Convex Hull

For *planar objects*, i.e., lying in the plane, the convex hull may be easily visualized by imagining an elastic band stretched open to encompass the given object; when released, it will assume the shape of the required convex hull.

It may seem natural to generalise this picture to higher dimensions by imagining the objects enveloped in a sort of idealised unpressurised elastic membrane or balloon under tension. However, the equilibrium (minimum-energy) surface in this case may not be the convex hull — parts of the resulting surface may have negative curvature, like a saddle surface. For the case of points in 3-dimensional space, if a rigid wire is first placed between each pair of points, then the balloon will spring back under tension to take the form of the convex hull of the points [7].

With all these steps the hand is detected and recognised by the web camera.

VIII. CONCLUSION

In this paper we have presented a real-time algorithm to track and recognise hand gestures for human computer interaction. We have proposed the methods to detect hand segmentation, hand tracking and gesture recognition from extracted hand features. The experiments have confirmed that the low cost web cam with low resolution is better for the capturing of the images [2].

We have also seen that the images captured cannot be absolutely perfect. The hand segmentation has to be fixed at a particular point in order to get the better image of the hand segmentation [1].

We can get the proper segmentation by capturing the image again and again so as to get the proper idea of surrounding. By using again and again we can test that in which light conditions we get the better images of the hand.

IX. REFERENCES

[1] This paper appears in: Information Technology Interfaces (ITI), 2010 32nd International Conference on Issue Date: 21-24 June 2010 On page(s): 289 – 294

[2] 2009 International Conference on Embedded Software and Systems Hand Gesture Recognition Based on MEB-SVM Hangzhou, Zhejiang P.R. China May 25-May 27 ISBN: 978-0-7695-3678-1

[3] Real –Time Hand Tracking and Gesture Recognition for Human-Computer Interaction By Cristina Manresa, Javier Varona, Ramon Mas and Francisco J. Perales.

[4] Vision-Based Skin-Colour Segmentation Of Moving Hands For Real-Time Applications by S.Askar, Y.Kondratyuk, K.Elazouzi, P. Kauff, O.Schreer Fraunhoer Institute of Telecommunications, Heinrich-Hertz-Institute Germany.

[5] http://en.wikipedia.org/wiki/ Gesture recognition.

[6] http://en.wikipedia.org/wiki/ Mathematical morphology.

[7] http://en.wikipedia.org/wiki/Convex_hull.

◈◈◈

Information Measurement

[1]Sarat K. Parhi & [2]L. Das

[1]Dept. of Mathematics, Vijayanjali Institute of Technology, Balasore
[2]Delhi Technological University, New Delhi, 110042

Abstract **-** This paper contains measurement of information function h(p) = -c log kp. As p tends to zero then the quantity of information is large and p tends to one then there is no information causes, when k is one. Also properties for monotonicity, additivity, grouping and inference are stated in the theorems. A case study on the application level is discussed in this paper.

Key Terms: *monotonicity,additivity,grouping,inference.*

I. INTRODUCTION

In this information measurement function h(p) = -c log kp, we observe that as p tends to zero then the quantity of information is large and this large number of information causes larger amount of uncertainty because the measurement of uncertainty involves with the measurement of probability distribution. To obtain expected information from a pool of information one has to design an uncertainty function and then he has to up date this function by examining some of the properties like monotonicity, additivity, grouping of the information with the others.

Let $X(x_1,x_2,\ldots x_m)$ and $Y(y_1,y_2,\ldots y_n)$ be two information sources and these can be modeled as two random variables with the probability distribution $P(p(x_1)=\frac{k}{m}, p(x_2)=\frac{k}{m},\ldots p(x_m)=\frac{k}{m})$ and $Q(p(y_1)=\frac{k}{n}, p(y_2)=\frac{k}{n},\ldots p(y_n)=\frac{k}{n})$ respectively. Where k, m and n are non negative integers.

Then the average uncertainty function f associated with the information $X(x_1,x_2,\ldots x_m)$ is defined as follows

$$H(p_1,p_2,\ldots p_m)=H\left(\frac{k}{m},\frac{k}{m},\ldots\frac{k}{m}\right)= f\left(\frac{k}{m}\right) = -\sum_{i=1}^{m} \log kp_i .$$

The above formulation of the uncertainty function obeys the properties of the entropy function because the uncertainty arises prior to the arrival of message, whereas the expected information entropy occurred after the arrival of the message

$$H(p_1,p_2,\ldots p_m)= -\sum_{i=1}^{m} \log p_i$$

Hence more uncertainty prior to the arrival of a message, implies the largest amount of information conveyed.

II. INFORMATION MEASUREMENT TECHNIQUES:

Suppose p the probability measure of the occurrence of an event E. Suppose the message "Occurrence of E" is obtained. Then the quantity of information conveyed in this message is log p. If p is close to one (say p=0.95),then message conveyed very little amount of information, because it is almost all determine that the occurrence of the event E. On the other hand if p closes to zero (say p=0.1), then it almost certain that E will not occur and consequently the message stating its occurrence is quite unexpected and hence contains a greater deal of information.

III. CHARACTERISTIC OF UNCERTAINTY FUNCTION:

The characteristic of uncertainty function satisfies the properties such as monotonicity, additivity and grouping.The following theorem explains this fact.

Theorem 3.1

The function $H(p_1,p_2,\ldots p_m)=H\left(\frac{k}{m},\frac{k}{m},\ldots\frac{k}{m}\right)= f\left(\frac{k}{m}\right) =-\sum_{i=1}^{m} \log kp_i$ satisfies monotonicity, additivity and grouping the positive parameter k and the arbitrary constant c.

Proof:

Monotonicity: This means the probability density function f(p) is monotonic. If $p_1 = \frac{k}{m}$ is the average probability then the function H(p)= f $(\frac{k}{m})$ is monotonically increasing in the other word, for m<m', f(m)<f(m'). It is obviously true , because $\frac{k}{m} > \frac{k}{m'}$ => -log $(\frac{k}{m})$ > -log $(\frac{k}{m})$ and f $(\frac{k}{m})$ = -log $(\frac{k}{m})$

Additivity: This can be verified by proving the relation f $((\frac{k}{m})^r)$ = r f $(\frac{k}{m})$ for r, m and k ≠0. Using method of induction, for r=2 the result is obvious, that means f $((\frac{k}{m})^2)$ = f $(\frac{k}{m})$ + f $(\frac{k}{m})$=2 f $(\frac{k}{m})$. Suppose the expression is true for r=n-1, that means f $((\frac{k}{m})^{(n-1)})$ = (n-1) f $(\frac{k}{m})$. This implies f $((\frac{k}{m})^n)$ = f $(\frac{k}{m})$. f $((\frac{k}{m})^{(n-1)})$ = f $(\frac{k}{m})$ + f $((\frac{k}{m})^{(n-1)})$ = f $(\frac{k}{m})$+ (n-1) f $(\frac{k}{m})$= nf $(\frac{k}{m})$.

Grouping: Let f $(\frac{k}{m})$ = c log $\frac{k}{m}$, for k = m,2m,3m…

For positive integer t>2,f(t-1)=c log (t-1) for k=(t-1)m and c>0.

$t^l \le 2^r \le t^{l+1}$ for some 0 < l ϵ R … (3.1)

Thus $f(t^l) \le f(2^r) \le f(t^{l+1})$ => $lf(t) \le rf(2) \le (l+1)f(t)$ and consequently

$(\frac{l}{r}) \le (\frac{f(2)}{f(t)}) \le (\frac{l+1}{r})$ ….(3.2)

Taking logarithm function over base 2 to the inequality (3.1) we get l log t ≤ r ≤ (l+1) log t

$(\frac{l}{r}) \le (\frac{1}{log\ t}) \le (\frac{l+1}{r})$ ….(3.3)

Subtracting the inequality (3.3)from the inequality (3.2)

$0 \le (\frac{f(2)}{f(t)}) - (\frac{1}{log\ t}) \le 0$

⇨ $(\frac{f(2)}{f(t)}) - (\frac{1}{log\ t}) = 0$

⇨ f(t) = f(2)log t = c log t for c = f(2)

⇨ f $(\frac{k}{m})$ = c log $\frac{k}{m}$

This establishes the proof of the theorem 3.1.

Axiom 3.1:

The amount of information can be quantified as the function of the probability measure of an event that is involved with that function.

Theorem 3.2: (Measurement of information characterization)

The function h(p)= -c log kp, satisfies Axiom 3.1

Proof:

Consider the following equations

l= - log kp for k>0 and p ϵ (0,1]…(3.4)

l_1= -log $kp_1 \Leftrightarrow e^{-l_1} = kp_1$ …(3.5)

l_2= -log $kp_2 \Leftrightarrow e^{-l_2} = kp_2$ …(3.6)

The right hand side of the equations 3.4 and 3.5 can be expressed as the function of g such that $h(kp_i) = h(e^{-l_i}) = g(l_i)$ for i=1,2 …(3.7)

And this is the solution of the differential equation

$(\frac{dl}{dx}) + (\frac{1}{x}) = 0$,where x=kp. Moreover l= -c log kp for c & k≠0 and p ϵ (0,1] is also a solution of the stated differential equation.

Thus we can rewrite it as either $G(x,l) = x - e^{-l}$ or G(x,l) = l+ log x. if l_1 & l_2 are two inferences of certain information then

$G(x_1.x_2,l_1+l_2) = G(x_1,l_1) + G(x_2,l_2)$ …(3.8)

For k ≠ 0, $l_1=l_2=0$, then $x_1=x_2$ and

$G(x_1^2,0)=2G(x_1,0)$ …(3.9)

If the parameter k is assumed as k = $\frac{1}{2p}$ then equation (3.4) become

L = log 2 = 1 and G(x,1)= 1+log $\frac{1}{2} \neq 0$.

Thus we can obtained a generating function for the function G(x,1). In the similar way one can verify the function G(x,1) is decreasing.

That means $G(x_1,l_1) > G(x_2,l_2)$ for $0<p_1\le 1$, Thus this establishes the theorem 3.2.

IV. APPLICATION : INFORMATION FLUCTUATION AND DAMPING.

$(\frac{dl}{dx}) + (\frac{1}{x}) = 0$ =>k= $-x\frac{dl}{dx}$ is the mathematical formula for information fluctuation. If one desire to select any information from the pool of information data then the primary assumption is that the selection of information are equally likely. The secondary task is to frequently subdivide the range space that is associated with the pool of information data into to mutually exclusive groups .Then the probability of selecting any group for the next search is p_j , for j=1,2. The sample set of the probability measure p_j is either s_1={1,2,3,…,r} or s_2={1,2,3…m}, r < m.

V. CONCLUSION:-

The Measurement of information is play a vital rule in the world information system. The properties has been discussed in this paper should give better clarification in the information measurement technique .This paper will help to be studied my theoretical problems in near future.

VI. REFERENCES:-

1. Swarup, Kanti ,Gupta,P.K. and Mohan,Man, Operation Research ,Sultan Chand and Sons,1994,P.657-690.
2. Parhi,S.K.,Study of Certain Mathematical Modeling and algorithm in the information Technology, Ph.D. Thesis of Utkal University,2002.

◈◈◈

Radio Access Network Requirement for New Deployment of WiMAX in Dhaka

Mohammad T. Kawser, Mohammed R. Al-Amin, Khondoker Z. Islam & Sifat-E-Mohammad

Dept of Electrical and Electronics Engineering, Islamic University of Technology, Gazipur, Bangladesh

Abstract **-** Mobile WiMAX is expected to be the next generation radio-interface, complementing WLAN and challenging EVDO/HSPA/LTE. High speed data rate, reduced latency, better Quality of service, and mobility can allow WiMAX to meet the rapidly growing demand of the users. A study of WiMAX Radio Network Planning (RNP) for an urban area like Dhaka city in Bangladesh is presented in this paper in order to help predetermine the radio access infrastructure requirements. A suitable radio planning tool has been used for this purpose. The simulation results of throughput and Carrier to Interference plus Noise Ratio (CINR) are provided.

Key words - *WiMAX; Network Dimensioning; Radio Network Planning; CINR.*

I. INTRODUCTION

This paper addresses the radio access network requirement for new deployment of WiMAX in a metropolitan area like Dhaka in Bangladesh based on the IEEE 802.16e air interface standards. The most important technical and business goal of radio access network is efficiently providing coverage and capacity, while avoiding the build-out of a large number of new base stations. This paper will focus on radio planning issues for new deployment of a cost-effective WiMAX radio access infrastructure using spectrum in the 3.3 GHz frequency bands. Current WiMAX deployments operate at 2.3 GHz in Bangladesh but 3.3 GHz is a likely carrier frequency for future spectrum allocation for WiMAX in Bangladesh. The paper organization is as follows: network dimensioning is explained in Section II; the detailedradio network planning is described in Section III; the simulation results for radio network planning are presented in Section IV; finally, the conclusions are highlighted in Section V.

II. NETWORK DIMENSIONING

Network dimensioning is the initial step of radio network planning for deployment of any generation technology. The target of network dimensioning is to estimate the number of required Base Stations (BSs) for the area of interest. The network dimensioning activities include radio link budget and coverage analysis, cell capacity estimation, determination of hardware configuration and equipment at different interfaces. The link budget determines the maximum cell radius for a given level of reliability. The result of this step depends on the propagation model used. With a rough estimate of the cell size and BS count, verification of coverage analysis is carried out for the required capacity.

The area of Dhaka city is approximately 1463 sq-km. This includes some areas where there are no dwellers (e.g. ditches). Also, in some areas, the number of potential users is not significantly high (e.g. slum areas). The estimated total coverage area can be 70% of the whole area, which is 1028 sq-km.

The total population of Dhaka city is approximately 15 million. A great numbers of the dwellers in Dhaka do not require access to internet facilities. The estimated number of target subscribers in first few years for a new operator can be about 1% of the whole population in Dhaka. The current operators claim to support about 100 thousand subscribers. The subscribers can be grouped based on their locations around main roads, secondary roads, small streets, railways and airports. The subscribers can also be classified based on their predominant data transfer in downlink, in uplink or in both downlink and uplink. Thus, the target subscribers are estimated as shown in Table I**.**

TABLE I. NUMBER OF SUBSCRIBERS FOR DIFFERENT CLASSES

Area	No. of users			
	DL	***UL***	***DL+UL***	***Total***
Main roads	***2857***	***747***	***268***	***3872***
Secondary roads	13804	3247	1393	18444
Small Streets	87628	21201	9157	118986
Airport	684	153	69	906
Railways	6	5	1	12
Total	**104979**	**25353**	**10888**	**142220**

The assumptions for link budget calculation are shown in Table III. Other assumptions are shown in Table II.

TABLE II. CERTAIN ASSUMED PARAMETER VALUES

Parameters	Value	Parameters	Value
Carrier frequency	3.3 GHz	Throughput per user	512 kbps
Bandwidth in a sector	10 MHz	Overbooking Factor	20
Frequency Reuse Ratio	1	No. of antennas at BS	2
Scheduling Algorithm	Proportional fair	No. of antennas at CPE	1

Rough estimates of the required number of BSs have been calculated to meet certain target capacity and target coverage for DHAKA city. The cell range and BS configurations have also been estimated. All these estimates have been later used as a baseline for detailed radio planning.

A. Dimensioning for Target Capacity

The capacity of a given network is measured in terms of the subscribers or the traffic load that it can handle. The former requires knowledge of the number of the subscribers and the types of their usage.

The estimated cell throughput for 10MHz bandwidth is 15Mbps. The assumed throughput per user is 512 kbps. Then the cell size should be such that it supports 15Mbps/512 kbps=30 active users simultaneously. Thus, a BS supports 90 active users. For overbooking factor 20, the number of total users under a BS can be 90 × 20 = 1800. Thus, the required number BS can be estimated as 142220/1800 ≈ 79.

B. Dimensioning for Target Coverage

Dimensioning for target coverage includes radio link budget and coverage analysis in both downlink and uplink. A link budget is developed assuming potential values for different parameters as shown in Table 3. The Maximum Allowable Path Loss (MAPL) is calculated based on the required CINR level at the receiver. The minimum of the maximum path losses in uplink and downlink is converted into cell radius, by using a propagation model appropriate to the deployment area.

TABLE III. RADIO LINK BUDGET

Parameters	Downlink	Uplink	Notes
Transmit power	49 dBm	30 dBm	A1
No of transmitting antenna	2	1	A2
Transmitter antenna gain	18 dBi	0 dBi	A4
Transmitter losses	3.0 dB	0 dB	A5
Effective Isotropic Radiated Power (EIRP)	67 dB	30 dB	A6=A1 + 10 log10 (A2)- A3+ A4 -A5
Channel bandwidth	10 MHz	10 MHz	A7
No of sub channels	16	16	A8
Receiver noise level	-104 dBm	-104 dBm	A9=-174+ 10 log10(A7*1e6)
CINR	8 dB	6 dB	A10
Macro diversity Gain	0 dB	0 dB	A11
Sub channelization Gain	0 dB	12 dB	A12 = 10 log10(A8)
Receiver sensitivity (dBm)	-96	-110	A13=A9+A10 + A11-A12
Receiver antenna gain	0 dBi	18 dBi	A14
System gain	163 dB	158 dB	A15=A6 –A13 +A14
Shadow -fade Margin	8.5 dB	8.5 dB	A16
Building penetration losses	0 dB	0 dB	A17 ; Assuming single wall
Path Loss	154.5dB	149.5 dB	A18=A15-A16- A17
Coverage range, d	3.17 km (1.97 miles)	2.34 km (1.45 miles)	Assuming Sakagami extended model

The minimum between downlink and uplink coverage ranges, d = 2.34 km is considered as the cell radius. The hexagonal cell site area is then calculated as, $3d2\times \sin(\pi/3) = 14.22$ sq-km. The required number of BS can be estimated as, $1028/14.22 \approx 72$. As a densely populated city, it is thus found that Dhaka requires little more number of BSs to fulfill capacity requirements than what is required for coverage requirements.

III. DETAILED RADIO NETWORK PLANNING

ATOLL, a Radio Network Planning (RNP) tool from FORSK has been used in the detailed radio network planning step. A digital map for Dhaka city is used which incorporates the terrain properties. The results from network dimensioning have been taken into account as initial estimation for number of BSs in order to meet capacity and coverage requirements. The location of BS sites, transmit power, antenna height, number of sectors, azimuth and down tilt have been carefully chosen based on results from dimensioning, terrain, subscriber densities, building densities, typical building heights, foliage, interference from neighboring cells and so forth. The number of BSs, their locations and BS configuration parameters are then adjusted through a good number of iterative simulations for optimum performance in terms of both coverage and throughput. This was a long manual process. This led to the establishment of 75 BSs in total in Dhaka with transmit power and antenna height configurations as shown in Fig.1. The frequencies of 10 MHz bandwidth have been allocated to sectors using automatic frequency planning feature. Preamble indexes have also been allocated using automatic allocation feature.

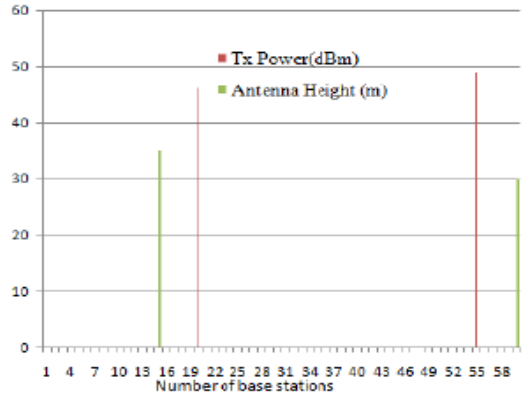


Fig. 1: Transmit power and antenna height vs. BS counts

IV. SIMULATION RESULTS

The simulation results from the radio planning for Dhaka using ATOLL are presented in this section. Coverage by Signal level and Downlink CINR distribution around BSs for a small area are shown in Fig.2 and Fig.3 respectively. As demonstrated, satisfactory coverage quality has been achieved for the presented network setup.

The Signal Level over the whole Dhaka city is demonstrated in Fig.4 using histogram. It may be noted that almost three-fourth of the whole Dhaka city have satisfactory coverage while subscribers are not prevalent everywhere .This histogram depicts signal level exceeding -90 dBm for pretty large amount area as shown by the end portion of the histogram. This confirms the achievement of good coverage quality.

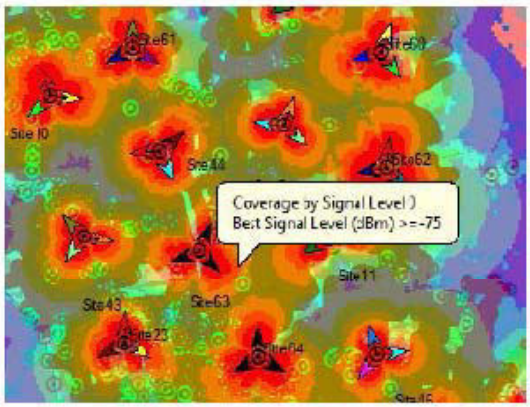


Fig. 2 : Coverage Signal level distribution around BS

Fig. 3 : Downlink CINR distribution around BS

The CINR distribution over the whole Dhaka city is demonstrated in Fig.5 using histogram. Since almost one-fourth of the whole Dhaka city is left out of coverage based on the presence of very few or no subscribers, this histogram leaves low CINR for a good amount of area. However, CINR exceeding 30 dBm exists for pretty large amount area as shown by the sharp rise at the end of the histogram. This confirms the achievement of good CINR quality.

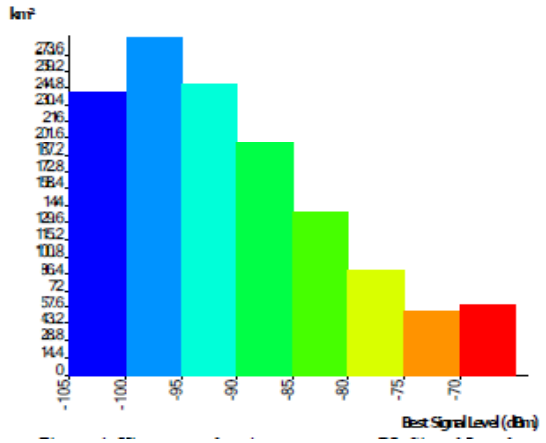


Fig.4 : Histogram showing area versus DL Signal Level

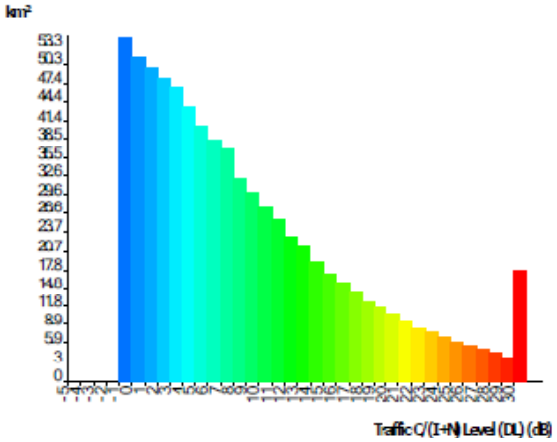


Fig. 5 : Histogram showing area versus DL CINR

The downlink throughput distribution around BSs over the whole Dhaka city area is shown in Fig.6. As demonstrated, satisfactory data rate has been achieved for the presented network setup.

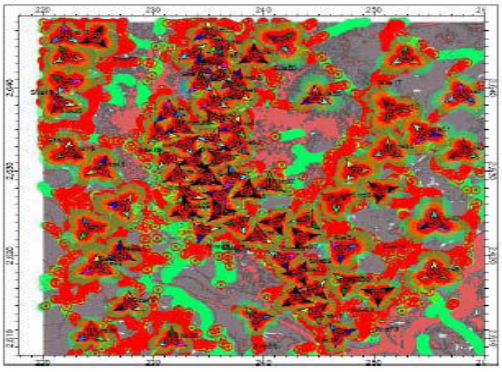


Fig. 6 : Overall throughput variation around BSs and distribution of subscribers.

The downlink throughput distribution over the whole Dhaka city is demonstrated in Fig.7 using histogram. It may be noted that the data rate is higher than 4 Mbps for most areas and it is higher than 9 Mbps for a large amount of area. This depicts the achievement of a satisfactory data rate for the users.

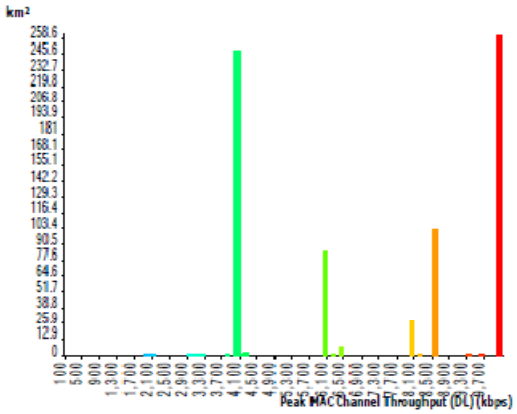


Fig. 7 : Histogram showing area versus downlink throughput

A sample downlink link budget at a cell edge location, generated by Atoll, is shown in Fig.8. This demonstrates that the presumed or target link budget condition shown in Table III conforms pretty well to the simulated results.

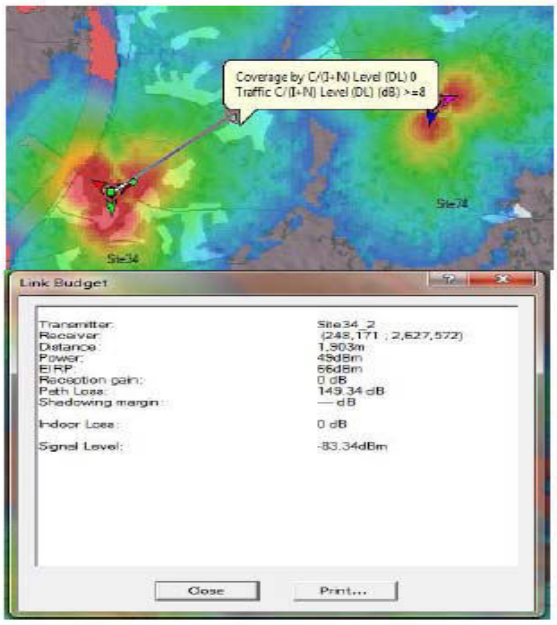


Fig. 8 : Sample link budget at a cell edge location for downlink

V. CONCLUSION

Radio network planning for new WiMAX deployment in Dhaka, Bangladesh has been performed. The simulation results show that satisfactory performance has been achieved in terms of coverage and throughput. The BS configurations presented in Section II and Fig.1 indicate radio access requirement for such a WiMAX deployment. This requirement analysis can function as a guideline for a new operator to meet the demand of the subscribers. It can help perform cost analysis and consider relevant issues at the outset. Of course, the operator can consider significant variations from the radio access infrastructure presented in this paper in order to support different number of subscribers or to bring about variations in target KPIs or configurations, but nevertheless, this analysis can help as a baseline.

REFERENCES

[1] G. M. Galvan-Tejada " WiMAX Urban Coverage Based on the Lee Model and the Deygout Diffraction Method". In:7th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2010) Tuxtla Gutiérrez, Chiapas, México. September 8-10, 2010.

[2] Prof. Walid Y. Ali-Ahmad, Mohamed Hasna, Ali Dabbous, and Adel Yammout, ImadAtwi, "Propagation Model Development and Radio Planning for future WiMAX Deployment in Beirut"American University of Beirut. In: Final Year Project, spring 2006.

[3] ShekharSrivastava, André Girard, and Brunilde Sansò "Capacity Planning and Design of WiMAX Access Networks". In: WiMAX Network Planning and Optimization, Auerbach Publications,2009.

[4] Teruya FUJII, Yoshichika OHTA and Hideki OMOTE" Empirical Time-Spatial Propagation Model in Outdoor NLOS Environments for Wideband Mobile Communication Systems" Wireless System Research Center, Research Division Tokyo, Japan.

[5] WiMAX FORUM "Mobile WIMAX – part I: A Technical Overview and Performance Evaluation" June 2006

◈◈◈

The Study of Detecting Replicate Documents Using MD5 Hash Function

Pushpendra Singh Tomar & Maneesh Shreevastava

Dept. of IT, LNCT, Bhopal, (M.P.) India

Abstract - A great deal of the Web is replicate or near- replicate content. Documents may be served in different formats: HTML, PDF, Text for different audiences. Documents may get mirrored to avoid delays or to provide fault tolerance. Algorithms for detecting replicate documents are critical in applications where data is obtained from multiple sources. The removal of replicate documents is necessary, not only to reduce runtime, but also to improve search accuracy. Today, search engine crawlers are retrieving billions of unique URL's, of which hundreds of millions are replicates of some form. Thus, quickly identifying replicate detection expedites indexing and searching. One vendor's analysis of 1.2 billion URL's resulted in 400 million exact replicates found with a MD5 hash. Reducing the collection sizes by tens of percentage points results in great savings in indexing time and a reduction in the amount of hardware required to support the system. Last and probably more significant, users benefit by eliminating replicate results. By efficiently presenting only unique documents, user satisfaction is likely to increase.

Key words - *unique documents, detecting replicate, replication, search engine.*

I. INTRODUCTION

The definition of what constitutes a replicate has somewhat different interpretations. For instance, some define a replicate as having the exact syntactic terms and sequence, whether having formatting differences or not. In effect, there are either no difference or only formatting differences and the contents of the data are exactly the same. In any case, data replication happens all the time. In large data warehouses, data replication is an inevitable phenomenon as millions of data are gathered at very short intervals.

Data warehouse involves a process called ETL which stands for extract, transform and load. During the extraction phase, multitudes of data come to the data warehouse from several sources and the system behind the warehouse consolidates the data so each separate system format will be read consistently by the data consumers of the warehouse. Data portals are everywhere. The tremendous growth of the Internet has spurred the existence of data portals for nearly every topic. Some of these portals are of general interest; some are highly domain specific. Independent of the focus, the vast majority of the portals obtain data, loosely called documents, from multiple sources [1]. Obtaining data from multiple input sources typically results in replication. The detection of replicate documents within a collection has recently become an area of great interest [2] and is the focus of our described effort.

Typically, inverted indexes are used to support efficient query processing in information search and retrieval engines. Storing replicate documents affects both the accuracy and efficiency of the search engine. Retrieving replicate documents in response to a user's query clearly lowers the number of valid responses provided to the user, hence lowering the accuracy of the user's response set. Furthermore, processing replicates necessitates additional computation

Replicates are abundant in short text databases. For example, popular mobile phone messages may be forwarded by millions of people, and millions of people may express their opinions on the same hot topic by mobile phone messages. In our investigation on mobile phone short messages, more than 40% short messages have at least one exact replicate. An even larger proportion of short messages are near-replicates. Detecting and eliminating these replicate short messages is of great importance for other short text processing, such as short text clustering, short text opinion mining, short text topic detection and tracking, short message community uncovering. Exact replicate short texts are easy to identify by standard hashing schemes. Informal abbreviations without introducing any additional benefit. Hence, the processing efficiency of the user's query is lowered. A problem introduced by the indexing of replicate documents is potentially skewed collection statistics. Collection statistics are often used as part of the similarity computation of a query to a document. Hence, the biasing of collection statistics may affect the overall precision of the entire system.

Simply put, not only is a given user's performance compromised by the existence of replicates, but also the

overall retrieval accuracy of the engine is jeopardized. The definition of what constitutes a replicate is unclear. For instance, a replicate can be defined as the exact syntactic terms, without formatting differences. Throughout our efforts however, we adhere to the definition previously referred to as a measure of resemblance [3]. The general notion is that if a document contains roughly the same semantic content it is a replicate whether or not it is a precise syntactic match. When searching web documents, one might think that, at least, matching URL's would identify exact matches. However, many web sites use dynamic presentation wherein the content changes depending on the region or other variables. In addition, data providers often create several names for one site in an attempt to attract users with different interests or perspectives. For instance, Fox4, Onsale-Channel-9, and Real-TV all point to an advertisement for real TV.

Some forms of replicated content, such as those appearing in publications of conference proceedings, important updates to studies, confirmation of contested results in controversial studies, and translations of important findings, may no doubt be beneficial to the scientific community. Replication is seen as unethical when the primary intent is to deceive peers, supervisors, and/or journal editors with false claims of novel data. Given the large number of papers published annually, the large diversity of journals with overlapping interests in which to publish, and the uneven access to journal publication content, it is not unreasonable to assume that the discovery of such replication is rare [4]. The recent development of algorithmic methods to systematically process published literature and identify instances of replicated/plagiarized text as accurately as possible should serve as an effective deterrent to authors considering this dubious path. Unfortunately, the methods in place now have a very limited reach, and are confined to abstracts and titles only.

Replicates: where they come from. One of the main problems with the existing geospatial databases is that they are known to contain many replicate points (e.g., [6] [7], [8]). The main reason why geospatial databases contain replicates is that the databases are rarely formed completely .from scratch., and instead are built by combining measurements from numerous sources. Since some measurements are represented in the data from several of the sources, we get replicate records.

Why replicates are a problem. Replicate values can corrupt the results of statistical data processing and analysis. For example, when instead of a single (actual) measurement result, we see several measurement results confirming each other, and we may get an erroneous impression that this measurement result is more reliable than it actually is. Detecting and eliminating replicates is therefore an important part of assuring and improving the quality of geospatial data, as recommended by the US Federal Standard [9].

The identification of exact replicate documents in the Reuters collection was the primary goal of Sanderson [10]. The method utilized correctly identified 320 pairs and only failing to find four, thus proving its effectiveness. In the creation of this detection method, they found a number of other replicate document types such as expanded documents, corrected documents, and template documents.

The efficient computation of the overlap between all pairs of web documents was considered by Shivakumar et al. [11]. The improvement of web crawlers, web archivers the presentation of search results, among others can be aided by this information. The statistics on how common replication is on the web was reported. In addition, the statistics on the cost of computing the above information for a relatively large subset of the web about 24 million web pages which correspond to about 150 gigabytes of textual information was presented.

Many organizations archiving the World Wide Web show more importance in topics dealing with documents that remain unchanged between harvesting rounds. Some of the key problems in dealing with this have been discussed by Sigurðsson [12].Subsequently, a simple, but effective way of managing at least a part of it has been summarized which the popular web crawler Heritrix [14] employed in the form of an add-on module. They discussed the limitations and some of the work necessitating improvement in handling replicates, in conclusion.

Theobald et al. [13] proved that SpotSigs provide both increased robustness of signatures as well as highly efficient replication compared to various state-of-the-art approaches. It was demonstrated that simple vector-length comparisons may already yield a very good partitioning condition to circumvent the otherwise quadratic runtime behavior for this family of clustering algorithms, for a reasonable range of similarity thresholds. Additionally, the SpotSigs replication algorithm runs "right out of the box" without the need for further tuning, while remaining exact and efficient, which is dissimilar to other approaches based on hashing. Provided that there is an effective means of bounding the similarity of two documents by a single property such as document or signature length, the SpotSigs matcher can easily be generalized toward more generic similarity search in metric spaces.

II. LITERATURE SURVEY

Recently, the detection of replicate and near replicate web documents has gained popularity in web mining research community. This survey extents and merges a wide range of works related to detection of replicate and near replicate documents and web documents. The detection techniques for identification of replicate and near replicate documents, detection algorithms, Web based tools and other researchers of replicate and near replicate documents are reviewed in the corresponding subsections.

III. PROPOSED TECHNIQUE

This standard specifies four secure hash algorithms, SHA-1 [5], SHA-256, SHA-384, and SHA-512. All four of the algorithms are iterative, one-way hash functions that can process a message to produce a condensed representation called a *message digest*. These algorithms enable the determination of a message's integrity: any change to the message will, with a very high probability, result in a different message digest. This property is useful in the generation and verification of digital signatures and message authentication codes, and in the generation of random numbers (bits).

In cryptography, SHA-1 is a cryptographic hash function designed by the National Security Agency and published by the NIST as a U.S. Federal Information Processing Standard. SHA stands for "secure hash algorithm". The three SHA algorithms are structured differently and are distinguished as *SHA-0*, *SHA-1*, and *SHA-2*. SHA-1 is very similar to SHA-0, but corrects an error in the original SHA hash specification that led to significant weaknesses. The SHA-0 algorithm was not adopted by many applications. SHA-2 on the other hand significantly differs from the SHA-1 hash function

Each algorithm can be described in two stages: preprocessing and hash computation. Preprocessing involves padding a message, parsing the padded message into *m*-bit blocks, and setting initialization values to be used in the hash computation. The hash computation generates a *message schedule* from the padded message and uses that schedule, along with functions, constants, and word operations to iteratively generate a series of hash values. The final hash value generated by the hash computation is used to determine the message digest.

The four algorithms differ most significantly in the number of bits of security that are provided for the data being hashed – this is directly related to the message digest length. When a secure hash algorithm is used in conjunction with another algorithm, there may be requirements specified elsewhere that require the use of a secure hash algorithm with a certain number of bits of security. For example, if a message is being signed with a digital signature algorithm that provides 128 bits of security, then that signature algorithm may require the use of a secure hash algorithm that also provides 128 bits of security (e.g., SHA-256).

Additionally, the four algorithms differ in terms of the size of the blocks and words of data that are used during hashing. Table 1 presents the basic properties of all four secure hash algorithms.

Algorithm	Message Size (bits)	Block Size (bits)	Word Size (bits)	Message Digest Size (bits)	Security 2 (bits)
SHA-1	<264	512	32	160	80
SHA-256	<264	512	32	256	128
SHA-384	<2128	1024	64	384	192
SHA-512	<2128	1024	64	512	256

Table 1. Basic properties of all four secure hash algorithms The performance numbers above were for a single-threaded implementation on an Intel Core 2 1.83 GHz processor under Windows Vista in 32-bit mode, and serve only as a rough point for general comparison. [15].

This function rapidly compares large numbers of files for identical content by computing the SHA-256 hash of each file and detecting replicates. The probability of two non-identical files having the same hash, even in a hypothetical directory containing millions of files, is exceedingly remote. Thus, since hashes rather than file contents are compared, the process of detecting replicates is greatly accelerated.

IV. TEST RESULTS AND ANALYSIS

It is important to mention that this process does not have to be sequential: if we have several processors, then we can eliminate records in parallel, we just need to make sure that if two record are replicates, e.g., $r1 = r2$, then when one processor eliminates $r1$ the other one does not eliminate $r2$.

To come up with a general algorithm for detecting and eliminating replicates under uncertainty, let us _rst consider an ideal case when there is no uncertainty, i.e., when replicate records $ri = (xi; yi; di)$ and $rj = (xj ; yj ; dj)$ mean that the corresponding coordinates are equal: $xi = xj$ and $yi = yj$.

In this case, to eliminate replicates, we can do the following. We _rst sort the records in lexicographic order, so that ri goes before rj if either $xi < xj$, or ($xi = xj$ and $yi \cdot yj$). In this order, replicates are next to each other. So, we _rst compare $r1$ with $r2$. If coordinates in $r2$ are

identical to coordinates in *r*1, we eliminate *r*2 as a replicate, and compare *r*1 with *r*3, etc. After the next element is no longer a replicate, we take the next record after *r*1 and do the same for it, etc.

After each comparison, we either eliminate a record as a replicate, or move to a next record. Since we only have *n* records in the original database, we can move only *n* steps to the right, and we can eliminate no more than *n* records. Thus, totally, we need no more than 2*n* comparison steps to complete our procedure.

Since 2*n* is asymptotically smaller than the time *O*(*n* ¢ log(*n*)) needed to sort the record, the total time for sorting and deleting replicates is *O*(*n*¢log(*n*))+2*n* = *O*(*n*¢log(*n*)). Since we want a sorted database as a result, and sorting requires at least *O*(*n* ¢ log(*n*)) steps, this algorithm is asymptotically optimal.

Algorithm

1. For each record, compute the indices

   pi = bxi=(C ¢ ")c; : : : ; qi = byi=(C ¢ ")c:

2. Sort the records in lexicographic order · by their index vector ~*pi* = (*pi; : : : ; qi*). If several records have the same index vector, check whether some are replicates of one another, and delete the replicates. As a result, we get an index-lexicographically ordered list of records:

   r(1) · *: : : :* · *r*(*n*0), where *n*0 · *n*.

3. For *i* from 1 to *n*, we compare the record *r*(*i*) with its following immediate neighbors; if one of the following immediate neighbors is a replicate to *r*(*i*), then we delete this neighbor.

V. CONCLUSION

We proposed a new replicate document detection algorithm called DRD and evaluated its performance using multiple data collections. The document collections used varied in size, degree of expected document replication, and document lengths. In terms of human usability, no similar document detection approach is perfect. The ultimate determination of how similar a document must be to be considered a replicate, relies on human judgment. Therefore, any solution must be easy to use. To support ease of use, all potential replicates should be uniquely grouped together.

Therefore, any match in even a single results in a potential replicate match indication. This results in the scattering of potential replicates across many groupings, and many false positive potential matches. DRD, in contrast, treats a document in its entirety and maps all potential replicate s into a single grouping. This reduces the processing demands on the user.

This paper has been felt necessary when the work on developing Replicate document detection is very hopeful, and is still in promising status. This survey paper intends to aid upcoming researchers in the field of Replicate document detection in web crawling to understand the available methods and help to perform their research in further direction.

REFERENCE

[1] BRODER, A., GLASSMAN, S., MANASSE, S., AND ZWEIG, G. 1997. Syntactic clustering of the web. In Proceedings of the Sixth International World Wide Web Conference (WWW6'97) (Santa Clara, CA., April). 391–404.

[2] SHIVAKUMAR, N. AND GARICA-MOLINA, H. 1998. Finding near-replicas of documents on the web. In Proceedings of Workshop on Web Databases (WebDB'98) (Valencia, Spain, March). 204–212.

[3] HEINTZE, N. 1996. Scalable document fingerprinting. In Proceedings of the Second USENIX Electronic Commerce Workshop (Oakland, CA., November). 191–200.

[4] SANDERSON,M. 1997. Replicate detection in the Reuters collection. Technical Report (TR-1997-5) of the Department of Computing Science at the University of Glasgow, Glasgow G12 8QQ, UK.

[5] The SHA-1 algorithm specified in this document is identical to the SHA-1 algorithm specified in FIPS 180-1

[6] McCain, M., and William C., 1998. Integrating Quality Assurance into the GIS Project Life Cycle, Proceedings of the 1998 ESRI Users Conference. http://www.dogcreek.com/html/documents.html

[7] Goodchild, M., and Gopal, S. (Eds.), 1989. Accuracy of Spatial Databases, Taylor & Francis, London.

[8] Scott, L., 1994. Identi_cation of GIS Attribute Error Using Exploratory Data Analysis, Professional Geographer 46(3), 378.386.

[9] FGDC Federal Geographic Data Committee, 1998. FGDC-STD- 001-1998. Content standard for digital geospatial metadata (revised June 1998), Federal Geographic Data Committee, Washington, D.C., http://www.fgdc.gov/ metadata/contstan.html

[10] Sanderson, M., 1997. "Duplicate Detection in the Reuters Collection", Technical Report (TR-1997-5), Department of Computing Science, University of Glasgow.

[11] Shivakumar, N., Garcia Molina, H., 1999. "Finding near-replicas of documents on the web",Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 1590, pp. 204-212.

[12] [52] Sigurðsson, K., 2006. "Managing duplicates across sequential crawls", proceedings of the 6thInternational Web Archiving Workshop.

[13] Theobald, M., Siddharth, J., Paepcke, A., 2008. "SpotSigs: Robust and Efficient Near Duplicate Detection in Large Web Collections", Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, pp. 563-570.

[14] Mohr, G., Stack, M., Ranitovic, I., Avery, D., and Kimpton, M., 2004. "An Introduction to Heritrix", 4th International Web Archiving Workshop.

[15] http://en.wikipedia.org/wiki/SHA-2

◈◈◈

Optimization of an Iterative Multiuser Detector for CDMA

Seema P Mishra, Suman P Wadkar & Bhosale J. D

Pillais Institute of Information Technology

Abstract **-** We utilize Extrinsic Information Transfer (EXIT) charts to optimize the power allocation in a multiuser CDMA system. We investigate two methods to obtain the optimal power levels: the first minimizes the total power; the second minimizes the area between the transfer curves of the interference canceller (IC) or turbo decoder. We show through simulation that the optimized power levels allow for successful decoding of heavily loaded systems. The optimal decoding schedule is derived dynamically using the power optimized EXIT chart and a Viterbi search algorithm. Dynamic scheduling is shown to be a more flexible approach which results in a more stable QoS for a typical system configuration than one-shot scheduling, and large complexity savings over a receiver without scheduling. We propose dynamic decoding schedule optimization to fix the problem, that is, on each iteration of the receiver derive the optimal schedule to achieve a target bit error rate using a minimum number of turbo decoder iterations.

I. INTRODUCTION

The advantage of the turbo decoding algorithm for parallel concatenated codes, a decade ago ranks among the most significant breakthroughs in modern communications in the past half century: a coding and decoding procedure of reasonable computational complexity was finally at hand offering performance approaching the previously elusive Shannon limit, which predicts reliable communications for all channel capacity rates slightly in excess of the source entropy rate. The practical success of the iterative turbo decoding algorithm has inspired its adaptation to other code classes, notably serially concatenated codes, and has rekindled interest in low-density parity-check codes, which give the definitive historical precedent in iterative decoding. The serial concatenated configuration holds particular interest for communication systems, since the "inner encoder" of such a configuration can be given more general interpretations, such as a "parasitic" encoder induced by a convolutional channel or by the spreading codes used in CDMA. The corresponding iterative decoding algorithm can then be extended into new arenas, giving rise to turbo equalization or turbo CDMA, among doubtless other possibilities. Such applications demonstrate the power of iterative techniques which aim to jointly optimize receiver components, compared to the traditional approach of adapting such components independently of one another.

Algorithms are often developed and tested in floating-point environments on GPPs in order to show the achievable optimal performance. Besides shortest development time, there are no requirements on, for example, processing speed or power consumption, and hence this platform is the best choice for the job. However, speed or power constraints might require an implementation in more or less specialized hardware. This transition usually causes many degradations, for example, reduced dynamic range caused by fixed-point arithmetic, which on the other hand provides tremendous reduction in implementation complexity.

II. CHANNEL CODING AND DECODING

This chapter deals with basics of channel coding and its decoding algorithms. Following is a brief description of the simple communication model that is assumed in the sequel. This model also helps to understand the purpose of channel coding. Then, two popular coding approaches are discussed more thoroughly: convolutional coding together with Gray-mapped signal constellations and set-partition coding. Decoding algorithms are presented from their theoretical background along with a basic complexity comparison.

Consider the block diagram of the simplified communication system in Figure 2.1. It consists of an information source (not explicitly drawn) that emits data symbols $\{u_k\}$. A channel encoder adds some form of redundancy, possibly jointly optimized with the modulator, to these symbols to yield the code symbol

sequence $\{c_k\}$, where ck denotes an Mary transmission symbol. Linear modulation is assumed, that is, modulation is based on a linear superposition of (orthogonal) pulses. The signal sent over the channel is therefore

$$s(t) = \sum_k c_k \cdot w(t - kT_s),$$

where $w(\,)$ is the pulse waveform and T_s is the symbol time. The waveform channel adds uncorrelated noise $n(t)$ to the signal, which results in the waveform $r(t)$ at the receiver. For the remainder, the disturbance introduced by the channel is assumed to be additive white Gaussian noise (AWGN). That is,

$$\mathcal{E}\{n(t)\} = 0$$
$$\mathcal{E}\{|n(t)|^2\} = N_0/2.$$

The received waveform $r(t)$ is demodulated to yield a discrete sequence of (soft) values $\{y_k\}$. Based on these values, the channel decoder puts out an estimate $\{\hat{u}_k\}$ for the data symbols $\{u_k\}$.

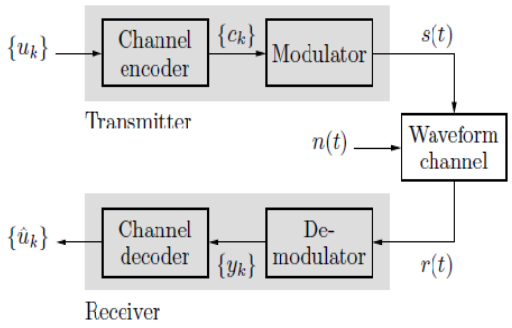


Fig. 2.1: A simplified communication system.

According to Shannon [85], reliable communication with arbitrarily low bit error rate (BER) in the AWGN channel can be achieved for transmission rates below

$$C = \frac{1}{2}\log_2\left(1 + \frac{2E_s}{N_0}\right) \text{ (bits/dimension)}.$$

If there are J orthogonal signal dimensions per channel use, the transmission rate of a (coded) communication system is defined as

$$R_d = \frac{\log_2 \mathcal{M}}{J} \cdot R_c \text{ (bits/dimension)}, \qquad (2.1)$$

where M is the number of possible symbols per channel use and $R_c < 1$ denotes the code rate of the channel code in data bits/code bits. For example, a communication system with a channel code of rate $Rc = 1/2$ per channel use and a 16-QAM constellation, that is, $M = 16$ and $J = 2$, has a transmission rate of $R_d = 1$ bit/dimension.

$$E_s = \frac{1}{\mathcal{M}}\sum_{i=1}^{\mathcal{M}} \|c_i\|^2,$$

For equiprobable signaling, the energy devoted to a transmission symbol is expressed as

or, alternatively, the energy per data bit is

$$E_b = \frac{E_s}{\log_2 \mathcal{M} \cdot R_c}. \qquad (2.2)$$

2.1 CHANNEL CODING:

A good channel code reduces the necessary E_b to achieve the same BER over a noisy channel as an uncoded transmission system of equal transmission rate R < C. This reduction is referred to as coding gain. The BER of many communication systems can be estimated in closed form based on the union bound [78]. Essentially, BER depends on the two-signal error probability, that is, the probability that one signal is mistaken for another upon decoding, and the minimum distance between signals. This probability resembles

$$p_e \sim \frac{2\mathcal{K}}{\mathcal{M}} Q\left(d_{\min}\sqrt{\frac{E_t}{N_0}}\right), \qquad (2.3)$$

where K is the number of signal pairs that lie at distance $d_{\min}$ apart from each other and $Q(\,)$ is the complementary error function defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-u^2/2)\,du.$$

In practice, BER is estimated by computer simulations of the underlying communication model. From Equation 2.3 the task of the channel code (together with the modulator) becomes apparent: either increase dmin, or decrease 2K/M, or both. Then, Eb can be lowered for the same BER.

There are two major classes of binary channel codes: block codes and convolutional codes. In the context of this thesis, only the latter codes are considered since they are widely applied in today's communication systems. Nevertheless, the rediscovery of low-density parity-check codes [49] might reclaim some share from convolutional-based coding in these systems in the near future.

2.2 DECODING ALGORITHMS:

From the considerations in Section 2.1.1, the trellis created by a convolutional encoder can be interpreted as finite-state discrete-time Markov source. Denote by *Xk* =[0,*N* − 1], *k*= Z, a possible state of the encoder at time *k*. At the receiver side, the probability of a trellis transition from state *Xk* to *Xk*+1 and

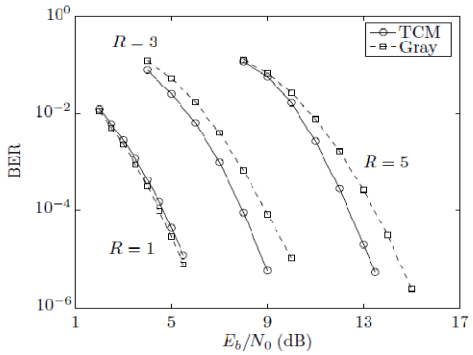


Fig. 2.1.1: Performance comparison of rate-R transmission schemes using TCM or convolutional coding with Gray-mapped constellations.

the outcome yk is given by

$$p(X_{k+1}, \mathbf{y}_k | X_k) = p(\mathbf{y}_k | X_k, X_{k+1}) \Pr(X_{k+1} | X_k) \tag{2.4}$$

Here $p(\mathbf{y}k|Xk,Xk+1)$ is the likelihood function of the received symbol $\mathbf{y}k$ given the transition ($Xk,Xk+1$) and $\Pr(Xk+1|Xk)$ is the transition's *a priori* probability. For convolutional codes, there are *c* code symbols along a trellis branch and thus $\mathbf{y}k = (y0,k \cdots yc-1,k)$. Depending on the code rate *Rc* and the transmission scheme, these *yi,k* stem from one or several i.i.d. code symbols. For TCM codes, there are subsets along the branches. These subsets consist of two-dimensional signals and $\mathbf{y}_k$ is a two-dimensional signal. When a demodulated noisy value yk is received from an AWGN channel with variance σ2 = N0/2, the likelihood function becomes

$$p(\mathbf{y}_k | X_k, X_{k+1}) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{|\mathbf{y}_k - c_k|^2}{N_0} \right).$$

One can take the logarithm of Equation 2.8 and scale with −N0 to yield the branch metric (BM)

$$\begin{aligned} \lambda(X_k, X_{k+1}) &\equiv -N_0 \log p(X_{k+1}, \mathbf{y}_k | X_k) \\ &= |\mathbf{y}_k - c_k|^2 - N_0 \log \Pr(X_{k+1} | X_k) - \underbrace{N_0 \log \frac{1}{\sqrt{\pi N_0}}}_{\text{constant}}. \end{aligned} \tag{2.5}$$

The first term in Equation 2.5 corresponds to the squared Euclidean distance between the received symbol y_k and the expected symbol ck along the branch (Xk,Xk+1). The second term is the weighted a priori probability of the branch. The constant can be neglected in the calculations since it contributes equally to all λ( ). Based on the previous notations, consider a received symbol sequence y = $\{y_k\}$. Since the channel is memoryless, maximum likelihood (ML) and maximum a posteriori (MAP) sequence estimates can be expressed as finding the I that achieves

$$\min_i \|\mathbf{y} - \mathbf{c}_i\|^2 \tag{2.6}$$

And

$$\min_i \left\{ \|\mathbf{y} - \mathbf{c}_i\|^2 - N_0 \sum_i \log \Pr(X_{i+1} | X_i) \right\}, \tag{2.7}$$

respectively. Clearly, ML and MAP decoders would estimate the same symbol sequence if all symbols were equally likely, that is, the a priori probability is equal for all branches. Then, the second term in Equation 2.11 is the same for all branches (Xi,Xi+1), and can thus be removed in calculating the branch metrics. If there is a priori information about the transition, though, the decoding might give different results for ML and MAP. In any case, ML minimizes the sequence error probability, whereas MAP can be set up so as to minimize the bit error probability [8].

III. TURBO CODES

In information theory, turbo codes (originally in French Turbo codes) are a class of high-performance forward error correction (FEC) codes developed in 1993, which were the first practical codes to closely approach the channel capacity, a theoretical maximum for the code rate at which reliable communication is still possible given a specific noise level. Turbo codes are finding use in (deep space) satellite communications and other applications where designers seek to achieve reliable information transfer over bandwidth- or latency-constrained communication links in the presence of data-corrupting noise. Turbo codes are nowadays competing with LDPC codes, which provide similar performance.

A) SOFT DECISION APPROACH:

The decoder front-end produces an integer for each bit in the data stream. This integer is a measure of how likely it is that the bit is a 0 or 1 and is also called *soft*

bit. The integer could be drawn from the range [-127, 127], where:

- -127 means "certainly 0"
- -100 means "very likely 0"
- 0 means "it could be either 0 or 1"
- 100 means "very likely 1"
- 127 means "certainly 1"
- etc.

This introduces a probabilistic aspect to the data-stream from the front end, but it conveys more information about each bit than just 0 or 1.

IV. SYSTEM DESCRIPTION

4.1 EXISTING SYSTEM:

The turbo decoding algorithm for error-correction codes is known not to converge, in general, to a maximum likelihood solution, although in practice it is usually observed to give comparable performance. The quest to understand the convergence behavior has spawned numerous inroads, including extrinsic information transfer (or EXIT) charts, density evolution of intermediate quantities, phase trajectory techniques, Gaussian approximations which simplify the analysis, and cross-entropy minimization, to name a few. Some of these analysis techniques have been applied with success to other configurations, such as turbo equalization. Connections to the belief propagation algorithm have also been identified, which approach in turn is closely linked to earlier work(6) on graph theoretic methods. In this context, the turbo decoding algorithm gives rise to a directed graph having cycles; the belief propagation algorithm is known to converge provided no cycles appear in the directed graph, although less can be said in general once cycles appear. Interest in turbo decoding and related topics now extends beyond the communications community, and has been met with useful insights from other fields; some references in this direction include which draws on nonlinear system analysis, which draws on computer science, in addition to (predating turbo codes) and (more recent) which inject ideas from statistical physics, which in turn can be rephrased in terms of information geometry. Despite this impressive pedigree of analysis techniques, the "turbo principle" remains difficult to master analytically and, given its fair share of specialized terminology if not a certain degree of mystique, is often perceived as difficult to grasp to the non specialist. In this spirit, the aim of this paper is to provide a reasonably self-contained and tutorial development of iterative decoding for parallel and serial concatenated codes. The paper does not aim at a comprehensive survey of available analysis techniques and implementation tricks surrounding iterative decoding, but rather chooses a particular ad- vantage point which steers clear of unnecessary sophistication and avoids approximations.

4.2 PROPOSED SYSTEM:

The project work focuses on joint optimization of the power and decoding schedule is prohibitively complex so we break the optimization in two parts and first optimize power levels of each user then optimize the decoding schedule using the optimized power levels. Large gainsin power efficiency and complexity can be achieved simultaneously. Furthermore, our optimized receiver has a lower convergence threshold and requires less iterations to achieve convergence than a conventional receiver. We show that our proposed optimization results in a more consistent quality of service (QoS).

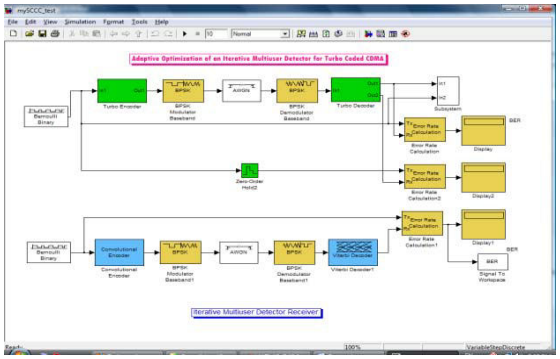


Fig.4. 1: IMUD receiver with control blocks

The major advantage of dynamic scheduling over static scheduling is that the method compensates for performance better/worse than expected (average) due to differences in channel conditions over decoding blocks, or differences in the decoding trajectory. Using dynamic scheduling we have a more reliable receiver or similar complexity.

V. IMPLEMENTATION

Implementation of any software is always preceded by important decisions regarding selection of the platform, the language used, etc. these decisions are often influenced by several factors such as real environment in which the system works, the speed that is required, the security concerns, and other implementation specific details. There are three major implementation decisions that have been made before the implementation of this project. They are as follows:

1. Selection of the platform (Operating System).

2. Selection of the programming language for development of the application.
3. Coding guideline to be followed.

5.1 IMPLEMENTATION REQUIREMENTS:

SOFTWARE REQUIREMENT:

- The language chosen for this project is Java Swing and software used is NetBeans 6.8.
- Operating System used: Microsoft windows XP

RESULTS:

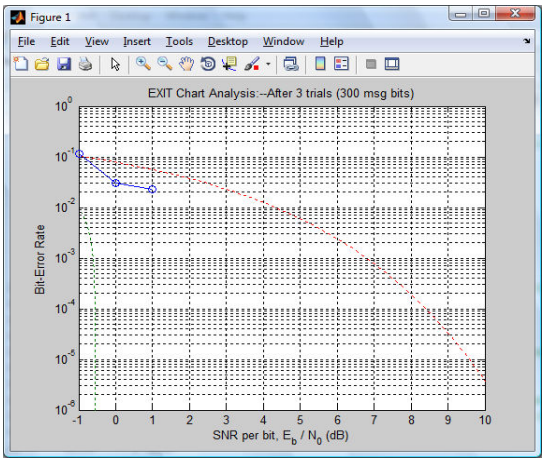


Fig 5.1 : EXIT Chart Analysis after 3 trial using 300 message bits

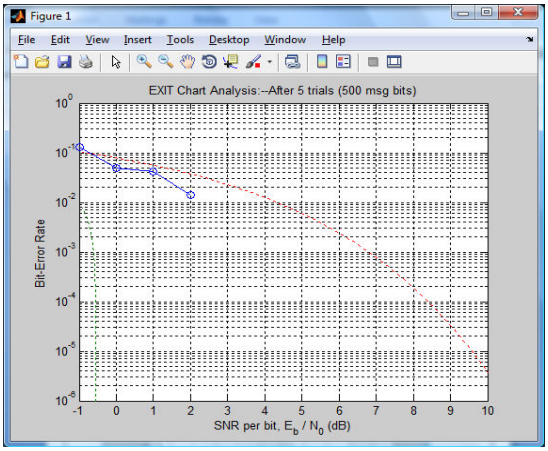


Fig. 5.2 : EXIT Chart Analysis after 5 trial using 500 message bits

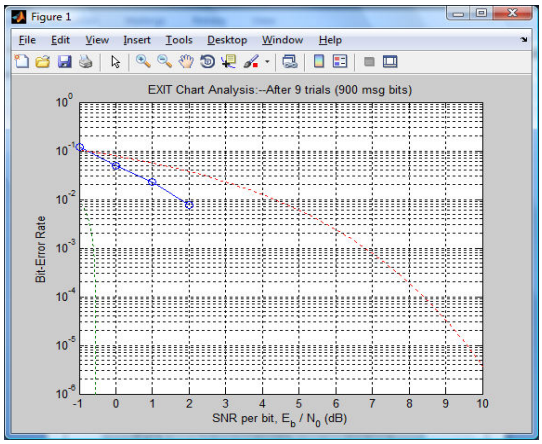


Fig 5.3 EXIT Chart Analysis after 9 trial using 900 message bits

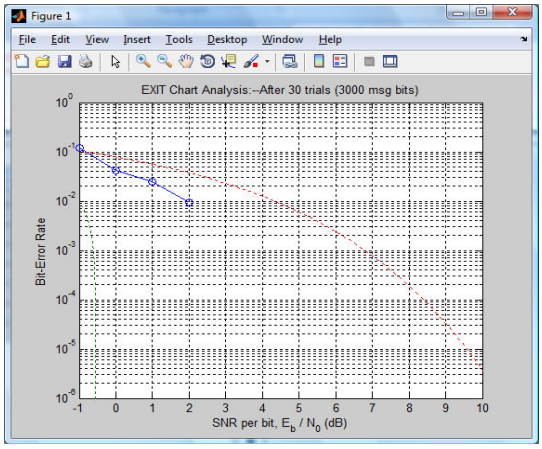


Fig. 5.4 : EXIT Chart Analysis after 30 trial using 3000 message bits

VI. CONCLUSION

We have optimized a turbo MUD receiver for unequal power turbo-coded CDMA system through EXIT chart analysis. The results in prior works were used to derive *effective* EXIT functions for FEC decoders and an interference canceller which enabled analysis of the system as in the equal power case. We utilized a nonlinear constrained optimization as in prior work to optimize the power levels of groups of users in the system. We modified the algorithm proposed in prior work to dynamically derive the optimal decoding schedule for the IMUD receiver. We then showed through simulation that this power optimized system using dynamic scheduling achieves similar BER performance as a conventional receiver with significant complexity savings. Furthermore it outperforms the statically derived optimal schedule through reducing the variance of the per packet BER. We also proposed a

method for estimating the SNR in an AWGN CDMA channel and showed that power and schedule may be optimized without any trade-off. Finally, we determined that a combination of static and dynamic scheduling offers the best benefit for the cost.

REFERENCES

[1] G. Caire, R. M¨uller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," IEEE Trans. Inform. Theory, vol. 50, no. 9, pp. 1950–1973, Sept. 2004.

[2] C. Schlegel and Z. Shi, "Optimal power allocation and code selection in iterative detection of random CDMA," in Proc. Zurich Seminar on Communications, Zurich, Switzerland, Feb. 2004, pp. 98–101.

[3] R. M¨uller and G. Caire, "The optimal received power distribution for IC-based iterative multiuser joint decoders," in Proc. Allerton Conf. on Commun., Control and Computing, Monticello, USA, Oct. 2001.

[4] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," IEEE Trans. Commun., vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[5] F. Br¨annstr¨om, L. K. Rasmussen, and A. J. Grant, "Convergence analysis and optimal scheduling for multiple concatenated codes," IEEE Trans. Inform. Theory, vol. 51, pp. 3354–3364, Sept. 2005.

[6] D. P. Shepherd, F. Br¨annstr¨om, and M. C. Reed, "Minimising complexity in iterative multiuser detection using dynamic decoding schedules," in Proc. IEEE Int. Workshop on Sig. Proc. Advanced in Wireless Communications, Cannes, France, 2006, pp. 1–5.

[7] K. Li and X. Wang, "EXIT chart analysis of turbo multiuser detection," IEEE Trans. Wireless Commun., vol. 4, no. 1, pp. 300–311, Jan. 2005.

[8] J. W. Lee and R. E. Blahut, "Convergence analysis and BER performance of finite-length turbo codes," IEEE Trans. Commun., vol. 55, no. 5, pp. 1033–1043, May 2007.

[9] D. P. Shepherd, F. Schreckenbach, and M. C. Reed, "Optimization of unequal power coded multiuser DS-CDMA using extrinsic information transfer charts," in Proc. Conf. Info. Sciences and Systems, Princeton, USA, Mar. 2006, pp. 1435–1439.

[10] D. Shepherd, F. Br¨annstr¨om, and M. Reed, "Dynamic scheduling for a turbo CDMA receiver using EXIT charts," in Proc. Aust. Commun. Theory Workshop, Adelaide, Australia, Feb. 2007, pp. 34–38.

[11] P. D. Alexander, A. J. Grant, and M. C. Reed, "Performance analysis of an iterative decoder for code-division multiple-access," European Trans. Telecom., vol. 9, no. 5, pp. 419–426, Sept./Oct. 1998.

[12] "3GPP TS 25.104 V5.9.0; 3rd generation partnership project; technical specification group radio access network;base station (BS) radio transmission and reception (FDD) (release 5)," Sept. 2004.

[13] D. P. Shepherd, Z. Shi, M. Anderson, and M. C. Reed, "EXIT chart analysis of an iterative receiver with channel estimation," in Proc. IEEE Global Telecommunications Conf., Washington D.C., USA, Nov. 2007, pp. 4010–4014.

[14] Z. Shi and C. Schlegel, "Performance analysis of iterative detection for unequal power coded CDMA systems," in Proc. IEEE Global Telecommunications Conf., vol. 3, Dec. 2003, pp. 1537–1542.

[15] D. P. Shepherd, F. Br¨annstr¨om, and M. C. Reed, "Fidelity charts and stopping/termination criteria for iterative multiuser detection," in Proc. 4th Int. Symp. on Turbo Codes and Related Topics, Munich, Germany, 2006.

[16] F. Br¨annstr¨om and L. K. Rasmussen, "Non-data-aided parameter estimation in an additive white gaussian noise channel," in Proc. IEEE Int. Symp. on Info. Theory, Adelaide, Australia, 2005, pp. 1446–1450.

[17] F. Br¨annstr¨om, "Convergence analysis and design of multiple concatenated codes," Ph.D. dissertation, Chalmers University of Technology, G¨oteborg, Sweden, 2004.

[18] M. Tuchler and J. Hagenauer, "EXIT charts of irregular codes," in Proc. Conf. Info. Sciences and Systems, Princeton, USA Mar. 2002.

[19] F. Schreckenbach and G. Bauch, "Bit-interleaved coded irregular modulation," European Trans. Telecommun., vol. 17, pp. 269–282, Mar. 2006.

[20] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," SIAM J. Optimization, vol. 6, pp. 418–445, 1996.

[21] "On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds," Mathematical Programming, vol. 67, no. 2, pp. 189–224, 1996.

[22] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced search BCJR algorithm," IEEE J. Select. Areas Commun., vol. 16, no. 2, pp. 186–195, Feb. 1998.

[23] U. Dasgupta and K. R. Narayanan, "Parallel decoding of turbo codes using soft output T-algorithms," IEEE Commun. Lett., vol. 5, no. 8, pp. 352–354, Aug. 2001.

◈◈◈

Critical Evaluation of Architectural Design Approaches Employed in Developing Software Application Systems

Yenugu. Padma

Department of Information Technology, PVPSIT, Knuru, Vijayawada, India

I. INTRODUCTION

In the last few decades, there have been great leaps of developments in the making and using of computing hardware and hence the types of enterprise applications one could dream of building. Increase in computing capacity and network abilities encouraged organizations to build complex applications. Naturally, such applications called for identifying the design patterns and leverage them using sound architectural principles. This array of application architectures has definitely introduced complexity in planning and building enterprise applications and their management.

The growth in the capacity of computing power and networking capacities enabled the application architects to venture into non-traditional ways of designing the application. This study attempts to understand and evaluate such architectural design strategies and propose a consolidated approach to use the pieces of all such architectural design strategies.

There are many aspects that make software development cumbersome and complex. The concept of software architecture has gained a wide popularity and is generally considered to play a fundamental role in coping with the inherent difficulties of the development of large-scale and complex software systems.

This study defines architectures and then develops a meta-model for architecture design methods. This model is used for classifying and evaluating various architecture design approaches. This study will focus on identifying the differences in methodology, advantages, and associated problems with these approaches.

The criterion for this classification will be the adopted basis for the identification of the key abstractions of architectures. General approaches employed in developing Architectural designs are one or more of the following resources:

1. Groupings of artefacts that are elicited from the requirement specification
2. Use case models that represents the system's intended functions
3. Architectural patterns from a pre-defined pattern catalogue, and
4. Domain models

Based on these we have *artefact-driven*, *use-case-driven*, *pattern-driven* and *domain-driven* architecture design approaches in practice.

This study proposes to complete the analysis and document the following for each approach of developing the architectural abstractions:

1. For each approach, we will catalogue the advantages, problems
2. Impact of these architectural designs on interoperability and Security of the system, and
3. Impact of a this architectural design on Project Management activities, and
4. Recommendations to ensure that the resources used are optimal

II. DESIGN THE ARCHITECTURE

Technical expertise is a critical success factor for designing the architecture. The architect must be:

- Familiar with existing architecture styles
- Knowledgeable about emerging technologies to consider as alternative design strategies
- Competent in mapping decisions about technology to the requirements

The architect must communicate with domain experts or subject matter experts in order to ensure that the business, quality, and functional requirements are being addressed by the candidate technologies.

This activity is the first step of an iterative process and is followed by documenting and analyzing the architecture as part of each iteration. Architecture

design may encompass combinations of other practices, such as those listed below, depending on the domain, scope, desired functionality, and complexity of the solution.

- Agreement on Interfaces
- Ensure Interoperability
- Leverage COTS/NDI
- Plan for Technology Insertion
- Formal Risk Management
- Assess Reuse Risks and Costs

III. DOCUMEENT THE ARCHITECTURE

This is the second step in the iterative process that is part of the Architecture-First Approach. The architecture itself is intangible; it is a concept, or collection of concepts, in the mind of the architect. The concept must be documented (or demonstrated through prototyping) in order to communicate the concept to all the stakeholders to ensure a common understanding of the system structure and behavior.

Documenting the architecture raises questions and forces clarification of issues not yet addressed.

Present multiple views. Each description (view) should allow a stakeholder to easily distinguish among the various design considerations presented. For a small standalone application, an architecture description can be as simple as a one-page drawing that illustrates the software components by name, and their connectors (how they relate to each other). Typically, architecture descriptions are collections of views of a system from several different perspectives (e.g. logical or functional, hardware, software, behavioral views, etc.). There should be a view to accommodate each type of stakeholder.

The larger the system the greater the need for focusing on architecture.

Use tools and "Architecture Description Languages (ADL)" when appropriate to facilitate and support use of the architecture for decision making. A new technology, ADL, is emerging that addresses languages for representing and analyzing architectures. ADLs can be used to document architectures in a way that allows analysis and exchange through other software tools.

Document each iteration of the architecture. There are no strict rules or sequence that must be followed with successive iterations. Some architects may choose to address the several architectural views concurrently but at a very high view. Then, with each successive iteration, the views would be refined. Others may elect to focus on a particular view until it is deemed sufficient and then, in successive iterations, focus on other views.

The most important aspect of this activity is that the architecture description is the basis for evaluating the architecture – and therefore, comes first and is refined with each iteration. Documenting the architecture AFTER assessing it through verbal discussion is not acceptable in the Architecture-First Approach. The description drives the analysis activity and therefore must precede it. Documenting an architecture to sit on a shelf as part of a project history is not the purpose of this activity.

IV. ANALYZE THE ARCHITECTURE

Step 3, analyzing the architecture, encompasses the following areas:

- **Ensure that stakeholders participate in the analysis**:

For each architectural view presented, the appropriate stakeholders need to be involved. The architect bears the burden of presenting the candidate architectures in a way that is meaningful to the "non-technical" stakeholder, as well as the "teckies". The domain expert (or user representative) may address usage scenarios and issues that were not previously considered during requirements elicitation. Substantial savings in time and costs are realized when problems are discovered and resolved during this "elaboration" phase, rather than waiting to be discovered during testing.

- **Assess the architecture with respect to functional goals, business goals, and quality goals**:

Does the proposed architecture address the goals that were originally identified as architectural drivers? Typically the act of asking the question reveals gaps, or identifies areas of conflict between the proposed architecture and business or quality goals. This provides the thrust for the next iteration. Eventually the architecture will be deemed to be responsive to all the goals, or the stakeholders will have negotiated trade-offs during the process. The architecture documentation includes capturing the rationale for choosing a particular architecture collection.

- **Identify decisions/action items that will drive the next iteration**:

Analysis and evaluation typically result in some discoveries that become the focus for refinement during the next iteration.

- **Conduct independent expert reviews by technically competent people**:

On large projects independent expertise is brought in to verify that the proposed architecture does, in fact, address the stated requirements and quality goals. Under a traditional development scenario, the architecture would not be scrutinized to this extent prior to approval.

- **Prove, through demonstration, that the proposed architecture will meet the stated requirements**:

This often involves creating prototypes for the proposed architecture in order to demonstrate that quality and performance goals can be met. Using independent technical experts helps to ensure the validity of the architecture prior to moving on to full-scale development.

V. REALIZE THE ARCHITETURE

This activity is about getting the most value from the time and effort expended on the architecture.

Use the architecture to drive decision-making and acquisition strategy for committing resources and structure for full-scale development. Decisions about software, operating systems, development languages, and tools, are made during architecting. Structuring and tasking development teams before the architecture is stabilized can adversely impact project schedule and cost. The teams may not be working on the right components. Team members may not have the expertise to implement the architecture that is ultimately selected. The team structure (from an organizational perspective) may not be well aligned for developing the components identified by the architecture.

The organizational structure of the development team must be easily mapped onto the software architecture and vice versa. Once the architecture has been agreed upon teams are allocated to work on the major components and the work breakdown structures that are created reflect those teams. For large systems, teams may belong to different subcontractors. Each group will need to establish liaisons and coordinate with the other groups. Team structure and controlling team interaction often turns out to be the largest single factor affecting a large project's success.

VI. MAINTAIN THE ARCHITECTURE

"Architecture-First" carries two meanings:

- Develop the architecture before committing resources for full scale development, and
- Focus on the architecture when making changes to a system in a maintenance environment.

There is a risk that the architecture may drift from its original precepts if a maintenance process is not addressed. If this happens, the original system (that was so carefully designed and analyzed) will be compromised and much of the value derived from the original "architecting" will be lost. Drifting may occur when changes are made by multiple developers that are not mutually consistent, or do not follow the original rationale for design decisions. The challenge is to ensure that the architecture of the "as-designed" and "as-built" systems remain congruent with the "as-maintained" system over time.

The architecture documentation should capture that rationale for the architecture.

Both the acquiring organization and the developing organization(s) need to ensure that the architecture description is well disseminated for use as a development reference, and both need to ensure that proposed changes to the system are reviewed for conformance to the architecture prior to implementation.

If architectural changes are necessary, then both organizations, (acquirer and developer), need to ensure that the architecture description is modified to reflect the new rationale.

Many architectural constructs have no realization in the development artifacts that programmers actually create and maintain. Without an architecture description, the maintainer (who may be a different organization from the developer) tries to "interpret" from the existing code what the architecture may have been, influenced by their own individual development experiences, which may be significantly different from what was in the mind of the architect. This is how the inconsistencies surface.

This sounds like a "simple" obvious thing to do – but very few organizations have actually done it. Consequently, over time, system designs are compromised because developers lose sight of the design rationale. Maintaining the architecture is really about maintaining a focus on "architecting" through the evolution of the system life cycle. It depends upon having a well-documented, well-disseminated architecture to start with, and having management support to maintain the architecture over the life of the system.

Many organizations may be maintaining systems that have no architecture descriptions. It is possible to initiate "architectural reconstruction" where an architect examines the code and associates naming conventions, file structures, and functions with architectural constructs, deriving useful abstractions that are relevant to the continued life cycle of the system. Architectural

reconstruction may be useful to organizations wanting to reengineer an existing system, or mine its existing software assets for reuse and product line development.

Some tools are now being developed to aid analysts in extracting information from systems in order to analyze and define patterns that ultimately define the architecture. See [Kasman, 1999] for further details on architectural reverse engineering and conformance testing.

REFERENCES:

[1] Object-Oriented Analysis and Design with Applications (3rd Edition) by Grady Booch, Robert A. Maksimchuk, Michael W. Engel, and Bobbi J. Young (Apr 30, 2007)

[2] Refactoring: Improving the Design of Existing Code by Martin Fowler, Kent Beck, John Brant, and William Opdyke (Jul 8, 1999)

[3] Patterns of Enterprise Application Architecture by, Martin Fowler et al.

[4] The Unified Modeling Language Reference Manual by James Rumbaugh, Ivar Jacobson, and Grady Booch (Jan 2, 1999)