Proceeding for International Conference

on

Emerging Trends

in

EngineeringTechnology

Chennai

25th October ’15

Institute for Engineering Research and Publication

(A Unit of VVERT)

4A, Girija Apartment, MMDA

Arumbakkam, Chennai-600106, India

www.iferp.in

Publisher: IFERP Explore

Editorial:

We cordially invite you to attend the International Conference on Emerging Trends in Engineering and Technology (ICET-15), which will be held in The Vijay park, Chennai on October 25, 2015. The main objective of ICET-15 is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Electronics, Mechanical, Electrical, Computer Science and Information Technology. This conference provides opportunities for the delegates to exchange new ideas and experience face to face, to establish business or research relations and to find global partners for future collaboration.

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

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

Since March 2015, the Organizing Committees have received more than 120 manuscript papers, and the papers cover all the aspects in Electronics, Computer Science and Information Technology. Finally, after review, about 9 papers were included to the proceedings of ICET- 2015.

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

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

Acknowledgement

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

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

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

CONTENTS

ORGANIZATION COMMITTEE

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

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

ORGANIZING SECRETARY
Asst. Prof. Bonia Mohan,
Department of CSE,
Dayananda sagar College of Engineering,
Bangalore,

PUBLICATIONS COMMITTEE
Dr. Shankar Narayanan
Dr.S. Sangeetha

PROGRAM CHAIR
Dr. Nalini Chidambaram
Professor
Bharth University

IMPROVED HARMONY SEARCH ALGORITHM FOR OPTIMAL PLACEMENT AND SIZING OF STATIC VAR COMPENSATORS IN POWER SYSTEMS

[1]A. Ashok Reddy,[2]P.B.Chennaiah,[3]Dr.M.PadmaLalitha
[1]PG Student,[2]Assistant Professor,[3]Professor & Head of the Department of Electrical & Electronics Engineering,Annamacharya Institute of Technology & Sciences, Rajampet, Andhra Pradesh , India.
E mail: ashokeee2012@gmail.com

***Abstract*—Voltage Instability is one of the major phenomena which has resulted a major obstruct to power system network. Use FACTS controllers possible that voltage stability status in a stressed power system could be improved with effective reactive power compensation. This paper proposes at multi-objective optimization problem such as minimization of real power loss, L-index and load voltage deviation. To find the critical buses in the system for optimal location of shunt connected FACTS controller known as Static Var Compensator (SVC) here used L-index. To find the optimal sizes of SVC for solving Multi-objective optimization problem a Meta-Heuristic Algorithm Known as Harmony Search Algorithm (HAS) will be applied. The Simulation works are performed on IEEE-14 bus test system. The results are shown that optimal location and sizing of SVC minimizes real power losses, load voltage deviation, L-index and also Voltage profiles are improved at different loading conditions. In this present paper work 125%, 150%, 175%, 200% overloading cases are considered.**

Index Terms—*Voltage stability, SVC, Harmonic Search Algorithm, L-index*

I. INTRODUCTION

Voltage stability can be defined as the ability of a power system to maintain acceptable voltage levels under normal operating conditions and after occurrence of disturbances [1]-[2].In current days an instability ,usually known as voltage instability which has been observed and been responsible for major network collapses in many countries. The voltage instability is mainly related with reactive power imbalance. The loading of a bus in the power system depends on the reactive power support that the bus can receive from the system. When the system reaches the maximum loading point (MLP) or the voltage collapse point both the real and reactive power losses increases rapidly. Therefore the Reactive power support must be local and adequate [3]-[5].Introducing sources of reactive power that is shunt capacitors and /or Flexible AC Transmission systems(FACTS)controllers at the suitable location is the most effective way for Utilities to enhance the voltage stability of the system. The rapid development of fast-acting and self-commutated power electronics converters, well known as Flexible Alternating Current Transmission system (FACTS)controllers, introduced in 1988 by Hingorani [6],are useful in taking fast control actions to ensure these unity of power systems. The Static VAR compensator (SVC) has been effectively used to provide voltage stabilization at critical buses amongst existing FACTS devices. In [7] the effects of SVC and TCSC on voltage collapse are studied by Canizares and Faur .The Voltage Stability Assessment of system with shunt compensation devices including shunt capacitors, SVC and STATCOM are compared in the IEEE 14 bus system [8]. FACTS devices are expensive and are not economical to place more devices in the system. In the literature there are different indices to find the weak bus for the location of FACTS devices [9]-[10]. Hence optimal placement and sizing of FACTS devices are the important issues. Appropriate placement of FACTS devices at suitable location with proper sizes would lead to maximum loading margin [11]-[13]. In [14] D.Thukaram and Abraham lomi proposed to select a suitable size and location of SVC in EHV network for voltage stability improvement based on L-Index of load bus. Four different objective functions namely, loss minimization, voltage profile improvement, Voltage stability Enhancement and Total cost minimization are proposed by S.Durairaj et al [15].

In this paper SVC is used for shunt compensation. It is a shunt-connected Static VAR Generator or absorber whose

output is adjusted to exchange capacitive or inductive current so as to provide voltage support. It can also reduce power losses in the system when it is installed in a proper location. Here L-index is used to find the weak bus in the system to place the SVC device and it is also for Voltage Stability analysis. L-index gives scalar number to each load bus. This index value ranges from 0 (no load system) to 1(Voltage collapse). The bus with highest L-index value will be the weakest bus in the system and hence this method helps in identifying the weak load bus which need critical reactive power support. Minimization of L-index is also one of the objectives of the optimization problem. A Meta-Heuristic algorithm Known as HSA is proposed to find the optimal sizes of SVC for multi objective of optimization such as Minimization of Real power loss, Voltage Deviation and Improvement of Voltage profile.

II. PROBLEM STATEMENT

The main objective function of this paper is to find the optimal rating of SVC for multi-objective optimization. This is mathematically stated as [16]-[18]:

Minimize F= [f1, f2, f3] (1)

Where f1 represents the real power losses as

$$f_1 = \sum_{k \in Ni} g_k \left(V_i^2 + V_j^2 - 2V_i V_j \cos \theta_{ij} \right) = P_{loss} \quad (2)$$

f2 represents the total voltage Deviation (VD) of all load buses from desired value of 1 p.u.

$$f_2 = VD = \sum_{k=1}^{N_{PQ}} \left(V_k - Vref_k \right)^2 \quad (3)$$

And f3 is the L-index of the jth bus and is given by:

$$f_3 = L_j = \left| 1 + \frac{V_{oj}}{V_j} \right| = \frac{S_j^*}{Y_{ji} V_j^2} \quad (4)$$

The minimization problem is subject to the following equality and inequality Constraints:

i) Load Flow Constraints:

$$P_i - V_i \sum_{j=1}^{N_B} V_j \left(G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij} \right) = 0, i = 1,2 \ldots\ldots N_B - 1 \quad (5)$$

$$Q_i - V_i \sum V_j \left(G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij} \right) = 0, i = 1,2 \ldots\ldots N_{PQ} - 1 \quad (6)$$

(ii) Voltage constraints:

$$V_i^{min} \le V_i \le V_i^{max}, i \in N_B \quad (7)$$

(iii) Reactive Power Generation Limit:

$$Q_{gi}^{min} \le Q_{gi} \le Q_{gi}^{max}, i \in N_g \quad (8)$$

(iv) Reactive Power Generation

$$Q_{ci}^{min} \le Q_{ci} \le Q_{ci}^{max}, i \in N_c \quad (9)$$

(v)Transformer Tap setting limit:

$$t_k^{min} \le t_k \le t_k^{max}, k \in N_t \quad (10)$$

(vi)Transmission line flow limit:

$$S_i \le S_i^{max}, i \in N_l \quad (11)$$

III. SVC IDEAL MODELLING

Static Var Compensator is shunt connected type FACTS device which output is adjusted to exchange capacitive or inductive current and is used to maintain reactive power in network. And SVC contains two main components. Thyristor controlled/switched reactor (TSR) and switched capacitor (TSC). To absorb reactive power TSR is used. And to provide the reactive power TSC is used under serious loading conditions of network. The Static Var Compensator (SVC), constructional details, characteristics and modelling are in [19]-[20]. Fig.1 shows the schematic diagram of SVC connected to an infinite bus. The operating range of SVC is -200Mvar to 200Mvar.

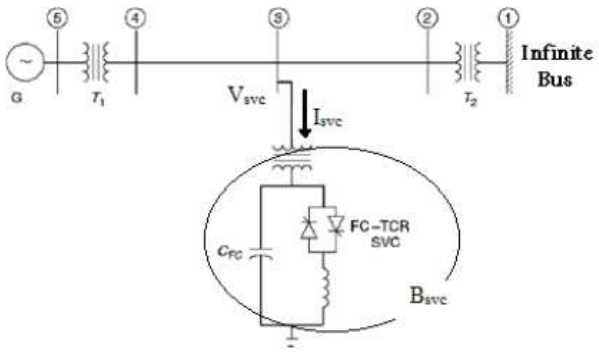


Fig 1: circuit diagram of svc connected to an infinite bus

From Fig.1, the current drawn and reactive power injected

By the SVC can be expressed as:

$$I_{SVC} = jB_{SVC} \times V$$

$$Q_{SVC} = -jB_{SVC} \times V^2$$

The reactive power generated by an SVC is given by

$$Q_{svc}^{min} \le Q_{svc} \le Q_{svc}^{max} \quad (12)$$

IV. VOLTAGE STABILITY INDEX

In [21], Kessel et al. was developed a voltage stability index based on the solution of the power flow equation. The L-index is a quantitative measure for the estimation of the distance of actual state of the system stability limit. It describes the stability of the complete system. Voltage stability index Lj for any load bus can be defined as given in equation (13)

$$L_j = \left|1 + \frac{V_{oj}}{V_j}\right| = \frac{S_j^*}{Y_{jj} V_j^2} \quad (13)$$

$$\text{Where } V_{oj} = -\sum_{i \in \alpha_j} F_{ji} V_i$$

Where the L-index varies between 0 (no-load) and 1(voltage collapse) and it gives scalar number to each load bus. When the L-index value approaches to 0 the voltage stability is assured.

V. POWER SYSTEM VOLTAGE STABILITY

At any point of time, a power system operating condition should be stable, meeting various operational criteria, and it should also be secure in the event of any credible contingency. Present day power systems are being operated closer to their stability limits due to economic and environmental constraints. Maintaining a stable and secure operation of a power system is therefore a very important and challenging issue. Voltage instability has been given much attention by power system researchers and planners in recent years, and is being regarded as one of the major sources of power system insecurity. Voltage instability phenomena are the ones in which the receiving end voltage decreases well below its normal value and does not come back even after setting restoring mechanisms such as VAR compensators, or continues to oscillate for lack of damping against the disturbances. Voltage collapse is the process by which the voltage falls to a low, unacceptable value as a result of an avalanche of events accompanying voltage instability . Once associated with weak systems and long lines, voltage problems are now also a source of concern in highly developed networks as a result of heavier loading.

HARMONY SEARCH ALGORITHM

The harmony search algorithm (HSA) is a new meta-heuristic algorithm [22] – [23] inspired by the operation of orchestra music to find the best harmony between components which are involved in the operation process, for optimal solution. It is simple in concept from natural musical performance processes. The musicians starting with some discrete musical notes based on player experience so finally HSA gives optimum value.

ALGORITHM TO FIND OPTIMAL SIZES OF SVC USING HARMONY SEARCH ALGORITHM

Step 1: Initialize all the parameters and constants of the Harmony search algorithm. They are QSVCminimum and QSVC maximum, hms, HMCR, PARmin and PARmax.
Step 2: Run the load flow program and find the total real power loss of the original system.
Step 3: Initialize the harmony memory i.e., generates [hms x n] number of initial solutions randomly within the limits, where hms is the harmony memory size and *n* is the number of static var compensators (SVC).
Step 4: obtain the loss reduction (fitness value) using equation (14)Fitness Value = Minimize F= [f1, f2, f3] (14)Repeat the same procedure for all the rows of the harmony vector to find Fitness values and obtain the best fitness value by comparing all the fitness values.
Step 5: Start the improvisation and iteration count is set to one.
Step 6: Improvisation of the New Harmony is generating a new harmony. A New Harmony vector is generated based on the following steps:
(i) **Random selection:** It is used to select one value randomly for a certain element of the new vector from the possible range (Qsvcmin, Qsvc max) of values.
(ii) **Memory consideration:** It is used to choose the value for a certain element of the new vector from the specified HM range.
xi=x'i_{x'1,x'2,…x HMS
i } with probability HMCR
(15)
x'i_ xi with probability (1-HMCR)
(16)
Step 7: Pitch adjustment: It is used to adjust the values of the New Harmony vector obtained in step7. (Between PARmin and PARmax). (bw - band width varies between a higher value and a lower value from first iteration to last iteration)x'i=x'i±rand(0,1)*bw
(17)
Step 8: Find the fitness values corresponding to the New Harmony generated and pitch adjusted insteps 6 and 7.
Step 9: Apply Greedy Search between old harmony and New Harmony by comparing fitness values.
Step 10: Update harmony memory, by replacing the worst harmony with the new best Harmony. Obtain the best fitness value by comparing all the fitness values.
Step 11: The improvisation (iteration) count is incremented and if iteration count is not reached maximum then go to step 7.
Step 12: The solution vector corresponding to the best fitness value gives the optimal SVC sizes in n optimal locations. In the present paper the HAS parameters are hms=

30, HMCR =85%, No of improvisations = 200, PARmin = 0.4 and PARmax = 0.9.

RESULTS AND DISCUSSION

IEEE 14 bus system [24] contains 5 generator buses (bus numbers: 1,2,3,6 and 8), 9 load buses (bus numbers: 4, 5, 7,9,10,11,12,13 and14) and 20 transmission lines including 3 transformers The details of the system data including 3 transformer nominal values are given in [24].The load has been increased from normal load by125%, 150%, 175% and 200% for IEEE 14-bus test system. As the load on the system increases L-index, real power losses and voltage Deviation at load buses also increases. The results of the corresponding are shown in tables 1- 4.However L-Index is used to find the Weak buses in the system to find the optimal location of Static VAR Compensator (SVC). When the load on the system increases buses 9 and 14 has more L-index and so these buses are the best locations to place the SVC. A Meta-heuristic algorithm known as Harmonic search Algorithm is used to find the optimal size of SVC to achieve multi-objectives. And finally when The SVC devices are placed at the buses 9 and 14 with optimal sizes and the corresponding results such as real power loss, voltage Profile and L-index with different loading

Conditions are shown in table1-4.

TABLE: 1 RESULT OF THE IEEE 14 BUS TEST SYSTEMS

LOADING CONDITION	REAL POWER LOSS WITHOUT SVC	SVC OPTIMAL LOCATION	HAS	
			RATING OF SVC	REAL POWER LOSS WITH SVC
NORMAL LOADING	13.3934	9 14	24.5573 6.9526	13.3336
125% LOADING	22.7259	9 14	36.1120 8.9587	22.1941
150% LOADING	35.5578	9 14	61.9958 15.2668	34.3739
175% LOADING	51.61	9 14	106.7474 15.6894	49.6396
200% LO ADING	70.8595	9 14	134.3521 19.2670	68.9691

TABLE: 2 L-INDEX AND VOLTAGE PROFILES AT BASECASE LOADING

BASE CASE LOADING				
BUS NO	WITHOUT SVC		WITH SVC	
	L-INDEX	VOLTAGE	L-INDEX	VOLTAGE
1	0.0000	1.0600	0.0000	1.0600
2	0.0000	1.0450	0.0000	1.0450
3	0.0000	1.0100	0.0000	1.0100
4	0.0116	1.0183	0.0115	1.0198
5	0.0020	1.0200	0.0020	1.0211
6	0.0000	1.0700	0.0000	1.0700
7	0.0000	1.0608	0.0000	1.0659
8	0.0000	1.0900	0.0000	1.0900
9	0.0123	1.0541	0.0120	1.0642
10	0.0061	1.0495	0.0060	1.0579
11	0.0038	1.0561	0.0038	1.0604
12	0.0084	1.0550	0.0084	1.0572
13	0.0106	1.0501	0.0105	1.0542
14	0.0248	1.0343	0.0240	1.0521

TABLE: 3 L-INDEX AND VOLTAGE PROFILE AT 125% LOADING

125% LOADING				
BUS NO	WITHOUT SVC		WITH SVC	
	L-INDEX	VOLTAGE	L-INDEX	VOLTAGE
1	0.0000	1.0600	0.0000	1.0600
2	0.0000	1.0250	0.0000	1.0350
3	0.0000	0.9800	0.0000	1.0000
4	0.0123	0.9883	0.0118	1.0079
5	0.0021	0.9926	0.0021	1.0097
6	0.0000	1.0400	0.0000	1.0700
7	0.0000	0.0302	0.0000	1.0623
8	0.0000	1.0700	0.0000	1.0900
9	0.231	1.0176	0.1920	1.0635

10	0.0066	1.0119	0.0061	1.0554
11	0.0040	1.0213	0.0038	1.0583
12	0.0090	1.0204	0.0084	1.0540
13	0.0113	1.0139	0.0106	1.0502
14	0.369	0.9925	0.0241	1.0483

TABLE: 4 L-INDEX AND VOLTAGE PROFILES AT 150% LOADING

150% LOADING				
BUS NO	WITHOUT SVC		WITH SVC	
	L-INDEX	VOLTAGE	L-INDEX	VOLTAGE
1	0.0000	1.0600	0.0000	1.0600
2	0.0000	1.0050	0.0000	1.0150
3	0.0000	0.9600	0.0000	0.9600
4	0.0130	0.9607	0.0123	0.9862
5	0.0023	0.9669	0.0022	0.9902
6	0.0000	1.0200	0.0000	1.0700
7	0.0000	1.0015	0.0000	1.0614
8	0.0000	1.0500	0.0000	1.0900
9	0.326	0.9845	0.224	1.0742
10	0.0070	0.9788	0.0060	1.0624
11	0.0043	0.9935	0.0038	1.0610
12	0.0094	0.9952	0.0084	1.0524
13	0.0120	0.9865	0.0106	1.0495
14	0.623	0.9559	0.528	1.0585

TABLE: 5. L-INDEX AND VOLTAGE PROFILES AT 175% LOADING

175% LOADING				
BUS NO	WITHOUT SVC		WITH SVC	
	L-INDEX	VOLTAGE	L-INDEX	VOLTAGE
1	0.0000	1.0600	0.0000	1.0600
2	0.0000	0.9950	0.0000	1.0050
3	0.0000	0.9600	0.0000	0.9600
4	0.0134	0.9448	0.0125	0.9789
5	0.0023	0.9514	0.0022	0.9806
6	0.0000	1.0200	0.0000	1.0700
7	0.0000	0.9851	0.0000	1.0726
8	0.0000	1.0400	0.0000	1.0900
9	0.426	0.9651	0.329	1.1034
10	0.0073	0.9606	0.0057	1.0848
11	0.0044	0.9833	0.0037	1.0715
12	0.0095	0.9898	0.0085	1.0511
13	0.0122	0.9785	0.0106	1.0492
14	0.792	0.9358	0.6289	1.0711

TABLE: 6 L-INDEX AND VOLTAGE PROFILES AT 200% LOADING

BUS NO	WITHOUT SVC		WITH SVC	
	L-INDEX	VOLTAGE	L-INDEX	VOLTAGE
1	0.0000	1.0600	0.0000	1.0600
2	0.0000	0.9950	0.0000	0.9950
3	0.0000	0.9600	0.0000	0.9600
4	0.0138	0.9326	0.0128	0.9658
5	0.0024	0.9390	0.0023	0.9655
6	0.0000	1.0200	0.0000	1.0600
7	0.0000	0.9740	0.0000	1.0730
8	0.0000	1.0400	0.0000	1.0900
9	0.5282	0.9494	0.4282	1.1127
10	0.0076	0.9453	0.0057	1.0889
11	0.0044	0.9745	0.0037	1.0678
12	0.0096	0.9845	0.0086	1.0393
13	0.0124	0.9708	0.0108	1.0383
14	0.829	0.9176	0.7326	1.0721

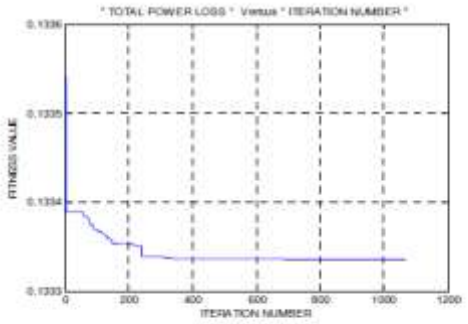


Fig: 2 Performance of algorithm of HAS algorithm

CONCLUSION

The performance of HSA optimization algorithm is

presented and applied to determine the rating of SVC which satisfies the multi-objectives such as minimization of real power loss, Voltage stability level index (L-Index), Load voltage deviation and improvement of voltage profile. The proposed multi-objective HSA algorithm has been validated on the IEEE-14 bus test system for all loads that is 125%, 150%, 175% and 200% of normal loading and it is observed the proposed algorithm using SVC the multi-objectives are achieved. L-index is used to find the weak buses in the system for optimal placement of SVC. Further it is possible to achieve better by using TCSC and UPFC.

REFERENCES

[1]. P.Kundur, Power system Stability and control, Newyork: McGraw-Hill, 1994.
[2]. IEEE/CIGRE, Joint taskforce on stability and Definitions, "Definition and classification of power system stability", IEEE transactions so power system", vol.19, no.2, pp.1387-1401, May 2004.
[3]. Dubson, H.D.chiang, "Towards a theory of Voltage collapse in electric power systems, Systems and control letters, Vol.13, pp.253-262, 1989
[4]. T.V.cutsem, "Voltage instability: phenomena, counter measures and analysis methods", proceedings of IEEE, Vol.88, pp.208- 227, Feb, 2000.
[5]. C.W.Taylor, Power system voltage stability, Newyork: McGraw-Hill 1994.
[6]. N.G. Hingorani, L. Gyugyi, and Understanding FACTS: Concepts and Technology of Flexible AC Transmission Systems, New York, IEEE Press, 2000.
[7]. C.A.Canizares, Z.T.faur, "Analysis of SVC and CSC Controllers in Voltage collapse, IEEE Transactions on power systems, vol.14, no.1, February 1999, pp.158-165.
[8].Sode-Yome and N.Mithulanathan, "comparison of Static voltage stability margin Enhancement" International journal of Engineering Education, UMIST, Vol.41, no.3, july2004.
[9]. Claudia Reis, and F.P.Maciel Barbosa, "A comparison of Voltage stability indices", IEEEmelcon 2006, may, 16-19, Benalmadena (Malaga),Spain.
[10]. VenkataramanaAjjarapu and Colin cristy, "The continuation power flow: A tool for steady state voltage stability analysis.", IEEE Transactions on Power systems, vol.7, no.1, pp.416-423, 1992.
[11]. J.G.Sing, S.N.singh, S.C.Srivastava, "Placement of FACTS controllers for enhancing power system load ability", in proceedings of IEEE Power India Conference, 2006, pp.10-17.
[12]. S.Gerbex, R.Cherkaoui, A.J. Germond, "optimal location of FACTS devices to enhance power System security", in proceedings of the IEEE power tech conference, 2003, Bologna,vol.3, pp.1-7.
[13]. F.Jurado, J.A. Rodriguez, "optimal location of SVC based on system load ability and contingency analysis", in proceedings of the emerging Technologies Factory automation Conference, 1999, vol.2, pp.1193-1199.
[14]. D.Thukaram, Abraham Lomi, Selection of static VAR Compensator Location and Size for System Voltage Stability Improvement, Electric Power Systems Research, Vol.54, 2000.

SELECTION AND SIZE OF MULTIPLE DGS USING KALMAN FILTER ALGORITHM FOR REDUCTION OF POWER LOSS AND VOLTAGE PROFILE IMPROVEMENT

[1]K. Babu Reddy, [2]K. Harinath Reddy,[3]P. Suresh Babu
[1]PG Student, [2]Assistant Professor, [3]Assistant Professor
Annamacharya Institute of Technology & Sciences, Rajampet, Andhra Pradesh , India
[1]Babureddy.kota@gmail.com,[2]harinathreddyks@gmail.com,[3]sureshram48@gmail.com

Abstract:Now a days, the consumption of electric power has been increased enormously which necessitates for the construction of new power plants, transmission lines, towers, protecting equipment etc. The environmental pollution is one of the major concerns for the power generation and also the cost of installing new power stations is high. Hence the distributed generation (DG) technology has been paid great attention as far as a potential solution for these problems.The beneficial effects of DG mainly depend on its location and size. Therefore, the selection of optimal location and size of the DG plays a key role to maintain the stability and reliability of existing system effectively before it is connected to the power grid. In this paper, a method to determine the optimal locations of DG is proposed by considering power loss. Also, their optimal sizes are determined by using kalman filter algorithm. The proposed KFA based approach is to be tested on standard IEEE-30 bus system.

Index Terms: Distributed Generation, Optimal location, Optimal Size, Power loss and Kalman filter algorithm.

I. INTRODUCTION

The structure, operation, planning and regulation of electric power industry will undergo considerable and rapid change due to increased prices of oil and natural gas. Therefore, electric utility companies are striving to achieve power from many different ways; one of them is distributed generation solution by an independent power producer (IPP) to meet growing customer load demand [1]. The renewable energy sources such as fuel cell, photovoltaic and wind power are the sources used by the distribution generation. In recent years, it becomes an integral component of modern power system for several reasons [2]. For example, the DG is a small scale electricity generation, which is connected to customer's side in a distribution system. The additional requirements such as huge power plant and transmission lines are reduced. So, the capital investments are reduced. Additionally, it has a great ability for responding to peak loads quickly and effectively. Therefore, the reliability of the system is improved. It is not a simple plug and play problem to install DG to an electric power grid. The non-optimal locations and non-optimal sizes of DG units may lead to stability, reliability, protection coordination, power loss, power quality issues, etc. [1]–[4].

First of all, it is important to determine the optimal location and size of a given DG before it is connected to a power system. Moreover, if multiple DGs are installed, an optimal approach for selection of their placement and sizing is imperative in order to maintain the stability and reliability of an existing power system effectively. This paper proposes a method to select the optimal locations of multiple DGs by considering total power loss in a steady-state operation. Thereafter, their optimal sizes are determined by using the Kalman filter algorithm.

II. SELECTION OF OPTIMAL LOCATIONS

A.Reduction of Power loss by connecting DG

In general, the power generated from the generating station is to be far from the consumers and they are severely effected by the low voltages. The IEEE 30-bus system is shown in Fig. 1 [5], where all loads can be classified under one of two classes. The first classification is the directly-connected-bus while the second is the load-concentration-bus. The directly-connected-bus is defined as a bus connected to a reference bus that does not pass through any other buses. For example, buses 12, 14, 18, and 23 in Fig. 1 are the directly-connected-buses if bus 15 is chosen as a reference bus. The load-concentration-bus handles relatively large loads, and is more connected to the other directly connected buses when compared to other nearby buses. In fig. 1, buses 10, 12, 27 and 5 can be

selected as the representative load concentration buses of Areas 1 through 4, respectively. When the DG is applied to this system, it is not desirable to connect each DG to every load bus to minimize power loss. Instead, the multiple representative DGs can be connected to the load concent-ration-buses. Then,they provide an effect similar to the case where there are allDGs on each load bus, but with added benefit of reduced power loss [6]–[9].

The power loss, P_{loss}between the two buses i and j is computed from the simplified unit circuit shown in fig. 2 by the following equation:

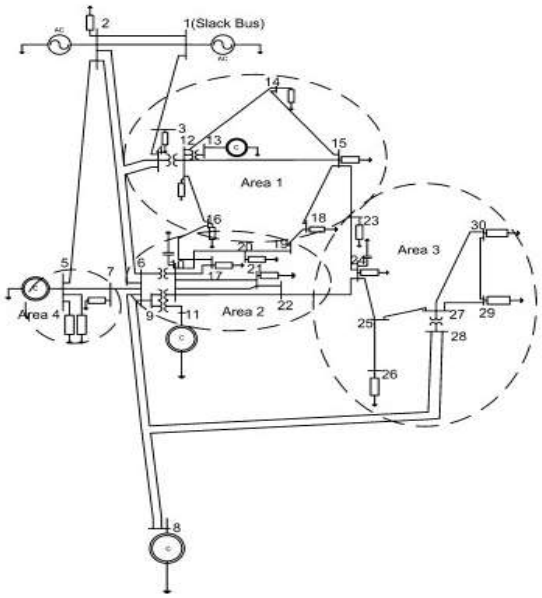


Fig. 1. IEEE 30-bus system

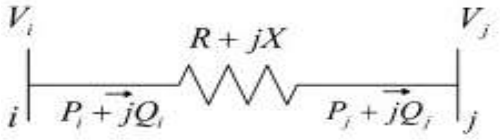


Fig.2. Simplified unit circuit between two buses

$$P_{loss,ij} = P_i - P_j = \frac{(P_i^2 + Q_i^2)r}{V_i^2} \qquad (1)$$

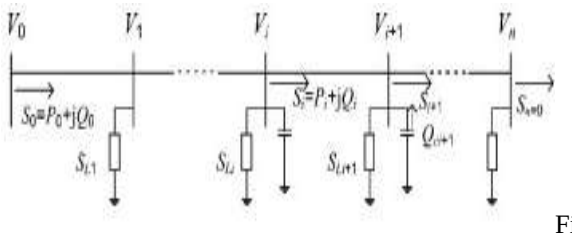


Fig.3. One line diagram of a distribution feeder

$$V_{i+1}^2 = V_i^2 - 2(r_{i+1}P_i + x_{i+1}Q_i) + (r_{i+1}^2 + x_{i+1}^2)(P_{i+1}^2 + Q_i^2)/V_i^2 \qquad (2)$$

Also, the one-line diagram of a distribution feeder with a totalof *n* unit circuits is shown in Fig. 3. When power flows in one of direction, the value of bus voltage, V_{i+1}, is smaller than that of V_i, and this associated equation can be expressed by (2). In general, the reactive power, Q_i, is reduced by connecting a capacitorbank on bus i in order to decrease the voltage gap between$V_{i+1}$and V_i . In other words, the capacitor bank at bus *i* makes it possible to reduce power loss and regulate the voltages by adjusting the value of$Q_i$in equation(2).

If a DG is installed at the location of the capacitor bank, theproper reactive power control of the DG has the same effecton the system as does the capacitor bank. Moreover, the mainfunction of the DG is to supply real supplementary power to therequired loads in an effective manner.The variation of power loss is relatively less sensitive to voltage changes when compared to the size of DG. In other words, the amount of real power supplied by the DG strongly influences the minimization of power loss. This means that the DG can control the bus voltage for reactive power compensation independently of its real power control to minimize power loss.

B. Selection of Optimal Location for DGs by Considering Power Loss

Before deriving the equations for the selection of optimal locations of DGs, the following terms are defined. The factor, D shown in the following is called the generalized generation distribution factor [10]:

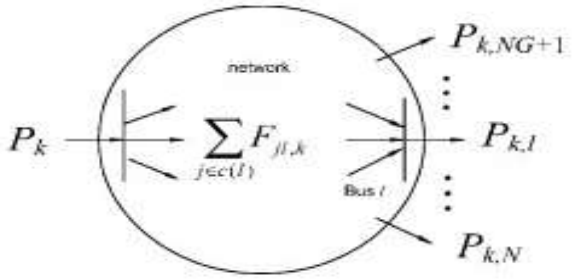


Fig.4. Power flow from the *k* th generator to the other several loads.

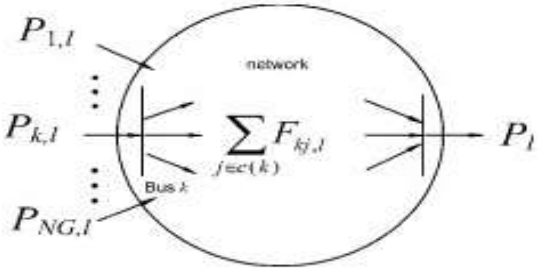


Fig.5. Power flow from the several generators to the*l*th load

P_k: power supplied by the *k* th generator in a power network

P_l: power consumed by the *l* th load in a power network

$P_{k,l}$: power flowing from the *k*th generator to the *l* th load

$F_{jl,k}$: power flowing from the *k* th generator to the *l* th load through bus j connected to the *l* th load.

$D_{jl,k}$: ratio of $F_{jl,k}$ to the power supplied by the *k* th generator

$P_{loss,k}$: power loss on transmission line due to the power supplied from the *k* th generator

$F_{kj,l}$: power flow from the *k* th generator to the *l* th load through bus j connected to the *k* th generator

$D_{kj,l}$: ratio of $F_{kj,l}$ to the power supplied by the *k* th generator

$P_{loss,l}$: power loss on a transmission line due to power supplied to the *l* th load

$P_{loss,ij}$: power loss between buses i and j

The IEEE 30-bus system in Fig. 1 is now analyzed for two different cases with respect to generator or load[11]. In other words, the first case is one where power flows from the *k* th generator to numerous loads. The second case isone where power is flowing from several generators to the *l* thload. These two conditions are shown in Figs. 4 and 5, respectively.

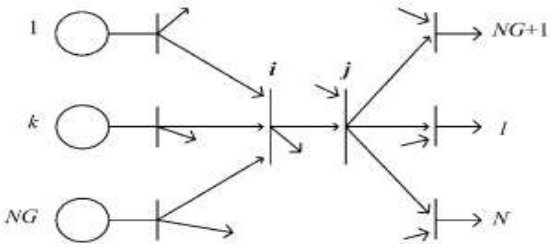


Fig.6. Simplified circuit with only power generations and consumptions.

For the first case (case-1), the power supplied from the *k*thgenerator to the *l* th load among several loads is calculated bythe following:

$$P_{k,l|case-1} = \sum_{j\epsilon c(l)} F_{jl,k} = \sum_{j\epsilon c(l)} D_{jl,k}\, P_k \qquad (3)$$

Where *c(l)* are the buses connected to the *l* th load. Then, the power loss associated with the *k* th generator is computed by the following, which is the difference between the power supplied from the *k*th generator and the sum of powers consumed in loads:

$$P_{loss,k} = P_k - \sum_{l=NG+1}^{N} P_{k,l} \qquad (4)$$

In the same manner, the power supplied from the *k* th generator among several generators to the *l*th load is calculated by the following for the second case (case-2):

$$P_{k,l|case-2} = \sum_{j\epsilon c(k)} F_{kj,l} = \sum_{j\epsilon c(k)} D_{kj,l}\, P_l \qquad (5)$$

Where *c(k)* are the buses connected to the *k*th generator. The power loss associated with the *l*th load is computed by the following:

$$P_{loss,l} = \sum_{k=1}^{NG} P_{k,l} - P_l \qquad (6)$$

The branch between buses i and j in Fig. 6 can become an arbitrary branch in Fig. 1. This means that the total power loss of the system can be calculated by summing the losses of all branches whenever the DG is connected to any bus. Each loss of the branch is thus simply computed by (1).

TABLE 1
Buses included in each area in fig.1

Area	Buses	Total amount of power consumption in loads
Area 1	3,4,12,13,14,15,16 and 18	45 MW
Area 2	10, 11, 17, 19, 20, 21, 22	44 MW
Area 3	23, 24, 25, 26, 27, 29, 30	28.4 MW
Area 4	5, 7	117 MW

TABLE 2
Buses with largest and smallest loads in each area

Area	Largest load bus	Smallest load bus
Area 1	12	16
Area 2	10	22
Area 3	27	26
Area 4	5	7

To minimize the total power loss, the largest load buses in each area, which are buses 12, 10, 27, and 5, can be selected as the optimal locations for the multiple DGs as the representative load-concentration-buses. Assume that the power losses between two adjacent buses in each area are

negligible. In this case, the multiple DGs with the same size as the total amount of power consumption at each area might be implicitly used to minimize the power loss. In other words, the total system power loss is 3.452 MW if each DG in Areas 1 through 4 supplies the real power of 45, 44, 28.4, and 117 MW, respectively. The resulting system power loss of 3.452 MW will be compared with the total power loss computed after the optimal size of multiple DGs is systematically determined by using the Kalman filter algotihm.

III. PROCEDURE TO SELECT THE OPTIMAL SIZE OF MULTIPLE DGs USING KALMAN FILTER ALGORITHM

The total amount of power consumption in Table I at each area could be chosen as the size of DGs to be placed. However, these are not optimal values for the DGs because the power loss in lines connecting two buses is ignored. To deal with this problem, the Kalman filter algorithm is applied to select the optimal sizes of multiple DGs by minimizing the total power loss of system. The Kalman filter algorithm [12], [13] has the smoothing properties and the noise rejection capability robust to the process and measurement noises. In practical environments (in which the states are driven by process noise and observation is made in the presence of measurement noise), the estimation problem for the optimal sizes of multiple DGs can be formulated with a linear time-varying state equation. Also, the error from interval of computation can be reduced during the estimation optimization process. In this study, the state model applied for the estimation is given as

$$\begin{aligned} X(n+1) &= \Phi X(n) + \Gamma\omega(n), x(0) = x_0 \\ y(n) &= c\,x(n) \\ z(n) &= y(n) + v(n) \end{aligned} \qquad (7)$$

Where the matrices $\Phi(\epsilon R^{n\times n})$ and $\Gamma(\epsilon R^{n\times m})$ and the vector, $c(\epsilon R^{1\times n})$, are known deterministic variables, and the identity matrix $I(\epsilon R^{n\times n})$ is usually chosen for the matrix Φ. The state vector, $x(\epsilon R^{n\times 1})$,, can represent the size of each of the multiple DGs or their coefficients. Also, $(\epsilon R^{m\times})$ is the process noise vector, is the measured power loss, and *v* is stationary measurement noise. Then, the estimate of the state vector is updated by using the following steps.

- Measurement update: Acquire the measurements, *z(n)* and compute a *posteriori* quantities:

$$\begin{aligned} k(n) &= P^-(n)\,c^T\,[cP^-(n)c^T + r]^{-1} \\ \hat{X}(n) &= \hat{x}^-(n) + k(n)[z(n) - c\hat{x}^-(n) \\ P(n) &= P^-(n) - k(n)cP^-(n) \end{aligned} \qquad (8)$$

Where $k(\epsilon R^{n\times 1})$ is the kalman gain, P is a positive definite symmetric matrix, and *r* is a positive number selected to avoid a singular matrix $P^-(0)$ is given as $P^-(0) = \lambda I(\lambda > 0)$, where I is an identity matrix.

- Time update:

$$\begin{aligned} \hat{x}^-(n+1) &= \Phi\hat{x}(n) \\ P^-(n+1) &= \Phi P(n)^T + \Gamma Q\Gamma^T \end{aligned} \qquad (9)$$

Where $Q(\epsilon R^{m\times m})$ is a positive definite covariance which is zero in this study because the stationary process and measurement noises are mutually independent.

- Time increment: Increment and repeat.

Thereafter, the estimated output (the total power loss of the system) is calculated as

$$y(n) = c\,\hat{x}(n) \qquad (10)$$

In Stage-1 of Fig. 7, the algorithm begins with the zero values for all DGs, and the index denotes the number of given DG. After adding the small amount of power, P_{step} of 10 MW to each DG, the initial power loss is obtained by a power flow computation based on the Newton -Raphson method [5]. Then, the information on the individual power loss, P_{loss}, corresponding to each DG increased by 10 MW is sent to Stage-2, where the values of P_{loss} are substituted with those of P_{temp}. After the minimum value of $P_{temp}$is selected, its value and the corresponding sizes of multiple DGs are stored in the memory of $P_{losses,n}$ and $DG_{i,n}$ in Fig. 7, respectively. This process is then repeated until the total sum of all DGs is the sameas the predefined value, P_{max}, in Stage-3 by increasing *n* to *n+1*. Finally, the accumulated data of the minimum power loss and sizes of DGs, which are $P_{losses,samples}$and $DG_{i,samples}$ respectively, are obtained.

The data samples obtained above might be different from the actual values due to the large sampling interval of 10 MW. If this sampling interval is reduced to find more accurate values,the computational requirement will be dramatically increased. To deal with this problem, the steps in Fig. 8 with two phases in application of the Kalman filter algorithm are taken to reduce the error between the estimated and actual values, and then the optimal sizes of multiple DGs are finally estimated.

In Phase-1 of Fig. 8, the estimated sizes of multiple DGs, $DG_{i,estimated}$, are determined by applying the Kalman filter algorithm with the data samples obtained from Fig. 7, which are $P_{losses,samples}$ and $DG_{i,samples}$. Its associated parameters are then given in the following:

$$\delta(n) = \sum_{i=1}^{4} DG_{i,samples}(n)/\max \times \left\{\sum_{i=1}^{4} DG_{i,samples}(n)\right\} \qquad (11)$$

$$c_{phase-1}(n) = [\delta(n), \delta^2(n), \delta^3(n), \delta^4(n)] \qquad (12)$$

$$z(n)|i = DG_{i,samples}(n) \qquad (13)$$

$$DG_{i,estimated}(n) = \hat{y}(n) = c_{phase-1}(n)\,.\,\hat{x}_{\text{phase1}}(n_{\max})|_i \quad (i=1, 2, 3, 4) \qquad (14)$$

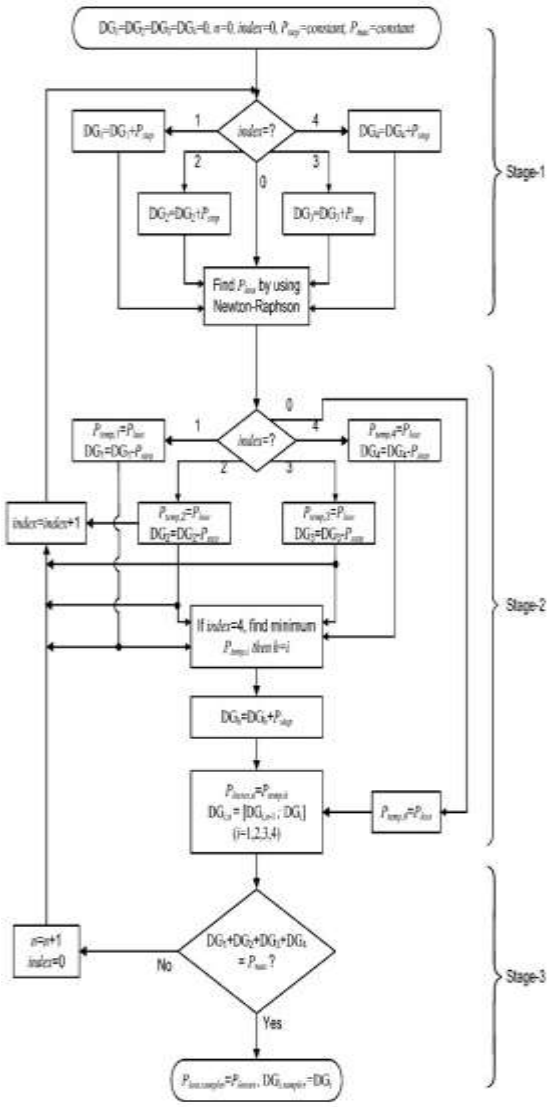


Fig.7. Procedure to obtain data samples of the multiple DGs and power loss required before applying the kalman filter algorithm

Where δ is the normalized value, and $n_{\max}$ is the number of last samples in $DG_{i,samples}$. To estimate the size of each DG ,thekalman filter algorithm is applied in sequence with different measurements of *z* in (13).

After estimating the optimal sizes of multiple DGs in Phase-1, the total power loss, $P_{loss,estimated}$, is estimated in Phase-2 of Fig. 8 with the power loss data samples, $P_{loss,samples}$. From Fig. 7 and the estimated DG sizes, $DG_{i,estimated}$, in phase-1. The associated parameters required to apply the Kalman filter algorithm are given in the following:

$$\beta_i(n) = DG_{i,estimated}(n) \quad (i = 1,2,3,4) \qquad (15)$$

$$c_{phase-2}(n) = [\beta_1(n), \beta_2(n), \beta_3(n), \beta_4(n)] \qquad (16)$$

$$z(n) = P_{loss,samples}(n) \qquad (17)$$

$$P_{loss,estimated}(n) = \hat{y}(n) = c_{phase-2}(n) \,.\, \hat{x}_{\text{phase-2}}(n_{\max}) \qquad (18)$$

Where β is the estimated size for each DG from (14).

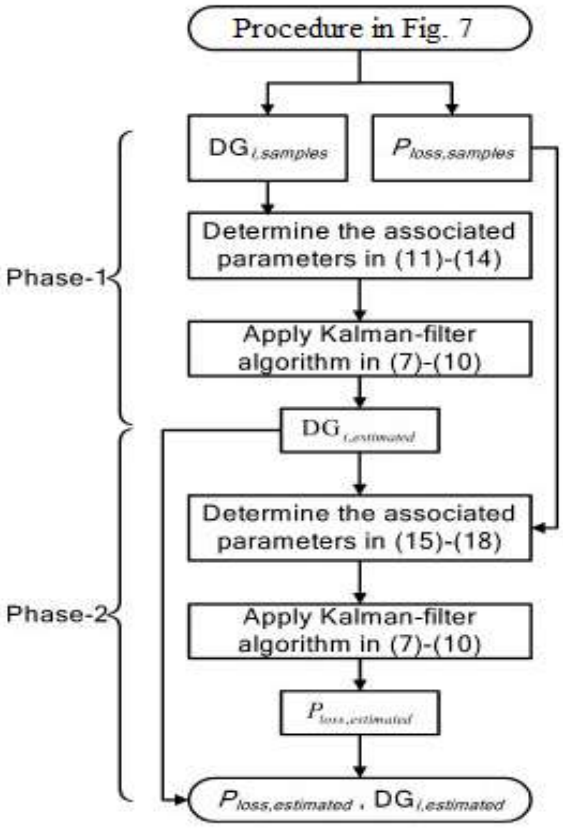


Fig.8. Steps to estimate the optimal size of multiple DGs in two phases by applying the kalman filter algorithm.

IV. SIMULATION RESULTS

TABLE 3
Comparison of total power loss

	DG_1	DG_2	DG_3	DG_4	Sum of DGs	Total power loss P_{loss}
Without DG and KFA	45 MW	44 MW	28.4 MW	117 MW	234.4 MW	19.016 MW
With DG and KFA	47.2 MW	67.7 MW	27.2 MW	91.8 MW	234.7 MW	1.907 MW

Finally, the total power loss is effectively reduced by the optimal size selection process. In particular, note that the size in Area 2 is required to increase from 44 to 67.7 MW, of which is a difference of 23.7 MW. In contrast, the size of in Area 4 is necessary to decrease significantly from 117 to 91.8 MW, which is difference of 25.2 MW.

CONCLUSION

This paper proposed the method for selecting the optimal locations and sizes of multiple distributed generations (DGs) to minimize the total power loss of system. To deal with this optimization problem, the Kalman filter algorithm was applied. When the optimal sizes of multiple DGs are selected, the computation efforts might be significantly increased with many data samples from a large-scale power system because the entire system must be analyzed for each data sample. The proposed procedure based on the Kalman filter algorithm took the only few samples, and therefore reduced the computational requirement dramatically during the optimization process.

Prior to the implementation and connection to an electric power grid, this study can be used as a decision-making process in the power system operation and planning for selecting the optimal locations and sizes of multiple DGs based on the renewable energy resources such as fuel cell, photovoltaic, micro-turbines, wind powers, etc.

REFERENCES

[1] A. A. Chowdhury, S. K. Agarwal, and D. O. Koval, "Reliability modeling of distributed generation in conventional distribution systems planning and analysis," IEEE Trans. Ind. bAppl., vol. 39, no. 5, pp.1493–1498, Oct. 2003

[2] M. F. AlHajri and M. E. El-Hawary, "Improving the voltage profiles of distribution networks using multiple distribution generation sources," in Proc. IEEE Large Engineering Systems Conf. Power Engineering, 2007, pp. 295–299.

[3] G. Carpinelli, G. Celli, S. Mocci, F. Pilo, and A. Russo, "Optimization of embedded eneration sizing and siting by using a double trade-off method," Proc. Inst. Elect. Eng. Gen., Transm.,Distrib., vol. 152, no.4, pp. 503–513, Jul. 2005

[4] T. Senjyu, Y. Miyazato, A. Yona, N. Urasaki, and T. Funabashi, "Optimal distribution voltage control and coordination with distributed generation," IEEE Trans. Power Del.,vol. 23, no. 2, pp. 1236–1242, Apr. 2008.

[5] H. Saadat, Power System Analysis, 2nd ed. , Singapore: McGraw- Hill, 2004, pp. 234– 227.

[6] J. J. Grainger and S. H. Lee, "Optimum size and location of shunt capacitors for reduction of losses on distribution feeders," IEEE Trans. Power App. Syst., vol. PAS-100, no. 3, pp.1105–1118, Mar. 1981.

[7] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," IEEE Trans. Power Del., vol. 4, no. 1, pp. 735–743, Jan. 1989.

[8] M. A. Kashem, A. D. T. Le, M. Negnevitsky, and G.Ledwich, "Distributed generation for minimization of power losses in distribution systems," in Proc. IEEE PES General Meeting, 2006, pp. 1–8.

[9] H. Chen, J. Chen, D. Shi, and X. Duan, "Power flow study and voltage stability analysis for distribution systems with distributed generation," in Proc. IEEE PES General Meeting, Jun. 2006, pp. 1–8.

[10] W. Y. Ng, "Generalized generation distribution factors for power system security evaluations," IEEE Trans. Power App. Syst., vol. PAS-100, no. 3, pp. 1001–1005, Mar.1981

[11] Y.-C. Chang and C.-N. Lu, "Bus-oriented transmission loss allocation," Proc. Inst.Elect. Eng., Gen., Transm.,Distrib., vol. 149, no. 4, pp. 402–406, Jul. 2002.

[12] R. A. Wiltshire, G. Ledwich, and P. O'Shea, "A Kalman filtering approach to rapidly detecting modal changes in power systems," IEEE Trans. Power Syst., vol. 22, no. 4, pp.1698–1706, Nov. 2007.

[13] E. W. Kamen and J. K. Su, Introduction to Optimal Estimation. London, U.K.: Springer- Verlag, 1999, pp. 149–183.

[14] S. Lee and J.-W. Park, "A reduced multivariate polynomial model for estimation of electric load composition", IEEE Trans. Ind. Appl., vol. 44, no. 5,pp 1333-1340 sep/oct Power Syst., vol. 22, no. 4, pp.1698–1706, Nov. 2007

MALWARE DETECTION IN DTN BASED ON BEHAVIOR OF NODES

[1]R.P.Kaaviya Priya,[2]G. Bhavani,[3]J.Jayapradha
[1]PG Student – CSE,[2][3]Assistant Professors –CSE,
Krishnasamy College Of EngineeringAnd Technology.

Abstract-In Delay Tolerant Network (DTN), behavioural characterization of malware is an effective alternative to pattern matching in detecting malware, especially when dealing with polymorphic or obfuscated malware. The DTN becoming a viable communication with mobile consumer electronics equipped with short range communication technologies such as Bluetooth, Wi-Fi Direct. This paper propose a general behaviour characterization of proximity malware based on Naive Bayesian model and was identified with two unique challenges for extending Bayesian malware detection to DTNs, "insufficient evidence versus evidence collection risk" and "filtering false evidence in sequential and distributed manner", and so propose a simple yet effective method "look ahead", to address those challenges with two extensions to look ahead, dogmatic filtering, and adaptive look ahead, they address the challenge of "malicious nodes sharing false evidence." Real mobile network traces are used to verify the effectiveness of the proposed methods.

Index terms: Proximity malware, Bayesian filtering, Dogmatic filtering

I. INTRODUCTION

The widespread adoption of the mobile devices, coupled with strong economic incentives, induces a class of malware that specifically targets DTNs. We call this class of malware proximity malware. An early example of proximity malware is the Symbian based *Cabir* worm[1] A later example is the iOS-based *Ikee* worm, which exploited the default SSH password on jail-broken[2] iPhones to propagate through IP-based Wi-Fi connections[3]. Previous researches[4],[5] quantify the threat of proximity malware attack in NFC and Wi-Fi direct[6][7]. Proximity malware based on the DTN model brings unique security challenges that are not present and also malware propagation cannot be detected by the cellular carrier in the traditional model. In this paper, we consider a general behavioural characterization of proximity malware. Behavioural characterization, in terms of system call and program flow has been previously proposed as an effective alternative
to pattern matching for malware detection [8], [9].Malware-infected node behaviours are observed by others during their multiple opportunistic encounters: Individual observations may be imperfect, but abnormal behaviours of infected nodes are identifiable in the long-run. The imperfection of a single, local
observation was previously in the context of distributed IDS against slowly propagating worms [10].Instead of assuming a sophisticated malware containment capability, such as patching or self-healing[11],[12] we consider a simple "cut-off" strategy. Our focus is on how individual nodes shall make such cut-off decisions against potentially malware-infected nodes, based on direct and indirect observations. In the context of DTNs, we face a dilemma when trying to detect proximity malware: Hypersensitivity leads to false positives, while hyposensitivity leads to false negatives. Our solution, *look ahead*, reflects individual node's intrinsic risk inclinations against malware infection, to balance between these two extremes.Essentially, we extend the naive Bayesian model, which has been applied in filtering email spams [13], [14], [15], detecting botnets [16], and designing IDSs [10], [17] and address two DTN specific, malware-related, problems on observation with individual nodes:

*1.Insufficient evidence versus evidence collection risk:*In DTNs, evidence (such as Bluetooth connection or SSH session requests) is collected only when nodes come into contact. But contacting malware-infected nodes carries the risk of being infected. Thus, nodes must make decisions (such as whether to cut off other nodes and, if yes, when) online based on potentially insufficient evidence.

*2.Filtering false evidence in sequential and distributed manner:*Sharing evidence among opportunistic acquaintances helps alleviating the aforementioned insufficient evidence problem; however, false evidence shared by malicious nodes (the liars) may negate the benefits of sharing. In DTNs, nodes must decide whether to accept received evidence sequentially and distributedly. This paper includes the following contributions:

1.A general behavioural characterization of proximity malware, which captures the imperfect nature in detecting proximity malware.

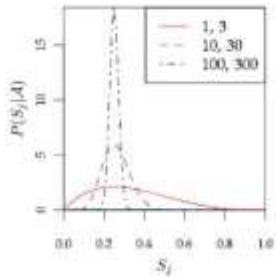
2.With a simple cut-off malware containment strategy, we formulate the malware detection process as a distributed decision problem.
3.Also the benefits of sharing assessments among nodes, and address challenges derived from the DTN model: liars (i.e., bad-mouthing and false praising malicious nodes) and defectors (i.e., good nodes that have turned rogue due to malware infections).
4. This paper presents two alternative techniques, *dogmatic filtering and adaptive look ahead*, that naturally extend look ahead to consolidate evidence provided by others, while containing the negative effect of false evidence. A nice property of the proposed evidence consolidation methods is that the results will not worsen even if liars are the majority in the neighbourhood. Real contact traces are used to verify the effectiveness of the methods.

II. MODEL

Consider a DTN consisting of n nodes. The neighbours of a node are the nodes it has (opportunistic) contact opportunities with. Proximity malware is a malicious program that disrupts the host node's normal function and has a chance of duplicating itself to other nodes during (opportunistic) contact opportunities between nodes in the DTN. When a duplication occurs, the other node is infected with the malware. In our model, we assume that each node is capable of assessing the other party for suspicious actions after each encounter, resulting in a binary assessment. For example, a
node can assess a Bluetooth connection or an SSH session for potential Cabir or Ikee infection. The watchdog components in previous works on malicious behavior detection in MANETs [18] and distributed reputation systems [19], [20] are other examples A node is either evil or good, based on if it is or is not infected by the malware. The suspicious action assessment is assumed to be an imperfect but functional indicator of malware infections: It may occasionally assess an evil node's actions as "non-suspicious" or a good node's actions as "suspicious," but most suspicious actions are correctly attributed to evil nodes. The functional assumption characterizes a malware infected node by the assessments of its neighbours. If node i has N (pair-wise) encounters with its neighbours and s_N of them are assessed as suspicious by the neighbours, its suspiciousness S_i is defined as

$$S_i = \lim_{N\to\infty} \frac{SN}{N}$$

$S_i \in [0, 1]$. A number $L_e \in (0,1)$ is chosen as the line between good and evil. Le depends on the quality of a particular suspicious-action assessment and, if the assessment is a functional discriminant feature of the malware
and the probabilistic distribution of the suspiciousness of both good and evil nodes are known, L_e can be chosen as the (Bayesian) decision boundary, which minimizes classification errors[21]. Node i is good if $S_i \leq L_e$, or evil if $S_i > L_e$: We draw a fine line between good and evil, and judge a node by its deeds. Instead of assuming a sophisticated malware coping mechanism, such as patching or self-healing, we consider a simple and widely applicable malware containment strategy: Based on past assessments, a node i decides whether to refuse future connections ("cut off") with a neighbour j.



III. HOUSEHOLD WATCH:

Consider the case in which i bases the cut-off decision against j only on i's own assessments on j. Since only direct assessments are involved, we call this model household watch Let $A = (a_1, a_2, \ldots a_A)$ be the assessment sequence (a_i is either 0 for "non-suspicious" or 1 for "suspicious") in chronological order, i.e., a_1 is the oldest assessment, and a_A is the newest one.Baye's theorem tells us

$$P(Sj \mid A) \propto P(A \mid Sj)\ X\ P(Sj)$$

where $P(S_j)$ encodes our prior belief on j's suspiciousness S_j; $P(A \mid S_j)$ is the likelihood of observing the assessment sequence A given Sj; $P(S_j|A)$ is the posterior probability, representing the plausibility of j having a suspiciousness of S_j given the observed assessment sequence A. Since the evidence $P(A)$ does not involve S_j and serves as a normalization factor in the computation, we omit it and write the quantitative relationship in the less cluttered proportional form. The structure of the behavioral malware characterization model (specifically, a single threshold L_e is used to distinguish the nature of a node) gives rise to a subtlety concerning i's prejudice against j in the distribution approach. The online supplemental material, if i makes no presumption on j's suspiciousness,when no assessment has been made yet. This leads to a discussion on whether such prejudices are warranted. The choice of Le depends on the assessment mechanism itself and, as mentioned previously, if the probabilistic distributions of suspiciousness of both good and evil nodes are known, can be determined by minimizing Bayesian decision errors. If $L_e > 0.5$, the assessment mechanism is biased toward false positive (good nodes actions being assessed as suspicious); if $L_e < 0.5$, the assessment mechanism is biased toward false negative (evil nodes' actions being assessed as nonsuspicious). However, before any assessment is made, i has no clue about the true nature of j. A bias in the assessment mechanism should not affect the i's neutrality on j's nature before the first assessment is made. Thus, we stipulate that the comparison

between P_g *(A)* and P_e *(A)* should be made only when $A \neq \emptyset$.Alternatively, in the maximizer approach, i uses the suspiciousness distribution's maximizer (see (4)) when making the cut-off decision against j. The justification for the maximizer approach is that the suspicious distribution's maximizer is the single most probable estimation of j's suspicisousness given the evidence. The maximizer approach precludes the prejudice problem, because the maximizer is undefined when $A=\emptyset$. Similar to the distribution

TABLE 1
Data Set Statistics

	nodes	entries	time span	avg. interval
Haggle	41	112, 295	15 days	12 secs
MIT reality	96	114, 046	490 days	371 secs

approac

h, i compares evidence that is both favorable and unfavorable to j. Evidence A is favorable to j if $s_A/|A| \leq L_e$ and is unfavorable to j if $s_A/|A| > L_e$. The maximizer approach significantly reduces the computation cost, in comparison with the distribution approach, while partially discarding information contained in the suspiciousness distribution derivable from the evidence collected so far. Whichever approach is taken, the cut-off decision problem has an asymmetric structure in the sense that cutting j off will immediately terminate the decision process (i.e., i will cease connecting with j; no further evidence will be collected), while the opposite decision will not. Thus, we only need to consider the decision problem when i considers cutting j off due to unfavorable evidence against j. The cut-off decision is made based on the risk estimation of such a decision. The key insight is that i shall estimate the cut-off decision's risk by looking ahead. More specifically, given the current assessment sequence $A=(a_1, \ldots, a_A)$, the next assessment a_{A+1} (which has not been taken yet) might be either 0 (nonsuspicious) or 1 (suspicious If the evidence is still unfavorable toward j, we say that i's decision of cutting j off is one-step-ahead robust. If the cut-off decision is one-step ahead robust, i is certain that exposing itself to the potential danger of infection by collecting one further assessment on j will not change the outlook that j is evil. Similarly, i can look multiple steps ahead. In fact, the number of steps i is willing to look ahead is a parameter of the decision process rather than a result of it. This parameter shows i's willingness to be exposed to a higher infection risk in exchange for a higher certainty about the nature of j and a lower risk of cutting off a good neighbor; in other words, it reflects i's intrinsic risk inclination against malware infection.

*Definition 1 (Look-Ahead λ):*The look-ahead _ is the number of steps i is willing to look ahead before making a cut-off decision. We can make a similar decision-robustness definition for
look-ahead λ.

Definition 2 (λ-Robustness): At a particular point in i's cutoff decision process against j (with assessment sequence $A = (a_1, \ldots, a_A)$),

TABLE 2
Neighbor Nature and Cut-Off Decision Combination

	...gets cut off.	...stays connected.
An evil neighbor...	True positive.	False negative.
A good neighbor...	False positive.	True negative.

i's

decision of cutting j off is said to beλ-step-ahead robust, or simply λ-robust, if 1) the current evidence A is unfavorable toward j; and 2) even if the next λ assessments ($a_{A+1}, \ldots, a_{A+\lambda}$) all turn out to be nonsuspicious (i.e., 0), the evidence against j is still unfavorable. Given the look-ahead λ, the proposed malware containment strategy is to cut j off if the cut-off decision is λ-robust, and not to cut j off otherwise. In Section 2 of the online supplemental material, we discuss how to adapt the look-ahead λ to individual nodes' intrinsic risk inclinations against the malware.

3.2 Neighborhood Watch

Besides using i's own assessments, i may incorporate other neighbors' assessments in the cut-off decision against j. This extension to the evidence collection process is inspired by the real-life neighborhood (crime) watch program, which encourages residents to report suspicious criminal activities in their neighborhood. Similarly, i shares assessments on j with its neighbors, and receives their assessments on j in return. In the neighborhood-watch model, the malicious nodes that are able to transmit malware (we will see next that there may be malicious nodes whose objective is other than transmitting malware) are assumed to be consistent over space and time. These are common assumptions in distributed trust management systems which incorporate neighboring nodes' opinions in estimating a local trust value. By being consistent over space, we mean that evil nodes suspicious actions are observable to all their neighbors, rather than only a few. If this is not the case, the evidence provided by neighbors, even if truthful, will contradict local evidence and, hence, cause confusions: Nodes shall discard received evidence and fall back to the household watch model. By being consistent over time, we mean that evil nodes cannot play strategies to fool the assessment mechanism. This is equivalent to the functional assumption in characterizing the nature of nodes by suspiciousness. The case in which the evil nodes can circumvent the suspiciousness characterization (such as by first accumulating good assessments, and then launch an attack through a short burst of concentrated suspicious actions) calls for game-theoretic analysis and design, and is beyond the scope of this paper. Instead, we propose a behavioural characterization of proximity malware; further gametheoretic analysis and design could base on this foundation.

3.2.1 Challenges

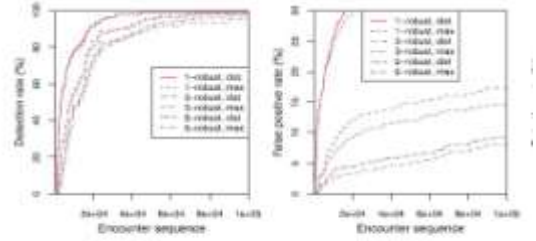
Two cases complicate the neighborhood watch model: liars and defectors. Liars are those evil nodes who confuse other nodes by sharing false assessments. A false assessment is either a false
praise or a false accusation. False praises understate evil nodes' suspiciousness, while false accusations exaggerate good nodes' suspiciousness. Furthermore, a liar can fake assessments on nodes that it has never met with. To hide their true nature, liars may do no evil other than lying, and, therefore, have low suspiciousness. Defectors are those nodes that change their nature due to malware infections. They start out as good nodes and faithfully share assessments with their neighbors; however, due to malware infections, they become evil. Their behaviours after the infection are under the control of the malware. These complications call for evidence consolidation. Two extremal, but naive, evidence-consolidation strategies are 1) to trust no one and 2) to trust everyone. The former degenerates to the household-watch model with the twist of the defectors (defectors change their nature and hence their behavioral pattern); the latter leads to confusions among good nodes.

3.2.2 Evidence

For a pair of neighboring nodes i and j, let N_i and N_j be the neighbors of i and j, respectively. At each encounter, i shares with j its assessments on the neighbor set Ni – {j}, and j shares with i its assessments on the neighbor set N_j –{i}.Since the cut-off decision only needs to be made against a neighbor, i only considers the assessments of its own neighbors $N_i \cap (N_j - \{i\})$ from the evidence provided by j.Without superimposed trust relationships among the nodes in the model, i and j only share their own assessments, instead of forwarding the ones provided by their neighbors.

3.2.3 Evidence Aging

The presence of defectors breaks the assumption when we characterize a node's nature by suspiciousness . A defector starts as a good node but turns evil due to malware infectio



(a) Haggle (b) Haggle

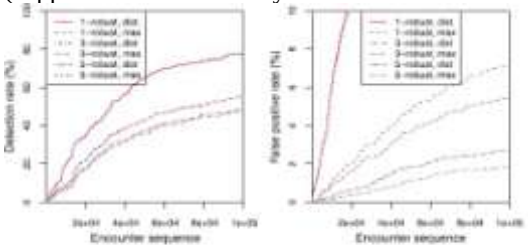
Fig. 2. Performance comparison between the λ-robust cut-off strategy approaches; λ = 1, 3, and 5.

ns; the assessments collected before the defector's change of nature, even truthful, are misleading. To alleviate the problem of outdated assessments, old assessments are discarded in a process called evidence aging. Each assessment is associated with a timestamp. Only assessments with timestamps less than a specific aging window T_E from now are included in the cut-off decision. To see that the aging window T_E alleviates the defector problem, consider a node that is infected at time T. Without evidence aging, all evidence before T mounts to testify that the node is good; if the amount of this prior evidence is large, it may take a long time for its neighbours to find out about the change in its nature. In comparison, with evidence aging, at time $T+ T_E$, all prior evidence expires and only those assessments after the infection are considered, which collectively testify against the node. However, in practice, the choice of the aging window T_E depends on the context. While a small T_E may speed up the detection of defectors by reducing the impact of stale information, TE must be large enough to accommodate enough assessments to make a sound cut-off decision. If T_E is too small, a node will not have enough assessments to make an λ-robust cut-off decision.

3.2.4 Evidence Consolidation

We propose two alternative methods, dogmatic filtering and adaptive look ahead, for consolidating evidence provided by other nodes, while containing the negative impact of liars. For exposition, we consider a scenario in which node i uses the assessments within the evidence aging window $[T-T_E, T]$ provided by i's neighbors (other than one of the neighbors, say, j) in making the cut-off decision against j. The implications are as follows:

1.Given enough assessments, honest nodes are likely to obtain a close estimation of a node's suspiciousness (suppose they have not



(c) MIT reality (d) MIT reality

with the distribution (dist) and maximizer (max) evidence weighing

cut the node off yet), even if they only use their own assessments.

2.The liars have to share a significant amount of false evidence to sway the public's opinion on a node's suspiciousness.

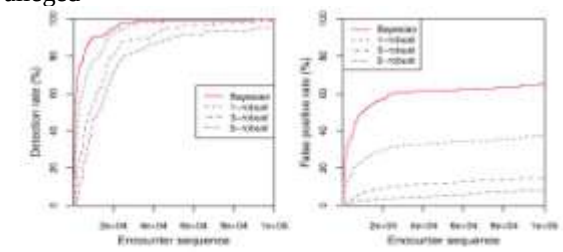
3. The most susceptible victims of liars are the nodes that have little evidence.

Dogmatic filtering. Dogmatic filtering is based on the observation that one's own assessments are truthful and, therefore, can be used to bootstrap the evidence consolidation process. A node shall only accept evidence that will not sway its current opinion too much. We call this observation the dogmatic principle. Our interpretation of the dogmatic principle depends on the following generalization of Definition 2.

Definition 3 (λ-Robust Judgment). Let A be the suspicious action assessments that i has on j. We say that i's judgment on j's nature is λ-robust (or (-λ)-robust) based on A, if 1) the evidence A is favorable (or unfavorable) toward j, 2) the evidence remains so even if the next λ assessments are all suspicious (or nonsuspicious), and 3) the evidence becomes unfavorable (or favorable) toward j if the next λ + 1 assessments are all suspicious (or nonsuspicious). As a special case, if a judgment is not even 1-robust (or (-1)-robust), we say that the judgment is 0-robust or not robust at all.λ-robust judgment reflects i's certainty of its judgment on j's nature (based on the evidence collected so far). The λ-robust cut-off decision against j (see Definition 2) is equivalent to the (-λ)-robust judgment on the (evil) nature of j. The sign of λ in Definition 3 represents j's nature: A negative number represents evilness, and a positive number represents goodness. i's cut-off decision against j works as follows with

dogmatic filtering.1. i will not consider cutting j off until i has at least one assessment on j

2. After its first encounter with j and with its own assessments A with the evidence aging window $[T+T_E, T]$, i considers whether or not to take another neighbor k's alleged



(a) Haggle (b) Haggle

Fig. 3. Performance comparison between the vanilla Bayesian (degenerat

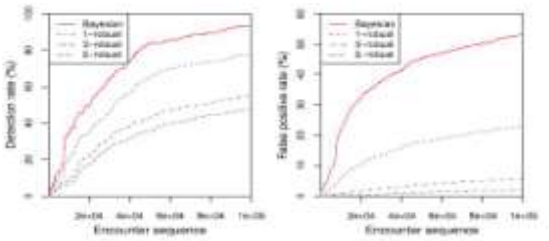
assessm

ents on j within the same window B when i and k meet. With dogmatic filtering, i is very conservative when its certainty about j's nature is still low (i.e., λA is small). At this early stage, i will accept the evidence provided by j only if the evidence would not significantly change its certainty on j's nature. In particular, if λ≤1, i will never accept a piece of evidence that would change its judgment on j's nature. Dogmatic filtering significantly contains the impact of liars on i while still allowing a change of certainty (on j's nature) comparable to its own. The afore mentioned observation that the liars have to fabricate a significant amount of false evidence to confuse honest nodes means that the evidence B provided by a liar k must have a high λB (albeit of the wrong sign) to be effective in confusing i. The liar's strategy will not work because i will refuse to take B when |λA| is small with dogmatic filtering, while λA and λB should be of different signs when λA is large (because by then, i should have a close estimation of j's true suspiciousness, and hence, λA is of the correct sign). The evidence filtering works even when the liars are the majority among i's neighbors. *Adaptive look ahead.* Adaptive look ahead takes a different approach toward evidence consolidation. Instead of deciding whether to use the evidence provided by others directly in the cut-off decision, adaptive lookhead indirectly uses the evidence by adapting the steps to look ahead to the diversity of opinion.

IV. SIMULATION

4.1 Data Sets

We verify our design with two real mobile network traces: Haggle [22] and MIT reality [23]. The raw data sets are rich in information, some of which is irrelevant to our study, for example, call logs and cell tower IDs in MIT reality. Therefore, we remove the irrelevant fields and retain the node IDs and time-stamps for each pair-wise node encounter. Since the



(c) MIT reality (d) MIT reality

ed 0-robust) cut-off strategy and the 3-robust look-ahead cut-off strategy.

Haggle data set has only 22,459 entries spanning over three days, we repeat it another four times to make it into a data set with 112,295 entries spanning over 15 days, and thus make it comparable to the MIT reality data set in quantity. Some statistics of the processed data sets are summarized in Table 1.

4.2 Setup

Without loss of generality, we choose Le =0.5 to be the line between good and evil. For each data set, we randomly pick 10 percent of the nodes to be the evil nodes and assign them with suspiciousness greater than 0.5, the rest of the nodes are good nodes and are assigned suspiciousness less than 0.5. For a particular pairwise encounter, a uniform random number is generated for each node; a node receives a "suspicious" assessment (by the other node) if the random number is greater than its suspiciousness and receives a "nonsuspicious" assessment otherwise. Thus, each assessment is binary, while the frequency of "suspicious" assessments for a particular node reflects its suspiciousness in the long term.

4.3 Performance Metric

The performance comparison is based on two metrics:detection rate and false positive rate. The categories of the"neighbor's nature" and "cut-off decision" combinationsare shown in Table 2. For each combination, we sum up all

the decisions made by good nodes (evil nodes' cut-off decisions are irrelevant) and obtain four counts: TP (true positives), FN (false negatives), TN (true negatives), and FP (false positives). The detection rate DR is defined as

$$DR = \frac{TP}{TP+FN} \text{ X } 100\%$$

and the false positive rate FPR is defined as

$$FPR = \frac{FP}{FP+TN} \text{ X } 100\%$$

V. RELATED WORK

Proximity malware and mitigation schemes. Su et al. [24] collected Bluetooth traces and demonstrated that malware could effectively propagate via Bluetooth with simulations. Yan et al. [25] developed a Bluetooth malware model. Bose and Shin [26] showed that Bluetooth can enhance malware propagation rate over SMS/MMS. Cheng et al. [27] analyzed malware propagation throughproximity channels in social networks. Akritidis et al. [4] quantified the threat of proximity malware in wide-area wireless networks. Li et al. [28] discussed optimal malware signature distribution in heterogeneous, resource-constrained mobile networks. In traditional, non-DTN, networks, Kolbitsch et al. [8] and Bayer et al. [9] proposed to detect malware with learned behavioral model, in terms of system call and program flow. We extend the Naive Bayesian model, which has been applied in filtering email spams [13], [14], [15], detecting botnets [16], and designing IDSs [10], [17], and address DTN-specific, malware-related, problems. In the context of detecting slowly propagating Internet worm, Dash et al. presented a distributed IDS architecture of local/global detector that resembles the neighborhood-watch model, with the assumption of attested/honest evidence, i.e., without liars [10]. Mobile network models and traces. In mobile networks, one cost-effective way to route packets is via the short-range channels of intermittently connected smartphones [29], [30], [31]. While early work in mobile networks used a variety of simplistic random i.i.d. models, such as random waypoint,

recent findings [32] show that these models may not be realistic. Moreover, many recent studies [33], based on real mobile traces, revealed that a node's mobility shows certain social network properties. Two real mobile network traces were used in our study. Reputation and trust in networking systems. In the neighborhood watch model, suspiciousness, defined in (1), can be seen as nodes' reputation; to cut a node off is to decide that the node is not trustworthy. Thus, our work can be viewed from the perspective of reputation/trust systems. Three schools of thoughts emerge from previous studies. The first one uses a central authority, which by convention is called the trusted third party. In the second school, one global trust value is drawn and published for each node, based on other nodes' opinions of it; eigenTrust [34] is an example. The last school of thoughts includes the trust management systems that allow each node to have its own view of other nodes. Our work differs from previous trust management work in addressing two DTNspecific, malware-related, trust management problems:1) insufficient evidence versus evidence collection risk and 2) sequential and distributed online evidence filtering.

CONCLUDING REMARKS

Behavioral characterization of malware is an effective alternative to pattern matching in detecting malware, especially when dealing with polymorphic or obfuscated malware. Naive Bayesian model has been successfully applied in non-DTN settings, such as filtering email spams and detecting botnets. We propose a general behavioural characterization of DTN-based proximity malware. We present look ahead, along with dogmatic filtering and adaptive look ahead, to address two unique challenging in extending Bayesian filtering to DTNs: "insufficient evidence versus evidence collection risk" and "filtering false evidence sequentially and distributedly." In prospect, extension of the behavioral characterization of proximity malware to account for strategic malware detection evasion with game theory is a challenging yet interesting future work.

REFERENCES

[1] Trend Micro Inc. SYMBOS_CABIR.A., http://goo.gl/aHcES, 2004.
[2] http://goo.gl/iqk7, 20 13.
[3] Trend Micro Inc. IOS_IKEE.A., http://goo.gl/z0j56, 2009.
[4] P. Akritidis, W. Chin, V. Lam, S. Sidiroglou, and K. Anagnostakis, "Proximity Breeds Danger: Emerging Threats in Metro-Area Wireless Networks," Proc. 16th USENIX Security Symp., 2007.
[5] A. Lee, "FBI Warns: New Malware Threat Targets Travelers, Infects via Hotel Wi-Fi," http://goo.gl/D8vNU, 2012.
[6] NFC Forum. about NFC, http://goo.gl/zSJqb, 2013.
[7] Wi-Fi Alliance. Wi-Fi Direct, http://goo.gl/fZuyE. 2013.
[8] C. Kolbitsch, P. Comparetti, C. Kruegel, E. Kirda, X. Zhou, and X. Wang, "Effective and Efficient Malware Detection at the End Host," Proc. 18th Conf. USENIX Security Symp.
[9] U. Bayer, P. Comparetti, C. Hlauschek, C. Kruegel, and E. Kirda, "Scalable, Behavior-Based Malware Clustering," Proc. 16th Ann. Network and Distributed System Security Symp. (NDSS), 2009.
[10] D. Dash, B. Kveton, J. Agosta, E. Schooler, J. Chandrashekar, A. Bachrach, and A. Newman, "When Gossip is Good: Distributed Probabilistic Inference for Detection of Slow Network Intrusions," Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI), 2006.
[11] G. Zyba, G. Voelker, M. Liljenstam, A. Me´hes, and P. Johansson, "Defending Mobile Phones from Proximity Malware," Proc. IEEE
INFOCOM, 2009.
[12] F. Li, Y. Yang, and J. Wu, "CPMC: An Efficient Proximity Malware Coping Scheme in Smartphone-Based Mobile Networks," Proc.
IEEE INFOCOM, 2010.

[13] I. Androutsopoulos, J. Koutsias, K. Chandrinos, and C. Spyropoulos, "An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-Mail Messages," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2000.
[14] P. Graham, "Better Bayesian Filtering," http://goo.gl/AgHkB, 2013.
[15] J. Zdziarski, Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification. No Starch Press, 2005.
[16] R. Villamarı´n-Salomo´n and J. Brustoloni, "Bayesian Bot Detection Based on DNS Traffic Similarity," Proc. ACMymp. Applied Computing (SAC), 2013.
[17] J. Agosta, C. Diuk-Wasser, J. Chandrashekar, and C. Livadas, "An Adaptive Anomaly Detector for Worm Detection," Proc. Second USENIX Workshop Tackling Computer Systems Problems with Machine Learning Techniques (SYSML), 2007.
[18] S. Marti et al., "Mitigating Routing Misbehavior in Mobile Ad Hoc Networks," Proc. ACM MobiCom, 2000.
[19] P. Michiardi and R. Molva, "Core: A Collaborative Reputation Mechanism to Enforce Node Cooperation in Mobile Ad Hoc
Networks," Proc. IFIP TC6/TC11 Sixth Joint Working Conf. Comm. and Multimedia Security, p. 107, 2002.
[20] S. Buchegger and J. Le Boudee, "Self-Policing Mobile Ad Hoc Networks by Reputation Systems," IEEE Comm. Magazine, vol. 43, no. 7, pp. 101-107, July 2005.
[21] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, second ed. Wiley-Interscience, Nov. 2001.
[22] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD Data Set Cambridge/Haggle (v. 2006-09-15)," http://goo.gl/RJrKN, Sept. 2006.
[23] N. Eagle and A. Pentland, "CRAWDAD Data Set MIT/Reality (v. 2005-07-01)," http://goo.gl/V3YKc, July 2005.
[24] J. Su, K. Chan, A. Miklas, K. Po, A. Akhavan, S. Saroiu, E. de Lara, and A. Goel, "A Preliminary Investigation of Worm Infections in a Bluetooth Environment," Proc. Fourth ACM Workshop Recurring Malcode (WORM), 2006.
[25] G. Yan, H. Flores, L. Cuellar, N. Hengartner, S. Eidenbenz, and V. Vu, "Bluetooth Worm Propagation: Mobility Pattern Matters!," Proc. Second ACM Symp. Information, Computer and Comm. Security
(ASIACCS), 2007.
[26] A. Bose and K. Shin, "On Mobile Viruses Exploiting Messaging and Bluetooth Services," Proc. SecureComm and Workshop, 2006.
[27] S. Cheng, W. Ao, P. Chen, and K. Chen, "On Modeling Malware Propagation in Generalized Social Networks," IEEE Comm. Letters, vol. 15, no. 1, pp. 25-27, Jan. 2011.
[28] Y. Li, P. Hui, L. Su, D. Jin, and L. Zeng, "An Optimal Distributed Malware Defense System for Mobile Networks with Heterogeneous Devices," Proc. IEEE Eighth Ann. Comm. Soc. Conf. Sensor, Mesh and Ad Hoc Comm. and Networks (SECON), 2011.
[29] A. Vahdat and D. Becker, "Epidemic Routing for Partially- Connected Ad Hoc Networks," technical report, Duke Univ., 2002.
[30] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks,"
[31] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation Forwarding," Proc. ACM MobiHoc, 2008..
[32] E. Daly and M. Haahr, "Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs," IEEE Trans.Mobile Computing, vol. 8, no. 5, pp. 606-621, May 2009.

● ● ●

LOW POWER ERROR CONTROL CODING IMPLEMENTATION FOR WIRELESS SENSOR NETWORKS

Mr. M.Vasanth, Assistant Professor, Mahendra college of Engineering, salem,
manickam.vasanth@gmail.com

***Abstract*- Wireless sensor networks (WSN) offer an increasinglyattractive mode of data gathering in distributed system architectures and dynamic access via wireless connectivity. ECC provides coding gain, resulting in transmitter energy savings, at the cost of added decoder power consumption. The main challenge in deploying WSN is to improve energy-efficiency and lifetime of the nodes while keeping communication reliability. The energy efficiency of error control schemes should be considered because of the strict energy constraints of wireless sensor networks. Wireless sensor networks require simple and facile error control schemes because of the low complexity request of sensor nodes. With the ever increasing data throughputs required by communication application, there is an actual need for new effective architectures (small area and high speed) for circuit parts dedicated to error detecting/correcting coding (EDC/ECC).The Convolutional Encoder and Decoder for Wireless Sensor Networks are studied and the right configurations for the encoder will be selected and a Parallel/Pipelined Architecture for the Convolutional Encoder will be explored and implemented in an FPGA platform in order to provide a low power and fast encoding scheme and the best architecture will presented.**

I. INTRODUCTION

In general, WSN nodes are made of battery-supplied small devices with reduced processing capability and a radio frequency transceiver unity, both operating in a collaborative way [1]. Therefore, much of the research in this area is concerned with energy conservation. The goal is to extend the Sensor node and network lifetimes, since the lost of a node can make the network unavailable. However, the aforementioned solutions are susceptible to channel impairments, because any radio signal is affected by random noise and channel fading [2, 3]. If a node receives a corrupted data packet, the data can be discarded and the node keeps waiting for a new transmission or the node employs an Automatic Repeat request (ARQ) procedure (a retransmission procedure). However, in both cases there is a waste of energy in the network. A particularly undesirable situation occurs when the channel condition is bad, causing successive retransmissions.

Another method to increase the energy conservation in WSN is to apply forward error correction (FEC) strategies, reducing the frame error rate and consequently the number of retransmissions. Basically there are two classes of error control codes: block codes and convolutional codes. The convolutional encoding technique is a strategy widely used inwireless communication environments like sensor networks, since they usually present a simpler implementation for the same performance of a competitor block code [4]. The convolutional encoder is implemented using a set of shift registers (memories) and module two adders. The efficiency of a convolutional code depends on its memory order and coding rate. Convolutional codes can be decoded by using the code trellis to find the most likely sequence of codes. The Viterbi Algorithm (VA) [4] simplifies the decoding task by limiting the number of sequences examined. In a preliminary analysis it seems to be adequate to apply a powerful error correcting code in all network nodes in a WSN, in order to obtain the maximum error correction capability. In our study, we considered FEC schemes employing convolutional codes with rate 1⁄2 for different complexities (memory orders). The ARQ scheme is assumed to follow a stop-and-wait protocol [6]. The main objective is to demonstrate the improvement in the network energy conservation through the use of optimized code rate selection.

In this paper, we introduce a new architectural scheme for the OTM convolutional encoders, in which parallel andpipelining techniques are used together. While these approaches are generally successful in matching the high throughput constraints, they generally tend to miss the low-cost constraints of end user applications.

II. RADIO CHANNEL MODEL

In a communication process, the transmitted signal suffers path loss attenuation and is corrupted by Additive White Gaussian Noise (AWGN). In a wireless environment there is an additional degradation generated by multipath fading, causing fluctuations in the received signal strength. Multipath Fading significantly decreases WSN performance

in terms of energy consumption, because in general the packets received with errors must be retransmitted. The fading process is classically modeled using the Rayleigh probability distribution [2]. The performance of a sensor node in terms of frame or bit error probability depends on the average Signal-to-Noise Ratio (SNR) at the receiver. By its turn the instantaneous SNR value depends on the channel gain during a symbol or block transmission, which follows the Rayleigh distribution. The path loss model considered in this work is definedby [2]

$$PL(d)\,[dB] = PL(d_0)\,[dB] + n\,10\log_{10}(d/d_0) + \chi\sigma \quad (1)$$

Where, *PL (d)* represents the path loss at a distance *d* from the transmitter. The parameter *PL (do)* defines the path loss in a reference distance *do* and *n* is the environment path loss exponent, usually between 2 and 4 [2]. The parameter $\chi\sigma$ is a random variable with log-normal distribution, which represents the shadowing effect in the received signal, i.e., fluctuations in the path loss value for the same distance between transmitter and receiver. The received average signal power, *Prx*, as a function of the distance *d* between transmitter and receiver is given by

$$Prx[\text{dBm}] = Ptx[\text{dBm}] - PL(d)\,[\text{dB}] \quad (2)$$

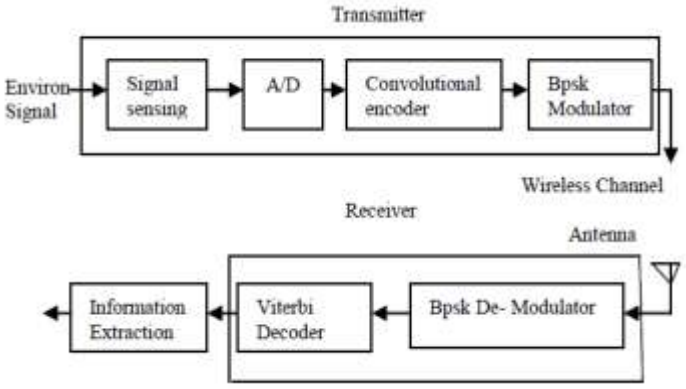
Where, *Ptx* is the transmitted power.

At the receiver, the SNR is defined as $\gamma = Prx/Pnoise$, Where *P*noise = *No. B* is the noise power, *No* is the unilateral Noise power spectral density (W/Hz), and *B* is the receiver bandwidth (Hz). This bandwidth depends on constructive characteristics of the receiver. The received power, *Prx*, may be expressed as *Prx = Eb. Rb*, where *Eb* is the average coded bit energy and *Rb* is the raw transmission rate. Therefore thereceived SNR can also be specified as

$$\gamma = Eb \cdot Rb / No \cdot B. \quad (3)$$

A sensor node generally employs a low cost and lowcomplexity radio transceiver. In our investigation, we consider Binary Phase Shift Keying (BPSK) modulation. The typical low transmission rate of a sensor node leads to a slowly-varying fading channel model [7]. The duration *b*size of a block-fading period is determined by the channel coherence time. A data frame transmission typically extends over multiple block-fading periods. The number of periods can be controlled through a slow frequency hopping systems at the transmitter. Our analysis of the bit error rate (BER) for the block fading channel was carried out through computer simulations. Then, the frame error rate (FER) can be estimatedfor a specific data frame size as $FER = 1 - (1 - BER)^{f^d_{size}}$,where f^d_{size} denotes the data frame size in bits. The coded data frame size is defined as $f^c_{\text{size}} = f^d_{size}/r$, where *r* is the coding rate.

III. SYSTEM FLOW MODEL



A) Signal Sensing:

A sensor node is a node in a wireless sensor network that is capable of performing some processing, gathering sensory information and communicating with other connected nodes in the network. The main components of a sensor node area microcontroller, transceiver, power source and one or more sensors.

B) Convolutional Encoder:

Convolutional codes are frequently used to correct errors in noisy channels. They have rather good correcting capability and perform well even on very bad channels (with error probabilities of about 10−3). Convolutional codes are extensively used in satellite communications.

i) Encoder Structure:

A convolutional code introduces redundant bits into the data stream through the use of linear shift register. A convolutional encoder [8] is specified by the parameters (*n,k,m, d*free*)*, where *n* is the number of output bits, *k* is thenumber of input bits, *m* is the memory order and *d*free is the free distance of the code, which is defined as the minimum Hamming distance between two coded sequences. The error correction capability is a function of the free distance. The information bits are input into shift registers and the output encoded bits are obtained by modulo-2 addition of the input information bits and the contents of the shift registers. The connections to the modulo-2 adders were developed heuristically with no algebraic or combinatorial foundation.The code rate r for a convolutional code is defined as r=k/n Where k is the number of parallel input information bits and n is the number of parallel Output encoded bits at one time interval. The constraint length K for a convolutional code is defined as K = m+ 1, m is the maximum number of stages (memory size) in any shift register. Convolutional code can become very complicated with various code rates and constraint lengths.

The encoder can be represented in several different but equivalent ways.

1) Generator Representation:

Generator representation shows the hardware connection of the shift register taps to the modulo-2 adders. A generator vector represents the position of the taps for an output. A "1" represents a connection and a "0" represents no connection.

2) Tree Diagram Representation:

The tree diagram representation shows all possible information and encoded sequences for the convolutional encoder. In the tree diagram, a solid line represents input information bit 0 and a dashed line represents input information bit 1. The corresponding output encoded bits are shown on the branches of the tree.

3) State Diagram Representation:

The state diagram shows the state information of a convolutional encoder. The state information of a convolutional encoder is stored in the shift registers. In the state diagram, the state information of the encoder is shown in the circles. Each new input information bit causes a transition from one state to another. The path information between the states, denoted as x/**c**, represents input information bit x and output encoded bits **c**.

4) Trellis Diagram Representation:

A convolutional encoder is often seen as a finite state machine. Each state corresponds to some value of theencoder's register. Given the input bit value, from a certain state the encoder can move to two other states. These state transitions constitute a diagram which is called a trellis diagram. The trellis diagram is basically a redrawing of the state diagram. It shows all possible state transitions at each time step.

C) BPSK Modulator:

In BPSK, individual data bits are used to control the phase of the carrier. During each bit interval, the modulator shifts the carrier to one of two possible phases, which are 180 degrees or π radians apart. This can be accomplished very simply by using a bipolar baseband signal to modulate the carrier's amplitude. The output of such a modulator can be represented mathematically as $x(t) = R(t)\cos(\omega ct + \theta)$, Where R (t) is the bipolar baseband signal, ωc is the carrier frequency, and θ is the Phase of the unmodulated carrier.

D) BPSK Demodulator:

The modulated signal is multiplied by the recovered carrier, and this product is integrated over a bit interval. If the integration result is positive, the received bit is deemed to be 1; if the integration result is negative, the received bit is deemed to be 0. The recovered carrier input to the model is in the form of a real-valued sinusoid, and the recovered clock input to the model is in the form of an integer-valued sequence that has zero values everywhere at the sampling instants corresponding to the end of each bit interval.

E) VITERBI DECODER:

The Viterbi algorithm is based on the principle of maximum likelihood decoding, which in the present case is equivalent to minimum distance decoding. Upon reception of a sequence of bits, the particular path through this diagram will be searched which is closest to this sequence in the sense of Hamming distance. Convolutional encoding is a simple procedure, decoding of a convolutional code is much more complex task. Viterbi decoding is an optimal (in a maximum-likelihood sense) algorithm for decoding of a Convolutional code. Its main drawback is that the decoding complexity grows exponentially with the code length. So, it can be utilized only for relatively short codes. A soft decision decoder is a decoder receiving bits from the channel with some kind of reliability estimate. Three bits are usually sufficient for this task.A hard decision decoder – a decoder which receives only bits from the channel (without any reliability estimate).A branch metric – a distance between the received pair of bits and one of the "ideal" pairs ("00", "01", "10", "11"). A path metric is a sum of metrics of all branches in the path.

IV. THEORETICAL DECODING COMPLEXITY

A convolutional code can be represented by a trellis diagram. This describes the permissible encoder states and transitions. The most employed convolutional decoding method is the Viterbi algorithm, which operates over the code trellis. As the transition between two states *(Si*,*Si*+1*)* defines the encoder output, the Viterbi algorithm evaluates at each instant the most likely next state, according to a previous metric of each initial state *Si* , and the cost of each encoderstate transition which is calculated based on the received data. The computational effort of the Viterbi algorithm is proportional to the density of the trellis module, or the trellis complexity. The trellis module has 2^m states. Each initial state is connected to 2^k final states. Therefore, there are a total of 2^{m+k} branches in a trellis module. Since each branch is labeled by *n* bits, in a trellis module there are a total of $n \cdot 2^{m+k}$ symbols [9].

$C = n/k \cdot 2^{m+k}$symbols/bit. (4)Figure.1shows the decoding complexity *C* for threedifferent Coding rates as a function of the code memory. The values are normalized by the decoding complexity of a convolutional code with parameters *(n, k, m) = (*2, 1, 1*)*. This complexity is proportional to the number of instructions (processing energy consumption) that must be executed by the receiver processing unit in order to decode a data frame. In our investigation we considered convolutional codes with rate *r* = 1/2 and different memory orders.

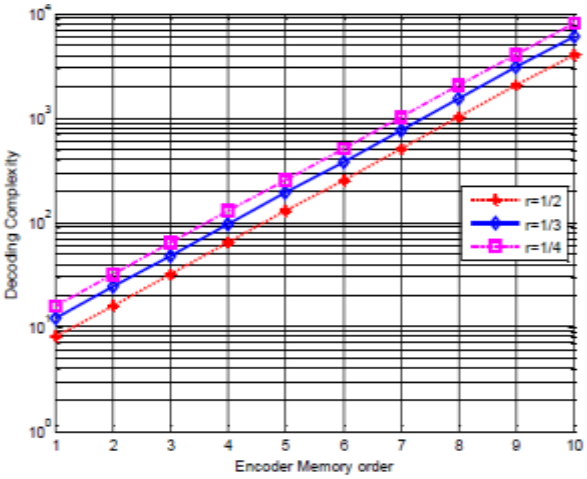


Fig. 1-Normalized theoretical decoding complexity of several convolutionalcodes

Figure.2 shows a comparison between the theoretical decoding Complexity metric (symbols/bit) given by $C = n/k \cdot 2^{m+k}$ symbols/bit (1). And the practical decoding complexity (instructions cycles/bit) given by

$IC_{total} = (5 \cdot 2^{m-1} + 2^{m-4} + n \cdot 2^{n-1} + 16.25)$ IC/bit (5)

It was assumed, without loss of generality, that the decoding of one symbol requires one instruction and the practical decoding complexity cycle (IC). In practice, the number of instruction cycles required per symbol depends on the specific platform implementation.

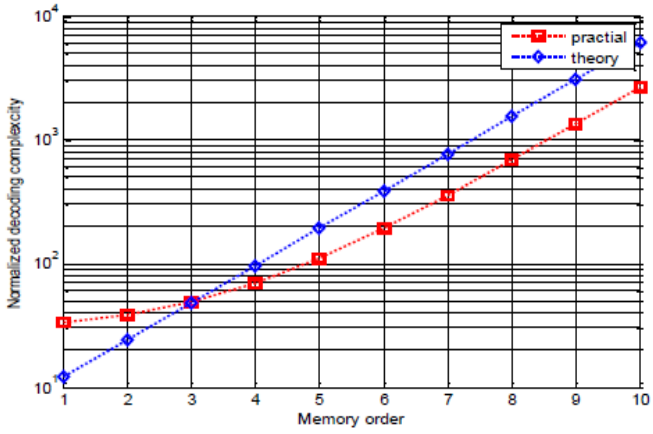


Fig. 2-Comparison between the theoretical decoding complexitiesGiven by (2) and the practical decoding complexity given by (2)

V. CONVOLUTIONAL CODING PERFORMANCES

Increasing the code complexity also increases the decoding Energy consumption (disadvantage) but results in a better error correction capability (advantage). In this sense, our study aims at determining the optimal choice for the convolutional code complexity used in the communication between two sensor nodes, in order to achieve the best trade-off between energy consumption and error correcting capability. We derived, through computer simulations, the performance of a rate $r = 1/2$ convolutional code with different memory orders for the block fading Rayleigh channel as defined in Sect. 2. Figure 3 shows the performance curves in terms of frame error rate (FER). From the figure it is clear that the performance improves with the memory order.

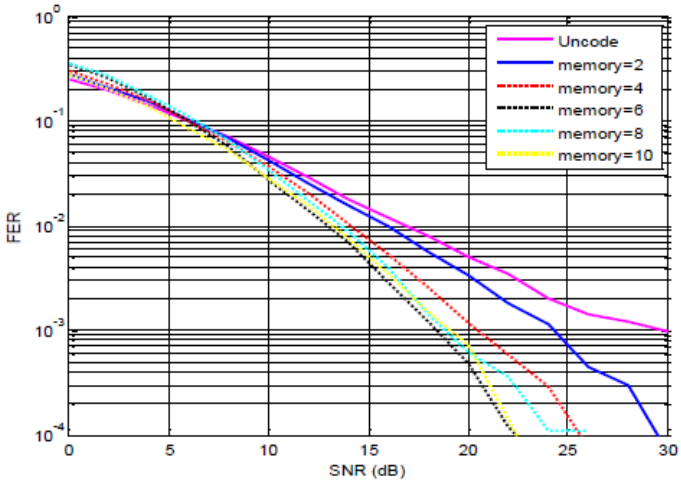


Fig. 3-Performance in terms of FER versus SNR for several rates 1/2Convolutional codes of different memory orders *m*

VI. SERIAL ARCHITECTURE

The serial form of a OTM convolutional encoder is shown in Fig. 4 for m = 3, where m is the encoder memory size. Inputs and outputs at time t are respectively equal to (X)tand (Y)t. The notational ()t is used here to denote the content of the referenced line at the beginning of clock cycle t. It will be used henceforth throughout the paper.

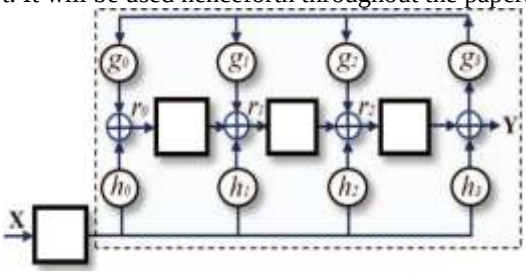


Fig 4-OTM serial encoder for m = 3.

Two types of OTM encoder exist: the Non Recursive Convolutional (NRC) and the Recursive Convolutional (RC). The encoder type (NRC or RC) is set according to the following equation [1]:

$$g_3 = 1 \Rightarrow \begin{cases} (g_0, g_1, g_2) = (0,0,0) \Rightarrow NRC \\ (g_0, g_1, g_2) \neq (0,0,0) \Rightarrow RC \end{cases} \quad (6)$$

Recursive convolutional (RC) codes differ from non-recursive Convolutional (NRC) codes by the fact that the

values in R = (r0, r1, r2) are not only driven by the input X, through the feed forward generator H = (h0, h1, h2, h3), but also by the output Y through a feedback loop controlled by the feedback generator G = (g0, g1, g2, g3).

VII. PARALLEL-PIPELINED ARCHITECTURE

The architecture presented hereafter, whose overall scheme is shown in Fig. 5, is a fast small-area parallel pipeline encoder to be used with convolutional codes. In the following, we will describe its operation principles and its application to OTM convolutional encoders. The CPS (Cumulated Pipeline State) acronym will be used down from here to designate its constitutive block, whose role is explained next.

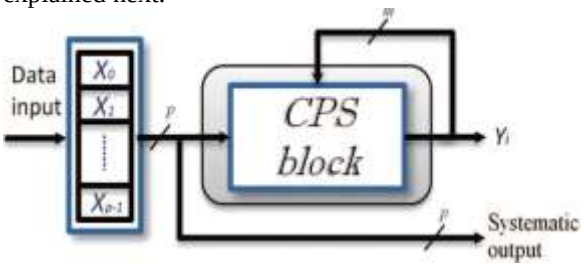


Fig.5- parallel-pipelined OTM encoder.

A parallel-pipelined approach is used to process p bits per clock cycle (p > 1), while preventing a slowdown of the clock frequency f_{clk}. As depicted in Fig. 5, the architecture is build up of just the CPS block. The structure of this block generates the p output bits of the OTM encoder combining delayed values (in the pipeline) of the data on the inputs X_i and the outputs Y_i using multiple feed-forward and feedback loops built over the coefficients of the H and the G generators, respectively.

CONCLUSION

The use of convolutional codes for error correction in wireless sensor networks can significantly increase the node and network lifetime, if the optimized code complexity is employed along each transmission hop. The use of the optimized complexity reduces the overall energy consumption generated by the local node processing and by the HARQ retransmission protocol. It is important to say that the savings in energy consumption depend on the network topology.

In this paper, we proposed new parallel-pipeline architecture for OTM (One To Many) convolutional encoders. Actually, in addition to global speed (throughput) acceleration, meaningful area savings were made possible. On 32-bit parallel-pipeline implementations, up to about 58% area savings have been achieved with data rates up to 8.10 Gbits/s. Practically, the new architectural approach proves to be an effective method in general to design small-area and high throughput convolutional encoders, able to satisfy the low cost and high-speed constraints characterizing modern digital communication systems.

REFERENCES

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., &Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer NetworksJournal*, *38*(4), 393–422.
- Rappaport, T. S. (1996). *Wireless communications: principles andPractice*. New York: Prentice Hall.
- Proakis, J. G. (1995). *Digital communications*. New York: McGraw-Hill.
- Hendrix, H. (2002). Viterbi decoding techniques for the TMS320C54x DSP generation. In *Texas instruments applicationreport SPRA071A*.
- Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions onInformation Theory*, *13*, 260–269.
- Wicker, S. B. (1995). *Error control systems for digitalcommunication and storage*. New York: Prentice Hall.
- Chuang, A., Guillen, A. F., Rasmussen, L. K., &Collings, I. B. (2006). Optimal rate-diversity-delay tradeoff in ARQ block-fading channels. In *Proceedings of the IEEE information theory workshop*, Uruguay (pp. 507–511).
- Lin, S., & Costello, D. J. (1983). *Error control coding:fundamentals and applications*. New York: Prentice Hall.
- McEliece, R. J., & Lin, W. (1996). The trellis complexity of Convolutional codes. *IEEE Transactions on Information Theory*, *42*(6), 1855–1864.
- Hendrix, H. (2002). Viterbi decoding techniques for the TMS320C54x DSP generation. In *Texas instruments applicationreport SPRA071A*.
- A. M'sir, F. Monteiro, A. Dandache, B. Lepley, "Design of a high speed parallel encoder for convolutional codes," Microelectronics Journal, vol. 35(2), pp. 151–166, Feb. 2004.
- Karvonen, H., Shelby, Z., &Pomalaza-Raez, C. (2004). Coding for energy efficient wireless embedded networks. In *Proceedingsof the IEEE international workshop on wireless ad-hoc networks*,Finland (pp. 300–304).
- KeshabK.Parhi.(1999), "VLSI Digital Signal Processing Systems" Design and Implementation, by John Wiley & Sons.

THE CERTAINTY OF BI SYSTEM FOR SME

[1]Govinda Rajulu Lanke, [2]Dr.T.Bhuvaneswari
[1]Research Scholar, Dept of Computer Scienece, SCSVMV University, Kanchipuram-TN , India-631 561.
[2]Asst Prof, Dept of Computer Scienece L.N. Govt College, Ponneri-TN, India-601 204.
[1]govinda.lanke@gmail.com,[2]t_bhuvaneswari@yahoo.com

Abstract— **Business Intelligence (BI) is the key attention of in-depth application of enterprise information. This paper analyzes the necessities and difficulties in putting business intelligence into SMEs information, proposes small and medium enterprises BI application solutions in the use of software distribution method.**

BUSINESS INTELLIGENCE (Bl) is one of the simpler technology terms to understand. However, getting it right has proved to be one of the more difficult tasks for companies, large and small, to achieve. In simple terms, Bl is all about taking the volume of data that every company collects on a daily basis and turning it into a form that can be understood by decision makers. The ability to use historical information to make business decisions is the fundamental basis of a business.By transforming data into information and bringing together disparate sources of valuable information so they can be analysed. BI can enable SMBs to gain timely access to high quality, reliable business data and metrics. That enables them to make better-informed decisions, whether dealing with customers, suppliers or internal processes. Fundamentally, Bi enables companies to more easily identify and reduce costs and to take guesswork out of decisions that help drive top-line revenue growth.

Keywords: Decision Support Systems (DSSs), Data Warehouse (DW), Business Intelligence (BI), Small and Mid-Sized Enterprises (SMEs), Smart Business Systems (SBSs), Business Intelligence Systems (BISs).

I. INTRODUCTION

During the last decade, data warehouses (**DWs**) have become an essential component of modern decision support systems in most companies of the world. In order to be competitive, even small and middle-sized enterprises (**SMEs**) now collect large volumes of information and are interested in business intelligence (**BI**) systems. SMEs are regarded as significantly important on a local, national or even global basis and they play an important part in the any national economy.

Although many studies have been conducted on the need of decision support systems (**DSSs**) for small businesses, most of them adopt existing solutions and approaches, which are appropriate for large-scaled enterprises, but are inadequate for small and middle-sized enterprises.

Difficulties to obtain a BI system for SMEs

In spite of multiples advantages, existing DSSs frequently remain inaccessible or insufficient for SMEs because of the following factors:

- HIGH PRICE
- HIGH REQUIREMENTS FOR A HARDWARE INFRASTRUCTURE
- COMPLEXITY FOR MOST USERS
- IRRELEVANT FUNCTIONALITY
- LOW FLEXIBILITY TO DEAL WITH A FAST CHANGING DYNAMICS
- BUSINESS ENVIRONMENT
- LOW ATTENTION TO DIFFERENCE IN DATA ACCESS NECESSITY IN SMES AND LARGE-SCALED ENTERPRISES.

In addition, many projects fail due to the complexity of the development process. Moreover, as the work philosophies of small and large-scaled enterprises are considerably different, it is not advisable to use tools destined to large-scaledenterprises. In short, “one size does not fit all”. Furthermore, there are a lot of problems in the identification of information needs of potential users in the process of building a data warehouse.

Thereby, SMEs require lightweight, cheap, flexible, simple and efficient solutions. To aim at these features, we can take advantage of light clients with web interfaces. For instance, web technologies are utilized for data warehousing by large corporations, but there is an even greater demand of such kind of systems among small and middle-sized enterprises. Usage of web technologies provides cheap

software, because it eliminates the necessity for numerous dispersed applications, the necessity of deployment and maintenance of corporate network, and reduces training time. It is simple for end-users to utilize web-based solutions. In addition, a web-based architecture requires only lightweight software clients (i.e., web browsers).

Thus, our objective is to propose original and adapted BI solutions for SMEs. To this aim, we first present and discuss web-based BI approaches, namely web data warehouses and web-based open source software for data warehousing. In Section. We finally conclude this paper and provide our view on how the research and technologies surveyed in this paper can be enhanced to fit SME's BI needs.

II. THE EMERGENCE OF SMART BUSINESS SYSTEMS(SBSs)

Today, the emergence of smart business systems (SBSs) is leading the way to optimizing a company's operations for changing times. Not only is computer technology changing more rapidly each day, but also are business requirements. Decision makers are being pressed to respond to customer needs and competitive threats in days and weeks instead of months or years.

In the past twentieth century, decision makers have utilized a wide range of information systems to improve their decisions. However, in this twenty-first century, there is need to take decision makers to a much higher level by providing them with the ability to "optimize the enterprise." More specifically, optimization refers to the ability to assess a myriad of possibilities in order to find the best one or near best one. At this time, the focus of smart business systems for the optimized organization centres on the functional areas of a typical company, in particular, corporate planning, marketing, manufacturing, and accounting. Such an approach can go beyond each functional area and tie in with the company's overall operations and its trading partners, thereby resulting in an integrated optimization approach. Essentially, the use of an optimization approach gives a typical company the necessary "smarts" to meet or possibly beat competition.

III. BUSINESS INTELLIGENCE SYSTEMS (BISS)

Fundamentally, business intelligence systems (BISs) make great use of data marts and data warehouses as well as operational databases for the purpose of measuring historical activity. Over time, however, business intelligence activities have been expanded to include other kinds of data, information, and knowledge that are future oriented. For example, software developers and their clients are integrating data mining tools to anticipate the future based on historical data, information, and knowledge, or visualization tools to quickly scan large amounts of relevant information and knowledge. Other companies are integrating text and images with data marts and data warehouses, using collocated document management systems or object relational databases. Also, there is a movement to "push" relevant information and knowledge to users in real time based on predefined business rules or collaborative arrangements among company personnel.

From this perspective, companies are looking at the organization holistically for a thorough understanding of its operations within a BIS operating mode. This generally means extending a company's functions, processes, and technology via E-commerce to its trading partners (i.e., customers and suppliers). A business intelligence system centres on managing internal and external information knowledge and their resulting intelligence in a proactive manner in order to create a competitive advantage that is linked to a company's achievable objectives and its measurable goals. It should be noted that an effective BIS operating mode centres on organizing and displaying business intelligence about important topical areas rather than trying to tell everything that is known. A business intelligence system can be looked upon as a set of tools and applications that allow decision makers to gather, organize, analyse, distribute, and act on critical business issues, with the goal of helping companies make faster, better, and more informed business decisions.

Business intelligence systems can be defined as systems for business that turn selected data, information, and knowledge into desired intelligence for business gain by decision makers. The type of system and software used is situational. Business intelligence systems employ various analytical and collaborative tools and utilize a database infrastructure—all within a global computer networking architecture. Overall, business intelligence systems provide decision makers with the ability to understand (i.e., the intelligence to gain insights into) the relationships of presented facts in the form of data, information, and knowledge in order to guide action toward a desired actionable goal. They provide decision makers with timely data, information, and knowledge for problem solving and, in particular, problem finding. As such, business intelligence systems are the forerunners of smart business systems.

IV. ESSENTIAL STEPS TO DEVELOPING AND IMPLEMENTING SMART BUSINESS SYSTEMS SUCCESSFULLY

Although there is no comprehensive approach nor is one anticipated in the near future to develop and implement smart business systems successfully, there are a number of suggested steps. Underlying all of these steps is the empowerment of a company's customers and employees to have a better understanding and more control over their total operations. The ultimate goal of a smart business system for

a typical decision maker is not to do a better job of understanding the company's inner workings, but rather to optimize its operations so that the company is run more effectively as well as efficiently in the short to long run.

Typically, there is an order that should be followed in undertaking these steps. They are as follows:

- **i.** Get support by starting at the very top of the company
- **ii.** Appoint a chief information officer to sponsor the smart technology
- **iii.** Select an experienced team to develop and implement the smart business system(s)
- **iv.** Select an effective smart business system design methodology
- **v.** Determine the appropriate data storage for optimizing results
- **vi.** Select appropriate software tools that canter on producing optimal or near-optimal results
- **vii.** Determine computer networking that ties in with smart technology
- **viii.** Develop important smart applications
- **ix.** Disseminate appropriate optimized results
- **x.** Focus on transforming optimized results into action

V. WEB-POWERED BI

The Web has become the platform of choice for the delivery of business applications for large-scaled enterprises as well as for SMEs. Web warehousing is a recent approach that merges data warehousing and business intelligence systems with web tech technologies. In this section, we present and discuss web data warehousing approaches, their features, advantages and possibilities, as well as their necessity and potential for SMEs.

5.1 Web Warehousing

There are two basic definitions of web warehousing. The first one simply states that web warehouses use data from the Web. The second concentrates on the use of web technologies in data warehousing. We focus on second definition in our paper.

Web-data warehouses inherit a lot of characteristics from traditional data warehouses, including: data are organized around major subjects in the enterprise; information is aggregated and validated; data is represented by times series, not by current status. Web-based data warehouses nonetheless differ from traditional DWs. Web warehouses organize and manage the stored items, but do not collect them. Web-based DW technology changes the pattern of users accessing to the DW: instead of accessing through a LAN (Local Area Network), users access via Internet/Intranet.

Specific issues raised by web-based DW include unrealistic user expectations, especially in terms of how much information they want to be able to access from the Web; security issues; technical implementation problems related to peak demand and load problems.

Eventually, web technologies make data warehouses and decision support systems friendlier to users. They are often used in data warehouses only to visualize information. At the same time, web technology opens up multiple information formats, such as structured data, semi-structured data and unstructured data, to end-users. This gives a lot of possibilities to users, but also creates a problem known as data heterogeneity management.

Another important issue is the necessity to view the Web as an enormous source of business data, without whose enterprises lose a lot of possibilities. Owing to the Web, business analysts can access large external to enterprise information and then study competitor's movements by analysing their web site content can analyse customer preferences or emerging trends. So, e-business technologies are expected to allow SMEs to gain capabilities that were once the preserve of their larger competitors. However, most of the information in the Web is unstructured, heterogeneous and hence difficult to analyse.

5.2 Cloud computing

Another, increasingly popular web-based solution is cloud computing. Cloud computing provides access to large amounts of data and computational resources through a variety of interfaces. It is provided as services via cloud (Internet). These services delivered through data centres are accessible anywhere. Besides, they allow the rise of cloud analytics.

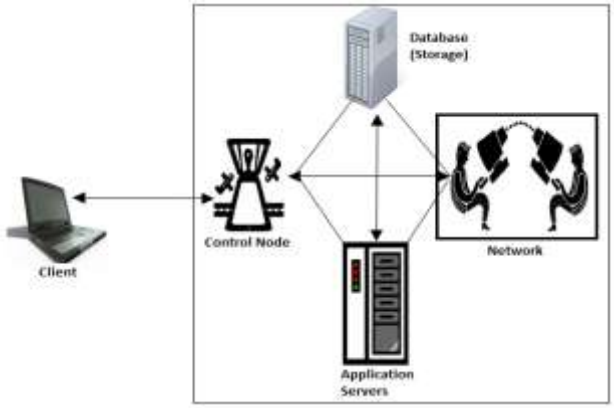
The main consumers of cloud computing are small enterprises and start-ups that do not have a legacy of IT investments to manage. Cloud computing based BI tools are rather cheap for small and middle-sized enterprises, because they provide no need of hardware and software maintenance and their prices increase according to required data storages. Contrariwise, cloud computing does not allow users to physically possess their data storage. It causes user dependence on the cloud computing provider, loss of data control and data security. In conclusion, most cloud computing-based BI tools do not fit enterprise requirements yet, researchers still finding new approaches on clouds to adopt Business Intelligence system in to the Enterprise market tis can shows the more benefits to Small and Mid Sized Enterprises in upcoming days.

5.2.1 How Cloud Computing Works

Let's say you're an executive at a large corporation. Your particular responsibilities include making sure that all of your employees have the right hardware and software they need to do their jobs. Buying computers for everyone isn't enough, you also have to purchase software or software licenses to give employees the tools they require. Whenever you have a new hire, you have to buy more software or make sure your current software license allows another user. It's so stressful that you find it difficult to go to sleep on your huge pile of money every night.

Soon, there may be an alternative for executives like you. Instead of installing a suite of software for each computer, you'd only have to load one application. That application would allow workers to log into a Web-based service which hosts all the programs the user would need for his or her job. Remote machines owned by another company would run everything from e-mail to word processing to complex data analysis programs. It's called cloud computing, and it could change the entire computer industry.

Fig. 1. How Cloud Computing Works



In a cloud computing system, there's a significant workload shift. Local computers no longer have to do all the heavy lifting when it comes to running applications. The network of computers that make up the cloud handles them instead. Hardware and software demands on the user's side decrease. The only thing the user's computer needs to be able to run is the cloud computing systems interface software, which can be as simple as a Web browser, and the cloud's network takes care of the rest.

There's a good chance you've already used some form of cloud computing. If you have an e-mail account with a Web-based e-mail service like Hotmail, Yahoo! Mail or Gmail, then you've had some experience with cloud computing. Instead of running an e-mail program on your computer, you log in to a Web e-mail account remotely. The software and storage for your account doesn't exist on your computer it's on the service's computer cloud.

5.3 Web-based open source software

In this section, we focus on ETL (Extraction Transformation Loading) tools, OLAP servers and OLAP clients. Their characteristics are summarized in Table 1.

TABLE I
WEB-BASED OPEN SOURCE SOFTWARE

		Tools	Platform	License
ETL	ROLAP	Clover ETL	Java	LGPL
		JasperETL	Java	GPL
	MOLAP	Palo ETL Server	Java	GPL
OLAP	Server	Mondrian	Java	CPL
		Palo	Java	GPL
	Clients	Free Analysis	Java	MPL
		JPalo	Java	GPL
		PoeOLPAP	Java	LGPL

5.3.1 ETL

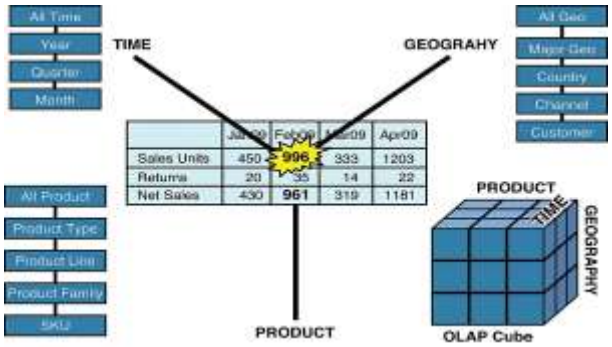
Web-based free ETL tools are in most cases ROLAP (Relational OLAP)- oriented. ROLAP-oriented ETL tools allow user to define and create data transformations in Java (JasperETL) or in TL (Clover.ETL)1. SingularMOLAP (Multidimensional OLAP)-oriented ETL Palo defines the ETL process either via web interfaces or via XML structures for experts. All studied ETL tools configure heterogeneous data sources and complex file formats. They interact with different DBMSs (DataBase ManagementSystems). Some of the tools can also extract data from ERP (Enterprise Resource Planning) and CRM (Customer Relationship Management) systems.

5.3.2 On-Line Analytical Processing (OLAP) Systems

Today, on-line analytical processing (**OLAP**) centres on systems that focus on asking and answering "what happened" to operations. A most important part of OLAP systems is their multidimensional analysis capabilities, that is, analysis that goes beyond the traditional two-dimensional analysis. Essentially, multidimensional analysis represents an important method for leveraging the contents of an organization's production data and other data stored in company databases and data warehouses because it allows users to look at different dimensions of the same data, say by business units, geographical areas, product levels, market segments, and distribution channels. As such, OLAP makes it easier to do analyse that cross departmental and even corporate boundaries. Another way of viewing OLAP is

getting a typical company out of the custom report-writing business and into the data-cube-server-building business. An OLAP data structure can be thought of as a Rubik's cube of data that users can twist and twirl in different ways to work through "what happened" scenarios to get at the real issues of the situation.

Fig. 2. Example of an OLAP Cube image



Current OLAP tools have proven their value in providing a multidimensional view of summarized data. Some of these tools are available within a business intelligence operating mode to further enhance a better understanding of a company's operations today as well as in the future. Although OLAP tools meet many needs, they do not allow for the analysis and understanding of individual customer behaviour at the transaction level. The reason is that OLAP tools, both those implemented on top of relational databases (ROLAP) and those implemented on top of multidimensional databases (MOLAP), centre on aggregating and summarizing data.

Although aggregated data can provide trend analysis information, it is not actionable at an individual level. For example, knowing that 5,000 products were sold does not help a company's decision makers to focus on individual customers. It knows who those 5,000 customers are that can help decision makers to get at the underlying profiles and possible motivation for buying a company's products or services. From this broader view, knowledge discovery is needed to complement that information found within an OLAP system that decision makers have found by "slicing and dicing" rapidly through reams of data. Overall, OLAP systems can be useful building blocks for the implementation of smart business systems.

In this section we review OLAP servers as well as OLAP clients. All OLAP severs use the MDX (Multi-Dimensional expression) language for aggregating tables. They parse MDX into SQL to retrieve answers to dimensional queries. All OLAP servers exist for Java, but a Palo exists also for .NET, PHP, and C.

CONCLUSION

Nowadays, BI becomes an essential part of any enterprise, even an SME. This necessity is caused by the increasing data volume indispensable for decision making. Existing solutions and tools are mostly aimed at large-scaled enterprises; thereby they are inaccessible or insufficient for SMEs because of high price, redundant functionality, complexity, and high hardware and software requirements. SMEs require solutions with light architectures that, moreover, are cheap and do not require additional hardware and software.

This paper discusses the importance of data warehousing for SMEs, presents the main characteristics and examples of web-based data warehousing, MOLAP systems, security issues in cloud computing systems. In this context, our research objective is to design BI solutions that are suitable for SMEs and avoid the aforementioned disadvantages.

REFERENCES

[1]. R. Kimball and M. Ross. The Data Warehouse Toolkit: the complete guide to dimensional modelling. Wiley Computer Publishing, 2002.

[2]. W. Chung and H. Chen. Web-Based Business Intelligence Systems: A Review and Case Studies. In G. Adomavicius and A. Gupta, editors, Business Computing, volume 3, chapter 14, pages 373–396. Emerald Group Publishing, 2009.

[3]. C. Hsieh and B. Lin. Web-based data warehousing: current status and perspective. The Journal of Computer Information Systems, 43:1–8, January 2002.

[4]. J. Staten. Is cloud computing ready for the enterprise? Forrester Research, March 2008. Retrieved September 1, 2010 from http://www.forrester.com/rb/Research/is cloud computing ready for enterprise/q/id/44229/t/2.

[5]. C. Thomsen and T. B. Pedersen. A Survey of Open Source Tools for Business Intelligence. International Journal of Data Warehousing and Mining, 5(3):56–75, jul-sep 2009.

[6]. Smart Business Systems for the Optimized Organization by Thereof, Robert J.; Hector, James J. Greenwood Publishing Group, isbn10 | asin: 1567205437

WEB BASED COMMUNICATION USING TEXT STEGANOGRAPHY

[1]Mr. M.Hareesh Babu,[2]Ms.M.Bharghavi
[1]M.TECH (CNIS) Scholar Dept. of CSE,Sree Vidyanikethan Engg College , Tirupathi, Andhra Pradesh, India
[2]Assistant Professor, Dept. of CSE,Sree Vidyanikethan Engg College, Tirupathi, Andhra Pradesh, India
[1]police.max11@gmail.com,[2]bhargavisvec@gmail.com

Abstract :**With the increase in Internet Technologies, great amount of information is following electronically everyday over the network. Information security is a way to protect information against its confidentiality, reliability and availability. Hiding exchange of information is an important factor in the field of security. Cryptography and Steganography are two very important methods for this purpose and are both used to ensure data confidentiality. In Steganography a cover media is used to hide the existence of data where cryptography is used to protect information by transferring plain text into cipher text. Here we discusses some proposed methods, implementations of different embedding techniques and two different ways for hiding data and also a comparative analysis is made based upon some security variables. Text Steganography is applied on XML files and is further encrypted using a cryptographic algorithm.**

Keywords:**Steganography, Cryptography, Encryption, Decryption, Cipher text, Plaintext**

I. INTRODUCTION

Secret communication has been a subject of interest for ages. With the vast expansion of Internet, massive web based information is travelling every day and securing data is a very important subject in this matter. For security reasoning, many different methods have been implemented and new methods are evolving every day. Cryptography, Steganography and Watermarking are well known ways of securing information but they all work under different mechanisms.Cryptography makes data unreadable by writing into secret code and it ensures authentication, confidentiality and integrity. Steganography hides the existence of data and it ensures transparency, robustness and capacity. Whereas, watermarking technique provides evidence for the intellectual property rights over certain content by hiding some information in it.

II. EXISTING SYSTEM

In the existing system we use the HTML file for carrying the input which is send to the receiver from the sender and the secret message is embedded by using the text-Steganographic techniques. Here we use the DES algorithm to perform cryptography. By this we encrypt the data by the 64-bit key. Then perform embedding on the encrypted message into the html file. The resultant data is kept in a secure web page and is transferred through a communication channel

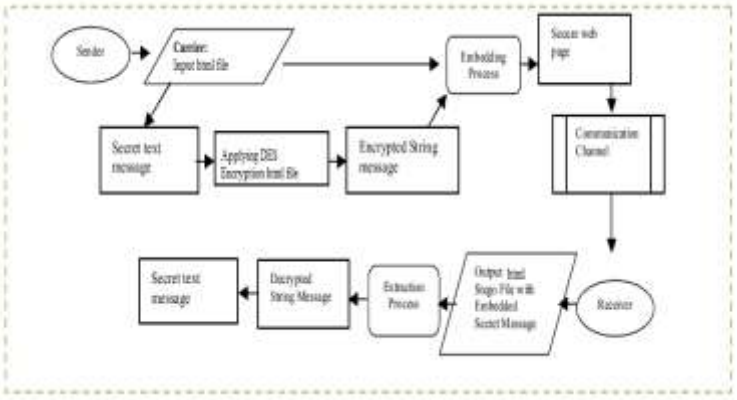


Fig.1.Existing system using DES

Disadvantages of Existing System:

HTML:

- In HTML there are no user defined tags.
- It cannot produce dynamic output alone, since it is a static language.
- Security features offered by HTML are limited.

DES :

- The 56-bit key size is the biggest defect of DES. Hardware implementations of DES are very fast. DES was not designed for software and hence runs relatively slowly

III. PROPOSED SYSTEM

In the proposed system we use the XML file as a carrier for carrying the secret message which is send to the receiver from the sender and that secret message is

embedded by using the text-Steganography techniques. Here we use the AES (Advanced Encryption Standard) algorithm to perform cryptography. In this we encrypt the data by using a key of 256-bit length. Then we perform embedding process on the encrypted message to store it in an xml file.

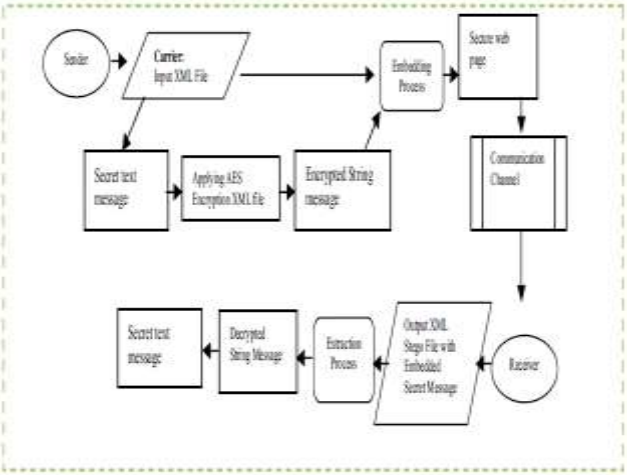


Fig.2.Proposed system using AES

Advantages of Proposed System:

XML file:

- It is a platform independent language.
- It is as easy as HTML.

AES algorithm:

- Advanced Encryption Standard not only assures security but also improves the performance in a variety of settings such as smartcards, hardware implementations etc.
- AES is federal information processing standard and there are currently no known non-brute-force direct attacks against AES.

Crypto module:

In crypto module we perform encryption and decryption. That is, encryption is done at the sender side and decryption is done at the receiver side.

- Encrypting: The secret message which sender wants to send is encrypted using the Advanced Encryption Standard (AES) algorithm.
- Decrypting: The received encrypted text which is in binary form is decrypted using the AES algorithm.

Stego module:

In stego module we perform embedding and extraction process. That is, embedding is done at the sender side and extraction is done at the receiver side.

- Embedding: In embedding process we apply the text Steganography techniques on xml file based on the cipher text that is obtained from AES algorithm.
- Extraction: This is the reverse process of Embedding at the receiver side in order to extract the cipher text from the xml file and that cipher text is given as input to the AES decryption algorithm.

Transmission module:

This module is used to send the xml file which is embedded using text Steganography techniques at the sender side and to receive that xml file at the receiver side.

Here the sender has to give the path of the xml file in order to send it to the receiver and the receiver has to give path to store the received xml file.

RESUITS

By creating a text file which contain plain text and here we named it as p.txt and the data which is written in it is the secret message which is to be transferred to the receiver. When we run the embedding process program, it asks for the plain text file path. We should specify the correct path. It takes the plain text present in that file and converts it into the cipher text. It displays cipher text and the length of the cipher text. If we give wrong path as input then it displays an error message saying that file does not exist and stops the execution process.

CONCLUSION & FUTURE ENHANCEMENTS

We have presented Text Steganography combined with Cryptography for hiding secret information using XML file to provide more security. There are nine different embedding techniques for the text Steganography, studied and applied on XML file. System has been implemented using java language for all nine methods combined with AES which has added another layer of security. All methods are measured with respect to different standards and it is analyzed that white space method, white space replacement method, color replacement method, line break method, synonyms method and acronyms methods are considered stronger and less vulnerable.

Furthermore, improvements in color, synonyms and acronyms are needed to make them more practical, efficient and stronger. Techniques discussed in our project have therefore been applied on textual information and hence could also be applied on other types of data in XML files, as XML does not only contain text but multimedia based information as well and the idea could be extended toward other parts.

REFERENCES

[1] Shingo, Kyoko, Ichiro, Osamu, “A Proposal on Information Hiding Methods using XML”, Mitsubishi Research Institute, Communication Research Laboratory, Yokohama National University and The University of Tokyo.

[2] Mohammad Laheen, Sun XingMing, “Techniques with Statistics for Web page Watermarking” 2005, NSFC No.60373062.

[3] Aasma, Sumbul, Asadullah, "Steganography: A New Horizon for Safe Communication through XML", 2005.
[4] http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
[5] http://www.infosecwriters.com/text_resources/pdf/Steganography_AMangarae.pdf

BIOGRAPHY

Mr.M Hareesh Babu is a M.Tech Scholar in the Computer Science & Engg Department, Sree Vidhyanikethan Engg College. He received Bachelor of Technology (I.T) degree in 2012 from JNTUA, Ananthapur,Andhra Pradesh, India. His research interests are Computer Networks (wireless Networks), HCI, Algorithms, web 2.0 etc.

Ms.B.Bargavi is a assistant professor in the Computer Science & Engg Department, Sree Vidhyanikethan Engg College. She received Master of Technology (CNIS) degree in 2012 from JNTUA, Ananthapur,Andhra Pradesh, India. Her research interests are Computer Networks (wireless Networks), Algorithms, Database Management Systems etc.

● ● ●

PREDICTIVE ACKS BASED CLOUD BANDWIDTH AND REDUCING COST IN THE SYSTEM

[1]Mr.B.Mahesh,[2]Mr.M.V.R.Purna Kumar
[1]M.TECH(CNIS)Scholar Dept. of CSE SreeVidyaikethanEnggCollege,Tirupathi,Andhra Pradesh
[2]Assistant Professor, Dept. of CSE SreeVidyanikethanEngg College, Tirupathi, Andhra Pradesh, India
b.mahesh552@gmail.com

Abstract: **In last number of years there is immense increase within the usage cloud computing as a result of cloud computing is growing category of IT-delivery within which applications, information and resources area unit are chop by chop provisioned, provided as standardized offerings to users with a versatile value. However it is necessary to produce the convenient evaluation model for the users of cloud. Hence we tend to present a traffic redundancy and elimination technique for which reduces the bandwidth of cloud and evaluation costs.**
Keywords: cloud computing, novel-TRE, pay-as-you-go.

I. INTRODUCTION

Now a day's the cloud computing archetype achieved extensive espousal in the field of computer science. Making customers' flexible in using the services on demand with a pay-as-you go [2] pricing model, which has proved convenient inmany respects made it successful. Low costs as well as increased flexibility made migration to the cloud undeniable. Cloud computing is that the long unreal vision of computing as a utility, wherever users will remotely store their information into the cloud therefore on relish the on demand high featured services as well as applications from a collective group of configured computing resources. By means of information outsourcing, users may be alleviated from the burden of native information storage space as well as maintenance. Traffic redundancy elimination approach is employed for minimizing the value.

Our new traffic redundancy elimination approach additionally referred to as novel-TRE or receiver primarily based TRE, that detects redundancy at the shopper facet and there's no want of server to ceaselessly. but for server specific TRE approach it's troublesome to handle the traffic expeditiously and it doesn't suite for the cloud setting owing to high process prices. Novel-TRE matches incoming chunks with a antecedently received chunk chain or native file and causation to the server for predicting the longer term information.

Packet level redundant content abolition is a well known primitive on all net routers, such a universal preparation would straight off scale back link hundreds everyplace. However, we tend to argue that way more vital network wide advantages may be derivative by remodeling network routing protocols to leverage the widespread preparation. However, we tend to believe that the numerous long run advantages of our approaches supply nice incentives for networks to adopt them. End-system redundancy elimination [4] provides quick, accommodative and stingy in memory usage so as to opportunistically leverage resources on finish hosts. EndRE is predicated on 2 modules server and therefore the shopper.

The server facet module is accountable for characteristic redundancy in network information by scrutiny against a cache of previous information and encoding the redundant information with minimized metadata. Module on the client side includes of a constant sized circular inventory accounting log of packets and easy logic to decrypt the metadata on "de-referencing" the server offsets. Thus, most of the complexness in EndRE is especially on the server facet. so it's server specific ineffective to keep up the complete synchronization between shopper and therefore the server. EndRE uses Sample Byte process theme that is faster than Rabin process. EndRE restricted for little redundant chunks of the order of 32-64 bytes. solely distinctive chunks area unit transmitted between file servers and shoppers, leading to lower information measure consumption. the essential plan underlying EndRE is that of associate degree object is divided into chunks in addition to that it is indexed by computing hashes. A limitation of this method chunk size is tiny and it's server specific. The tactic and equipment for reducing network traffic over low information measure links [5] describes the way to get aside with triangular shake

between the sender and consequently the receiver is at complete synchronization is maintained. The technique revealed for plummeting network traffic at a sender, is an information chunk is known for transmission to a receiver, which is associated to the sender over a communication linkage. The sender analyses a signature of the information chunk and determines whether the date chunk has been antecedently transmitted by trying up the signature. The sender index table associates the signatures of antecedently transmitted information chunks with distinctive index values. A message is transmitted to the receiver, wherever if the information chunk has antecedently been transmitted then the message includes associate degree index worth from the sender index table that's related to the signature of the information chunk. At the receiver, the information chunk is located during a receiver cache that stores the antecedently transmitted information chunks by trying up the index worth enclosed within the message during a receiver index table. The receiver index table associates the distinctive index values with the locations within the receiver cache of the antecedently transmitted information chunks.

To prevent chunks from being overlarge or too tiny, minimum and most chunk sizes may be such additionally. Since Rabin process determines chunk boundaries by content, instead of offset, localized changes within the information stream solely have an effect on chunks that area unit close to the changes.. Limitations of this approach are(1) End-to- finish encrypted traffic don 't cope well middle boxes.(2) It creates latency for non cached information and middle-boxes won't improve the performance.

II. SYSTEM DESIGN

System design of traffic redundancy and elimination approach is shown in figure1. The on top of figure [1] shows the design of a novel-TRE. so as to evolve to existing firewalls and minimize Over heads, we tend to use the protocol choices field to hold the PACK wire protocol. it's clear that novel-TRE may be enforced on top of the transmission level whereas victimization similar message varieties and control fields. The Figure one illustrates manner the novel-TRE operates beneath the idea that the information is redundant.

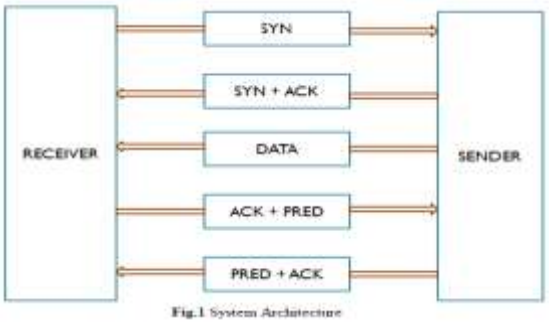


Fig.1 System Architecture

The on top of figure [1] shows the design of a novel-TRE. so as to evolve to existing firewalls and minimize Over heads, we tend to use the protocol choices field to hold the PACK wire protocol. it 's clear that novel-TRE may be enforced on top of the transmission level whereas victimization similar message varieties and control fields. The Figure one illustrates manner the novel-TRE operates beneath the idea that the information is redundant. First, each side alter the PACK possibility throughout the initial protocol shake by adding a PACK permissible flag to the protocol choices field. Then, the sender sends the (redundant) information in one or additional protocol segments, and therefore the receiver identifies that a presently received chunk is just like a bit in its chunk store. The receiver, in turn, triggers a protocol ACK message and includes the prediction within the packet's choices field. Last, the sender sends a confirmation message PRED-ACK commutation the particular information.

Traffic redundancy and elimination approach for reducing price of cloud computing consists 3 modules namely:

- *Data owner:* during this module we'll be concentrating on the choosing the infrastructure of cloud associate degreed making an account. Uploading the file into cloud server so obtaining the main points of the cloud price.
- *Cloud server module:* during this module we tend to area unit storing the received file into the server, maintaining the account details of the information owner and computing the mathematical modeling of the value.
- *Receiver module:* during this module we tend to area unit getting into all the valid details like file name, secret key, and infrastructure choice and causation predictions for the server for the longer term information.

III. TRAFFIC REDUNDANCY ELIMINATION

A typical cloud design consists of a large set of physical machines (host), every of them equipped with some virtualization mechanisms, from hardware virtualization up to micro-partitioning, OS virtualization, software system virtualization. These mechanisms permit every machine to host a instruction execution of many virtual machines (guest) every with its own OS and applications. To accommodate varied demands for various varieties of process, the foremost fashionable cloud infrastructures embrace dynamic management capabilities and virtual machine quality that's, the power to maneuver transparently virtual machines from one host to a different. By migrating a guest from associate degree full host to a different not important host, it's potential to enhance resource utilization and higher load sharing. Severally of the migration techniques, they share a typical management model: any call formula for migration has got to choose one or additional sender hosts from that some virtual machines area unit touched to different destination hosts, specifically receivers.

This paper addresses the most problems associated with migration choices, that is, it aims to answer to the subsequent questions: once it 's necessary to activate a migration, that guests of a sender host ought to migrate, and wherever they must be touched.

The load state of a bunch is obtained through a periodic assortment of measures from server monitors. These measures area unit generally characterized by noises and non stationary effects within the short-medium term, whereas there 's some periodic behavior during a long run vision (day, we tend toek) that we don't think about during this paper. Figure one shows four load profiles (concerning host electronic equipment utilizations) during a cloud design wherever physical machines host any variety of virtual machines and applications, like websites, databases, access controls, CMSes, mail servers, management software system. during a similar context, the standard threshold based approach [4] that classifies a bunch as a sender or receiver as a result of its load is on the far side or below some given lines cannot work. This downside is even additional serious during a cloud context with thousands of hosts wherever, at a stop, a threshold might signal many senders and, at the sequential stop, the quantity of senders will become few dozen or, even worse, stay within the order of lots of however wherever most servers area unit completely different from those of the previous set. the choice concerning that guests is beneficial to migrate from one server to a different is laid low with similar issues if we tend to adopt some threshold-based technique.

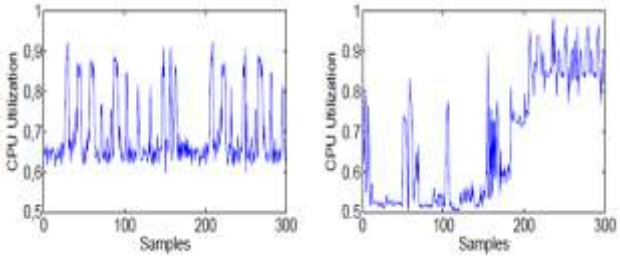


Fig.2 Load equalization Profiles in cloud Architectures

The projected management formula is activated sporadically (typically within the order of few minutes) and, at every stop, it aims at shaping 3 sets: sender hosts, receiver hosts, and migrating guests, wherever their cardinalities area unit denoted as S, R, and G, severally. Let additionally N be the full variety of hosts. we've to ensure that $N \geq S + R$, which the intersection between the set of sender hosts and of receiver hosts is null. The formula is predicated on the subsequent four phases.

Phase 1: Choice Of Sender Hosts the primary action needs the choice of the set of sender hosts that need the migration of a number of their guests. We tend to describe our strategy that's supported the CUSUM models in Section four. Six Sara Casolari et al. the thought is to possess a selective and strong formula so the cardinality S of the set of senders is way smaller than the full variety of hosts that's, $S \ll N$.

Phase 2: Choice Of Guests. Once elite the senders, we've to judge what number and that guests it 's convenient to migrate. to the present purpose, in Section five we tend to propose associate degree formula that 's ready to choose the foremost important guests for every server on the idea of a load trend-based model rather than ancient approaches supported instant or average load measures. Even for this part, the goal is to limit the quantity of guests for every host that ought to migrate, so $G < (N - S)$. If this doesn 't occur once the primary analysis, the guest choice takings iteratively till the constraint is happy. (It is value to look at that no experiment needed associate degree iteration)

Phase 3: Choice Of Receiver Hosts. Once elite the guests that got to migrate, we've to outline the set of receiver hosts. to the present purpose, we tend to don 't propose any specific innovative formula. From our past expertise in different geographically distributed architectures and initial experiments on cloud architectures, we will conclude that the most important risk we wish to avoid could be a dynamic migration that tends to overload some receiver hosts so at the sequential stop a receiver might become a sender. Similar fluctuations devastate system performance and stability. Hence, our plan is to line $R = G$ so every receiver host receives at the most one guest. the chosen receivers area unit the R hosts that exhibit all-time low load computed on the idea of the trend model delineated .

Phase 4: Assignment of guests. The guests elite within the part a pair of area unit allotted to the receivers through a classical greedy formula wherever we start to assign the foremost taxing guests to all-time low loaded hosts. (It is value to look at that in actual cloud architectures there area unit different discipline and application constraints that ought to be happy within the guest migration part. These constraints limit the mixtures of potential assignments to completely different sets therefore reducing the process price of sorting.)

CONCLUSION

Dynamic migrations of virtual machines is changing into a motivating chance to permit cloud infrastructures to accommodate dynamical demands for various varieties of process with heterogeneous workloads and time constraints. nonetheless, there area unit several open problems concerning the foremost convenient alternative concerning once to activate migration, the way to choose guest machines to be migrated, and Dynamic Management the foremost convenient destinations. These classical issues area unit even additional severe in {a very} cloud context characterized by a very sizable amount of hosts. We tend to propose novel algorithms and models that area unit ready to determine simply the important important host and guest devices, by considering the load profile of hosts and therefore the load trend behavior of the guest rather than

thresholds, instant or average measures that area unit generally utilized in literature.

REFERENCES

[1] E. Zohar, I. Cidon, and O. Mokryn, "The power of prediction: Cloud
bandwidth and cost reduction," in *Proc. SIGCOMM*, 2011, pp. 86–97.
[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R.Katz, A. Konwinski,
G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of
cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.
[3] U. Manber, "Finding similar files in a large file system," in *Proc.*
USENIX Winter Tech. Conf., 1994, pp. 1–10.
[4] N. T. Spring and D. Wetherall, "A protocol-independent technique for
eliminating redundant network traffic," in *Proc. SIGCOMM*, 2000, vol.
30, pp. 87–95.
[5] A. Muthitacharoen, B. Chen, and D. Mazières, "A low-bandwidth network
file system," in *Proc. SOSP*, 2001, pp. 174–187

BIOGRAPHY

Mr.B.Mahesh is aM.Tech Scholar in the Computer Science &Engg Department, SreeVidhyanikethanEngg College. He received Bachelor of Technology (C.S.E) degree in 2012 from JNTUA, Ananthapur,Andhra Pradesh, India. His research interests are Computer Networks (wireless Networks), HCI, Algorithms, web 2.0 etc.

Mr.M.V.R.PurnaKumarisa Assistant Professor in the Computer Science &Engg Department, SreeVidhyanikethanEngg College. He received Master of Technology degree from NIT Kerala, India. His research interests are Computer Networks (wireless Networks).

● ● ●

Fusion of MRI and CT Images using DTCWT and SOFM

[1]C.Karthikeyan, [2]Dr.B.Ramadoss
[1]Research Scholar, Jawaharlal Nehru Technological University, Hyderabad, India.
[2]Professor, Dept. of Computer Applications, National Institute of Technology, Trichy, India.
[1]ckarthik2k@gmail.com

Abstract: **Medical imaging is the technique of making visual representations of the inner part of a body of clinical and medical analysis. Medical imaging tries to reveal inner structures covered up by the skin and bones, and in addition to diagnose and treat disease. Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) techniques in medical image has revolutionized for medical diagnosis. To make a medical diagnosis easier and accurate, image fusion combines MRI and CT images into a single image, which contains the relevant information from the original source images. Image fusion technique is based on Dual Tree Complex Wavelet Transform (DT - CWT) and Robust Self Organizing Feature Mapping (Robust SOFM) neural network. DT_CWT decomposes the images in a multi-scale and multi-direction to extract important features. The Robust SOFM neural network is utilized to recognize the complementary features. These features are integrated using a criteria based on activity level. Fused of both MR, CT images are made from a fused feature set. Finally the fused image is attained by executing Inverse Dual Tree Complex Wavelet Transform (IDTCWT). The experimental results show that the proposed algorithm can significantly outperform image fusion technique and has a better Peak Signal to Noise Ratio (PSNR) value as compared with other transforms, DWT, FDCT, NSCT, and DTCWT.**

Keywords: **Dual Tree Complex Wavelet Transform, Robust Self Organizing Feature Mapping.**

I. INTRODUCTION

Multimodality medical images are required to support more accurate clinical data for doctors to manage with medical diagnosis, such as X-ray, CT, MRI, and Magnetic Resonance Angiography (MRA) images [1]. These medical images generally provide correlative and occasionally conflicting data. The CT images can yield information about bones, implants with less distortion, however it cannot identify physiological changes. The MR images can yield information about typical and neurotic soft tissue data; however, it cannot help the bones information. In this situation, fusion of the multimodal medical images only sufficient to provide accurate clinical necessities to the doctors. Thus, the fusion of the medical images is essential and very challenging in research areas.

Image fusion is defined as the methodology of combing multiple input images or their features into a single image without loss of information [2-3]. Thus, the fused image contains a more accurate and more suitable for human visual [4]. The fusion of medical images not only lead to additional clinical information compare with the individual images, but also reduce the storage cost of storing single fused image rather than multi source images. Generally, the image fusion methods can be grouped into three levels, such as pixel, feature, and decision levels.

Image fusion based on Discrete Wavelet Transform (DWT) developed faster. It has good time-frequency
characteristics. The drawback of DWT is that problem of filling missing data occur. Fast Discrete Curvelet Transforms (FDCT) [5] is simpler, faster, and less redundant than DWT. Non-Subsampled Contourlet Transform (NSCT) can be partitioned into two stages incorporates Non-Subsampled Pyramid (NSP) and Non-Subsampled Directional filter bank (NSDFB). The main disadvantage of curvelet method is that it has the disadvantage of poor directional specificity. In this paper, the proposed fusion technique based on DTCWT and robust SOFM is used.

II. DUAL TREE COMPLEX WAVELET TRANSFORM

The DTCWT is an enhancement to the discrete wavelet transform (DWT), with main properties: It is close to shift invariant and directionally selective in two or higher dimensions. It achieves with a redundancy factor of two dimensional signals, which is substantially lower than DWT. The DWT is predominantly acquired by a perfect reconstruction (PR) filter bank (FB) on its lowpass output, and decomposes a discrete-time signal according to octave-band frequency decomposition. Though, the DWT is not only far

away from shift invariant, but also not generate a geometrically oriented decomposition in multiple dimensions. Additionally the wavelet FB used by the DWT, the DT-CWT [7-8] uses a second wavelet FB. Particularly, the second wavelet FB is decomposed so that itsimpulse responses are almost same to the discrete Hilbert transforms (DHT) of the first wavelet FB. At that point, the first FB as the real part and the second FB as the imaginary part of a complex transform, The frequency response of each one channel is important for the DT-CWT to posses its desirable properties.

As compared with real valued DWT, DT-CWT has two advantages. 1. It has better edge representation and

2. Approximate shift-invariant property. The DT-CWT utilizes two genuine DWTs; 1. The real part of the DWT transforms 2. The imaginary part of the DWT transforms. The two real DWTs utilize two separate sets of filters, with each one fulfilling the PR conditions. These sets of filters are designed together; hence the overall tra nsform is almost analytic.

III. Robust Self-Organizing Feature Map (Robust SOFM)

SOFM is based on Artificial Neural Network (ANN), the nodes of which become particularly tuned to different input signal patterns through an unverified learning process. Any dimension of the SOFM input signals can be transformed into a one-dimensional or a two-dimensional discrete grid. And it depends upon the human cerebral cortex function imitation. SOFM has the characteristics of automatic identification and classification. The SOFM neural network [6] algorithm is an unsupervised learning, can automatically classify the samples and achieve better classification results in less number of trained samples. Most importantly, as compare with SOFM, Robust SOFM networks can parallel processing to improve the speed of the operation twice.

The SOFM neural network structure is demonstrated in Fig. 1. SOFM network consists of two layers: 1. Lower layer for input, 2. Upper for the output layer.

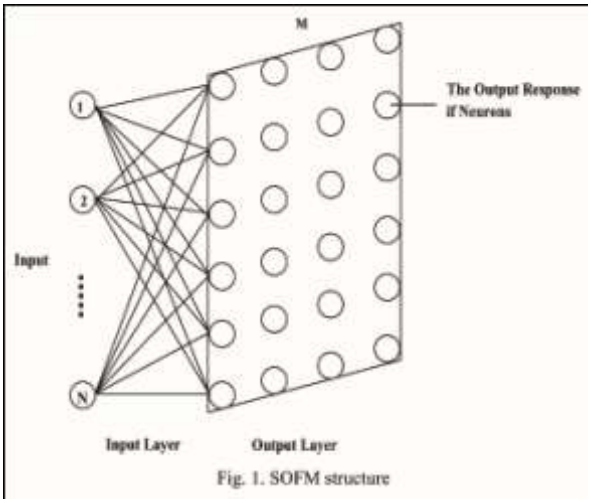


Fig. 1. SOFM structure

SOFM network is connected fully, each one of input neuron nodes is connected to all output neuron nodes, Whenever the Euclidean Distance (ED) is the minimum, input neurons, input vectors and the weights of output neurons are activated. At this point the connection weights are modified by the network at termination condition. The output neuron is called as the competitive winning neuron. Although in a few regions SOFM algorithm gets problematic. The SOFM network convergence speed is slow.

In SOFM, few neurons in the output layer weights regularly win probability and adjust, but few neurons weights infrequently adjust effetely. And the network's learning process and results are unfair at different initial conditions and input samples. The main advantage is that the following features are dealing with the unidentified image. Robust SOFM is used to enhance the convergence speed of the SOFM neural network and by utilizing of limited samples train neural network and also enhance the classification accuracy of neural network. Hence, Robust SOFM successfully reduces the calculation time of SOFM. This Robust SOFM algorithm enhances the compression ratio as well as reducing the search range and computation time.

IV. FUSION TECHNIQUE BASED ON DTCWT AND ROBUST SOFM

The following steps describe the proposed fusion technique based on DTCWT and Robust SOFM algorithm.

1. Select Medical images; MRI and CT. And obtained images are progressively decomposed by DT-CWT. Select Medical images; MRI and CT. The intensity based image registration done in MRI to position same coordination. And obtained

images are progressively decomposed by DT-CWT.

2. DT-CWT generates sets of coefficients (Approximation and Details) at each level of decomposition. The coefficients at each level represent sets of the image feature vector.
3. The robust SOFM neural network is utilized to recognize and extract the features. This can be done by training and simulating the network for the resultant coefficients (approximation and detail) of each level of MR and CT images.
4. By using the fusion technique, merge the approximation and detail coefficients.
5. Apply IDTCWT to the fusion result to get the final fused image.

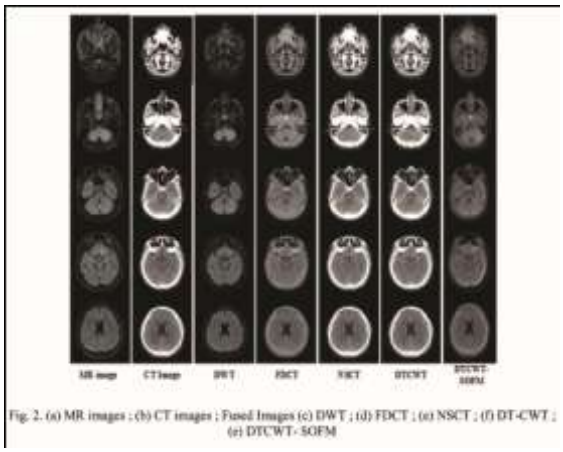


Fig. 2. (a) MR images ; (b) CT images ; Fused Images (c) DWT ; (d) FDCT ; (e) NSCT ; (f) DT-CWT ; (e) DTCWT- SOFM

V. RESULT AND DISCUSSION

Quantitative evaluation results demonstrate that the proposed method gives better performance for multiclass object classification in comparison to other state-of-the-art methods. Experiment results show that the PSNR of the fusion method based on DT-CWT with Robust SOFM is better than the fusion methods of DWT, FDCT, NSCT, and DTCWT, shown in Fig. 2. Table 1 describes that the PSNR of fused images are compared with the PSNR of both MR and CT images of different transform methods, shown in Fig. 3 and Fig. 4 respectively.

Peak Signal to Noise Ratio (PSNR): To get better fusion results, the PSNR value will be high.

$$PSNR = 20\log_{10}\left[\frac{L^2}{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_r(i,j)-I_f(i,j)\right)^2}\right],$$

Where L = No. of gray levels in the image

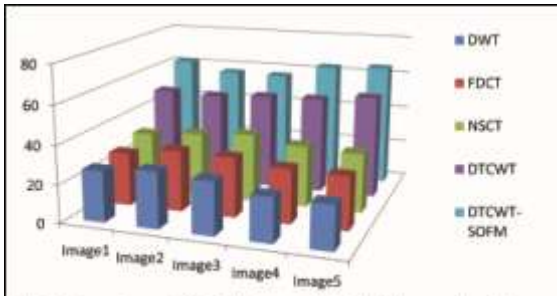


Fig. 3. Comparison on PSNR of different methods for MR Image vs Fused Image

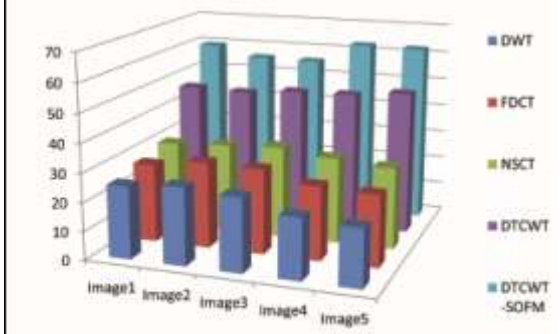


Fig. 4. Comparison on PSNR of different methods for CT Image vs Fused Image

Table. 1. Comparison between DTCWT Fusion Techniques and Proposed Method

Image Set	PSNR : MR Image vs Fused Image					PSNR : CT Image vs Fused Image				
	DWT	FDCT	NSCT	DTCWT	DTCWT-SOFM	DWT	FDCT	NSCT	DTCWT	DTCWT-SOFM
1	26.73	27.84	31.32	49.42	61.64	25.32	27.53	30.14	46.88	59.47
2	29.36	31.9	33.73	48.57	56.74	27.05	30.08	31.16	46.28	55.87
3	27.74	31.32	35.78	50.31	56.63	25.84	29.65	32.13	47.89	55.62
4	23.28	28.49	32.27	51.02	63.36	21.35	25.91	30.15	48.2	62.73
5	22.97	27.77	31.08	53.96	64.73	20.14	25.05	28.61	50.01	62.42

CONCLUSION

In medical image processing applications, specifically in MRI and CT images, the edge preserve is an important in complementary details of input images. DT-CWT is implemented using two real wavelet transforms. Robust SOFM networks can be parallel processed to improve the speed of the operation and for the characteristics of automatic identification. Thus, the fused image contains a more accurate and more suitable for human visual. Through the results, as compared with the DWT, FDCT, NSCT and DTCWT fusion methods we found that image fusion method DTCWT with Robust SOFM gives better PSNR value.

REFERENCES

[1] Maes F, Vandermeulen D. Medical image registration using mutual information. *IEEE Proceedings*, 2003; 1699–1721.

[2] VS Petrovic, CS Xydeas. Gradient-based multiresolution image fusion. *IEEE Transactions on Image Processing*. 2004; 228–237.

[3] Z Zhang, RS Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *IEEE proceedings*. 1999; 1315–1326.

[4] ST Shivappa, BD Rao, MM Trivedi. An iterative decoding algorithm for fusion of multimodal information. *EURASIP Journal on Advances in Signal Processing*. 2008.

[5] Emmanuel Candes, Laurent Demanet, David Donoho and Lexing Ying. Fast Discrete Curvelet Transforms. 2006; 1-44.

[6] Teuvo Kohonen. The self-organizing map. *IEEE Proceedings*. 1990;1464-1480.

[7] Sruthy.S, Parameswaran, L. Image Fusion Technique using DT-CWT. *International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*. 2013; 160-164,

[8] Shiyvan Yang, Huimin Lv, Yvjie Li, Seiichi Serikawa. Proposal of A Multi-Frame Images Fusion Model On Dual Tree Complex Wavelet Transform Domain. *IEEE international Conference on Machine Learning and Cybernetics*. 2013. ; 952-956.

Real time traffic control using occupancy estimation with camera

[1]Swaathikka Karthikeyan, [2]Kiran Sudhir, [3]Jerry George Thomas
[1] [2]Dept. of Computer Science and Engineering, [3]Dept. of Electronics and Communications Engineering
[1] [2] [3]SSN College of Engineering
[1] [2] [3]Chennai, India.
[1]swaathikka@gmail.com, [2]kiransudhir95@gmail.com, [3]jerrythomas.ssn@gmail.com

Abstract: **The objective was to create an integrated system which mediates traffic based on traffic density and real time movement of vehicles. Traffic density is quantified from a live video relay processed by MATLAB. The data is uploaded to a main server which consolidates data from different points and uses it to take real time decisions to regulate the flow of traffic. The system will prove to be an improvement from the present systems and provide for an efficient means to control traffic.**

Keywords: **Traffic control, real time, traffic density, quantify, image processing, Internet of Things (IoT)**

I. INTRODUCTION

With the number of vehicles on Indian roads doubling every year, there exists a need for a better traffic management system especially in urban areas. Present systems use preset timers to control traffic signals or in some metropolitan cities, infra red sensors are used to assess traffic. This system is inadequate due to its assumption that traffic patterns are static rather than dynamic. Very often we see streets with no traffic given the green signal while congested roads are made to wait resulting in traffic build up. In our system we assume that traffic patterns are dynamic and hence we identify the need to make real time decisions to control traffic. Image processing techniques are adopted to quantify the traffic on the street and to classify them into three categories – High low and No traffic which is uploaded to the server and used to dictate the flow of traffic.

II. DESCRIPTION

A. Processing a static image

The aim of the image processing module is to quantify the traffic on the street. This is done by edge detection using canny algorithm. We chose canny algorithm due to its robust nature and higher accuracy in detecting all the edges in the image. It also has a low error rate and to a fair extent, does not detect false edges. This is of great importance since the edges are going to be used to quantify the amount of traffic in our system. The process of Canny edge detection [1] involves smoothing of the image by a low pass Gaussian Filter, followed by non-maximum suppression and hysteresis to identify the strong edges. The higher accuracy of canny algorithm compared to other algorithms is demonstrated in the pictures below. [2]

Fig. 1(a) Prewitt Algorithm
Fig. 1(b) Canny algorithm

our system, we use a reference image of an empty street which is stored and taken as an input. A real time image of the street with traffic is captured and taken as another input using the same function. Both images are resized and converted to grayscale to improve the time required for computation. Another reason for converting the image to grayscale is to increase the accuracy of edge detection since it is easier to detect edges in grayscale. Edge detection using canny algorithm is applied on both using the *edge()* function in MATLAB.

Further, if the need arises, we can increase the accuracy of edge detection by specifying the *threshold* and *sigma* attributes of the algorithm. *Threshold* specifies the sensitivity threshold of the

algorithm and the function will ignore all edges below this threshold. *Sigma* specifies the standard deviation of the filter.

B. Processing a live video relay

The next step was to perform the same image processing on a live feed. For this purpose two webcams were used. It was found that higher the resolution of the webcam, higher was the accuracy indicating is a direct relationship with the sensitivity of the webcam. Objects were created for each webcam and their properties were set. This involved specifying the *FramesPerTrigger*, *TriggerRepeat* and *ReturnedColorSpace* attributes. Once the object is initalised, MATLAB starts capturing the video. To obtain a single image from the video, the *getsnapshot()* function is used. This provides us the necessary image for processing.

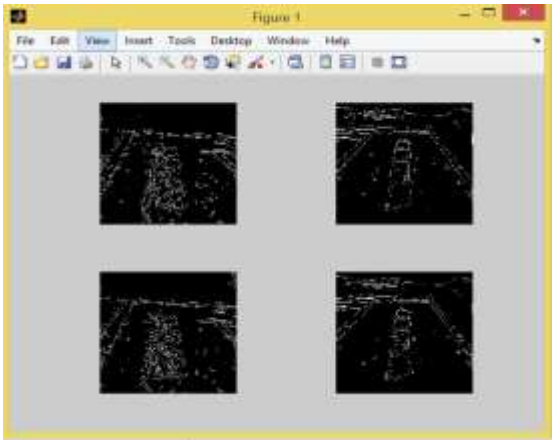


Fig 2(a) The top row contains reference images of an empty street

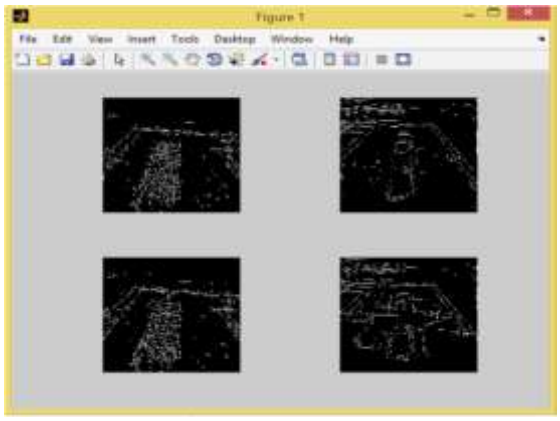


Fig 2(b) The bottom right image is the processed image of a road with traffic on it.

C. Quantifying the density of traffic

Once edge detection is performed on the images, they are compared. First the number of white dots and their positions in the reference image is computed and then the number of white points and their positions in the real time image with traffic is calculated. This is compared with that of the reference image. Any new vehicle will introduce new edges in the picture. The new edges will show up as a dissimilarity during comparison. This is quantified in the form of percentage. The traffic level is decided based on this percentage. We have assigned a value of 0 for 0-30%, a value of 1 for 30%-60% and a value of 2 for any percentage above 60%.

Fig 3 : Real time quantified traffic output

D. Uploading MATLAB output to Remote Database

MATLAB writes the output traffic densities into text files corresponding to the roads. A java code connects to a hosted Remote Database using the following connection code [3].

CONNECTION CODE

```
con=DriverManager.getConnection("jdbc:mysql://
216.12.194.32/acrossim_traffic",  "acrossim_user",
"useruser");
```

The code also reads the data from the files using *InputStreamReader [4]*, and inserts this data into the table corresponding to the road. The text file ‘data1’ corresponds to traffic level in road 1 and ‘data2’ corresponds to traffic level in road 2. The data are inserted into tables info and info2.

CODE FOR CURRENT TIME

```
Date dnow= new Date();
SimpleDateFormat dform= new
SimpleDateFormat("yyyy- MM-dd hh:mm:ss");
t=dform.format(dnow);
```

READING FROM FILE

```
//For road 1
InputStream in =new FileInputStream("data1.txt");
BufferedReader r = new BufferedReader(new
InputStreamReader(in));
int lev = r.readLine();
//For road 2
InputStream in2 =new
FileInputStream("data2.txt");
BufferedReader r2 = new BufferedReader(new
InputStreamReader(in2));
int lev2 = r2.readLine();
```

INSERTING INTO TABLES

```
//Table1 – info : id , time level
Statement s1;
String state="insert into acrossim_traffic.info
(time,level) values ('"+t+"','"+lev+"')";
s1.execute(state);
//Table2- info2 : id,time,level
String state="insert into acrossim_traffic.info2
(time,level) values ('"+t+"','"+lev2+"')";
s1.execute(state);
```

The tables each have three columns- id, time and level. 'id' is a unique identifier that is generated automatically for every insertion. The 'time' value is assigned using the system time and storing it in a format suitable for the *timestamp* data type. 'level' value is obtained from the text file written into by MATLAB. These values are inserted into the database for the corresponding road.

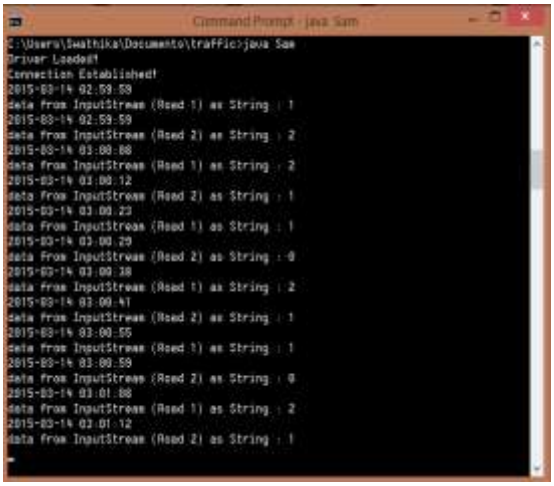


Fig 4 : Uploading data into database

id	time	level
78	2015-03-14 10:13:45	1
79	2015-03-14 10:14:00	0
80	2015-03-14 10:14:14	1
81	2015-03-14 10:14:34	0
82	2015-03-14 10:14:53	2
83	2015-03-14 10:15:07	1
84	2015-03-14 10:15:31	0
85	2015-03-14 10:15:56	1
86	2015-03-14 10:16:12	2
87	2015-03-14 10:16:45	2

info : Lane 1

id	time	level
35	2015-03-14 10:16:45	0
34	2015-03-14 10:16:00	1
33	2015-03-14 10:15:34	0
32	2015-03-14 10:15:10	1
31	2015-03-14 10:14:56	0
30	2015-03-14 10:14:38	1
29	2015-03-14 10:14:19	2
28	2015-03-14 10:14:03	1
27	2015-03-14 10:13:50	0
26	2015-03-14 10:13:25	1

info2 : Lane 2

Fig 5 : Uploaded table

E. Using an Arduino to control the traffic

An Arduino board was used to control the traffic signal for the junction corresponding to the level so as to ease and equalize the flow. The Arduino receives the current traffic level output of the two roads from MATLAB through serial port and compares their values, and changes the traffic light accordingly.

Setup

The system was developed to consider a two road setup with cameras monitoring each road. The video feed from this recording was processed to obtain the traffic level at the road. An Arduino board is used to control the traffic lights at the junction. The software is run on a laptop for this prototype.

Fig 6 : Setup and traffic control

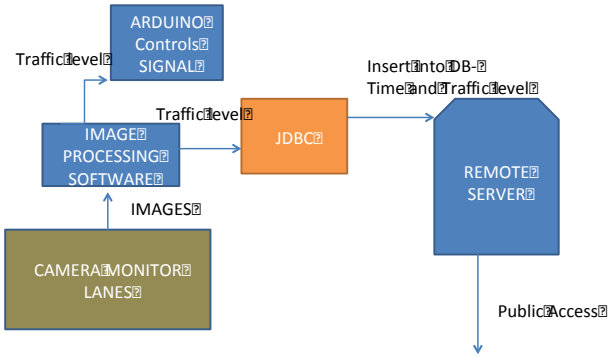


Fig 7 : Working system representation

CONCLUSION

The above system overcomes several disadvantages faced by the current traffic signals being implemented today. As the traffic is checked and compared in real time, unwanted stalls are avoided and traffic can be controlled very effectively. This would lead to significant improvement of traffic especially during the busy hours of the day and avoid congestion to a great extent. As the traffic density is uploaded into a remote database in real time, it can be used by the public to find out which routes are crowded and plan their trips accordingly, choosing alternative routes if necessary. The above system can be extended over the entire road network of a city, by using data from interconnected roads. Furthermore, this system has a lot of scope to be further improved by implementing machine learning techniques to intelligently control the traffic.

REFERENCES

[1] Canny, John, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol. PAMI-8, No. 6, 1986, pp. 679-698.

[2] http://in.mathworks.com/help/images/ref/edge.html?searchHighlight=edge

[3] http://www.tutorialspoint.com/jdbc/jdbc-db-connections.htm
http://docs.oracle.com/javase/7/docs/api/java/io/InputStreamReader.html