IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 2 Issue 9

September 2011

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI

www.ijacsa.thesai.org

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS

A Publication of
The Science and Information Organization

IJACSA Editorial

From the Desk of Managing Editor…

It is a pleasure to present our readers with the September 2011 Issue of International Journal of Advanced Computer Science and Applications (IJACSA).

The renaissance stimulated by the field of Computer Science is generating multiple formats and channels of communication and creativity. IJACSA is one of the most prominent publications in the field and engaging the ubiquitous spread of subject knowledge with effectiveness in all classes of audience. Nevertheless, the promise of increased engagement requires that we consider how this might be accomplished, delivering up-to-date and authoritative coverage of advanced computer science and applications.

The journal has a wide scope ranging from the many facets of methodological foundations to the details of technical issues and the aspects of industrial practice. It includes articles related to research findings, technical evaluations, and reviews. In addition it provides a forum for the exchange of information on all aspects.

The editorial board of the IJACSA consists of individuals who are committed to the search for high-quality research suitable for publication. These individuals, working with the editor to achieve IJACSA objectives, assess the quality, relevance, and readability of individual articles.

The contents include original research and innovative applications from all parts of the world. This interdisciplinary journal has brought together researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to computer science and application with its convergence strategies, and to disseminate the most innovative research. As a consequence only 31% of the received articles have been finally accepted for publication.

Therefore, IJACSA in general, could serve as a reliable resource for everybody loosely or tightly attached to this field of science.

The published papers are expected to present results of significant value to solve the various problems with application services and other problems which are within the scope of IJACSA. In addition, we expect they will trigger further related research and technological improvements relevant to our future lives.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications.

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
Volume 2 Issue 9, September 2011
editorijacsa@thesai.org
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

IJACSA Reviewer Board

CONTENTS

Evaluation of the Segmentation by Multispectral Fusion Approach with Adaptive Operators: Application to Medical Images

Lamiche Chaabane
Department of Computer Science
University of M'sila
Algeria

Moussaoui Abdelouahab
Department of Computer Science
University of Setif
Algeria

Abstract— **With the development of acquisition image techniques, more and more image data from different sources of image become available. Multi-modality image fusion seeks to combine information from different images to obtain more inferences than can be derived from a single modality. In medical imaging based application fields, image fusion has emerged as a promising research area since the end of the last century. The paper presents the evaluation of the segmentation of MR images using the multispectral fusion approach in the possibility theory context . Some results are presented and discussed.**

Keywords- image fusion; possibility theory; segmentation; MR images.

I. INTRODUCTION

Magnetic resonance (MR) imaging has been widely applied in biological research and diagnostics, primarily because of its excellent soft tissue contrast, non-invasive character, high spatial resolution and easy slice selection at any orientation. In many applications, its segmentation plays an important role on the following sides: (a) identifying anatomical areas of interest for diagnosis, treatment, or surgery planning paradigms; (b) preprocessing for multimodality image registration; and (c) improved correlation of anatomical areas of interest with localized functional metrics [1].

Segmenting MR images has been found a quite hard problem due to the existence of image noise, partial volume effects, the presence of smoothly varying intensity inhomogeneity, and large amounts of data to be processed. To handle these difficulties, a large number of approaches have been studied, including fuzzy logic methods [2], neural networks [3], Markov random field methods with the maximum expectation [4], statistical methods [4], and data fusion methods [5], to name a few.

As one typical data fusion problem, the segmentation of multi-modality brain MR images aims at achieving improved segmentation performance by taking advantage of redundancy and complementariness in information provided by multiple sources. There have existed many data fusion methodologies, which are capable of reasoning under various types of uncertainty. Typical ones include probability theory based approaches, possibility theory based approaches, and Dempster-Shafer evidence theory based approaches [6].

Traditionally probabilities theory was the primary model used to deal with uncertainty problems, but they suffer from drawbacks which are still a matter of discussion. Whereas the Dempster-Shafer theory also allows to representing these two natures of information using functions of mass but the set of operators used by this theory in fusion step is very restricted.

Alternative to this approach is the possibility theory where uncertainty and imprecision are easily modeled, in this article we will focus on this last one for two essential reasons : this theory allows to combining information coming from various sources by the use a wide range of available combination operators. In addition, this theory seems to us the most adapted to the considered problem in the modeling step [1][6].

In this work we aim to evaluate the segmentation of the human brain tissues using a multispectral fusion approach. This approach consists of the computation of fuzzy tissue maps in each of two modalities of MR images namely T2 and PD as an information source, the creation of fuzzy maps by a combination operator and a segmented image is computed in decision step.

The remainder of the paper is organized as follows : In section II, some previous related works are briefly cited. section III describes the FPCM algorithm. In section IV, we present the principals of possibility theory reasoning . Section V outlined the fusion process. Steps of fusion in medical image processing are presented in section VI. Section VII present some numerical experiments. We finally provide concluding remarks in Section VIII.

II. PREVIOUS RELATED WORKS

Many works have been done in the field of fuzzy information fusion in the literature. A brief review of some of them is presented in this section. Waltz [10] presented three basic levels of image data fusion : pixel level, feature level and decision level, which correspond to three processing architectures. I. Bloch [1] have outlined some features of Dempster-Shafer evidence theory, which can very useful for medical image fusion for classification, segmentation or recognition purposes. Examples were provided to show its ability to take into account a large variety of situations. Registration-based methods are considered as pixel-level fusion, such as MRI-PET (position emission tomography) data fusion[11]. Some techniques of knowledge-based segmentation can be considered as the feature-level fusion such as the methods proposed in [15].

Some belief functions, uncertainty theory, Dempster-Shafer theory are often used for decision-level fusion such as in [13]. In [16], I. Bloch proposed an unified framework of information fusion in the medical field based on the fuzzy sets, allow to represent and to process the numerical data as well as symbolic systems, the fuzzy sets theory is applied to three levels: at the low level to treat the basic numerical information contained in the images, as well as possible ambiguity between the classes; on the level object, to represent objects or structures in the images such as a fuzzy objects. at the higher level, to take into account a structural information and some characteristics as the distance, adjacency, and the relative position between objects.

V. Barra and J. Y. Boire [8] have described a general framework of the fusion of anatomical and functional medical images. The aim of their work is to fuse anatomical and functional information coming from medical imaging, the fusion process is performed in possibilistic logic frame, which allows for the management of uncertainty and imprecision inherent to the images. They particularly focus on the aggregation step with the introduction of a new class of operators based on information theory and the whole process is finally illustrated in two clinical cases : the study of Alzheimer's disease by MR/SPECT fusion and the study of epilepsy with MR/PET/SPECT. The obtained results was very encouraging.

V. Barra and J. Y. Boire [14] proposed a new scheme of information fusion to segment intern cerebral structures. The information is provided by MR images and expert knowledge, and consists of constitution, morphological and topological characteristics of tissues. The fusion of multimodality images is used in [12]. In [7], the authors have presented a framework of fuzzy information fusion to automatically segment tumor areas of human brain from multispectral magnetic resonance imaging (MRI) such as T1-weighted, T2-weighted and proton density (PD) images; in this approach three fuzzy models are introduced to represent tumor features for different MR image sequences. They allow to create corresponding fuzzy feature space of tumor. All the t-norm or fuzzy intersection operators can be used as fusion operators for this fuzzy features. the geometric mean is chosen using experiments allowing us to take correctly into account the three fuzzy spaces in a simple way. The fuzzy region growing is used to improve the fused result.

Maria del C. and al [9] proposed a new multispectral MRI data fusion technique for white matter lesion segmentation, in that a method is described and comparison with thresholding in FLAIR images is illustrated. Recently, The authors in [38] have presented a new framework of fuzzy information fusion using T2-weighted and proton density (PD) images to improve the brain tissue segmentation.

III. THE FPCM ALGORITHM CLUSTERING

Clustering is a process of finding groups in unlabeled dataset based on a similarity measure between the data patterns (elements) [16]. A cluster contains similar patterns placed together. One of the most widely used clustering methods is the FPCM algorithm. The FPCM algorithm solves the noise sensitivity defect of Fuzzy C-Means algorithm and overcomes the problem of coincident clusters of Possibilistic C-means algorithm. Given a set of N data patterns $X=\{x_1, x_2, x_3, \ldots, x_n\}$ the Fuzzy Possibilistic C-Means (FPCM) clustering algorithm minimizes the objective function [31][32]:

$$J(B,U,T,X) = \sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij}^{m} + t_{ij}^{\lambda})d^2(x_j, b_i) \tag{1}$$

Where x_j is the j-th P-dimensional data vector, b_i is the center of cluster i, m>1 is the weighting exponent, $\lambda\in[3,5]$ is the typicality exponent, $d^2(x_j,b_i)$ is the Euclidean distance between data x_j and cluster center b_i, $[U]_{CxN}$ is the fuzzy matrix and $[T]_{CxN}$ is the typicality matrix.

The minimization of objective function $J(B,U,T,X)$ can be brought by an iterative process in which updating of membership degrees u_{ij} , typicality degrees t_{ij} and the cluster centers are done for each iteration by :

$$u_{ij} = \left[\sum_{k=1}^{C}\left(\frac{d^2(x_j,b_i)}{d^2(x_j,b_k)}\right)^{2/(m-1)}\right]^{-1}. \tag{2}$$

$$t_{ij} = \left[\sum_{k=1}^{C}\left(\frac{d^2(x_j,b_i)}{d^2(x_j,b_k)}\right)^{2/(\lambda-1)}\right]^{-1}. \tag{3}$$

$$b_i = \frac{\sum_{k=1}^{N}(u_{ik}^{m} + t_{ik}^{\lambda})x_k}{\sum_{k=1}^{N}(u_{ik}^{m} + t_{ik}^{\lambda})}. \tag{4}$$

Where :

$$\forall i \in \{1..C\}, \quad \forall j \in \{1..N\} \quad \begin{cases} u_{ij} \in [0,1] \\ 0 < \sum_{i=1}^{N} u_{ij} < N \end{cases}. \tag{5}$$

$$\forall j \in \{1..N\} \quad \sum_{i=1}^{C} u_{ij} = 1 \cdot \tag{6}$$

$$\forall i \in \{1..C\} \quad \sum_{j=1}^{N} t_{ij} = 1 \cdot \tag{7}$$

The algorithm of the FPCM consists then of the reiterated application of (2), (3) and (4) until stability of the solutions.

IV. THE POSSIBILITY THEORY

Possibilistic logic was introduced by Zadeh (1978) following its former works in fuzzy logic (Zadeh, 1965) in order to simultaneously represent imprecise and uncertain knowledge. In fuzzy set theory, a fuzzy measure is a representation of the uncertainty, giving for each subset Y of the universe of discourse X a coefficient in [0,1] assessing the degree of certitude for the realization of the event Y. In possibilistic logic, this fuzzy measure is modeled as a measure of possibility Π satisfying:

$$\Pi(X)=1 \quad et \quad \Pi(\phi)=0$$

$$(\forall(Y_i))\Pi(\bigcup_i Y_i) \;=\; \underset{i}{Sup}\;\; \Pi(Y_i)$$

An event Y is completely possible if $\Pi(Y)=1$ and is impossible if $\Pi(Y)=0$. Zadeh showed that Π could completely be defined from the assessment of the certitude on each singleton of X. Such a definition relies on the definition of a distribution of possibility π satisfying :

$$\pi : X \rightarrow [0,1]$$
$$x \rightarrow \pi(x) / \underset{x}{Sup}\{\pi(x)=1\}$$

Fuzzy sets F can then be represented by distributions of possibility, from the definition of their characteristic function μ_F :

$$(\forall x \in X)\mu_F(x)=\pi(x)$$

Distributions of possibility can mathematically be related to probabilities, and they moreover offer the capability to declare the ignorance about an event. Considering such an event A (e.g., voxel v belongs to tissue T, (where v is at the interface between two tissues), the probabilities would assign $P(A)=P(\overline{A})=0.5$, whereas the possibility theory allows fully possible $\Pi(A)=\Pi(\overline{A})=1$. We chose to model all the information using distributions of possibility, and equivalently we represented this information using fuzzy sets [22].

The literature classically distinguishes three modes for combination of uncertainty and imprecise information in a possibility theory framework [27] :

The conjunction: gather the operators of t-norms (fuzzy intersection), this mode of combination must be used if measurements are coherent, i.e. without conflict.

The compromise: gather the median operator and some average operators, it must be used when measurements are in partial conflict.

The Disjunction: gather the operators of t-conorms (fuzzy union), it must be used when measurements are in disaccord, i.e. in severe conflict.

In introduction, we underlined the inopportunity to combining information in a fixed mode: if observations are in accord, it is legitimate to combine them in a conjunctive mode or compromise in order to extract a more relevant information. But if a serious conflict appears, it is better to combining in a disjunctive mode. For example, if two measurements of the same parameter prove completely different, it is not judicious to make an average of it, better is worth to say than one or the other is true [28].

V. The Fusion Process and Type of Architectures

A general information fusion problem can be stated in the following terms : given l sources S_1, S_2,...S_l representing heterogeneous data on the observed phenomenon, take a decision d_i on an element x, where x is higher level object extracted from information, and D_i belongs to a decision space $D=\{d_1, d_2, d_3,..., d_n\}$ (or set of hypotheses). In numerical fusion methods, the information relating x to each possible decision d_i according to each source S_j is represented as a number M_{ij} having different properties and different meanings depending on the mathematical fusion framework. In the centralized scheme , the measures related to each possible decision i and provided by all sources are combined in a global evaluation of this decision, taking the form, for each i : $M_i = F(M_{i1}, M_{i2}, M_{i3}, ..., M_{in})$, where F is a fusion operator. Then a decision is taken from the set of M_i, $1 \leq i \leq n$. in this scheme, no intermediate decision is taken and the final decision is issued at the end of the processing chain. In decentralized scheme decisions at intermediate steps are taken with partial information only, which usually require a difficult control or arbitration step to diminish contradictions and conflicts [6][8].

The three-steps fusion can be therefore described as :

- Modeling of information in a common theoretical frame to manage vague, ambiguous knowledge and information imperfection. In addition, in this step the M_{ij} values are estimated according to the chosen mathematical framework.
- Combination : the information is then aggregated with a fusion operator F. This operator must affirm redundancy and manage the complementarities and conflicts.
- Decision : it is the ultimate step of the fusion, which makes it possible to pass from information provided by the sources to the choice of a decision d_i.

VI. Data Fusion in Image Processing Using Possibility Theory

A. Modeling Step

In the framework of possibility theory and fuzzy sets [17][18][19], the M_{ij}'s represent membership degrees to a fuzzy set or possibility distribution π , taking the form for each decision d_i and source S_i :. $M_{ij}=\pi_j(d_i)$. Particularly, in our study this step consists in the creation of WM, GM, CSF and background (BG) fuzzy maps for both T2 and PD images using the FPCM algorithm then $u_{ij}=\pi_j(d_i)$

B. Fusion step

For the aggregation step in the fusion process, the advantages of possibility theory rely in the variety of combination operators, which must affirm redundancy and manage the complementarities. And may deal with heterogeneous information [20][21][22]. It is particular interest to note that, unlike other data fusion theories like Bayesian or

Dempster-Shafer combination, possibility theory provides a great flexibility in the choice of the operator, that can be adapted to any situation at hand [6]. If $\pi_T^{T2}(v), \pi_T^{PD}(v)$ are the memberships of a voxel v to tissue T resulting from step 1 then a fusion operator F generate a new membership value $\pi_T(v)=F(\pi_T^{T2}(v),_T^{PD}(v))$ and can managing the existing ambiguity and redundancy. The possibility theory propose a wide range of operators for the combination of memberships. I. Bloch [24] classified these operators in three classes defined as:

- Context independent and constant behavior operators (CICB);

- Context independent and variable behavior operators (CIVB);

- Context dependent operators (CD).

For our MR images fusion, we chose a context-based conjunctive operator because in the medical context, both images were supposed to be almost everywhere concordant, except near boundaries between tissues and in pathologic areas [20]. In addition, the context-based behavior allowed to take into account these ambiguous but diagnosis–relevant areas. Then three operators of this class are chosen [23][24]:

FOP1 : $\pi_T(v) = \min(\pi_T^{T2}(v), \pi_T^{PD}(v)) + 1 - h$

FOP2 : $\pi_T(v) = \max\left(\frac{\min(\pi_T^{T2}(v), \pi_T^{PD}(v))}{h}, 1 - h\right)$

FOP3 : $\pi_T(v) = \min\left(1, \frac{\min(\pi_T^{T2}(v), \pi_T^{PD}(v))}{h} + 1 - h\right)$

Where h is a measure of agreement between π_T^{T2} and π_T^{PD} :

$$h = 1 - \sum_{v \in Image} \left|\pi_T^{T2}(v) - \pi_T^{PD}(v)\right| / \left|Image\right|$$

C. Decision step

A segmented image was finally obtained using the four maps computed in step 2 by assigning to the tissue T any voxel for which it had the greatest degree of membership (i.e maximum of possibility rule)[6][23].

The general algorithm using for fusion process can be summarized as follows :

General algorithm

Modeling of the image
For i **in** {T2,PD} **do**
FPCM (i) { *Computation of membership degrees for both images T2 and PD*}
End For
Fusion
Possibilistic fusion {*Between each class of T2 image and the same one of PD image using FOP1, FOP2 or FOP3 operator*}
Decision
Segmented image {*maximum of possibility rule*}

It should be noted that the stability of this algorithm depend to the stability of the algorithm used in the modeling step[31][32].

VII. Experimental Results

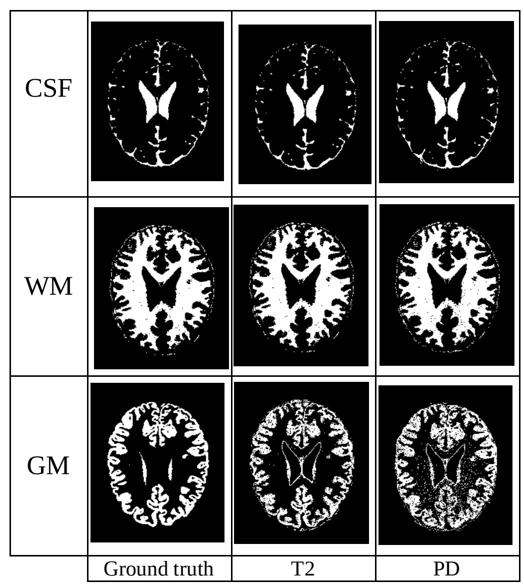
Since the ground truth of segmentation for real MR images is not usually available, it is impossible to evaluate the segmentation performance quantitatively, but only visually. However, Brainweb[35] provides a simulated brain database (SBD) including a set of realistic MRI data volumes produced by an MRI simulator. These data enable us to evaluate the performance of various image analysis methods in a setting where the truth is known [34][35][36].

to have tests under realistic conditions, one volume was generated with a thickness of 1 mm and a level of noise of 3%. We fixed at 20% the parameter of heterogeneity.

The results of each step of fusion on a noisy 94th brain only slice are shown in figure 1, 2 and 3. This noisy slice was segmented into four clusters: background, CSF, white matter, and gray matter using FPCM algorithm, however the background was neglected from the viewing results.

94th simulated T2 slice 94th simulated PD slice

(a)



(b)

Figure 1. (a) Simulated T2, PD images illustrate the fusion. (b) Fuzzy maps of CSF, WM and GM obtained by FPCM algorithm accompagnied by the ground truth for T2 and PD image.

The fused maps produced in fusion step using FOP1, FOP2 and FOP3 operators respectively are presented in figure 2 below :

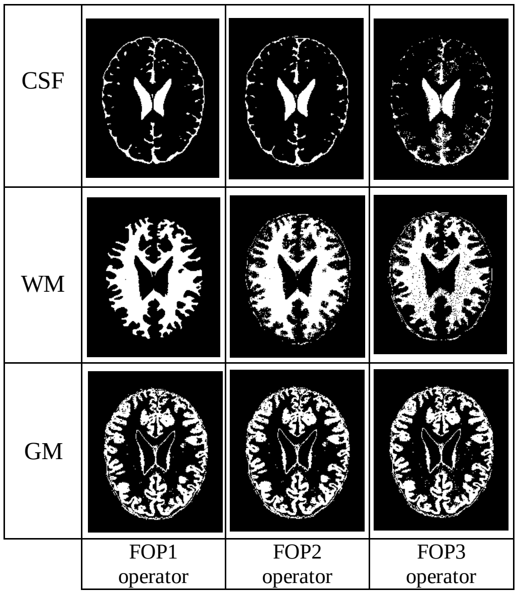


Figure 2. Fused maps obtained with different operators.

And the results of final segmentation with the different operators FOP1, FOP2 and FOP3 are shown in figure 3 below.

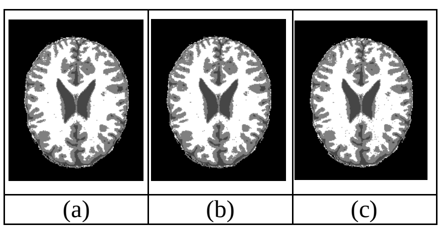


Figure 3. Final segmented images obtained by fusion. (a) FOP1 operator (b) FOP2 operator, (c) FOP3 operator.

The fusion with FOP1 operator improve significantly the CSF map of T2 and PD image.

The WM fused map is strongly improved compared to that obtained by the T2 only and the PD only.

Information in GM fused map with FOP1 operator is reinforced in area of agreement (mainly in the cortex). And the fusion showed a significant improvement and reduces the effect of noise in images.

To compare the performance of these various final segmentation produced by different operators, we compute different coefficients reflecting how well two segmented volumes match. We use a different performance measures[38]:

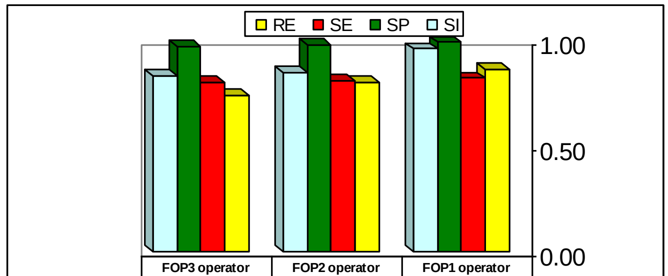
$$Overlap(RE) = \frac{TP}{TP + FN + FP}.$$

$$Sensitivity(SE) = \frac{TP}{TP + FN}.$$

$$Specificity(SP) = \frac{TN}{TN + FP}.$$

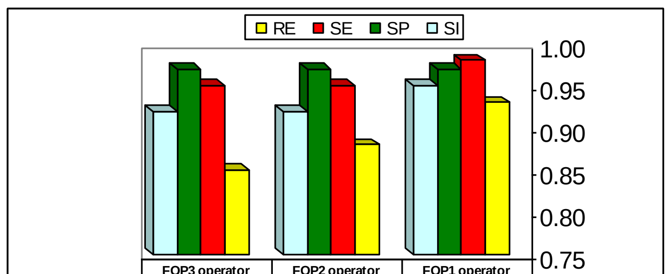
$$Similarity(SI) = \frac{2.TP}{2.TP + FN + FP}.$$

Where *TP* and *FP* stand for true positive and false positive, which were defined as the number of voxels correctly and incorrectly classified as brain tissue by the automated algorithm. *TN* and *FN* stand for true negative and false negative, which were defined as the number of voxels correctly and incorrectly classified as non-brain tissue by the automated algorithm. The results obtained by all operators are presented in figures 4, 5 and 6 below :



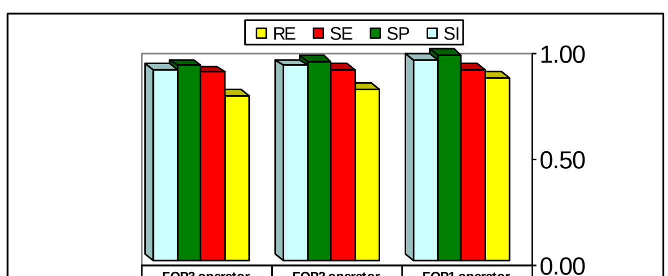
	FOP3 operator	FOP2 operator	FOP1 operator
RE	0.74	0.80	0.86
SE	0.80	0.81	0.82
SP	0.97	0.98	0.99
SI	0.83	0.85	0.96

Figure 4. Performance measures of CSF mater.



	FOP3 operator	FOP2 operator	FOP1 operator
RE	0.85	0.88	0.93
SE	0.95	0.95	0.98
SP	0.97	0.97	0.97
SI	0.92	0.92	0.95

Figure 5. Performance measures of WM mater.



	FOP3 operator	FOP2 operator	FOP1 operator
RE	0.78	0.81	0.86
SE	0.89	0.90	0.90
SP	0.92	0.94	0.97
SI	0.90	0.92	0.95

Figure 6. Performance measures of GM mater.

The results showing in figures 4, 5 and 6 demonstrate easily the predominance of the FOP1 operator compared to both FOP2 and FOP3 operators, then the best segmentation is obtained by this one.

A. Comparison with other methods :

To validate the interest of fusion produced by operator FOP1 in terms of segmentation of the cerebral tissues, we compared the results obtained on fusion T2/PD with a fuzzy segmentation computed by the algorithm of classification FPCM on the T2 image alone and the PD image alone. An example of segmentation result for the slice 94 of Brainweb is presented in the figure 7 below:

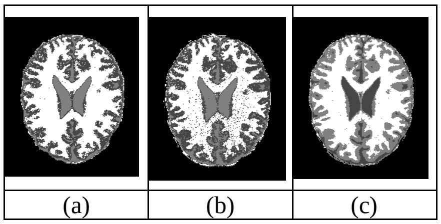
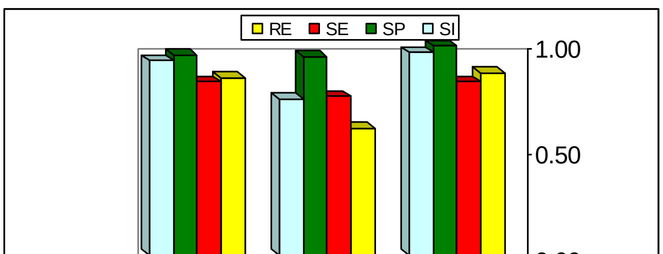


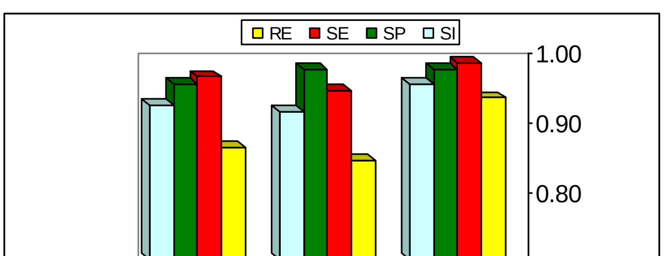
Figure 7. (a) T2 segmeneted with FPCM algorithm, (b) PD segmeneted with FPCM algorithm , (d) Image of fusion with FOP1 operator.

For each one of the segmentation, we calculated four measurements of validation: overlap, sensitivity, specificity and the similarity. For all tissues CSF, WM and GM. The results are reported in the figure 8, 9 and 10 below :



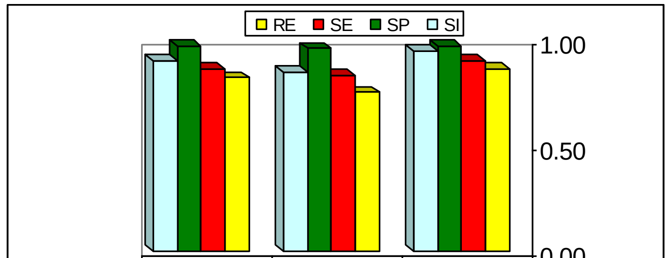
	T2 alone	PD alone	Fusion T2/PD
RE	0.84	0.60	0.86
SE	0.82	0.75	0.82
SP	0.95	0.94	0.99
SI	0.92	0.74	0.96

Figure 8. Performance measures of CSF mater.



	T2 alone	PD alone	Fusion T2/PD
RE	0.86	0.84	0.93
SE	0.96	0.94	0.98
SP	0.95	0.97	0.97
SI	0.92	0.91	0.95

Figure 9. Performance measures of WM mater.



	T2 alone	PD alone	Fusion T2/PD
RE	0.82	0.75	0.86
SE	0.86	0.83	0.90
SP	0.97	0.96	0.97
SI	0.90	0.85	0.95

Figure 10. Performance measures of GM mater.

The graphics of figures 8, 9 and 10 underline the advantages of the multispectral fusion images within the fuzzy possibilistic framework to improve the segmentation results clearly. Indeed all performance measurements obtained with fusion of T2 and PD images for all tissues CSF, WM and GM are greater than ones obtained when to taking into account of only one weighting in MR image segmentation.

VIII. CONCLUSION

This paper mainly focus on the study and evaluation of the segmentation of MR images with multispectral fusion approach using three different adaptive operators. We outlined in here some features of possibility theory, which can be very useful for medical images fusion. And which constitute advantages over classical theories. Our study demonstrate the superior capabilities of fusion approach compared to the taking into account of only one weighting in MR image segmentation.

As a perspective of this work other adaptive operators or more robust algorithms to representing a data are desired. In addition, we can integrate other numerical, symbolic information or images coming from other imaging devices include computer tomography(CT), the newer positron emission tomography (PET) or a major functional modality SPECT in order to improve the segmentation of the MR images or to detect anomalies in the pathological images.

ACKNOWLEDGMENT

The authors thank the reviewers for their valuable comments and suggestions that helped us to make the paper in its present form.

REFERENCES

[1] I. Bloch, "Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," Pattern Recognition Letters, vol. 17, pp. 905–919, 1996.

[2] Y. Hata, S. Kobashi, and S. Hirano, "Automated segmentation of human brain MR images aided by fuzzy information granulation and fuzzy inference", IEEE Trans. SMC, Part C, vol. 30, pp. 381–395, 2000.

[3] D. Goldberg-Zimring, A. Achiron, and S. Miron, "Automated detection and characterization of multiple sclerosis lesions in brain MR images", Magnetic Resonance Imaging, vol. 16 pp. 311–318,1998.

[4] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain", IEEE Trans. Medical Imaging, vol. 18, pp. 897–908, 1999.

[5] Y. Wang, T. Adali, J. Xuan, and Z. Szabo. "Magnetic resonance image analysis by information theoretic criteria and stochastic models", IEEE Trans, Information Technology in Biomedicine, vol. 5, pp. 150-158, 2001.

[6] I. Bloch, and H. Maitre, "Data fusion in 2D and 3D image processing: An overview," X Brazilian symposium on computer graphics and image processing, Brazil, pp. 127–134, 1997.

[7] W. Dou, S. Ruan, Y. Chen, D. Bloyet, J. M. Constans, "A framwork of fuzzy information fusion for the segmentation of brain tumor tissues on MR images," Image and vision Computing, vol. 25, pp. 164–171, 2007

[8] V. Barra and J. Y. Boire, "A General Framework for the Fusion of Anatomical and Functional Medical Images," NeuroImage, vol. 13**,** 410–424, 2001.

[9] D. C. Maria, H. Valdés, J. F. Karen, M. C. Francesca, M. W. Joanna, "New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images," Eur Radiol, vol. 20, 1684–1691, 2010.

[10] E.D Waltz, the principals and practice of image and spatial data fusion in : Davis L, Hall, James llinas (Eds), Proceedings of the Eight, national Data Fusion Conference ,(Dalls, TX. March 15-17, 1995) Handbook of Multisensor Data Fusion, CRC pess, West Bam Beach, FL,1995, pp. 257–278. (p4–1–4–18).

[11] F. Behloul, M. Janier, P. Croisille, C. Poirier, A. Boudraa, R. Unterreiner, J. C. Mason, D. Revel, "Automatic assessment of myocardial viability based on PET-MRI data fusion," Eng. Med. Biol. Mag., Proc. 20th Ann. Int. Conf. IEEE, vol. 1, pp. 492–495, 1998.

[12] M. Aguilar, R. Joshua, "New fusion of multi-modality volumetric medical imagery," ISIF, pp. 1206–1212, 2002.

[13] E. Lefevre, P. Vannoorenberghe, O. Colot, "About the use of Dempster-Shafer theory for color image segmentation," First international conference on color in graphics and image processing, Saint-Etienne, France, 2000.

[14] V.Barra, and J-Y Boire, "Automatic segmentation of subcortical brain structures in MR images using information fusion," IEEE Trans. on Med. Imaging., vol. 20, n7, pp. 549–558, 2001.

[15] M. C. Clarck, L.O. Hall, D.B. Goldgof, R Velthuizen, F.R. Murtagh, M.S. Silbiger, "Automatic tumor segmentation using knowledge-based techniques," IEEE Trans. Med. Imaging, vol. 17, pp. 187–201, 1998.

[16] I. Bloch, "Fusion of numerical and structural image information in medical imaging in the framework of fuzzy sets," Fuzzy Systems in Medicine, ed. par P. Szczepaniak et al., pp. 429–447, Springer Verlag, 2000.

[17] L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, pp. 338–353, 1965.

[18] L. A. Zadeh, "fuzzy sets as a basic for a theory of possibility," Int. Jour. of Fuzzy Sets and Systems, vol. 1, pp. 3–28, 1978.

[19] D. Dubois and H. Prade. Fuzzy sets and systems : Theory and Application, New-York: Academic Press, 1980.

[20] D. Dubois and H. Prade, "A review of Fuzzy Set Aggregation Connectives," Information Sciences, vol. 36, pp. 85–121, 1985.

[21] R. R. Yager, "Connectives and quantifiers in fuzzy sets," Int. Jour. of Fuzzy sets and systems, vol. 40, pp. 39–75,1991.

[22] D. Dubois and H. Prade, "Combination of information in the framework of possibility theory," In M. AL Abidi et al., editor, Data Fusion in Robotics and Machine Intelligence: Academic Press, 1992.

[23] V. Barra, Fusion d'images 3D du cerveau : Etude de modèles et Applications. Thèse de doctorat, Université d'Auvergne, 2000.

[24] I. Bloch, "Information combination operators for data fusion : A Comparative Review with Classification," IEEE Transactions en systems, Man. and Cybernitics, vol. 1, pp. 52–67, 1996.

[25] W. Dou, Segmentation des images multispectrales basée sur la fusion d'information: application aux images IRM. Thèse de doctorat Université CAEN/Basse-Normandi , 2006.

[26] A.Bendjebbour, W.Pieczynski, "Segmentation d'images multisenseurs par fusion évidentielle dans un contexte markovien," Traitement du Signal, vol. 14, n15, pp. 453–463, 1997.

[27] H.J. Zimmermann. Fuzzy Sets Theory and its Applications. 2nd ed. Boston: Kluwer Academic Publishers, 1991.

[28] S. Deveughelle, B. Dubuisson, "Adaptabilité et combinaison possibiliste: application à la vision multi caméras," Traitement du Signal, vol. 11, n6, pp. 559–568, 1994.

[29] C.Barillot, J.C.Gee,L.Le Briquer,G.Le Goualher, "Fusion intra et inter individus en imagerie médicale appliquée à la modélisation anatomique du cerveau humain," Traitement du Signal, vol. 11, n6, 1994.

[30] J. Bezdek, "A convergence theorem for the fuzzy data clustering algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence TPAMI, vol. 2, pp. 1–8,1980.

[31] J. C. Bezdek, J. Keller, R. Krishnapuram and N. R. Pal, "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing", Kluwer Academic Publishers, TA 1650.F89, 1999.

[32] N. R. Pal, K. Pal and J. C. Bezdek, "A mixed c-means clustering model", Proceedings of the Sixth IEEE International Conference onFuzzy Systems, Vol. 1, pp. 11-21, 1997.

[33] V.Barra, J. H. Boire, Segmentation floue des tissus cérébraux en IRM 3D: une approche possibiliste versus autres méthodes, Rencontres Francophones sur la logique floue et ses applications, Valenciennes, Editions Cépaduès, pp. 193–198, 1999.

[34] I.Bloch, H. Maitre, "Fusion de données en traitement d'images : modèles d'informations et décisions," Traitement du Signal, vol. 11, n6, pp. 435–446, 1994.

[35] BrainWeb [Online]. Available: www.bic.mni.mcgill.ca/brainweb/

[36] C. A. Cocosco, V. Kollokian, R. K. -S. Kwan, and A. C. Evans, "BrainWeb: Online interface to a 3D MRI simulated brain database," NeuroImage, pt. 2/4, vol. 5, n4, p. S425, 1997.

[37] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," IEEE Trans. Med. Imaging, vol. 17, n3, pp. 463–468, 1998.

[38] C.Lamiche, A.Moussaoui, "Improvement of brain tissue segmentation using information fusion approach", International Journal of Advanced Computer Science and Applications, vol 2, n6, pp. 84–90, 2011.

[39] N. Richard, M. Dojat, and c. garbay, "Automated segmentation of humain brain MR images using a multi-agent approach", Artificial intelligence in Medecine, 2004.

AUTHORS PROFILE

Lamiche Chaabane received his computer science engineer degree in 1997 from the Department of Computer Science from Ferhat Abbas University, Algeria. He also received a Magister degree in Computer Science in 2006 from University of M'sila. He has 4 years experience in teaching. His areas of interests include data mining and warehousing, artificial intelligence, image processing and operational research. His current research interests include the data mining techniques and medical image analysis.

Moussaoui Abdelouahab received his BSc in Computer Science in 1990 from the Department of Computer Science from the University of Science and Technology of Houari Boumedienne (USTHB), Algeria. He also received a MSc in Space Engineering in 1991 from University of Science and Technology of Oran (USTO). He received also a MSc degree in Machine Learning from Reims University (France) since 1992 and Master's degree in Computer Science in 1995 from University of Sidi Bel-abbes, Algeria and PhD degree in Computer Science from Ferhat Abbas University, Algeria. He is IEEE Member and AJIT, IJMMIA & IJSC Referee. His researches are in the areas of clustering algorithms and multivariate image classification applications. His current research interests include the fuzzy neuronal network and non parametric classification using unsupervised knowledge system applied to biomedical image segmentation. He also works from a long time on pattern recognition's algorithm, complex data mining and medical image analysis.

Towards an Adaptive Learning System Based on a New Learning Object Granularity Approach

Amal Battou
IRF-SIC Laboratory, Faculty of science,
Ibn Zohr University
Agadir, Morocco

Ali El Mezouary
IRF-SIC Laboratory, Faculty of science
Ibn Zohr University
Agadir, Morocco

Chihab Cherkaoui
IRF-SIC Laboratory, ENCG
Ibn Zohr University
Agadir, Morocco

Driss Mammass
IRF-SIC Laboratory, Faculty of science
Ibn Zohr University
Agadir, Morocco

***Abstract*—To achieve the adaptability required in ALS, adaptive learning system (ALS) takes advantage of granular and reusable content. The main goal of this paper is to examine the learning object granularity issue which is directly related with Learning Object (LO) reusability and the adaptability process required in ALS. For that purpose, we present the learning objects approach and the related technologies. Then, we discuss the fine-grained as a fundamental characteristic to reach the adaptability and individualization required in ALS. After that, we present some learning object granularity approaches in the literature before presenting our granularity approach. Finally, we propose an example of implementation of our approach to test its ability to meet the properties associated with fine-grained and adaptability.**

Keywords- aaptability; learning Content; adaptive learning systems, learning object; granularit;y learning content.

I. INTRODUCTION

The Learning Object granularity is one of the most critical properties of the Learning Object. Wiley [1] ensures that the two most important instructional properties of learning objects are reusability and granularity.

Regarding the first property, reusing LOs is believed to generate economical and pedagogical advantage over the construction of LO from scratch [2]. For the second property, researches and works on the concept of granularity in the literature are almost modest except for some works such as [3, 4, 5, 6, 7, 8]. The most used technique is the concept of aggregation.

The relationship among reusability and granularity of learning objects is straightforward. The relationship can be depicted in a simple but effective equation: the more granular a learning object is, the more reusable it becomes [9].

As the impact of granularity is linked directly to reusability, the most difficult problem facing the designers of LO is how big a learning object should be to ensure the reusability. This problem should be at the focus of the designers of LO.

We are interested in this paper to the issue of learning content granularity and its impact on the ability to adapt, aggregate and to arrange content suiting the learner needs and preferences.

The aim of the remainder of this paper is structured as follow. We will firstly begin by "demystifying" the concept of LO and the instructional design principles. We discuss next, the granularity concept as a fundamental characteristic to achieve adaptability and individualization in the field of ALS. Afterwards, we explore the main approaches defending LO granularity from diverse angles and theirs limits. In the next section, we propose a new vision of the granularity concept and study next its ability to meet the supposed objectives of adaptability. Finally, we present the architecture of our system called ALS-CPL (Adaptive Learning System - C Programming Language) that implements the proposed approach. Then we discuss the ability of this work to achieve the adaptability required in ALS in line with other works [10, 11].

II. BACKGROUND

The Learning content is the main component of Adaptive Hypermedia System. Its development raises various issues related to the approach used. The most used one is the LO Approach.

A. What is a Learning Object ?

The "Learning Object" is a new name that emerges in the field of educational resources and learning, which does not escape to ambiguity. Most proposed definitions focus on the general principles governing concept of LO such as: reusability in different situation for learning and the independence of context [12].

Balatsoukas [9] gives a typical example of the Polsani definition [8]. This author defines a LO as a unit of content Learning independent and autonomous, which is predisposed to be reuse in multiple learning contexts. Other authors such as

Bibeau [13] considers LO as the smallest information unit or the smallest processing tool information (or applications software) used in an educational context with an intention teaching for learning through the media technology. Flamand [14] identifies three categories of LO. He distinguishes objects with little media complex and context-free (video speech of a head of state radio interview, etc.) utilitarian (modeling software, etc.) and LOs consisting of elements basic information (facts, ideas, concepts, principles, processes).

Finally, other approaches such as those of Downes [7] consider the LO size as important. Barron [5] trying to consider this approach, suggests that five to nine information objects (text, image, video, photos, etc.) can be combined to form a LO. Other works of Mortimer [6] undertaken in this direction approach the LO size in terms of time. A LO takes no longer than 15 minutes to complete.

In addition to these theoretical conceptualizations and sometimes ambiguous, other definitions emerge from various works on standardization (SCORM, LOM, IMS, etc.). For IEEE Learning Technology Standards Committee, LO are defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. Normetic [15] adds to this definition the technological support that covers the multimedia content, content instruction, educational software and software tools mentioned in a learning context to support technology. Finally, the center of Wisconsin online resources [16] defines a LO as "small learning units with a duration between 2 and 15 minutes".

The next section is designed to provide information about instructional design principles and how they relate to teaching and learning.

B. *The instructional design principles*

The issue of designing LO in the field of education has often been mentioned by several researchers [4, 7], etc.

Designing a LO with the features mentioned above (reusability, interoperability, durability etc.) needs to think and reflect on how to design and develop these objects. Studies of the best practices and scientific research on the LO design, reveals a series of principles and guidelines for such design [6].

We note that standards in the field of ALS, do not give steps to follow specifying methods of designing and creating LO. However, we found in the literature, a few principles that can be used in the design process and execution of LO. We will then present the most cited:

- A LO must be initially developed as "bricks" of a relatively small and designed in a way that facilitates reuse in a new educational context.
- A LO must be independent and separable from one context of use. It should contain generic information as possible.
- A LO must be indexed by metadata, based on a standard to provide information (size, author, type of interactivity, etc.) on that object. This metadata facilitates search and retrieval of LO stored in repositories.
- A LO must follow a standard format of instruction. The current standard of education facilitates the creation of uniform LOs with a clear educational strategy.
- A sequence of LOs must have a context. To build an educational unit (module, courses, etc.) from a LO, We must specify the context of (re) use of those LOs or leave the choice to learners to specify their own context.

However, even if these principles revealed some stability on the LOs design, this progress remains probably insufficient to capture the essence of the LO approach. We can cite for example, the confusion surrounding "granularity" as important attribute of LOs and which is apparent in the literature as we will present later in this paper. The granularity, as mentioned in [17], has a crucial impact on the ability to adapt, aggregate and to arrange content suiting the learner needs and preferences.

In the next section we are interested in the influence of fine-grained content on the adaptability and individualization required by ALS.

III. Granularity and Adaptability

The LO granularity is a key factor to allow aggregating and organizing content, to adapt the instruction to the preferences of a given learner. On one hand, an insufficient granularity (using for example *large blocks of contents*), probably prevents the possibility of integrating educational content in new contexts and new ALS. On the other hand, the fact of splitting up contents in several LOs of small size with a main idea, allows several options for adaptation [12].

The first possibility is to aggregate and arrange multiple objects to create other more consistent and reusable objects. The second possibility is to build and customize a LO by proposing several presentations with different computer interfaces. Another possibility implies a classification of LOs into classes of objects (for example theorems, definitions, etc.), which makes it possible to filter them more easily, improve research and thus to individualize the content.

In addition, the granularity combined with the indexing plays an important role to facilitate the adaptability. Indeed, instead of adding meta-data to big blocks of contents, learning objects of "fine granularity" are indexed, which increases thus the research space. This distinction also helps to increase the possibilities of finding the most adapted elements to a specific situation. It also allows annihilating the research silence, which can be due to an insufficient granularity.

The adaptability here then consists in choosing between the various grains those who are appropriate to a given situation.

To show the correlation between the granularity and adaptability, we propose the following learning objects granularity approach to allow a flexible representation, respecting the standards and capable of building contents in a dynamic way, from basic fragments, from the representation of the learning domain and the learner model. But before this, we study first the main used approaches of LO granularity in the literature and discuss their limits.

IV. Review of Granularity in the Literature

There are two main approaches defending LO granularity from two different angles.

The first one, supported by several authors like [3, 4, 5, 6, 7, 8].This approach focuses on the content. The granularity of a LO is related to the number of concept combination in this LO. A LO is called granular and therefore has great potential for reuse if it contains one basic concept.

A. Approach focus on content

For Wiley [4], the granularity of a LO depends closely on the context in which the granule will be inserted.

Similarly, South [18] defines the granularity in terms of content domain of LO which suggests that the objects have the greatest potential for reuse when they are focused on a single concept.

Other authors such as Polsani [5] argues that the granularity depends on the size of a LO. But the size designated by the author cannot be expressed in terms of bytes or duration of a LO. Size here refers to the number of ideas that a LO can transmit.

Generally, a LO must transmit one or few ideas. In the case where a LO consists of several ideas, one of these ideas may be primary and the others derive or depend directly on this one. The "fine grained" is then to considers concept as unifying principle that frees the LO of any consideration related exclusively to the size such as time or the subjectivity of the designer.

B. Approach focus on media

Concerning the second approach, it's supported by several specifications and standards bodies (IMS Global Learning Consortium, IEEE LTSC, etc.). This approach is based on a definition of granularity focused on the media. The LO granularity is directly related to the media that will be combined to create larger LOs. It uses the concept of level of aggregation instead of the notion of granularity and provides models of educational content that provide a means for defining structures.

SCORM defines an associated structure with three levels of aggregation, indicating three main components. ***the asset*** presents the smallest piece of reusable educational content that may be Web pages, animations, pictures, videos, etc, the ***Sharable Content Object*** (SCO) can be composed of several assets and the ***Content Organization*** (COs) bound to a tree structure that acts like a table of contents.

For LOM, this model consists of four levels of aggregation or "functional granularity". The first level is the lowest level of aggregation; it consists of raw media or fragments. The second level includes a collection of learning objects of level 1, such as a lesson. A collection of LOs of level 2 such as a course constitutes the third level. The fourth level of granularity is composed by a set of courses which lead to a certificate or a diploma.

Cisco Systems has published a strategy based on the concept "RLO / RIO". The content structure is composed of two basic levels: the RIOs and the RLOs. A RIO (Reusable Information Object) is a reusable granule independent of the publishing format. It is presented by five various types of knowledge, including concepts, facts, procedures, processes or principles and associated with assessments (usually two) to evaluate the learner's assimilation of different concepts, facts, etc. A RLO (Reusable Learning Object) is the result of a combination of five to nine (7 ± 2) RIOs, attached to an overview and summary, to meet a clearly defined educational objective.

We note that other approaches are also mentioned in the literature, we cite, for example, the approach based on the execution time as a criterion of LO granularity. This approach is supported by the Wisconsin Center [19] which states that the execution time / consultation of LO must not exceed 15 minutes. Another approach is also cited in the literature that is based on the LO size in terms of bit. We believe that these two approaches are a bit outdated as it refers neither to the LO content or the presentation.

C. Discussion

The different approaches presented in the previous parts, tried to break down the content into a set of items or building blocks that make sense, also called grain teaching. Defined grains, although they can be reused in other contexts, probably don't fully satisfy the concept of fine grained LO that allows several option for adaptation (as we present above).

For the first approach, it defines the granularity of the LO based on the number of concept and idea combined in this LO. It does not put any restriction on the number of concept or idea or their types (complex, easy uncomplicated). Indeed, the concept of a given area may contain several ideas with different levels of complexity. This allows several other possibilities for cutting this concept on sub concepts or simple idea. Moreover, this approach does not take into consideration the presentation of a LO that can contain different format that is to say that LO can hold together text, image, video, etc. There is no restriction on either the presentation or the content of the idea presented. In our opinion this approach does not define the fine-grained as we seek to reach for better adaptability of learning content.

Regarding the second approach, in all the presented models, the first level (Asset, RIO multimedia object) cannot really match the criteria of sense unless the grain is associated with an educational objective. The size criterion is not necessarily considered and generally depends on the designer. Indeed, in most of these models there is no information about the size or the semantic density (number of idea) of o LO. In addition to the standard definition of a LO differs from one model to another. Each model is a specific profile. The LO use by one model can't be reused in another model.

V. A New Formulation of Granularity

Confusion surrounding granularity as an important attribute of LOs is apparent in the literature [15].

The notion of granularity, we propose, is based on several approaches of the LO granularity. In our opinion a fine grained LO is to combine the concept of meaning in terms of ideas carried by the grain, size in term of time of execution/consultation and media type as unifying principle. In addition, we propose a semantic structuring educational content.

For our approach, the main concept is fragment that corresponds to the notion of learning object. It can be an introduction, an assessment, exercise, a synthesis, an observation, motivation, definition, example, etc. Each of these fragments is presented by bricks multimedia: text, image, sound, video, simulation, animation, etc. A fragment has a size in terms of execution/consultation time of the corresponding brick media. The fragment size must not exceed 15 minutes. In addition, it must send a simple idea and it is described by a single brick multimedia.

The notion of fragment we use is an abstract concept. It corresponds to a multimedia brick related to an educational objectives and a semantic description.

The approach we propose is between several approaches. Indeed, it offers a functional aggregation of four main levels: courses, documents, fragments, and multimedia brick Fig. 1.

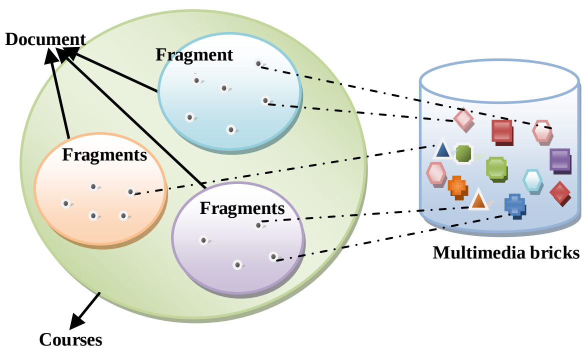


Figure 1. The proposed content model

The multimedia bricks in our approach correspond to Assets of multimedia objects and SCORM LOM, but by removing the proceeding related to a web page, which is not considered fine enough since it can present several ideas, has a large size (in term of execution time) and can contains different presentation(text, image, video, etc). The fragments in our model can be an introduction, a definition, an example, an exercise, a paragraph, a comment, an evaluation, a synthesis, or an illustration, etc. It May corresponds to the notion of RIO defined in the model CISCO.

Taking as example the fragment of the introduction, it can be presented by the text, contains one idea that is to introduce the topic of the concept. Also the time of playback time of the introduction should not exceed 15 minutes at worst (if we have a slow learner in the reading). In addition, our grains meet the test range of ideas and meaning. Indeed, the multimedia bricks are closely associated with fragments. Thus, a definition can be materialized by a picture, video, sound, etc. This increases the capacity of adaptability because the brick that will be chosen for the presentation depends on the preferences of learners. In addition, the multimedia bricks are associated with the metadata defined. The choice of a multimedia brick depends on the progress of a learner (acquired concepts, etc.), Preferences and such prior knowledge. Also linked to the semantic level that facilitates storage and improving grain research in response to a given situation.

The next part presents the implementation of our granularity approach to validate its capacity to achieve adaptability required in ALS.

VI. A CASE OF STUDY : ALS-CPL SYSTEM

In this section we illustrate some functionalities of our system called Adaptive Learning System-the C Programming Language (ALS-CPL) which implements the LO granularity approach we proposed above.

The architectural design of the proposed system is composed by three main components Fig. 2.

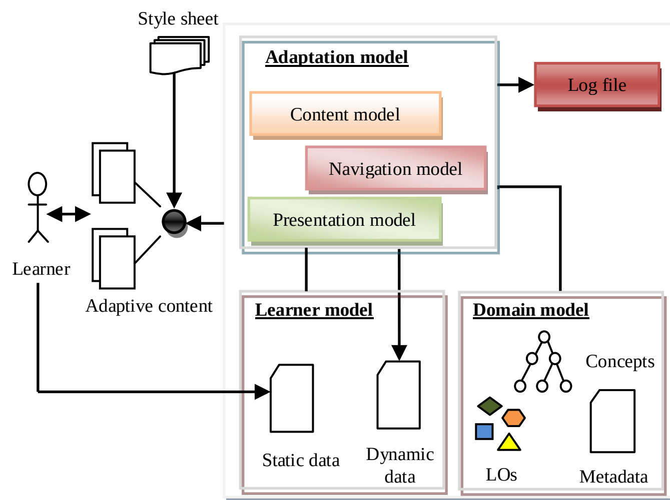


Figure 2. General architecture of ALS-CPL

From this figure we identify directly the main components of an ALS-CPL and their contents in terms of subcomponents. In the following, we present these components, their descriptions, their features and interactions between them

A. *The Domain model*

The domain model is characterized by its competence in terms of representation of concepts to learn, the resources available to learners and the structuring of various elements of the field

We have separated the domain model into two parts: one that includes all domain concepts that the learner can learn, regardless of the different kind resources that enable the acquisition. The second one, the most important for our work, concerns the LOs used for the acquisition of these concepts accompanied by their metadata. We create LOs according the criterions of the proposed granularity approach and the principles instructional design we present above.

This part consists of an editor of LOs Fig. 3 and reflects the semantic model presented above. This component uses PHP code to load one of the LO that we created. This editor permits the editing of the metadata to qualify LO by exploiting metadata and the necessary descriptors so important to characterize each LO.

The multimedia bricks and concepts (Learning goal) can be associated to the LO. Other items for indexing LOs are added in this step. As output, of this form, an XML file is generated e.g Fig. 4.

Figure 3. Editing Learning Object metadata

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<metadata>
<general>
<identifiant>introduction_scanf</identifiant>
<intitule>introduction au concept scanf</intitule>
<langue>fr</langue>
<description>cette ressources introduit les concept de la
fonction d'entrée scanf()</description>
<keyword> scanf, fonction d'entrée, introduction,
concept</keyword>
</general>
<lifecycle>
<version>1</version>
<statut>final</statut>
<author>Masha Nikolski</author>
<mail></mail>
<organisation>CS Department, Technion</organisation>
<date> April, 2006</date>
```

Figure 4. Example of an XML file of metadata

B. The learner model

The learner model represents the information Fig. 5 known by the system about the learner. Three strands of information are considered: personal (name, email address, phone, etc.), preferences (language preferences, favorite colors, the preferred type of educational content, etc.) and Knowledge of learner that is described in relation to each domain model. This component evolves dynamically as the student progress in his course.

The learner model we make is open for editing and viewing by both the learners themselves and the system. The aim of this choice is to involve the learner in the construction of his model. So that it contains information and makes it more reliable and more representative of the learner.

For our model, it consists of two main parts:

- Static data: the data is indicated by the student during his first access to the system and can be updated by the learner at any time of his learning.
- Dynamic Data: This data is updated only by the system and highly dependent on the results and interactions of the learner with the content presented.

Figure 5. Editing the learner model information in ALS-CPL

As output, an XML file is generated. Fig. 6 presents an example of a model learner.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type='text/xsl' href='transf.xsl'?>
<!DOCTYPE apprenant SYSTEM "apprenant.dtd">
<apprenant>
    <statique>
        <identification>
            <nom>battou</nom>
            <prenom>amal</prenom>
            …
            <email>amal.battou@gmail.com</email>
        </identification>
        <securité
            <password>16b5480e7b6e68607fe48815d16b5d6d
</password>
                                ….
    </statique>
    <dynamique>
        <concept> </concept>
        <note></note>
        <etat_emotionnel></etat_emotionnel>
                    …
    </dynamique>
</apprenant>
```

Figure 6. Example of an XML file of a learner model

C. The adaptation model

The adaptation model deals with the generation of adaptive content that will be subsequently presented to the learner. This component has three sub-component: the navigation model, the presentation model and the content model. Each sub-component contains a set of rules to achieve the adaptation.

- The model navigation: the navigation model defines the structure of the hypermedia system and describes how to traverse the various nodes of the system.

- The presentation model: it is used to adapt the layout for the visual line with the preferences or needs of the learner.
- The content model: this model is used to provide additional content, similar content, alternative content, or hide content.

The process of how these sub-components Fig. 7 and an example of a content interface Fig. 8 are presented as follow.

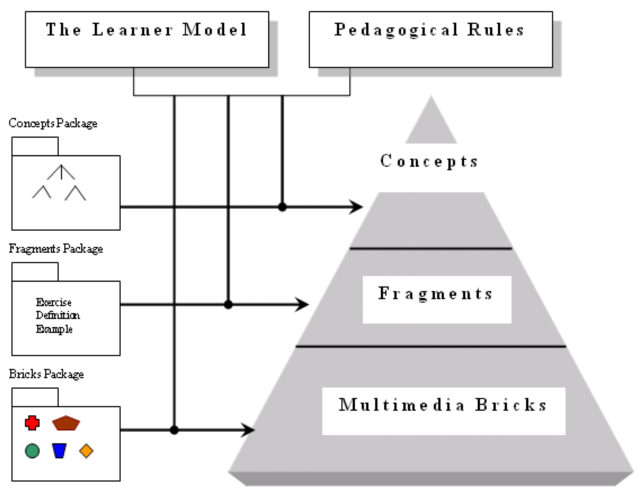


Figure 7. The assembling courses process.

Figure 8. Example of a content interface

The simplest case, when a learner interacts for the first time with the system, the list of the acquired concepts is empty. The concepts that have no pre-requisites in the graph of the concepts and have not been acquired will initialize the list of the active concepts, which enable to choose the objective of the session.

Some elements of the learner model can influence this decision. These considerations come from the background knowledge and skills of the learner represented in the learner model like the level concerning the programming languages (beginner, intermediate, Expert), or the background knowledge composed of a set of programming concepts (variables, decision-making code, loop structures, procedures and functions, data bases, etc.). Some pedagogical rules for such decision are applied.

The choice of one or more concept(s) associated with other information coming in particular from representations of the learner, determines a sequence that will then be derived in fragments. If, for example, the model of the learner indicates that he (she) prefers to learn by examples, the sequence will consist of more examples. For exercises, the difficulty level will depend on information extracted from the model of the learner corresponding to his level (Beginner, Intermediate, and Advanced).

This sequence corresponds to a prototypical sequence of fragments to achieve the selected learning concept. For each fragment of this sequence, the system associates a multimedia brick, still according to the learner model. If the learner model indicates for example that learner prefer pictures and videos, the system will promote anything that is multimedia. If he (she) prefers reading on the screen, the text associated with fragments will be used to create a course document.

VII. Discussion

The fine granularity concept that we have presented above, meets the criterion of scope ideas and meaning. Indeed, the bricks are closely associated with fragments. Thus, a definition can be materialized by a picture, video, sound, etc. This increases the capacity of adaptability because the choice of bricks depends on the learners preferences. We note that each fragment corresponds to a single brick multimedia depending on the preference of the learner (video, audio, text, etc.)

In addition, the bricks are associated with domain concepts through metadata set. The choice of a brick depends on the progress of a learner (acquired concepts, etc.), preferences and those acquired earlier. Also, the semantic level linked to the indexing, allows easy storage and improving grain research responding to a given situation.

Finally, we can highlight some correlation of our model compared with recent work focusing on adaptability in dynamic adaptive hypermedia. We emphasize in particular the work on the project Medyna [10] and work related to the assembly of existing resources by using graphs and operators of decision [11]. Our model is also inspired by the work of Brusilovsky [20] on the graphs of concepts and the relationship between concepts, brick fragments and multi-media, with a distinction related to meta-data used and techniques for adaptability and assembly course.

VIII. Conclusion

In this paper, we have proposed a model of granular LO for the adaptability and the re-use of the learning contents. This model of content is designed respecting to the various characteristics of the stated granularity. The first advantage of this model is its hierarchical structure in the form of “grains” of contents which respect the specifications of the existing standards (LOM, SCORM, etc.). Another advantage lies in the fact that the same fragment or a multimedia brick could easily be re-used in several documents or then directly in another context of learning. We can also note that the model suggested is open. It can indeed employs the proprietary format of the contents, or import it from the web. Moreover, the granularity combined with indexing plays an important role in facilitating the search mechanism and adaptability. Indeed, instead of adding metadata to large blocks of educational content, small size granules are indexed, which enlarges the search space.

As a second point, we have showed the system architecture able of integrating the LOs infrastructure, the domain concept structure and the learner model interface. Different interfaces are presented.

It is clear that several issues remain to be addressed to arrive at the expected system. Our work continues along these lines to try to finish a first functional prototype which will be tested and validated.

References

[1] D.A. Wiley, “Learning Objects and the New CAI: So what do I do with a learning object?”, 1999. Retrieved in http://opencontent.org/docs/instruct-arch.pdf.

[2] X. Ochoa, E. Duval, Measuring Learning Object Reuse In Times of Convergence. Technologies Across Learning Contexts, Vol. 5192, 2008, pp. 322-325.

[3] P. Balatsoukas, A. Morris, A. & O’Brien, “Learning Objects Update: Review and Critical Approach to Content Aggregation”, Educational Technology & Society, 2008, Vol. 11 (2), pp-119-130.

[4] D.A. Wiley, “Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy”. In D.A. Wiley (Ed.), The instructional use of learning objects, 2000.

[5] P.R. Polsani, “Use and abuse of reusable learning objects”, Journal of Digital Information, 2003, 3(4). Retrieved in http://jodi.tamu.edu/Articles/v03/i04/Polsani/.

[6] S. Downes, “Design and reusability of learning objects in an academic context: A new economy of education”, JSDLA Journal, 2003, 17(1). Retrieved in http://www.usdla.org/html/journal/JAN03_Issue/article01.html.

[7] T. Barron, “Learning object approach is making inroads”, In Learning Circuits, 2002. Retrieved in http://www.learningcircuits.org/2002/may2002/barron.html.

[8] L. Mortimer, “Learning objects of desire: Promise and practicality”. Learning Circuits, 2008, Retrieved In http://www.astd.org/LC/2002/0402_mortimer.htm.

[9] I.F. Silveira, “Reusability and Interoperability of Adaptive Learning Objects Repositories” Symposium de l’informatique en education, 2005. Retrieved in http://www.br-ie.org/pub/index.php/sbie/article/viewFile/407/393.

[10] A. Behaz, “Création dynamique de documents hypermédias adaptatifs”, Available : http://lodel.irevues.inist.fr/cide/index.php?id=281.

[11] A. Bouzeghoub, “Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine”, Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation, Vol. 12, (2005).

[12] A. Battou, A. El Mezouary, C. Cherkaoui, D. Mammass, The granularity approach of learning objects to support adaptability in adaptive learning systems. *Journal of Theoretical and Applied Information Technology*. Vol 18. No. 1 – 2010 Available :http://www.jatit.org/volumes/eighteenth_volume_1_2010.php.

[13] R. Bibeau, “Les TIC à l'école : proposition de taxonomie et analyse des obstacles à leur intégration”, *EpiNet* n° 79, novembre 2005. Available : http://www.epi.asso.fr/revue/articles/a0511a.htm.

[14] P. Flamand, A. Gervais, “Les objets d'apprentissage : au-delà de la technologie, la pédagogie”, *Le bulletin Clic*, 2004, N°54, pp 1-5. Available : http://www.clic.ntic.org/clic54/objets.html.

[15] Normetic, “La description normalisée des ressources : vers un patrimoine éducatif”, *CREPUQ Novasys* Inc., 2003. Available : http://www.profetic.org:16080/normetic/pdf/normetic.pdf.

[16] WISC, (2009) “Visconsin on line Ressource center”. Available : http://www.wisconline.com.

[17] M. Memmel, E. Ras, P.Klaus. J. and M. Yacci. “Approaches to learning Object Oriented Instructional Design”, Learning Objects and Instructional Design, pages 281-326. Informing Science Press, 2007, Retrieved in http://www.dfki.uni-kl.de/~memmel/literature/2007/MemmelRasJantkeYacci+07.pdf.

[18] J.B. South, D.W. Monson, (2000). A university-wide system for creating, capturing, and delivering learning objects. In D. A. Wiley (Ed.), The Instructional Use of Learning Objects [Online] Available : http://reusability.org/read/chapters/south.doc.

[19] WISC, (2009) “Visconsin on line Ressource center”. Available : http://www.wisconline.com.

[20] P. Brusilovsky, “Adaptive Hypermedia: From Intelligent Tutoring Systems to Web-Based Education”, 2000, Available: http://www2.sis.pitt.edu/~peterb/papers/ITS00inv.html.

Authors Profile

Mrs A. Battou received her Master's degree in Applied Mathematic and Computers in 2008 in the university Ibn Zohr, Faculty of Science Agadir. Morocco. She is pursuing her Ph. D degree. Her doctorate study is interested by adaptability in Adaptive Learning Systems, using Learning Objects. She is currently a state engineer obtained from a high school of of Engineering and works at the University Ibn Zohr since five years.

Mr A. EL Mezouary received his Master's degree in Applied Mathematic and Computers in 2008 from the university Ibn Zohr, Faculty of Science Agadir. Morocco. He is a research student pursuing his Ph. D degree. His thesis is part of a general problem of the Educational Semantic Web. He is interested specifically by the adaptability in Adaptive Learning Systems, using Learning Objects and the Semantic Web. He served as professor of computers and data processing at the secondary school.

Dr. C. Cherkaoui is currently research professor at the School of Business and Management Agadir - Morocco. He holds a doctorate in computers and data processing at a French university since 1996. The research area is part of continued discipline of Artificial Intelligence. He has worked on various topics such as: planning and action theory, adaptive systems for human learning, multi-agent systems, e-learning, knowledge engineering, knowledge representation including ontologies. He obtained his habilitation to direct research in 2006. He has published various articles and papers in journals and international conferences. He served as Head of Department of Computer Science and is currently team leader of the research group of Intelligent and Communicating Systems.

Dr. D. Mammass is Professor of higher education at the Faculty of Science of Agadir since a score of years. He obtained a state doctorate in Mathematics and Image Processing in 1999. He currently directs several theses in varied disciplines of mathematics and data processing such as: the geographical Information Systems, the Digital Processing of the images and pattern recognitions, the geographical databases, the knowledge management, the semantic Web, etc. He is currently Director of the School of Technology Agadir and head of the laboratory IRF-CLC (Image Recognition of Forms, Intelligent Systems and communicating).

An Efficient Scheme for MANET Domain Formation (ESMDF)

Adwan AbdelFattah
Computer Science Department,
The Arab American University of Jenin
Jenin, Palestine

***Abstract*—Mobile Ad hoc Network (MANET) has random topology as MANET devices leaving or joining to the network at anytime. The dynamic nature of MANETs makes achieving secrecy, connectivity and high performance, a big challenge and a complex task. In this paper, we proposed an efficient technique for Dynamic construction of large MANET based on division the network into interoperable domains. This technique is a hybrid of centralized and distributed control of packets forwarding that balances power consumption, minimizes the routing tables and improves the security features. The principles of domain formation based on joining adjacent devices into one group which controlled by one capable device called domain controller. The presented scheme enhances the throughput and the stability of large MANET by minimizing the flooding of messages for keeping track of Devices and during the domain formation.**

Keywords- MANET; ESMDF; Domain; Domain Router; Domain Formation.

I. INTRODUCTION

Mobile Ad-hoc Network (MANET) is a wireless network with dynamic topology. In MANET each node is free to move randomly, and is considered to be equal to other nodes. Each node is capable of transferring the data between the arbitrary source and destination. Thus, each node in MANET can act as a source or destination or router [1].

MANETs are in areas where rapid deployment and dynamic reconfiguration are necessary and wired network is not available. These include military battlefields, emergency search, rescue sites, classrooms and conventions, where participants share information dynamically using their mobile devices.

MANET routing and topology management has become an important issue. Many efficient routing protocols have been developed which ensure the connection of sending and receiving nodes with minimum delay and unnecessary control overhead. Existing routing protocols for MANET can be classified into four different basic categories namely flooding, proactive routing, reactive routing and dynamic cluster based routing [2]. However none of these routing schemes guarantees constant network connectivity during the movement and each of these schemes has constant route maintenance overhead. A particular node may even be disconnected in the worst case. Centralized topology management schemes [3, 4, and 5] discuss a self-adaptive movement control algorithm, which ensures the retention of network connectivity even during the nodes movement. But in this case, the coordinator has to be elected and all other nodes should follow the instructions from the coordinator. The main disadvantages of the centralized topology management scheme are increase in control overhead and non-scalability.

Distributed topology management schemes [6, 7] are generally scalable and adaptive to mobility due to the fact that each node relies on local information collected from nearby nodes. The information obtained by each node is limited, and the strong connectivity of the node is not achieved in this approach.

MANET's characteristics create challenges in several areas. The hosts in the MANET have a limited battery power. In the case of large MANET's, a flat structure may not be the most efficient organization for routing between nodes. Instead, many clustering schemes have been proposed that organize the MANET into a hierarchy, with a view to improve the efficiency of routing. It is important that cluster formation and maintenance should not be costly, in terms of resources used such as bandwidth, battery power etc. Otherwise, the purpose of clustering is defeated.

II. RELATED WORKS

In this section, we describe some of the most important protocols and clustering schemes that have been proposed to enhance the quality of service (QoS) and many features of MANETs.

R. Braden, D. Clark, and S. Shenker [8] proposed Integrated Services (IntServ protocol which provides a high level of assurance in fixed network, limited QoS support for mobile networks infrastructure. It requires a high processing power, and this protocol does not support fast QoS changes.

Yu-Xuan Wang[9] proposed an entropy-based WCA (EWCA) which can enhance the stability of the Network for the high mobility of nodes but this leads to high frequency of rejoining which will increase the network overhead. The authors discussed that in the revised algorithm (EWCA-TS) in which performance has been improved with respect to the original WCA, especially on the number of clusters and the rejoining frequency.

S. Blake, et al [10] proposed Differentiated Services (DiffServ) Protocol which can be easily implemented with MANET but has low level of assurance. It does not guarantee service on end to end basis.

J. Wu and H. L. Li [11] proposed scheme, for routing based on a set of dominating nodes which function as the cluster heads and relay routing information and data packets. The vertices of a Dominating Set (DS) act as cluster heads and each node in a MANET is assigned to one cluster head that dominates it. A DS is called a Connected Dominating Set (CDS) if all the dominating nodes are directly connected to each other.

Qi Xue and Aura Ganz [12] Ad-Hoc QoS On-demand Routing(AQOR) protocol AQOR deals with bandwidth and end to end delay. This protocol includes three main steps: on demand route discovery, signaling function and hop to hop routing. This protocol does not deal with the latency delay.

Taewook Kang, et al [13] proposed a new method for selecting cluster heads to evenly distribute cluster heads and they show that their scheme reduces energy dissipation and prolongs network lifetime as compared with others. They tried to evenly distribute cluster heads over the whole network and avoid creating redundant cluster heads within a small range so that it can increase the network lifetime.

M. Mirhakkak, et al [14] proposed Dynamic Source Routing RSVP (DSRRSVP) protocol which can be easily implemented with DSR routing, but this protocol applicable to a small network with low mobility.

Y.Z.P Chen and A.L Liestman [15] proposed Weakly Connected Dominating Set (WCDS) which relaxed some of the rules of Wu's Algorithm to form a Weakly Connected Dominating Set. There are many disadvantages with the CDS algorithm. The cluster head in CDS algorithm dissipates more power as compared to other nodes in the cluster since all inter-cluster routing and forwarding happen through it alone. Hence it has a shorter lifespan than the other nodes in the cluster. The cluster head re-election is done after the cluster head dies or moves out of the range of the cluster. This re-clustering incurs a large communication overhead and power dissipation.

Hannan XIAO, et al [16] proposed Flexible QoS Model for MANET(FQMM). FQMM is the first QoS model for MANET. This model is hybrid of both IntServ and DiffServ. Problems of DiffServ and IntServ are present.

Tzay-Farn Shih And Hsu-Chun Yen [17] developed A Location-Aided Cluster-Based Routing Protocol Called Core Location-Aided Cluster-Based Routing Protocol (CLACR). They show that CLACR can be extended as a Geo-Casting Routing Protocol easily, the location Server and Cluster Head can provide location services for different applications. The performance of their routing protocol is better than other protocols.

SWAN Project [18] Service Differentiation in Wireless Ad-hoc Network (SWAN) protocol which uses rate control of TCP and UDP traffic to maintain manageable levels of congestion in the network. It uses admission control for real time traffic and varies the rate of TCP traffic based on feedback from MAC layer to maintain delay and bandwidth bounds for real time traffic. The throughput of this protocol is very low. This protocol does not scale well with high mobility.

Vikas Kawadia and P.R.Kumar [19] proposed CLUSTERPOW algorithm in which dynamic and implicit clustering is done on the basis of transmit power level. The transmit power level is the power level required to transmit each packet. The transmit power level to a node inside the cluster is the less as compared to the level required to send a node outside the cluster. So here the clustering is done keeping the nodes with lower transmit power level together. The primary drawback of their scheme is that there is no cluster head or cluster gateway. Each node here has routing tables corresponding to different transmit power levels. The routing table for a power level in a node is built by communicating with the peer routing table of the same power level at another node. The next hop to route the packet is determined by consulting the lowest power routing table through which the destination is reachable.This approach suggests that each node should know the route to other nodes and also know the transmission power level at which a destination node is reachable. This leads to the overhead of collecting the power level state information and building many routing tables in each node.

P.Basu, et al [20.] proposed clustering scheme which takes aggregate of local mobility as the metric for cluster formation. In such schemes, a cluster is formed by grouping mobile nodes moving with the same velocity. Each node broadcasts two hello packets, separated by a time interval, to its neighbors.

Every node calculates the relative mobility for each of its neighbors using the signal strength of the hello packets received from each adjacent node. Each node then calculates its aggregate mobility as the average of the relative mobility of its neighbors and broadcasts it to the other nodes. The node with the lowest aggregate mobility is chosen as the cluster head. This requires larger communication overhead and a higher latency in cluster formation.

III. PROPOSED DOMAIN FORMATION SCHEME

We define the following terminologies that are used in the remaining sections.

Domain: Is a set of related devices that can be connected directly or through Domain Router.

Domain Router (DR): Domain Server that performs the following functions:

- Routing the messages among its clients.
- Resolve the problem of Domains merging.
- Forwards the messages to neighbor domains.
- Resolve the IP conflict.
- Limiting the Domain boundaries.

Co-Domain Router (CDR): it is the nearest device to DR, it is selected by DR. CDR should be capable of performing the same function as DR. CDR used to enhance the stability of MANET topology and to minimize the domain formation processes messages.

Inter Domain Router (IDR): A gateway among domains, this device can forward messages to the neighboring DRs and their clients.

Client Router (CR): A Domain member.

Client fringe (CF): A device that has only one connection with CR, its domain id **(DI) =0.**

A. Domain properties

Each domain consists of two or more devices and each individual device interacts directly with other devices in a peer-to-peer fashion. At any time each device can be only in one state (DR, CDR, IDR CR or CF). The state of device is determined by its capability and its location. See fig. 1.

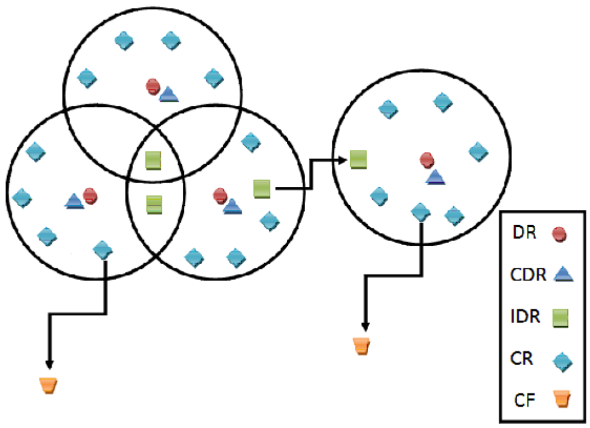


Figure1. Domains representation in MANET

The device mobility changes the domains structure of MANET and consequently the state of devices. Fig.2 describes the devices transition state diagram .

B. Improving Domain Properties

The proposed scheme of domain formation satisfies the goal of building energy conserving and adaptable domains. This scheme tries to distribute the responsibility among the individual entities. No single entity is in charge of the overall organization. In our proposed scheme, joining new nodes to the domain can be achieved smoothly because of obtaining the information directly from DR or CR.

In our scheme each DR has a CDR that can operates as DR in case of unavailability or over load of DR, and this approach enhances connectivity and decreases power consumption in DR. The process of messages forwarding is very easy, it doesn't need to maintain a huge routing table, as each device store a table of addresses of its neighbors. Each DR can communicate with one or more IDR which is very useful in balancing the load of IDR and in maintaining the connection among domains in case of unavailability of one IDR.

We assume that during device movements there is no far jumps from one domain to the remote one and this mean that only adjacent domains can be affected by these movements which confines the flooding of state change messages and routing information among them.

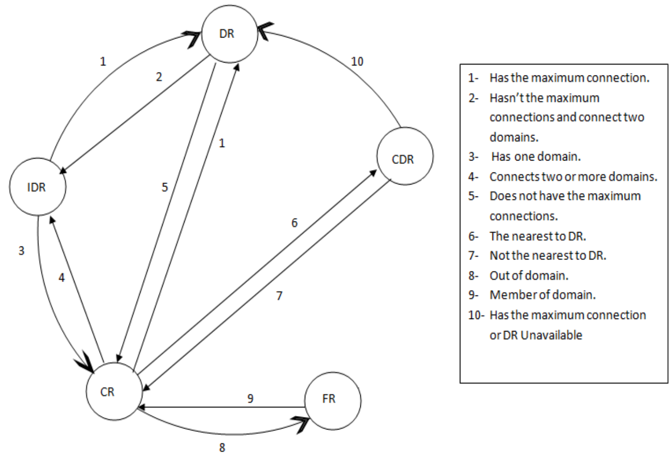


Figure2. Transition state diagram of MANET Devices

C. DR and CR selection

The selection criteria of DR is an important issue as it affects over all network .generally there are many aspects that affect the selection of DR as Power level(**P**) , connectivity(**C**) and Mobility(**M**) .

Generally the DR is selected based on any one of the above mentioned factors, but there are some approaches that considered all of these factors, by getting summation of **P**, **C** and **M**, and the selection of DR based on the greatest sum.

In our work we didn't consider the mobility factor as it is difficult to determine and specially at the begging of device state determination process as this factor is unpredictable because of its dynamic characteristics which leads to incorrect DR selection or frequent DR reformation which increase network overhead.

The proposed work considers the following factors:

- Power level (**P**)–The remaining power in the battery of Mobile device, and this factor changed during the running process.
- The Ability (**A**)-The ability of device depends on many factors: Processor Power, Bandwidth and the installed systems. This factor is predefined and doesn't change during the transaction process with others. This factor contains logical value (True or false).
- Connectivity(**C**) – Number of connected devices, the maximum **C**, the maximum Connectivity.

The first two factors are very important and we couldn't include them in a quantitative calculation as the absences of one of them will violate the role of these devices. Instead we replace factors P and A by one parameter which we will call it Readiness factor(**R**).

This factor mean that the device is capable of being DR, IDR or CDR. The value of **R** can be false or true and simply computed by using algorithm1.

DR selection consists of two phases:

1) Domain refinement.

At this step the devices that are not capable to be DR

Will be excluded from DR candidates list and this can be accomplished by using algorithm1

Algorithm1. Finding the Readiness factor(R) of device

//AP- Acceptable Power level
//MP- maximum power
AP=MP/2;
R=false
If A=true then
Begin
Repeat
If P>= AP then R=true
Else AP=AP- 0.1.AP
Until R=True or AP<0.1.MP or Ready message received
If R=true Send ready message.
End.

2) DR Determination

After determination of R, the process of **DR** selection became very simple and can be described by the following algorithm2.

Algorithm2. DR Selection

- Each device sends Hello Message (HM) that informs neighbors about its existence. Then starts to compute **R** by using algorithm1. At the end of this step each Device maintain a list of neighbors, and DR candidates.
- Each capable Device(R=true) Multicast the number of connected devices(C) to DR candidates.
- DR is the device with Maximum C, and changes its current state to DR State.
- Wining DR Broadcasts a message which we call DR Wining Message (DRWM) to inform all domain members.
- Domain members change their state to CR and send Change state Message (CSM) to the neighbors, this step prevents inclusion of CR into more than one Domain and help the others members to correct their **C** factor.

The states are arranged into a priority manner to enhance the state selection criteria of each member which based on the capability, number of connections and location. The state priority of DR is the highest one as shown in table1.

TABLE1. STATE PRIORITY OF DOMAIN MEMBERS.

Domain Member	DR	IDR	CDR	CR	CF
Priority	1	2	3	4	5

C. IDR Selection

The selection of IDR consists of two phases:

The first phase is finding the readiness factor(**R**) of the device by using algorithm1, and the second phase can be described in algorithm 3.

Algorithm3. IDR Determination
// R[CR i]- R factor of i^{th} CR.
// C- number of Client Routers.
For i=1 to C
Begin
If (R[CRi] =true)and (CRi connected to another(DR(s)or CR(s) or IDR(s)) from different domain) Then
Begin
Inform all available DR
If CR received an acceptance message then
CRi changes its state to IDR
End
End.

Before changing CRs state to IDR state it has to receive Acceptance message from all available DRs to enhance security measures and maintenance issues

D. CDR Selection

The selection of CDR is initiated by DR by the following conditions:

- CDR should pass readiness factor test.
- CDR is the nearest CR to DR.

The first condition can be achieved by using algorithm1, and the second condition can be accomplished by using algorithm4.

Algorithm4. Determination the Nearest CR
//CDRTP- CDR Transmission Power
//DRTP- DR Transmission Power
// MXTP- Maximum Transmission Power
// MNTP-Minimum Transmission Power
Ω =0.05.MXTP
Δ=MNTP
CDRFound=false
Repeat
Send CDR Discovery Message
If (CDR Response Message(s) received)
Then
Begin
Select CR with Maximum C.
DR sends CDR selection Message to CR
CR changes its state to CDR
CR broadcast Change status Message
End
Else
Δ=Δ+Ω.
Until CDRFound=True
CDRTP =DRTP + Δ

Δ- is the transmission power among DR and CDR.

DR informs the selected CDR by including this value in

CDR Selection Message, this value is very important for CDR. By using the Δ value. CDR can determine its transmission power that guarantees the connectivity to all domain members. See figure3.

After the selection of DR, IDR, CDR and CR the status of other devices can be simply determined by algorithm5.

Algorithm5. CF Determination
//C- Connections number
If (Device status not determined) then
If (Device C=0) Then Device Status= Unconnected
Else Begin
Device status=CF
Send Device Status Message to Neighbor
End.

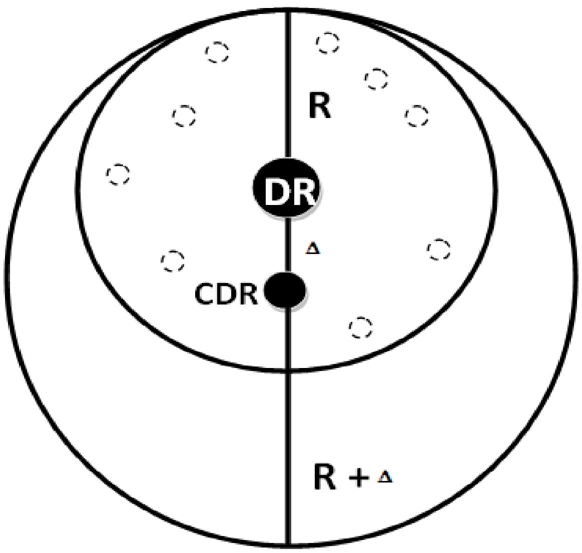


Figure3. CDR Transmission Range.

IV. ROUTING THE MESSAGE

Each member in the domain maintains a list of neighbors addresses s. Each Device in the domain also stores the address of the DR and CDR. DR also maintains members and IDR Addresses. Whenever a device generates a request to transfer the data to a particular device, it checks the destination address in its list. If the matching device is found in the address list, message is transferred to that device. If no match is found, then the Message will be sent to DR. DR will again check for the match in its Address table. If no match is found, DR will forward Message to IDR. This process will continue till the destination device is reached. If Destination device not found or hops exceed maximum acceptable hops count, unreachable error message will be reported. See fig.4.

V. CONCLUSION AND FUTURE WORK

Dynamic and unpredictable topology, limited bandwidth, limited resources in terms of battery and storing capacity are the major characteristics of MANET. The proposed scheme of domain formation improves many characteristics of MANET. This scheme is applicable and does not depend on a specific architecture or topology of MANET, it doesn't require central device to start domain formation.

The proposed Scheme reduces the storage space requirement of domain members by storing only the addresses of neighbors. Each device does not require maintaining the addresses of all MANET members.

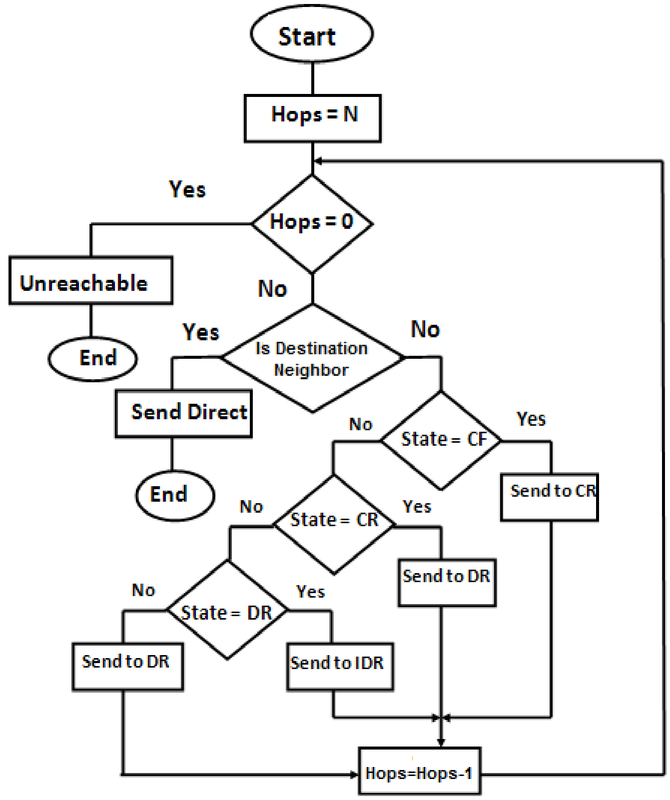


Figure4. Flow chart of the Message Routing.

It reduces broadcasts messages during the message routing or during domain formation which reduces the power consumption and enhances performance.

The proposed scheme is generally scalable and adaptive to mobility due to the fact that each node relies on local information collected from nearby nodes.

The stability achieved by Selecting CDR and by the way the domain formed and the messages forwarded.

Our future work aims to find Security and Maintenance solutions to our proposed Scheme.

REFERENCES

[1] "Mobile ad-hoc network" fromhttp://en.wikipedia.org/wiki/ Mobile_ad-hoc_network.

[2] Elizabeth M.Royer and Chai-Keong Toh, "A Review of current routing protocols for Ad-HocMobile Networks", *IEEE Personal Communications,Vol. 6, No. 2, pp. 46-55, April 1999.*

[3] S.Samanta, S.S.Ray, S.SenGupta, M.K.Naskar, "A Novel Algorithm for Managing Network Configuration", Asian International Mobile Computing Conference 2006, Kolkata, Jan 04-07, pp. 51-58.

[4] S.S.Basu and A.Chaudhari, "Self-adaptive Topology Management for Mobile Ad- hoc Network", *IE(I) Journal-ET, Vol.84, July 2003.*

[5] Soumya Sankar Basu, Atal Chaudhari, "Self- Adaptive MANET: A CentralizedApproach", *Foundations of Computing and Decision Sciences, Vol.29, 2004*

[6] R. Wattenhofer, L. Li, P. Bahl, and Y.-M.Wang, ªDistributed Topology Control for Power Efficient Operation in Multihop Wireless Ad Hoc Networks,º *Proc. Infocom 2001*, Apr. 2001.

[7] Errol L. Lloyd, Rui Liu,Madhav V. Marathe Ram Ramanathan S. S. Ravi Algorithmic Aspects of Topology Control Problems for Ad hoc Networks *IEEE Mobile Ad Hoc Networking and Computing (MOBIHOC)*, June 2002.

[8] R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet architecture: an overview," 1994, IETF RFC 1633.

[9] Yu-Xuan Wang, Forrest Sheng Bao "An Entropy-based Weighted Clustering Algorithm and Its Optimization for Ad hoc Networks" Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2007) 0-7695- 2889-9/07 2007

[10] S. Blake, D. Black, M. Carlson, E. Davies, Z.Wang, andW. Weiss, "An architecture for differentiated services," 1998, IETF RFC 2475.

[11] J. Wu and H.L.Li, "On Calculating Connected Dominating Set for Efficient Routing in Adhoc Wireless Networks",*Proc 3rd Int'l. Wksp. Discrete Algorithms and Methods for Mobile Comp. and Commun*. 1999 pp. 7 – 14

[12] Qi Xue , Aura Ganz "Ad hoc QoS on-demand routing (AQOR) in mobile ad hoc networks ". Journal of Parallel and Distributed Computing,(63), 154-165, 2003.

[13] Taewook Kang, Jangkyu Yun, Hoseung Lee,Icksoo Lee,Hyunsook Kim,Byunghwa Lee,Byeongjik Lee,Kijun Han "A Clustering Method for Energy Efficient Routing in Wireless Sensor Networks", Proceedings of the 6th WSEAS Int. Conf. on Electronics, Hardware, Wireless and Optical Communications, Corfu Island, Greece, February 16-19, 2007 133.

[14] M. Mirhakkak, N. Schult, D. Thomson, "A new Approach for providing Quality-of-Service in a Dynamic Network Environment", The MITRE Corporation, Network and Communications Engineering Department, Reston, Virginia.

[15] Y.Z.P Chen and A.L Liestman, "Approximating Minimum Size Weakly – Connected Dominating Sets for ClusteringMobile Adhoc Networks", *in Proc 3rd ACM Int'l. Symp. Mobile Adhoc Net. & Comp*, June 2002, pp. 165 – 72.

[16] Hannan XIAO, Winston K.G. SEAH, Anthony LO and Kee Chiang CHUA " Flexible Quality Service Model for Ad-HOC Networks". In proceedings of the IEEE Vehicular Technology Conference, Tokio, Japan, May 2000.

[17] Tzay-Farn shih and Hsu-Chun Yen "Core Location- Aided Cluster-Based Routing Protocol for Mobile Ad Hoc Networks Proceedings of the 10th WSEAS International Conference on Communications, " Vouliagmeni, Athens, Greece, July 10-12, 2006 (pp223-228).

[18] SWAN Project: http://comet.columbia.edu/swan, 2002.

[19] Vikas Kawadia and P.R.Kumar, "Power Control and Clustering in Adhoc Networks", *IEEE Proc. INFOCOMconference* 2003, pp 459 - 469.

[20] P. Basu, N. Khan, and T.D.C. Little, "A Mobility Based Metric for Clustering in Mobile Adhoc Networks", *in Proc.IEEE ICDCSW'01*, Apr 2001, pp. 413 – 18.

AUTHOR PROFILE

Adwan AbdelFattah is an assistant professor at the Computer Science Department of the Arab American University of Jenin, Palestine. Previously he worked at Philadelphia and Zarka Private University, Jordan. He received his PhD degree from the National Technical University of Ukraine in 1996. His research interests include Computer Networks, Computer Architecture, Cryptography and Networks Security.

Backup Communication Routing Through Internet Satellite, WINDS for Transmitting of Disaster Relief Data

Countermeasure for Round Trip Delay which occurs in between Satellite and Ground with Network Accelerator

Kohei Arai
Dept.of Information Science,
Graduate School of Science and Engineering
Saga University
Saga city, Japan

Abstract—**A countermeasure for round trip delay which occurs in between satellite and ground with network accelerator is investigated together with operating system dependency on effectiveness of accelerator. Also disaster relief data transmission experiments are conducted for mitigation of disaster together with acceleration of disaster related data transmission between local government and disaster prevention center. Disaster relief information including remote sensing satellite images and information from the disaster occurred areas to local government for creation of evacuation information is accelerated so that it becomes possible to send them to the residents in the suffered areas due to disaster through data broadcasting in the digital TV channel.**

Keywords-internet satellite; disaster mitigation; TCP/IP protocol; accerelator.

I. INTRODUCTION

Satellite communication is widely used [1], [2]. In particular, Internet communication is also widely used through geostationary satellites [3], [4].

Japanese first Internet communication satellite: WINDS[1] was launched and put into the geostationary orbit in 2010. Since then, many experiments with the satellite have been conducted to demonstrate usefulness and effectiveness of the satellite for e-learning among the Asian countries, disaster mitigation, and distance medicine etc.

One of the major problems on the WINDS satellite is delay time in the TCP/IP communications. Round trip time between the geostationary orbit altitude and the ground is 0.6 seconds. Other than this, there is 0.2 seconds of delay time in the repeater of ATM Switcher [2] onboard WINDS satellite. Therefore, the delay time may affect to the throughput, in particular, due to the IP communication protocol (Acknowledge communications). There is another influence due to the packet window size. When the window size is fixed and small, then influence on throughput is large.

In order to overcome such influences, hardware accelerator gives a solution. Hardware accelerator allows adjustment of the window size then throughput is recovered in somehow. Some experiments have been conducted for confirmation of the effect of hardware accelerator with WINDS satellite. Also an attempt has been made for creation of software accelerator[3] in order to provide more flexibility as well as reduce the manufacturing cost.

Next section describes the configuration and procedure of the experiments with WINDS satellite followed by scientific purpose of the experiment of disaster relief data transmissions for disaster mitigation. Then the proposed software accelerator is described with conclusion and some discussions.

II. EXPERIEMTS WITH WINDS SATELLITE

A. System Configuration

Fig.1 shows the configuration of the experiments with WINDS satellite. There are three sites of transmitting and receiving stations. Any station may transmit and receive data through WINDS satellite. SkyX[4] of hardware accelerator is employed at the stations as shown in Fig.2.

Under the antenna, there is low noise amplifier and down convertor. After that the received data are sent to the SkyX through outdoor unit and indoor unit. Then work station or PC receives the data through router and switching HUB.

---

[1] http://www.jaxa.jp/countdown/f14/special/column_j.html

[2] http://www.cisco.com/en/US/products/hw/switches/ps718/products_command_reference_chapter09186a00800f025b.html

[3] http://www.brothersoft.com/downloads/network-accelerator.html
http://www.softpedia.com/get/Tweak/Network-Tweak/Cloud-Drive-Network-Accelerator.shtml

[4] http://www.netone.co.jp/seminar/tfa9q100000031eb-att/06DSE_SkyX.pdf

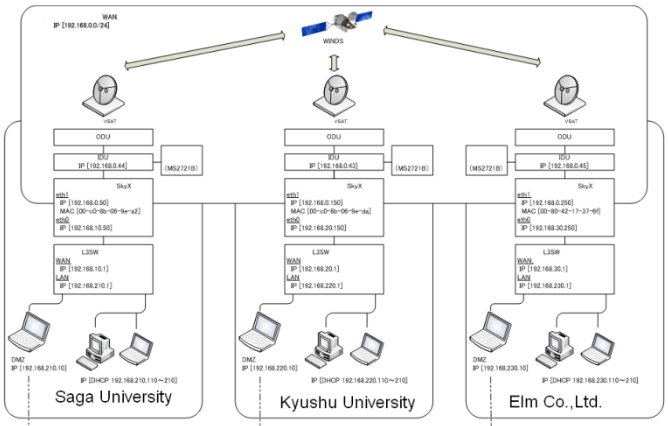


Figure 1. Configuration for the experiments with WINDS satellite.

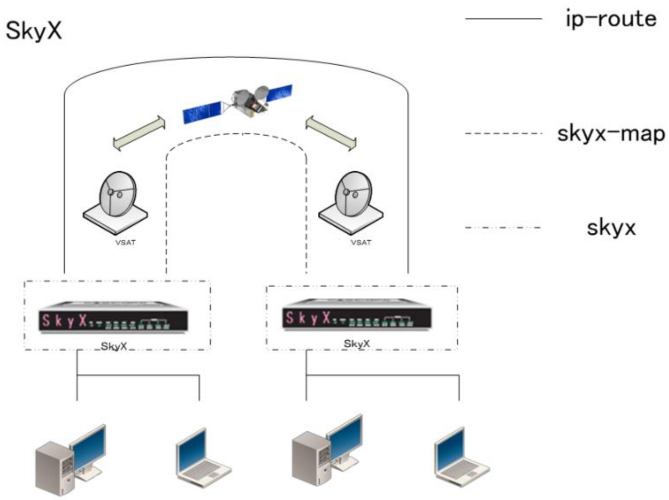


Figure 2. SkyX setup configurations

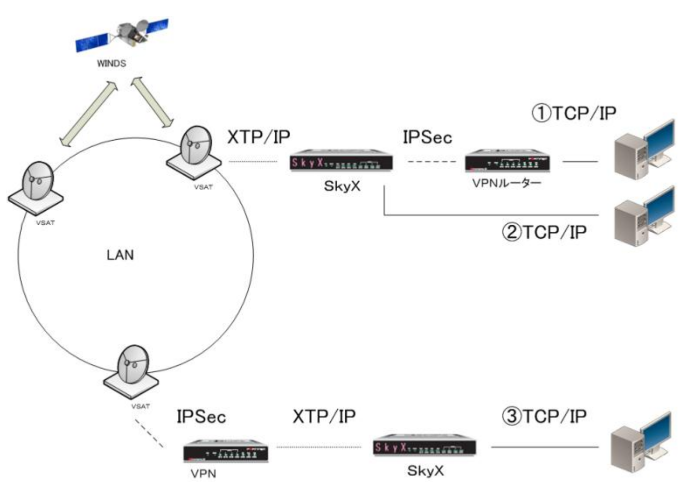


Figure 3. Some configurations do not work.

SkyX does not work because TCP/IP packets is sent to receiver through VPN router with tunneling on IPsec[5], ① and ② does not work. On the other hand, SkyX does work for this configuration. It is meaningless because VPN[6] is constructed Under the WINDS-VSAT[7] network, SkyX is valid though.

B. Experimental Results on ftp Transmission

Ftp data transmission experiments are conducted with the following conditions, (1) Data rate for uplink: 51 MBps, (2) Output power is set for data transmission of 30 MBps without any packet losses. The experimental results are shown in Table 1. As the result, it is found that ftp transmission with SkyX is125 time faster than that without SkyX

TABLE 1. FTP DATA TRANSMISSION EXPERIMENTAL RESULTS

SkyX ON File: WindsTestSmall.dat (1079064 bytes)		
Sender/Receiver	Time(s)	Data rate(Kbytes/sec)
Kyushu U.←Saga U.	1.06	1016.07
Saga U.←Kyushu U.	1.08	1000.99
SkyX OFF File: WindsTestSmall.dat (1079064 bytes)		
Sender/Receiver	Time(s)	Data rate(Kbytes/sec)
Kyushu U.←Saga U.	129.92	8.31
Saga U.←Kyushu U.	146.63	7.36
SkyX ON File: WindsTest.dat (310677846 bytes)		
Sender/Receiver	Time(s)	Data rate(Kbytes/sec)
Kyushu U.←Saga U.	84.61	3671.88
Saga U.←Kyushu U.	75.94	4091.20

C. OS Dependency

In order to check OS dependency, Microsoft Windows OS of VISTA and XP are tested for comparison with the ftp transmission of WindsTestSmall.dat (1079064 bytes). Table 2 shows the results.

It is confirmed that Auto Tuning function included in the Windows Vista does work for acceleration of ftp data transmission through TCP/IP protocol. Data transmission rate is improved by 4.97 time for Windows VISTA in comparison to the Windows XP.

D. Data Transmission Experiments for Disatser Mitigation

Data transmission experiments for disaster mitigation are conducted with MODIS[8] satellite data and disaster relieved data with MODIS through TCP/IP and UDP[9] protocols with and without SkyX.

---

5 http://searchmidmarketsecurity.techtarget.com/definition/IPsec

6 http://strongvpn.com/GC_packages_japan.shtml?gclid=CMu4k56k6KoCFQZhgwodJF9S9g

7 http://en.wikipedia.org/wiki/Very_small_aperture_terminal

8 http://modis.gsfc.nasa.gov/

9 http://tsunami-udp.sourceforge.net/

TABLE 2. COMPARISON OF FTP DATA TRANSMISSION PERFORMANCE BETWEEN WINDOWS VISTA AND XP.

Vista-Vista:SkyX ON 1.33sec 813.77 kbyte/sec
Vista-Vista:SkyX OFF 29.56sec 36.54 kbyte/sec
XP-XP:SkyX OFF 146.85sec 7.4 kbyte/sec
Vista-Vista:SkyX OFF 29.56sec 36.54 kbyte/sec

Fig.4 shows examples of MODIS data and MODIS derived disaster relieved data.

Figure 4. Examples of MODIS data and MODIS derived disaster relieved data

Fig.5 (a) shows the transmitting file size, the data size which has to be transmitted, and transmission rate during the ftp transmission of MODIS data without SkyX. The horizontal axis shows the different types of MODIS data. Meanwhile, Fig.5 (b) shows the file size, the data size, and the transmission rate during the ftp transmission of MODIS data with SkyX. In comparison between Fig.5 (a) and (b), the transmitting file size for both is quite different. The transmitting file size is followed by the data size which has to be transmitted in the case of with SkyX. On the other hand, Fig.5 (c) shows those of data transmission performances together with the number of packet losses in the UDP protocol of data transmission. In the data transmission through UDP protocol, the transmitting file size does not follows the data size which has to be transmitted due to loss of packet.

E. Internet Connectivity

Internet connectivity is checked with the configuration shown in Fig.6. Trace rout command is delivered to Yahoo homepage in the Internet from Saga University to Kyushu University through WINDS satellite. Fig.7 shows trace route window display when Saga University access to the Yahoo home page.

F. Influence Due to Rainfall Attenuations

Influence due to rainfall on data transmission with Ka band[10] of WINDS satellite frequency channel is confirmed. There are two chances of rainfall, August 22 2010 and August 28 2010 during our two weeks experiment as shown in Table 3.

TABLE 3. RAINFALL RATE ON 22 AND 28 AUGUST 2010.

8/22	Kyushu university	around 8:40	Rmax=8.0mm
	Saga university	around 12:00	Rmax=13.5mm
8/28	Kyushu university	7:50-10:50	Rmax=0.5mm
	Saga university	11:00-12:00	Rmax=0.5mm

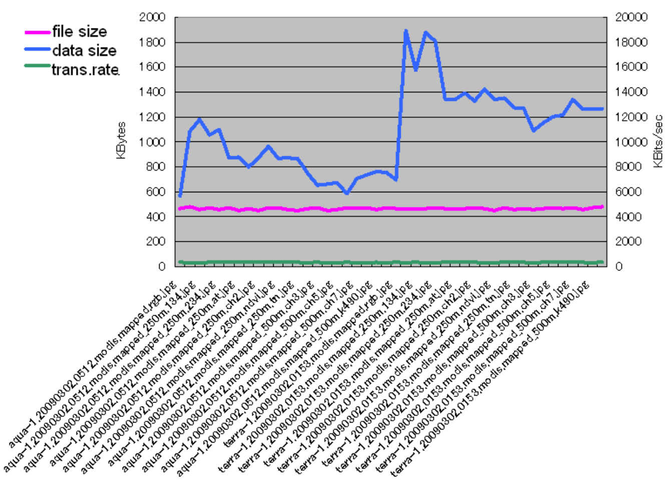
The most severe case (received signal is getting down to -75db and C/No[11] is also getting down to 81db) is occurred at Saga University on around noon on August 22 2009. Fig.8 shows the attenuation due to rainfall which was observed at Saga University on 22 August 2010.

Rainfall influence compensator is equipped in WINDS satellite. Sending power would be better to suppress for consideration of affection to the others (C/No has to be up to 95 dB). The most preferable compensation factor has to be calculated in advance. This can be done through experiments under rainfall

G. Software Accelerator

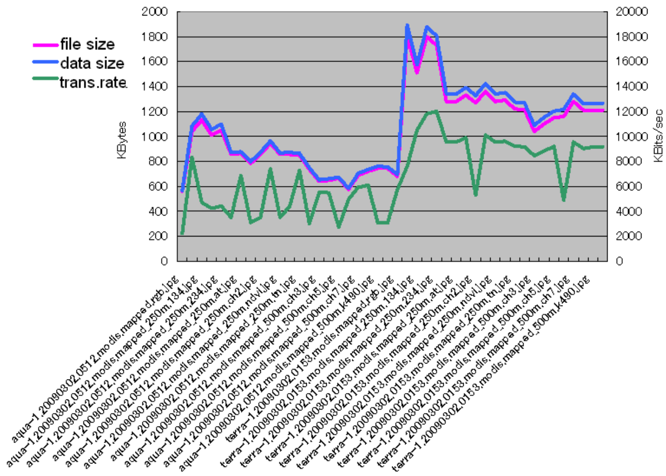
Network accelerator is developed with (1) a shortened acknowledge process, (2) maximum buffer size information transfer.

(1) The required time for transferring acknowledge is shortened. Actually acknowledge is returned immediately after against TCP packet. After that data packers are transferred to the receiver with the other protocol. Although the time required for acknowledge transfer is almost zero of delay time, data packets have to be sent to the receiver after all.

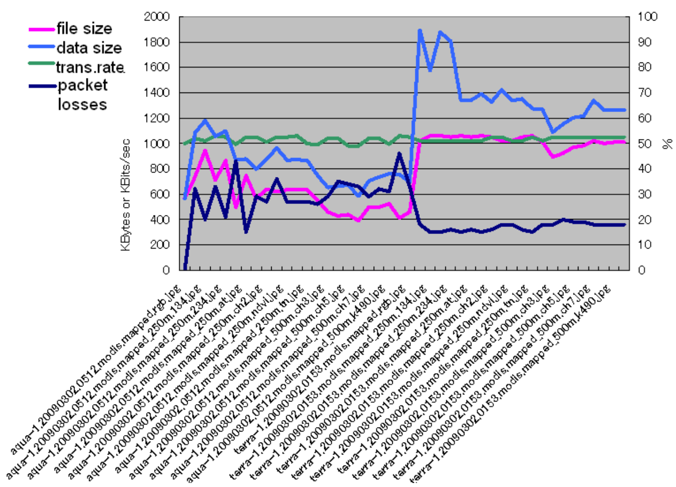


(a) TCP/IP protocol of data transmission without SkyX

[10] http://www.tech-faq.com/ka-band.html
[11] Carrier noise ratio



(b) TCP/IP data transmission with SkyX



(c)UDP data transmission

Figure 5. Data transmission performance through TCP/IP protocol with and with SkyX as well as UDP protocol

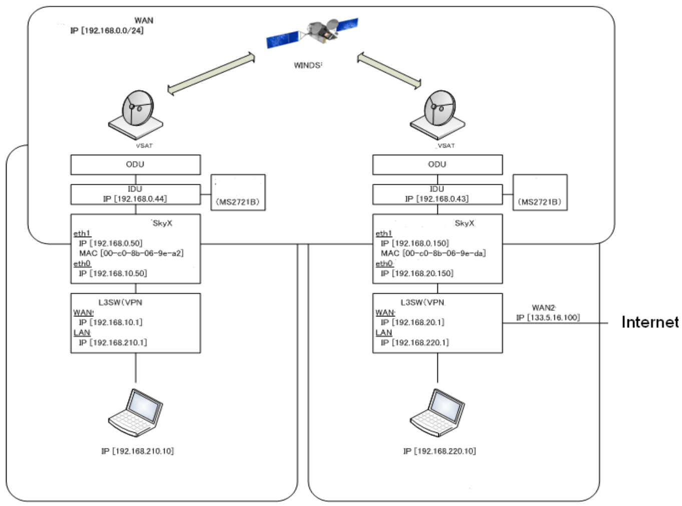
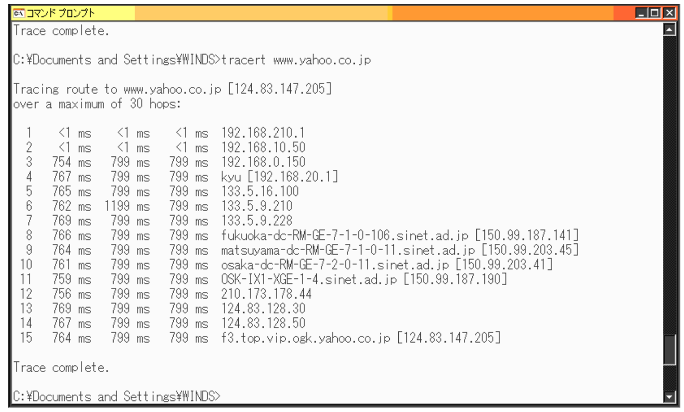
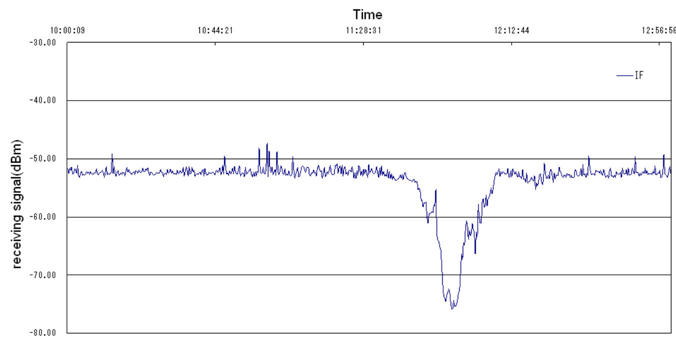


Figure 6. Configuration of internet connectivity

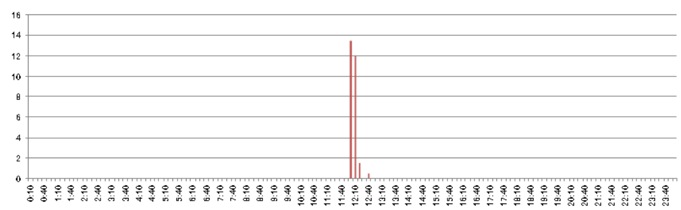


```
コマンド プロンプト
Trace complete.

C:¥Documents and Settings¥WINDS>tracert www.yahoo.co.jp

Tracing route to www.yahoo.co.jp [124.83.147.205]
over a maximum of 30 hops:

  1    <1 ms    <1 ms    <1 ms  192.168.210.1
  2    <1 ms    <1 ms    <1 ms  192.168.10.50
  3   754 ms   799 ms   799 ms  192.168.0.150
  4   767 ms   799 ms   799 ms  kyu [192.168.20.1]
  5   765 ms   799 ms   799 ms  133.5.16.100
  6   762 ms  1199 ms   799 ms  133.5.9.210
  7   769 ms   799 ms   799 ms  133.5.9.228
  8   766 ms   799 ms   799 ms  fukuoka-dc-RM-GE-7-1-0-106.sinet.ad.jp [150.99.187.141]
  9   764 ms   799 ms   799 ms  matsuyama-dc-RM-GE-7-1-0-11.sinet.ad.jp [150.99.203.45]
 10   761 ms   799 ms   799 ms  osaka-dc-RM-GE-7-2-0-11.sinet.ad.jp [150.99.203.41]
 11   759 ms   799 ms   799 ms  OSK-IX1-XGE-1-4.sinet.ad.jp [150.99.187.190]
 12   756 ms   799 ms   799 ms  210.173.178.44
 13   769 ms   799 ms   799 ms  124.83.128.30
 14   767 ms   799 ms   799 ms  124.83.128.50
 15   764 ms   799 ms   799 ms  f3.top.vip.ogk.yahoo.co.jp [124.83.147.205]

Trace complete.

C:¥Documents and Settings¥WINDS>
```

Figure 7. The result of the trace route command when Saga University access to the Yahoo home page through WINDS satellite.



(a) Receiving signal from WINDS satellite to Saga University on August 22 2009.



(b) Coincidence between dip of receiving signal and rainfall at Saga

Figure 8. Attenuation due to rainfall which was observed at Saga University on 22 August 2010.

(2) TCP automatic window size tuning based on RFC 1323[12] and Windows VISTA are to adjust the buffer size of the receiver. Windows VISTA also adjust the buffer size of transmitter. Although the receiver's buffer size is known, data amount which has not received yet by the application software is unknown. During the delay time of 0.8 seconds for satellite communications, most of data are not finished to process if the current PC capability is taken into account. Namely, actual buffer size considering the PC capability and application software processing speed is not transferred to the transmitter and the receiver. Therefore, software acceleration is attempt to replace the buffer size of receiver to the maximum receivable buffer size.

[12] Request for Comments

Fig.9 shows the software accelerator control panel. With this panel, all the buffer size can be monitored and all the operation modes and parameters are selected.

Fig.10 shows Microsoft Network Monitor 3.3[13] utilized packet monitor free software. Using this software, all the packet on the Local Area Network is monitored together with network performance evaluation. It can be monitored the flags of SYN and FIN on the TCP segment which are corresponding to the ftp connection and disconnection with the packet monitor software. Then the time required for transmission is evaluated. Actually, the time required for data transmission evaluated with the packet monitor software is shorter than that of the required time which measured on the DOS windows. Therefore, the additional time is taken into account for getting close to the actual required time.

Experiment on software accelerator is conducted between Saga University and Kyushu University with up-link data rate of 51 Mbps on 27 August. The experimental results are as follows,

(1) Acknowledge process, Maximum window size is transfer, data used: WindsTestSmall.dat(1079064Byte): required time is 1.426757sec

(2) Acknowledge process, without transfer the maximum window size, data used: WindsTestSmall.dat(1079064Byte): required time is 1.426757sec

(3) Without acknowledge process, without transfer the maximum window size, data used: WindsTestSmall.dat(1079064Byte): required time is 367.4569725sec

(4) Acknowledge process, Maximum window size is transfer, data used: WindsTest.dat(310677846Byte): required time is 64.688477sec

(5) SkyX, data used: WindsTestSmall.dat(1079064Byte): required time is 1.08sec

(6) SkyX, data used: WindsTestl.dat(310677846Byte): required time is 75.94sec

Consequently, the proposed software does work as SkyX. SkyX converts TCP protocol to XTP: Xpress Transport Protocol[14]. The proposed software accelerator has almost same functionality as SkyX, XTP protocol specification does not clear though.

III. CONCLUSION

As the experimental results, it is found that

(1) Antennas for communication to the internet satellite, WINDS can be set-up within a 10minutes,

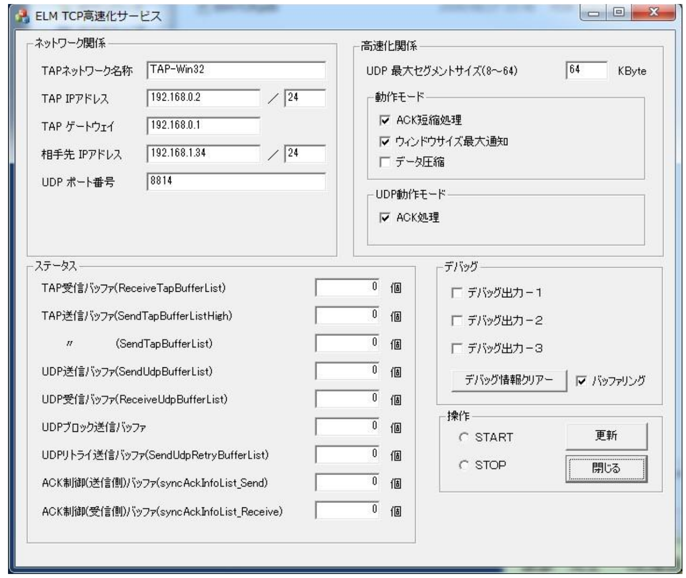


Figure 9. Software accelerator control panel

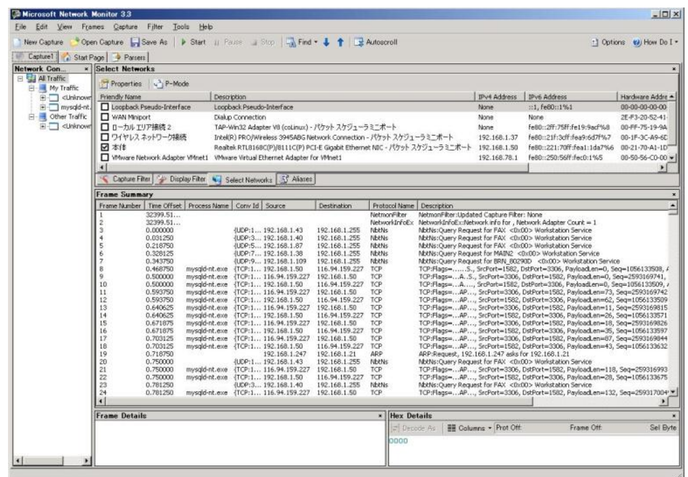


Figure 10. Microsoft Network Monitor 3.3 utilized packet monitor free software.

(2) Network accelerator shows an enough performance for transmission of disaster relief data of MODIS satellite data and moving pictures of disaster relief,

(3) Also acceleration of transmission of disaster information from the residents to local government and evacuation information from local government to the residents

(4) Software accelerator functions are confirmed. It does work effectively and is almost similar effectiveness of the hardware accelerator.

ACKNOWLEDGMENT

The author would like to thank to Prof.Dr.Kiyotaka Fujisaki pf Kyushu University and Mr.Miyahara and Miyauchi of ELM Co.Ltd. for their great effort to conduct the experiments.

REFERENCES

[1] Data Coomunications Protocol Performance on Geo-stationary Satellite Links (Hans Kruse, Ohio University, 1996)

13 http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=4865

14 http://www.weblio.jp/content/XTP

[2] Takashi Iida Satellite Communications: System and Its Design Technology, IOS Press, 2000, ISBN 4274903796, 9784274903793

[3] "WildBlue: High Speed Satellite Internet Provider". *Official web site*. http://get.wildblue.com/pricing.html. Retrieved July 17, 2011.

[4] Martyn Williams (December 27, 2010). "European broadband-Internet satellite launched". *Network World*. http://www.networkworld.com/news/2010/122710-european-broadband-internet-satellite.html. Retrieved July 17, 2011.

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

Embedded Object Detection with Radar Echo Data by Means of Wavelet Analysis of MRA: Multi-Resolution Analysis

Kohei Arai
Department of Information Science
Saga University
Saga City, Japan

***Abstract*—A method for embedded object detection with radar echo data by means of wavelet analysis of MRA: Multi-Resolution Analysis, in particular, three dimensional wavelet transformations is proposed. In order to improve embedded object detecting capability, not only one dimensional radar echo data but also three dimensional data are used. Through a comparison between one dimensional edge detection with Sobel operator and three dimensional wavelet transformation based edge detection, it is found that the proposed method is superior to the Sobel operator based method**

Keywords-Three dimensional wavelet transformation; Radar echo data; wavelet MRA; Edge detection.

I. INTRODUCTION

Measuring instrument of Pulsed Radar for Underground object detections and object shape estimations: PRU data is used to use for detecting embedded objects underground in general. The problems of the PRU utilized embedded object findings are (1) range and azimuth resolution is not good enough for small object detections, (2) poor resolution of radar echo signal receivers in the case that there is small difference of dielectric characteristics between embedded objects and its surrounding materials. Other than these, (3) multiple reflections from embedded objects and surrounding materials are another problems together with (4) interference due to multiple echo from embedded objects and surrounding materials such as boundary layers, tiny stones, etc.

Through human perceptions with PRU radar echo data, in general, embedded objects are used to find usually. Therefore, it requires huge experiences and knowledge for the person in charge on human perceptions. Meanwhile, automatic detections with differentiation operator of embedded objects from the radar echo data are used to attempt. The method for estimation of the shape of embedded objects and medium of material parameters from radar echo data is proposed [1]. Also the method utilizing filtering and correlation analysis for noise removals is proposed together with the preprocessing of wavelet analysis based method [2]. On the other hand, the method for feature extractions in the frequency domain which uses Fourier transformation is proposed. Meanwhile, Hough transformation based method is proposed for extraction of embedded objects utilizing the well known feature of which cylindrical shape of embedded objects show bi-polar function of the curves theoretically [4].

The conventional embedded object finding methods are based on the edge detection algorithms utilizing the differentiation operator and wavelet based Multi Resolution Analysis [5]: MRA. In the methods, radar echo data are treated as one dimensional data. One dimensional differential operators and MRA are used to use in the conventional methods. Radar echo data are essentially three dimensional data. Therefore, it is possible to improve edge detection performances by using three dimensional differentiation operators and three dimensional MRA.

The following section describes the proposed three dimensional MRA for detection of embedded objects with radar echo data followed by a comparison of edge detection performance between the conventional one dimensional differentiation operators (The well-known Sobel operator in the paper) and the proposed three dimensional MRA. Then conclusions and some discussions are followed.

II. PROPOSED METHOD

A. *Pulsed Rdar for Underground Object Detections and Object Shape Estimations*

Fig.1 shows the principle of the pulsed radar for underground object detections and object shape estimations. Radar echo data is the function of time, *f(t)* at the position in concern and is scanned in two dimensional ground surface, *f(t:x,y)*.

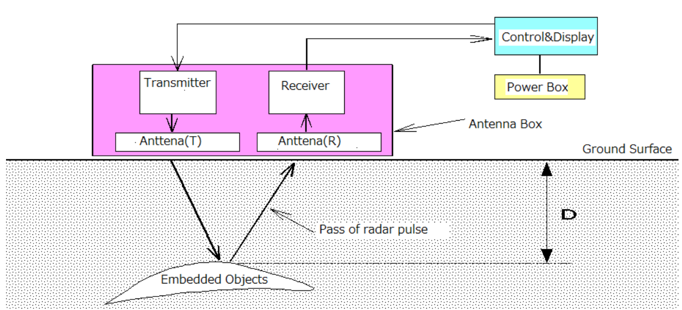


Figure 1. Principle of the Pulsed Radar for Underground object detections.

In the figure, the depth from the ground surface is expressed as following equation (1)

$$D = \frac{V \cdot T}{2} = \frac{C \cdot T}{2\sqrt{\varepsilon_r}} \qquad (1)$$

where V and C denotes, respectively, electric magnetic wave propagation speed of the underground and the air. T denotes time interval of the pulse repeat cycle while ε_r denotes di-electric constant of the soils of the underground.

Cylindrical or spherical shapes of embedded objects in the underground show the edges as a parabolic function in the pulsed radar echo image as shown in Fig.2.

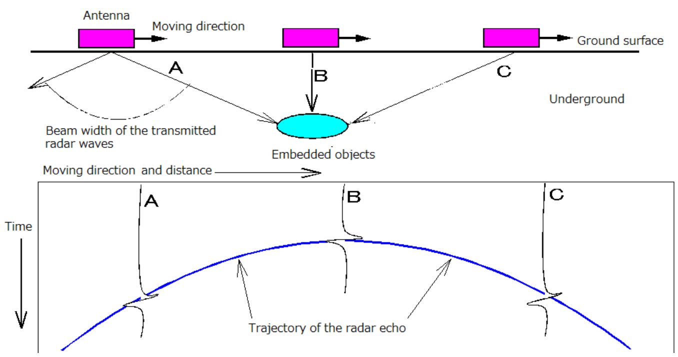


Figure 2. Radar echo waveforms from cylindrical and spherical shapes of embedded objects in the underground

If the embedded objects are such cylindrical and spherical shapes of objects, object shapes and three dimensional location of the object can be estimated with a prior knowledge of the radar echo waveforms. It, however, is not always appropriate of the assumption of the object shapes. For instance, there are some embedded cluster pipes and box type of pipes. It have to be estimated and determined the object shapes and the three dimensional locations without such assumption.

B. The Proposed Method based on Wavelet Multi Resolution Analysis: MRA

In order to estimate and to determine the object shapes and the three dimensional locations, edges in the radar echo is extracted. In the conventional method for edge detection, one dimensional differentiation in time domain is used to apply to the acquired radar echo. Also one dimensional wavelet transformation or Multi Resolution Analysis: MRA is applied to the radar echo data then inverse wavelet transformation or reconstruction with high frequency component is applied to the wavelet transformed data thus edges are extracted, in general. Radar echo data, in essential, is three dimensional data.

Therefore, edge detection performance of the three dimensional MRA or wavelet transformation based edge detection method is superior to the one dimensional methods. Namely, two dimensionally scanned radar echo signals shows three dimensional data. Then three dimensional MRA is applied to the three dimensional radar echo data. After that, reconstruction is made without low frequency components results in three dimensional edges are detected. It could be superior to the conventional one dimensional edge detection with one dimensional differentiation or differential operators.

C. Daubechies base function based Wavelet Transformation

One dimensional wavelet transformation is expressed with the equation (2).

$$F = Cn\, f \qquad (2)$$

where F, f denotes wavelet frequency component and radar echo signal as a function of time. Cn denotes wavelet transformation matrix which is expressed as a bi-orthogonal function based on base functions. Cn can be determined with a reference to the appendix. Therefore, $CnCn^t=I$. Then f is converted to $F_1=(L_1,H_1)$, $F_2=CnL_1=(L_2,H_2)$, $F_3=CnL_2=(L_3,H_3)$, and $F_m=CnL_{m-1}=(L_m,H_m)$. Also f is reconstructed as $Cn^{-1}F_m=Cn^{-1}(L_m,H_m)=L_{m-1},\ldots\ldots,Cn^{-1}F_2=L_1$, $Cn^{-1}F_1=f$.

There are some based functions such as Haar, Daubechies[1], etc. Through the experiment with preliminary simulation of radar echo data, Daubechies base function is selected. Daubechies base function is one of bi-orthogonal functions. Then $F=(L_1,H_1)$.where L_1, H_1 denotes low and high frequency components, respectively. If inverse wavelet transformation is applied to the H_1 component and zero filled L_1 component, then one dimensional edge (high frequency component) can be extracted.

One dimensional wavelet transformation can be expanded to two dimensional transformations easily. Namely,

$$F=[Cn[Cmfxy]^t] \qquad (3)$$

where Cm denotes another wavelet transformation matrix while fxy denotes two dimensional scanned radar echo data. As the result, $F=(LL_1,LH_1,HL_1,HH_1)$ with four frequency components in x and y directions, LL_1,LH_1,HL_1,HH_1. If the two dimensional inverse wavelet transformation is applied to the F with zero filled LL_1 and the other frequency components, then two dimensional edges are extracted.

Pulsed radar echo data is expressed with three dimensional data essentially, $fxyz$. x and y directions are corresponding to the scanned directions of the ground surface while z direction is corresponding to time. Then three dimensional wavelet transformations are defined with the equation (4).

$$F=[Cn[Cm[Clfxyz]^t]^t] \qquad (4)$$

Thus eight frequency components are derived. Then the three dimensional inverse wavelet transformations is applied to F with zero filled LLL_1 and the other frequency components results in three dimensional edges are extracted.

[1] Daubechies base function is defined as $\{\alpha_k\}$ satisfying the following conditions,

$$\phi(x) = \sum_k \alpha_k \sqrt{2}\, \phi(2x-k) \qquad (1)$$

$$\beta_k = (-1)^k \alpha_{1-k} \qquad (2)$$

$$\phi(x) = \sum_k \beta_k \sqrt{2}\, \phi(2x-k) \qquad (3)$$

III. EXPERIMENTS

A. Data Used

PRU of SIR-2 which is manufactured by GSSI: Geophysical Survey Systems, Inc. with the center frequency of 400MHz is used to acquire the pulsed radar echo data of embedded objects of the road from the Kami-Tozu to the Shimo-Tozu, Kokura ward, Kita-Kyushu City, in Japan on 2nd of April 2004. Fig.3 shows the structure in the underground of this road.

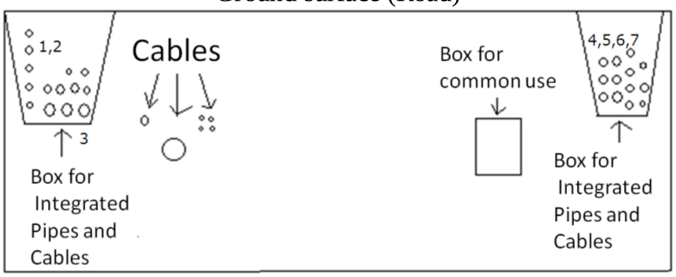


Figure 3. Cross section of the underground structure of the embedded objects.

The embedded objects as shown in Table 1 are found from the underground of the road in concern.

TABLE 1. OBJECTS WHICH ARE EMBEDDED IN THE UNDERGROUND OF THE ROAD IN CONCERN. DISTANCE DENOTES THE DISTANCE FROM THE LEFT END OF THE ROAD. DEPTH 1 SHOWS THE PREDICTION RESULTS WHILE DEPTH 2 DENOTES THE ACTUAL DEPTH OF THE EMBEDDED OBJECTS THROUGH HUMAN PERCEPTIONS WITH PRU RADAR ECHO DATA.

Object No.	Object name	Distance(m)	Depth1(m)	Depth2(m)
1	Cables	1.0-2.0	0.5-1.2	0.7-1.3
2	Pipes	1.25	1.5	1.3
3	Common use boxes	4.95	Unknown	Unknown
4	Old pipes	17.3	1.2	1.1
5	Cables	18-18.3	0.75-1.5	1.2
6	Common use boxes	19.3-20.5	Unknown	Unknown
7	Pipes	21	1.1	1.1

The common use boxes include some pipes and cables for commonly used for the local governments and public agents and the others. The distances of the common use boxes are known while their depth is unknown. These data are not found in the old list. Also it is difficult to estimate the depth and dimensions of the common use boxes because it is not identifiable in the PRU radar echo data as well. The cables and pipes are situated at the two locations, respectively. Their dimensions are known. The old pipes are not useable situations.

The depth and the dimensions of the embedded objects derived from human perceptions are coincident to the predicted depth and dimensions which are estimated from the record which is created at the construction of the road. Although human perceptions with PRU radar echo data seems reliable, it takes huge experiences and knowledge.

B. Attempt for Finding Unknown Objects of Common Use Boxes

As it is mentioned above, the common use boxes are difficult to find the location and shapes (dimensions) through human perceptions with PRU radar echo data. An attempt is made to estimate the location and shapes of the common use boxes with the proposed method. Fig.4 shows the PRU radar echo data (reflected echo signals).

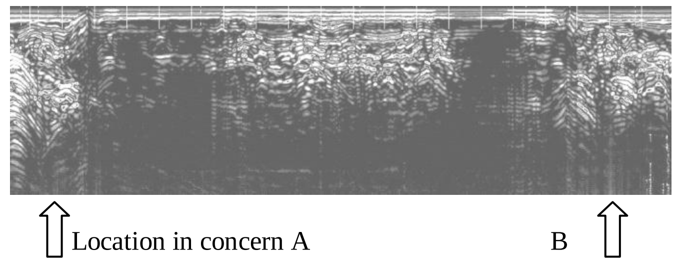


Figure 4. Received scanned PRU radar echo data which corresponds to cross section of the road. The top row numbers ranges from 1 to 7 are corresponding to the embedded objects which are listed in the Table 1.

In the Fig.4, the locations of two unknown embedded objects are indicated with location in concern A and B, respectively. It is hard to identify by human perceptions even for the expertise persons. On the other hand, the other objects are recognizable and are identified the location and the shapes.

Three dimensional representations of PRU radar echo data is illustrated in Fig.5.Three dimensional edges are detected with the proposed MRA based method. Namely, if the reconstruction process is made without low frequency component, LLL1, then high frequency component of edges can be detected. Three dimensional edges have much information rather than two dimensional edges and also rather than one dimensional edge.

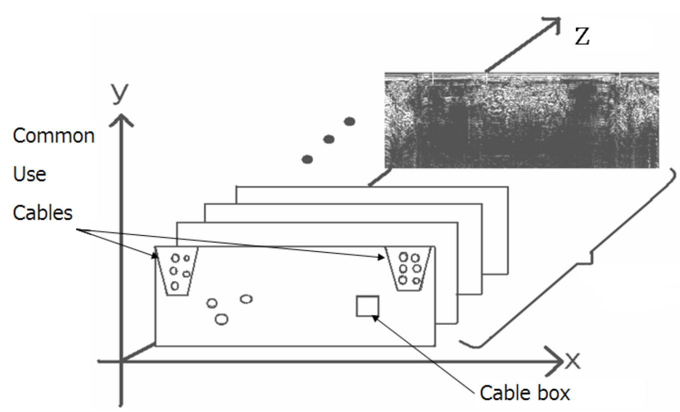
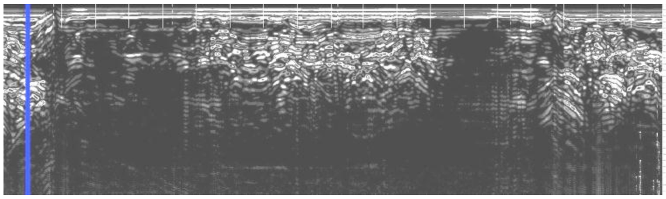


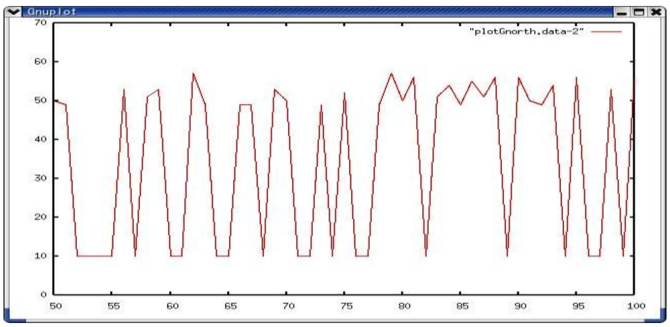
Figure 5. Conceptual three dimensional PRU of radar echo of the underground of the road.

C. Experimental Results

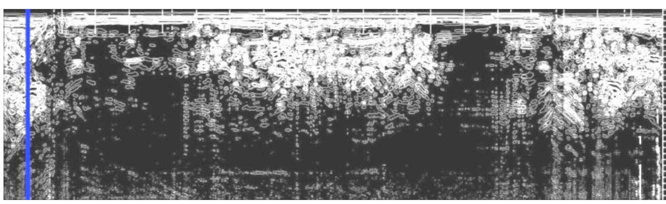
One of the examples of the experimental results is shown in Fig.6. Fig.6 (a) and (g) shows PRU radar echo data and shows the location of embedded objects No.3 and No.6. In the figure, not only PRU radar echo but also one dimensional profile in the direction of depth repeatedly.



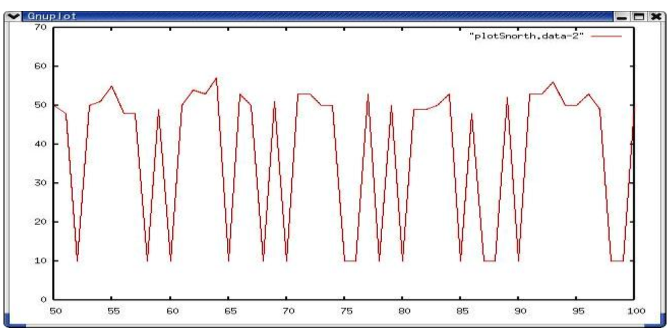
(a) The location in concern A in the PRU radar echo data



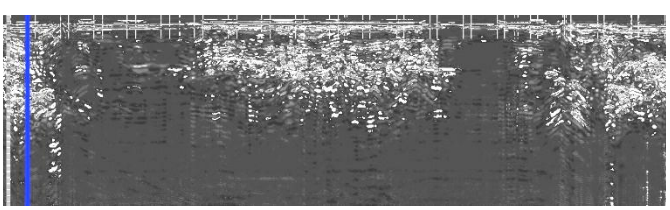
(b)One dimensional profile at the location A in the depth direction



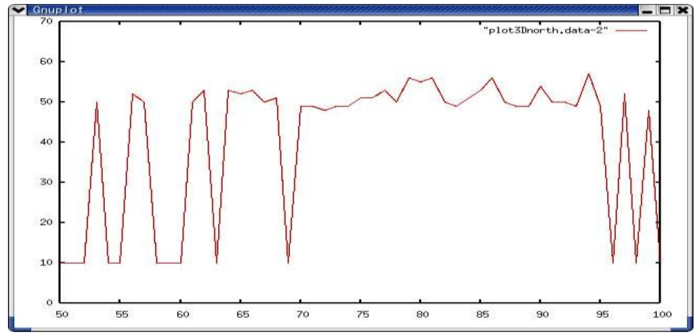
(c)Detected edges of the location A based on Sobel differentiation operator



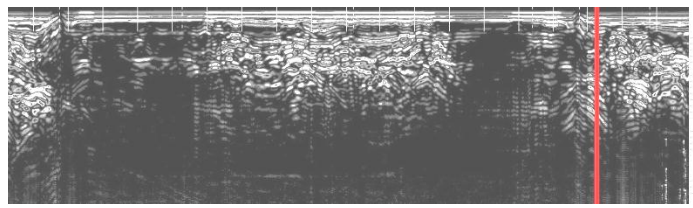
(d)One dimensional profile of the location A in the depth direction(edges are detected with Sobel differentiation operator)



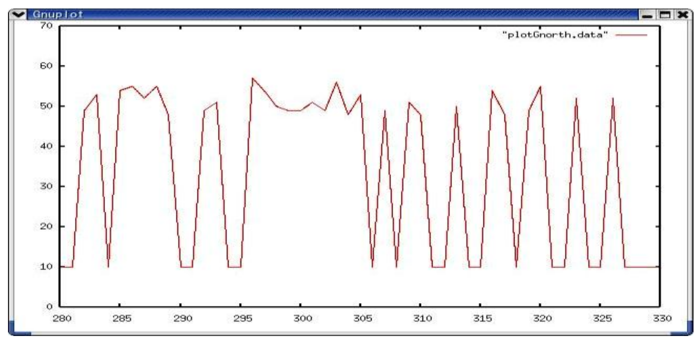
(e)Detected edges of the location A based on the proposed three dimensional MRA based method with three dimensional data of PRU radar echo data



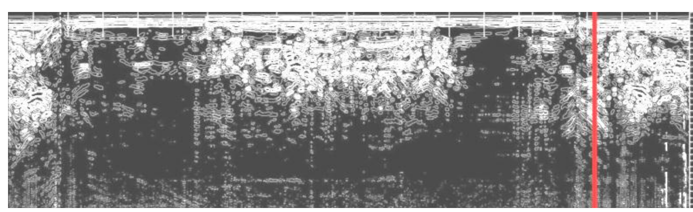
(f)One dimensional profile of the detected edges of the location A based on the proposed three dimensional MRA based method with three dimensional data of PRU radar echo data



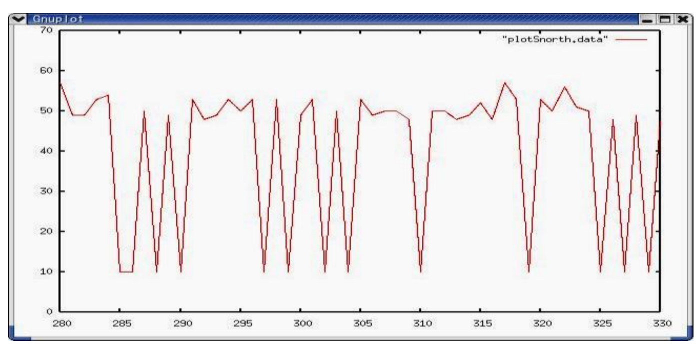
(g)The location B in the PRU radar echo data



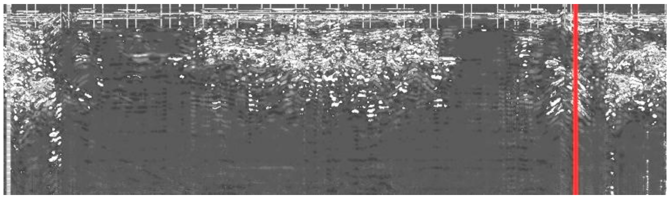
(h)One dimensional profile of the location B in the depth direction



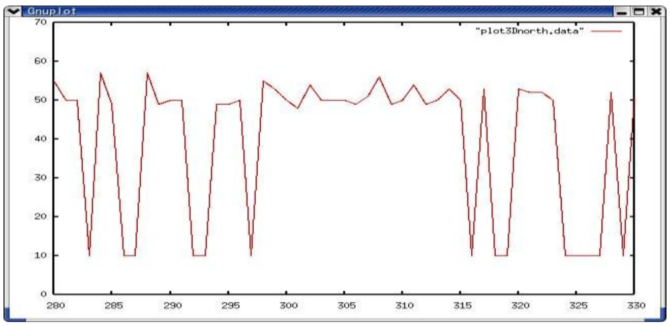
(i) Detected edges of the location B by using Sobel operator



(j) One dimensional profile of the location B of the detected edges with Soble operator



(k)Detected edges of the location B based on three dimensional MRA based method with three dimensional PRU radar echo data



(l)One dimensional profile of the detected edges of the location B based on three dimensional MRA based method with three dimensional PRU radar echo data

Figure 6. Comparisons of the detected edges by Sobel operator and the proposed three dimensional MRA based method

D. Discussionss

In the case of Sobel operator, undesired edges are detected results in valuable and desirable edges are hidden by the detected undesired edges. Meanwhile, the proposed three dimensional MRA based edge detection method allows detection of desirable edges depending on frequency components of edges. In this paper, although one level, single stage of MRA is used, variety frequency components of edges can be detected by changing the level. MRA does work as filter bank.

Fig.6 (f), (i) shows the edges which are corresponding to the common cable and pipes boxes with the proposed three dimensional MRA based edge detecting method. The location at which the edges are detected is totally equal to the boxes locations. Meanwhile, the resultant images of Sobel operator of the boxes does not show clear edges which are corresponding to the boxes.

IV. Conclusion

It may concluded that

(1) the proposed three dimensional MRA based edge detection is superior to the conventional edge detection method with differentiation operator such as Sobel, Laplasian, etc. in terms of sensitivity to undesired edges,

(2) The proposed three dimensional MRA based edge detection method is superior to the one dimensional MRA based edge detection method in terms of three dimensional edge detection,

(3) Detected edges are coincident to these of which the expertise persons' results, in particular, the edges which are difficult to detect such as the common cable and pipe boxes in the experiment of the roads.

Acknowledgement

The author would like to thank to GIS Kyushu Co. Ltd. for providing the PRU radar data as well as the information of the embedded objects in the underground of the road in concern. Also the author would like to thank to joint research member of the cooperated research was made between GIS Kyushu Co. Ltd. and the Saga University.

References

[1] Tanaka, et al., Surveillances of the embedded objects by means of time domain analysis, Proceedings of the 5th Workshop on the Underground Measurements with Electro Magnetic Waves, 67-72, 1999.

[2] Tada, et al., Investigation of characteristics with the noise removal method utilizing two dimensional parabolic function of wavelet for PRU radar data, Proceedings of the 5th Workshop on the Underground Measurements with Electro Magnetic Waves, 73-78, 1999.

[3] Nagashima, et al., Technology of embedded object findings by means of pattern recognition method in frequency domains, Proceedings of Space Aeronautics Electronics Symposium, SANE-87-54, 1987

[4] Yamauchi, H., Image processing methods for underground object identifications with PRU radar images, Ritsumeikan University Symposium on Image Processing and Analysis, pp.21-30,1999.

[5] Arai, K. Fundamental Theory of Wavelet Analysis, Morikita Shuppan Publicing Co. Ltd., 2000.

Appendix (*Wavelet transformation matrix*)

For instance, the 8th order[2] of Daubechies base function based Cn with the support length of two can be expressed with the equation (A1).

$$
C_8^{[2]}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix}=\begin{bmatrix}p_0 & p_1 & & & & & & \\ q_0 & q_1 & & & & & & \\ & & p_0 & p_1 & & & & \\ & & q_0 & q_1 & & & & \\ & & & & p_0 & p_1 & & \\ & & & & q_0 & q_1 & & \\ & & & & & & p_0 & p_1\\ & & & & & & q_0 & q_1\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix}
=\begin{bmatrix}p_0x_1+p_1x_2\\ q_0\eta_1+q_1\eta_2\\ p_0\eta_3+p_1\eta_4\\ q_0\eta_3+q_1\eta_4\\ p_0\eta_5+p_1\eta_6\\ q_0\eta_5+q_1\eta_6\\ p_0\eta_7+p_1\eta_8\\ q_0\eta_7+q_1\eta_8\end{bmatrix} \quad \text{(A1)}
$$

The 8th order of Cn with the support length of four is also expressed with the equation (A2).

[2] Scalar data consists eight numerical data, η_1 to η_8 is assumed.

$$C_8^{[4]}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix}=\begin{bmatrix}p_0 & p_1 & p_2 & p_3 & & & & \\ q_0 & q_1 & q_2 & q_3 & & & & \\ & & p_0 & p_1 & p_2 & p_3 & & \\ & & q_0 & q_1 & q_2 & q_3 & & \\ & & & & p_0 & p_1 & p_2 & p_3\\ & & & & q_0 & q_1 & q_2 & q_3\\ p_2 & p_3 & & & & & p_0 & p_1\\ q_2 & q_3 & & & & & q_0 & q_1\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\\ \eta_5\\ \eta_6\\ \eta_7\\ \eta_8\end{bmatrix}$$

$$=\begin{bmatrix}p_0\eta_1+p_1\eta_2+p_2\eta_3+p_3\eta_4\\ q_0\eta_1+q_1\eta_2+q_2\eta_3+q_3\eta_4\\ p_0\eta_3+p_1\eta_4+p_2\eta_5+p_3\eta_6\\ q_0\eta_3+q_1\eta_4+q_2\eta_5+q_3\eta_6\\ p_0\eta_5+p_1\eta_6+p_2\eta_7+p_3\eta_8\\ q_0\eta_5+q_1\eta_6+q_2\eta_7+q_3\eta_8\\ p_0\eta_7+p_1\eta_8+p_2\eta_1+p_3\eta_2\\ q_0\eta_7+q_1\eta_8+q_2\eta_1+q_3\eta_2\end{bmatrix} \qquad \text{(A2)}$$

pi and qi in the equations (A1) and (A2) is also expressed with equation (A3) and (A4), respectively.

$$\begin{aligned}&\left(C_n^{[2]}\right)^T C_n^{[2]}=I_n\\ &p_0+p_1=\sqrt{2}\\ &q_0=p_1\\ &q_1=-p_0\\ &0^0q_0+1^0q_1=0\end{aligned} \qquad \text{(A3)}$$

$$\begin{aligned}&\left(C_n^{[4]}\right)^T C_n^{[4]}=I_n\\ &p_0+p_1+p_2+p_3=\sqrt{2}\\ &q_0=p_3\\ &q_1=-p_2\\ &q_2=p_1\\ &q_3=-p_0\\ &0^0q_0+1^0q_1+2^0q_2+3^0q_3=0\\ &0^1q_0+1^1q_1+2^1q_2+3^1q_3=0\end{aligned} \qquad \text{(A4)}$$

These equation can be expanded to the general support length of Cn as shown in equation (A5).

$$\begin{aligned}&\left(C_n^{[\sup]}\right)^T C_n^{[\sup]}=I_n\\ &\sum_{j=0}^{\sup-1}p_j=\sqrt{2}\\ &q_j=(-1)^j p_{((\sup-1)-j)} \quad \left(j=0,1,2,\ldots,(\sup-1)\right)\\ &\sum_{j=0}^{\sup-1}j^r q_j=0 \quad \left(r=0,1,2,\ldots,(\frac{\sup}{2}-1)\right)\end{aligned} \qquad \text{(A5)}$$

Therefore, the coefficients of the Daubechies base function can be determined from the solution of the equation (A5).

AUTHORS PROFILE

Kohei ARAI
Saga University
Saga, Japan

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

An improved Approach for Document Retrieval Using Suffix Trees

N. Sandhya
Associate Professor,
CSE Department,
GRIET,
Hyderabad, Andhra Pradesh, 500 072, India

Y. Sri Lalitha
Associate Professor, CSE Department,
GRIET,
Hyderabad, Andhra Pradesh, 500 072, India

Dr. A. Govardhan
Principal,
JNTUH College of Engineering
Jagityal, Andhra Pradesh 500 501, India

Dr. K. Anuradha
Professor and Head,
CSE Department,
GRIET,
Hyderabad, Andhra Pradesh, 500 072, India

***Abstract*—Huge collection of documents is available at few mouse clicks. The current World Wide Web is a web of pages. Users have to guess possible keywords that might lead through search engines to the pages that contain information of interest and browse hundreds or even thousands of the returned pages in order to obtain what they want. In our work we build a generalized suffix tree for our documents and propose a search technique for retrieving documents based on a sort of phrase called word sequences. Our proposed method efficiently searches for a given phrase (with missing or additional words in between) with better performance.**

Keywords-Document retrieval; Frequent Word Sequences; Suffix tree; Traversal technique.

I. INTRODUCTION

With the growth of web, hundreds of millions of people engage in information retrieval process every day when they use web search engine or search their emails. IR is fast becoming the dominant form of information access, overtaking traditional database style searching. IR process begins when user enters a query like search strings in web search engines, phrases etc. to identify the related documents or URLs.

Now almost all the documents have electronic copies. With the development of WWW it is an efficient technique to retrieve the documents using the web search engines based on a query. But this should not be time consuming. That is the reason precision of the retrieval of related documents for a given query is vital for the search engine. Cluster based information retrieval techniques also exist [11].

The next section deals with the Information Retrieval and its related work on text documents. Section 3 describes Suffix Tree. Section 4 deals with building generalized suffix tree. Section 5 explains traversal technique Algorithm used for quick retrieval of documents. Section 6 shows the experiment setting, results and analysis. Section 7 concludes and discusses future work.

I. RELATED WORK

Information Retrieval for a given query is retrieving relevant documents efficiently. An application needs to be developed that facilitates the user with an efficient retrieval of the information that is needed. Search engines are the keys to find specific information on the World Wide Web. Without sophisticated search engines, it would be virtually impossible to locate anything on the Web. A search engine is a program that searches documents for specified keywords and returns a list of the documents where the keywords are found based on certain algorithms. Document clustering has initially been investigated in Information Retrieval mainly as a means of improving the performance of search engines by pre-clustering the entire corpus [12].

The assumption (implicitly or explicitly) upon which most commercial information retrieval systems are based is that if a query and a document have a keyword in common, then the document is about the query to some extent and if there are more key words in common, then the document is about the query. In this respect, an IR System represents the documents and also query in separate vector space models as document - terms matrix, where each column indicates terms in documents and rows correspond to documents in IRS. This representation is also called “bag of words” mechanism. An IR system matches the bag of keywords in the user’s query with the bag of keywords representing the documents to identify related documents and this approach suffers from a number of problems.

- It does not handle lexical variation, i.e. different words are used to represent the same meaning or concept in queries and documents.
- It cannot deal with semantic variation, where a single word has multiple meaning.

- It does not handle properly syntactical variation, i.e. words that co-occur in multiple documents are probably related.
- Morphological variations i.e. words appearing in different numbers (singular or plural) and in different cases (Active, passive cases)

All the above problems hurt the retrieval system in terms of precision and recall [8].

To overcome the bag of words problems, here we choose to treat text documents as sequence of words and to retrieve documents that share frequent word sequences from text databases. The sequential relationship between the words and documents is preserved using suffix tree data structure.

II. SUFFIX TREE

A suffix tree is a data structure that admits efficient string matching and querying. Suffix trees have been studied and used extensively, and have been applied to fundamental string problems such as finding the longest repeated substring [6], strings comparing [4], and text compression [5]. The suffix trees became useful as the search time is independent of the length of the string. The following description of the suffix tree was taken from Dan Gus field's book on Strings, Trees and Sequences [7]. A suffix tree of a string is simply a compact trie of all the suffixes of that string. Here we treat documents as sequences of words, not characters. The main purpose of using Suffix trees is that it is used to identify the document IDs that contain the suffixes efficiently because the leaf nodes of suffix tree stores documents IDs. Suffix trees are useful in clustering [10].

A. Definition

A suffix tree T for an m-word string S is a rooted directed tree with exactly m leaves numbered 1 to m. Each internal node, other than the root, has at least two children and each edge is labeled with a nonempty sub-string of words of S. The label of a node is defined to be the concatenation of the edge-labels on the path from the root to that node. No two edges out of a node can have edge labels beginning with the same word. For each suffix s of S, there exists a suffix node whose label equals s.

The suffix trees are fast, incremental and are constructed in linear time of the suffixes generated.

III. CONSTRUCTION OF SUFFIX TREES FOR DOCUMENTS

A text document D is viewed as a sequence of words, so that it can be represented as D = (w1, w2, w3 . . .), where w1, w2,w3, . . . are words appearing in D. Like a frequent itemset in the association rule mining of a transaction data set [9], a word set is frequent when at least the specified minimum number (or percentage) of documents contains this word set. A frequent word set containing k words is called frequent k-word set.

A frequent k-word sequence is an FS with length k, such as FS = (w1, w2. . . wk), and it has two frequent subsequences of length k - 1, which are (w1, w2, . ., wk-1) and (w2,w3, . . ,wk). [2].

In our work finding the frequent word sequences has two steps: finding frequent 2-word sets first, then finding frequent word sequences of all length by using the Generalized Suffix Tree (GST) data structure.

A. Finding frequent 2-word sets

The goal of this step is to reduce the dimension of the database (i.e. the number of unique words) by eliminating those words that are not frequent enough to be in a frequent k-word sequence, for k >= 2. This step is simple and straight forward. We use an association rule miner to find the frequent 2-word sets that satisfy the minimum support. All the words in frequent 2-word sets are put into a set. After finding the frequent 2-word sets, we remove all the words in the documents that are not in WS. After the removal, the resulting documents are called compact documents. Let us consider an example database

D = {d1, d2, d3}:

d1: Almost all children eat chocolates.
d2: Some of the children eat dry fruits.
d3: Children like to eat dry fruits and chocolates.

There are 13 unique words in this database D: {all, and, almost, children, chocolates, dry, eat, fruits, like, of, some, to, the}. If we specify the minimum support as 60%, the minimum support count is 2 for this case. The set of frequent 2-word sets is {{children, eat}, {children, chocolates},{eat, chocolates}, {children, dry}, {children, fruits}, {dry, fruits},{eat, fruits}, {eat, dry}, }; and WS = {children, eat, chocolates, dry, fruits}. After removing those words not in WS, the database D becomes D'={d1', d2', d3'} as follows, where the removed words are shown in parentheses.

- d1': (Almost all) children eat chocolates.
- d2': (Some of the) children eat dry fruits.
- d3': children (like to) eat dry fruits (and) chocolates.

Thus reducing the dimensions of documents, this has a considerable impact in the next step i.e. building of a generalized suffix tree. To find frequent word sequences of the database, we adopted the suffix tree [6], a well known data structure for sequence pattern matching, to find all the frequent word sequences. Each compact document is treated as a string of words and inserted into a generalized suffix tree (GST) one by one. Finally, by collecting the information stored in all the nodes of the GST, we can find all the frequent word sequences of the database.

A suffix tree for a string S is actually a compressed trie for the non-empty suffixes of S. A GST is a suffix tree that combines the suffixes of a set of strings. In our case, we build a GST of all the compact documents in the text database.

Each suffix node has a box attached, and it contains the document id set of the suffix node. After building the GST, we traverse it by depth-first. On the way down, the labels of the edges are concatenated to become the string L of each node. On the way up, each child node sends its document id set to its parent. The support count of the label (i.e. string L) of this parent node is the size of the union of all the document id sets of its children. By checking the support count and the length of the label of each node, we can get the information about all the frequent word sequences in the database. In our example shown above, we have seven nodes in the GST, and the details are given in Table I.

Since the minimum support for frequent words is set to 60% in this example, the minimum support for frequent word sequences could not be smaller than 60%. Only those words whose support is at least 60% are kept in the compact documents, so that we can find only the frequent word sequences with that minimum support.

TABLE I: WORD SEQUENCES ASSOCIATED WITH THE NODES IN FIGURE 1

Node no	***Word sequence***	***Length of word sequence***	***Document Ids***
1	chocolates	1	D1.txt D3.txt
2	dry fruits	2	D2.txt D3.txt
3	dry fruits chocolates	3	D3.txt
4	fruits	1	D2.txt D3.txt
5	fruits chocolates	2	D3.txt
6	children eat	2	D1.txt D2.txt D3.txt
7	children eat chocolates	3	D1.txt
8	children eat dry fruits	4	D2.txt D3.txt
9	children eat dry fruits chocolates	5	D3.txt
10	eat	1	D1.txt D2.txt D3.txt
11	eat chocolates	2	D1.txt
12	eat dry fruits	3	D2.txt D3.txt
13	eat dry fruits chocolates	4	D3.txt

After building the suffix tree as mentioned above, we traverse the tree for a given word sequence " eat chocolates ", which should retrieve all the documents that contain **"children eat chocolates", "children eat dry fruits and chocolates" " children of four years eat many chocolates".** To perform this we propose SuffixTree algorithm for sequence of words. Here we used level order traversal to find the word "eat", after getting to the node with word "eat", perform depth first search to get all the strings that start with node "eat" applying k-mismatch method to get the document ids of all the documents in which the word sequence occurs.

IV. ALGORITHM

Step 1: Given word sequence to be searched is tokenized first.

```
Step 2: Initialize n to 1
        Search the nodes of root for first token
        If there is a match
                Perform step3
        Else
                If n is last node of root +1
                        Return search failed
        Else
                Increment n and again search with next node of root

Step 3: Consider only the Level1 nth node sub tree
        Compare next token with Level2 first node
        do
        {
           If there is match
              (i) perform depth first search traversal on the tree
                  comparing with remaining tokens
             (ii) applying K-mismatch retrieve the documents
                  that contain the word sequence
           Else
              If all the Level1 nth node sub tree nodes are
              traversed word sequence is not present
              Else
                 apply K-mismatch, perform DFS traversal and
                 compare with the next token
        }
        While all the tokens of word sequence are not
        completed or entire Level1 nth node sub tree nodes
        are not traversed
```

V. EXPERIMENTAL SETUP

Suffixes of the phrases are generated [1, 10]. We treat the documents as a sequence of words instead of bag of words. Then the similarity measurements are done based on the shared frequent word sequences among the documents. Each document is reduced to a compact document by keeping only the frequent words [1]. A generalized suffix tree for all the compact documents is built. The frequent word sequences and the documents sharing them are found. We proposed an approach to improve the precision of retrieval. We used level order traversal with depth first traversal of the GST to search for the related documents based on word sequence.

A. Cleaning of documents and generating suffixes

Preprocessing of documents involves removal of all the special symbols called nonword tokens (such as numbers, HTML tags, and most punctuation) from each of the documents, splitting the document contents line wise, removing unwanted characters, removal of stop words and stripping other text. Sentence boundaries are marked. Identifying sentence boundaries is an important task as the approach used here is word sequence based method for finding suffixes.

B. Generating Suffixes and Building GST

The suffix of each line of the document is generated. After finding all the suffixes of all the documents we extracted unique suffixes and built compact suffix tree with these unique suffixes. All Suffixes and Unique Suffixes for the given example are shown in the Table II. We pruned the tree with nodes that do not satisfy the user specified threshold, thus the tree contains only those nodes that are frequent word sequences. A Generalized Suffix Tree (GST) for the given example is shown in Fig 1. Then we apply our algorithm for traversing the GST in order to retrieve documents when a phrase with missing or additional words is given.

TABLE II. RESULTS

All suffixes	***Unique suffixes***
children eat chocolates	children eat chocolates
children eat dry fruits	children eat dry fruits
children eat dry fruits chocolates	children eat dry fruits chocolates
chocolates	chocolates
chocolates	dry fruits
dry fruits	dry fruits chocolates
dry fruits chocolates	eat chocolates
eat chocolates	eat dry fruits
eat dry fruits	eat dry fruits chocolates
eat dry fruits chocolates	fruits
fruits	fruits chocolates
fruits chocolates	

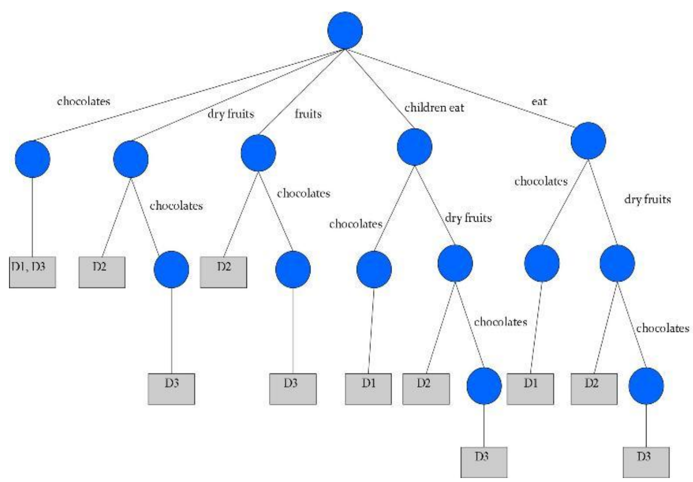


Figure 1. Generalized Suffix Tree:

VI. CONCLUSIONS

Our method is an efficient method over phrase based retrieval. Phrase based retrieval requires the complete phrase to appear in documents. We relaxed this condition and matched related phrases using k-mismatch method. Here by using k-mismatch the words sequence is maintained with some additional words/missing words in between. Our method proved efficient compared to simple phrase based retrieval in two aspects: a) we pruned the infrequent terms thus reduced the dimensions and b) proposed algorithm for efficient suffix tree traversal. Our results have shown better performance than simple phrase matching. Suffix trees can be constructed incrementally. In our future work we would like to apply this technique on Opinion Mining. Also we would like to extend this technique for concept retrieval. We also would like to investigate the application of this algorithm in cluster based information retrieval with hierarchal, hybrid and incremental clustering.

REFERENCES

[1] Yanjun Li, Soon M. Chung, John D. Holt Text document clustering based on frequent word meaning sequences.

[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th VLDB Conference, 1994, pp.

[3] Horatiu Mocian Text Mining with suffix trees.

[4] A. Ehrenfeucht and D. Haussler. A new distance metric on strings computable in linear time. Discrete Applied Math, 1988.

[5] M. Rodeh, V. R. Pratt, and S. Even. Linear algorithm for data compression via string matching. In Journal of the ACM, pages 28(1):16–24, 1981.

[6] P. Weiner. Linear pattern matching algorithms. In The 14th Annual Symposium on Foundations of Computer Science, pages 1–11, 1973.

[7] D. Gusfield. In Algorithms on strings, trees and sequences: computer science and computational biology. Cambridge University Press, 1997.

[8] A.T. Arampatzis, T.Tsoris, C.H.A. Koster, Th.P. Vander Weide, RIAO 97 Proceeding, Phrase based information retrieval.

[9] Wikipedia. (2006). *English on the Internet*. http://www.wikipedia.org Accessed on 1st December 2006.

[10] Oren Eli Zamir. Clustering Web Documents: A Phrase-Based Method for Grouping Search Engine Results

[11] Anastasios Tombros. The effectiveness of query-based hierarchic clustering of documents for information retrieval

[12] Jardine, N. and van Rijsbergen, C. J. The use of hierarchical clustering in information retrieval. *Information Storage and Retrieval*, 7:217-240, 1971.

AUTHORS PROFILE

N.Sandhya B.Tech, M.Tech (Ph.D).
Working as an associate professor in the CSE Dept of GRIET. Has 11years of experience in teaching. My areas of interest are Databases, Information Retrieval, Data Mining and Text Mining.

Y.Srilalitha M.Tech (Ph.D).
Working as an associate professor in the CSE Dept of GRIET. Has 16 years of experience in teaching. My areas of interest are Information Retrieval, Data Mining and Text Mining.

Dr.K.Anuradha M.Tech, Ph.D.
Working as a professor and Head of the CSE Dept in GRIET.
Has an experience of 25 years in teaching. My areas of interest are Information Retrieval, Data Mining and Text Mining.

Dr.A.Govardhan M.Tech, Ph.D.
Working as a professor and Principal of JNTU, Jagityal. Has an experience of 20 years in teaching. My areas of interest are Information Retrieval, Databases, Data Mining and Text Mining.

Polarimetric SAR Image Classification with High Frequency Component Derived from Wavelet Multi Resolution Analysis: MRA

Maximum Curvature of Trajectory of Back Scattering Cross Section Converted from Eigen Space of Ellipticity and Orientation Angle Space of Polarization Signature

Kohei Arai
Department.of Information Science,
Graduate School of Science and Engineering
Saga University
Saga city, Japan

***Abstract*—A method for polarimetric Synthetic Aperture Radar: SAR image classification with high frequency component derived from wavelet Multi-Resolution Analysis: MRA is proposed. Although it is well known that polarization signature derived from fully polarized SAR data is useful for SAR image classifications, it is still unknown how to utilize the polarization signature in the image classification. High frequency component of the polarization signature calculated with the fully polarized SAR data is taken as one of the features utilizing in the classification into account. Thus improvement of classification performance is achieved for the proposed classification method of which such feature is included in the feature space for the Maximum Likelihood based classification method.**

Keywords- polarimetric SAR; polarization signature; wavelet MRA.

I. INTRODUCTION

Radar polarimetry allows measurement the physical characteristics such as di-electric constant, slope of the ground cover targets as well as directionality of artificial objects by using scattering mechanism between electromagnetic (EM) wave and the targets [1],[2]. Polarimetric SAR image classification with the following three components of the polrimetric SAR data, (1) transmit Electro-magnetic wave with Horizontal Polarization(H-Pol) and receive the echo from the ground with H-Pol(HH), (2) transmit Electro-magnetic wave with H-Pol and receive the echo with Vertical Polarization(V-Pol)(HV) and (3) transmit Electro-magnetic wave with V-Pol and receive the echo with V-pol(VV) is widely available [3],[4]. On the other hand, the extraction of the scattering characteristics of the targets of interest by applying eigen value decomposition to the covariance matrix derived from the scattering matrix which is calculated from the three components are proposed [5]. Furthermore, the classification methods with the single / double /multiple, odd / even / diffuse, and odd / even / Bragg / multiple scattering components derived from the eigen value decomposition were proposed [6] while the classification methods with the sphere / deplane / helix, and sphere / Bragg / double of scattering components which are based on the spherical polarization which are derived from the scattering matrix were also proposed [7],[8].

Aforementioned proposed methods were reviewed [9]. Moreover, the classification method with the entropy (H) which is defined with the sum of the first to third eigen values and the ratio of each eigen values, the anisotropy (A) which is defined as the ratio of sum and subtraction of the second and the third eigen values and cosine α (cos(α)) which is defined with the elements of the eigen vector corresponding to the first eigen value which is called coherency matrix(3 by 3) was proposed by E.Pottier [10].

The application of these methods to sea ice discrimination (such as thin ice (TI), smooth first year ice (SF), rough first year ice (RF) and open water (OW)) with the polarimetric SAR were attempted by using H, A, and cos(α) [11]. Classification performance, however, were not satisfactory (20-40% of classification errors were occurred for the classification of sea ice into four classes, ridged, compressed, new forming and smooth surface due to the fact that scattering mechanism based features were not used effectively. Meanwhile polarimentirc SAR image classification with polarization signature which are derived from Stokes or Muller or scattering matrix is widely available [12]. Polarization signature represents the scattering mechanism, in particular, surface roughness of the targets in concern.

One of the problems on the classification with polarization signature is classification performance. The method for effective utilization of polarization signature is still unclear to improve classification performance. The method proposed here is for extraction of effective information from the polarization signature by transforming the polarization signature onto an eigen space (eigen value decomposition). As the results from the eigen value decomposition which corresponds to the largest eigen value, a trajectory can be drawn. The trajectory represents the scattering mechanism in concern so that the largest curvature of the trajectory represents the most effective representatives of the scattering mechanism of the target of interest [13]. This is the theoretical background to propose the utilization of maximum curvature of the trajectory in an eigen space which is derived from the polarization signature to the sea ice classification.

Firstly, the proposed method is introduced with a theoretical background followed by experimental data and the results from the experimental are described together with the results from a comparative study between the proposed method and the existing conventional methods. Finally, concluding and remakes with some discussions is followed.

II. PROPOSED METHOD

A. Polarization Signature

Received echo signal $(E_h, E_v)^T_{rec}$ of polarimetric SAR is expressed as follows,

$$\begin{pmatrix} E_h \\ E_v \end{pmatrix}_{rec} = \frac{e^{-jkR}}{kR}\begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix}\begin{pmatrix} E_h \\ E_v \end{pmatrix}_{ill} \tag{1}$$

$$S = \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix}$$

where S is a scattering matrix composed with HH, HV, VH and VV components of S_{hh}, S_{hv}, S_{vh} and S_{vv}, $(E_h, E_v)^T_{ill}$ denotes incident Electro-magnetic wave, k is wave number of the incident Electro-magnetic wave, R denotes range, respectively. From these elements of S, Muller matrix is calculated while Stokes vector at the receiver J_r is calculated with $(E_h, E_v)^T_{rec}$ and also Stokes vector at the transmitter J_t is calculated with $(E_h, E_v)^T_{ill}$ thus the polarization signature σ^0 is represented by the following equation,

$$\sigma^0 = KJ_r^T\langle M_s\rangle J_t = \sigma^0(\chi_r, \ \phi_r, \ \chi_t, \ \phi_t) \qquad K: const. \tag{2}$$

where the χ and φ denote the ellipticity angle and the orientation angle of the orientation of the electric field vector, respectively of which the polarization is described with the two parameters. The polarization signature describes the scattering coefficient as a function of any assumed transmit and receive antenna polarization and allows measure the variation of the scattering coefficient with polarization so that the different targets show the different polarization signature as shown in Fig.1.

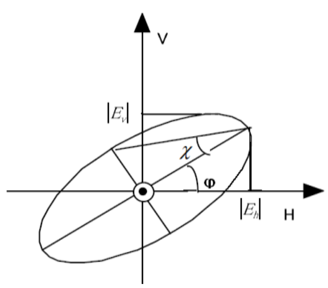


Figure 1. Definition of polarization signature

The method for effective utilization of polarization signature is still unclear to improve classification performance.

Let X be matrix with three scattering components as follows,

$$X = [S_{hh} S_{hv} S_{vv}]^T \qquad C = < XX^{*T} >$$

$$C = \ \lambda_1 K_1(K_1^*)^T \ \ +\lambda_2 K_2(K_2^*)^T \ \ +\lambda_3 K_3(K_3^*)^T \tag{3}$$

where C denotes covariance matrix of X and $\lambda_1,\lambda_2,\lambda_3$ are eigen values of C. These are corresponding to odd times scattering, even times scattering and diffuse scattering, respectively. Also * and T denotes complex conjugate and transpose, respectively. Thus

$$\frac{\lambda_i}{\lambda_1+\lambda_2+\lambda_3} \qquad (i = 1, 2, 3)$$

is called a contribution factors. If the equation (4) are formulated,

$$S_{RR} = \frac{1}{2}(S_{hh} - S_{vv} + j2S_{hv})$$
$$S_{LL} = \frac{1}{2}(S_{hh} - S_{vv} - j2S_{hv})$$
$$S_{LR} = \frac{1}{2}(S_{hh} + S_{vv}) \tag{4}$$

then spherical, di-plane and helix components are defined as follows,

$$K_s = |S_{LR}| \quad K_d = |S_{LL}| \quad K_h = |S_{RR}| - |S_{LL}| \quad for|S_{RR}| > |S_{LL}| \tag{5}$$

Thus contribution factor for each is defined as follows,

$$\frac{K_i}{K_s + K_d + K_h} \qquad (i = s, d, h) \tag{6}$$

Polarization signature represents the polarization feature of the ground cover target, obviously.

B. Proposed Method

Classification method proposed here is to use wavelet frequency components as features for classification. In particular, high frequency components derived from wavelet Multi Resolution Analysis: MRA is one of spatial feature of the target class. One dimensional wavelet transformation is expressed with the equation (7).

$$F=Cn\,f \tag{7}$$

where F, f denotes wavelet frequency component and radar echo signal as a function of time. Cn denotes wavelet transformation matrix which is expressed as a bi-orthogonal function based on base functions. Cn can be determined with a reference to the appendix. Therefore, $CnCn^t=I$. Then f is converted to $F_1=(L_1,H_1)$, $F_2=CnL_1=(L_2,H_2)$, $F_3=CnL_2=(L_3,H_3)$, and $F_m=CnL_{m-1}=(L_m,H_m)$. Also f is reconstructed as

$Cn^{-1}F_m=Cn^{-1}(L_m,H_m)=L_{m-1}$,.......,$Cn^{-1}F_2=L_1$, $Cn^{-1}F_1=f$. The suffix of 1 to m is called "level". Level m implies that wavelet transformation is applied m times. MRA ensure that the original signal can be reconstructed with the wavelet coefficients or frequency components of level m. The frequency components derived from MRA are corresponding to the level m. Therefore, MRA does work as filter bank.

There are some based functions such as Haar, Daubechies[1], etc. Through the experiment with preliminary simulation of radar echo data, Daubechies base function is selected. Daubechies base function is one of bi-orthogonal functions. MRA is applied to the polarization signature then the extracted high frequency component is added to the typical features of received polarization power signals, HH, HV, and VV components where HH denotes the horizontal polarization of electro-magnetic wave is transmitted then the returned echo signal is received in horizontal polarization. Thus spatial feature of the polarization of feature of the ground cover target is taken into account in the classifications.

Typically, three polarization power signals are acquired with polarimetric SAR system.

C. Methods for Comparison

The classification performance of the proposed method is compared to the Maximum Likelihood classification with only three components of the polarimetric SAR data of HH, HV, VV, with the three components + H + A + $cos(\alpha)$, with the three components + C_l, and with the three components + Odd + Even + Diffuse for discrimination among the following three classes, Urban, Vegetation, and Paddy field. Also classification performance is compared with the previously proposed classification method with information from the polarization signature by using eigen value decomposition of the polarization signature. As the results from the eigen value decomposition which corresponds to the largest eigen value, a trajectory can be drawn. The trajectory represents the scattering mechanism in concern so that the largest curvature (C_l) of the trajectory represents the most effective representatives of the scattering mechanism of the target of interest. If the largest curvature is large, then the polarization signature is steeply while the largest curvature is small, then the polarization signature is calm. The proposed classification method is based on the well known maximum likelihood classification with the received signal of the three different polarizations, co-polarization (HH and VV) and cross-polarization (HV) as well as the maximum curvature of the trajectory in the eigen space through the eigen value decomposition from the polarization signature.

III. EXPERIMENTS

A. Data Used

The PI-SAR (Polarimetric and Interferometric SAR) data of Tsukuba in Japan which was acquired by CRL (Communication Research Laboratory, current NICT: National Institute of Communication Technology) and NASDA (National Space Development Agency of Japan, current JAXA: Japan Aeronautics Exploration Agency) on 23 Feb. 1999 was used. The major characteristics of the PI-SAR are in Table 1.

TABLE 1. MAJOR CHARACTERISTICS OF PI-SAR

Instrument	NASDA/L-band SAR
Center frequency	1.27GHz
Peak power	3.5kW
Band width	50MHz
Antenna size	1.6m x 0.7m
Polarization	HH/HV/VH/VV(Full Pol.)
Incident angle	20-60degrees(Fixed)
Swath width	42.5km
Spatial resolution	3m
Quantization bit	8bits(I and Q)

From the data of SSC (Single-look Slant-range Complex) of the data,

B. Polarization Signature

Fig.2 shows an example of SSC (Single-look Slant-range Complex) data (one pixel is corresponding to 3 by 3 meters)of Tuskuba city, Japan which is acquired on 30 September 1997. Blue, yellow, and green square shows the training areas of Paddy fields (After harvest), Urban areas, and Vegetated areas, respectively.

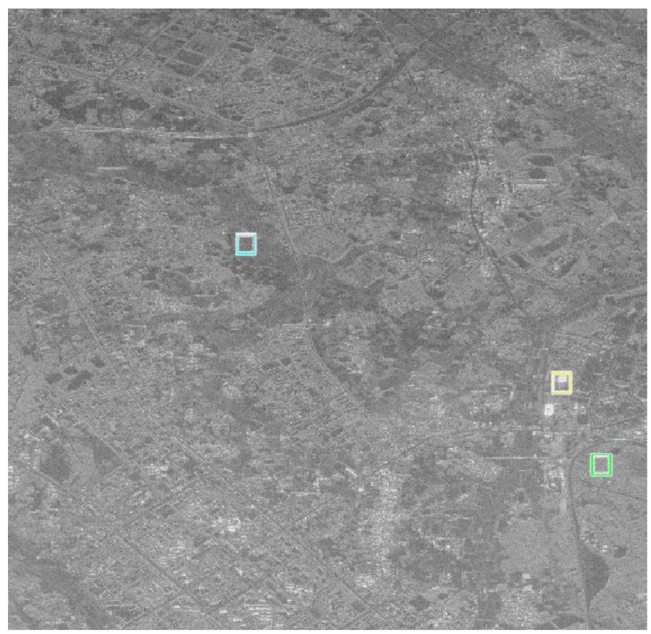


Figure 2. An example of SSC (Single-look Slant-range Complex) data of Tuskuba city, Japan which is acquired on 30 September 1997.

Using equation (1), polarization signature can be calculated by pixel by pixel because the parallel polarization, Co-pol. and cross polarization, Cross-Pol. are acquired by pixel by pixel. Fig.3 shows the calculated polarization signature for each class. The calculated polarization signatures are different each other. Therefore, it can be classified. The problem of the conventional classification method which uses the polarization signature and the returned echo signal powers, HH, HV, and VV is poor classification performance. The proposed method utilizes

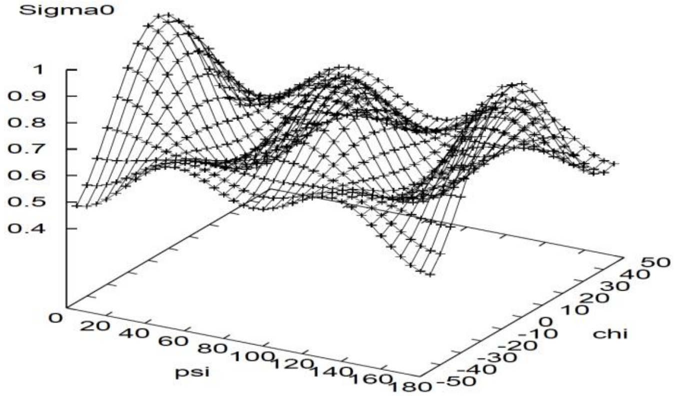
[1] Daubechies base function is defined as { α_k} satisfying the following conditions,

$$\phi(x) = \sum_k \alpha_k \sqrt{2}\, \phi(2x-k) \qquad (1)$$

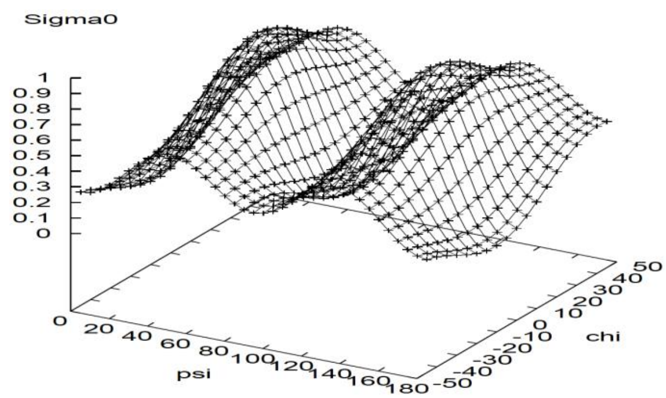
$$\beta_k = (-1)^k \alpha_{1-k} \qquad (2)$$

$$\phi(x) = \sum_k \beta_k \sqrt{2}\, \phi(2x-k) \qquad (3)$$

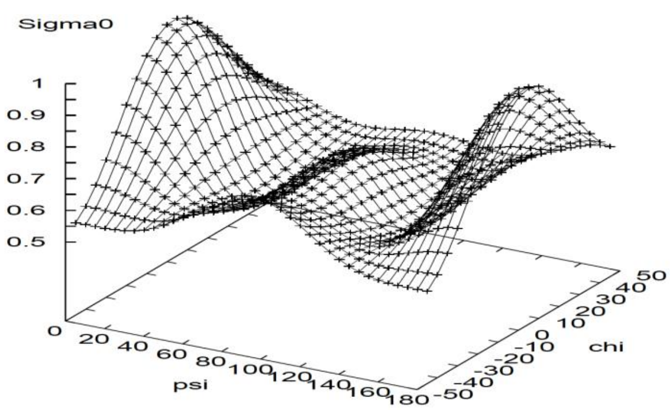
high wavelet frequency components other than these polarization signature and returned echo signal powers.



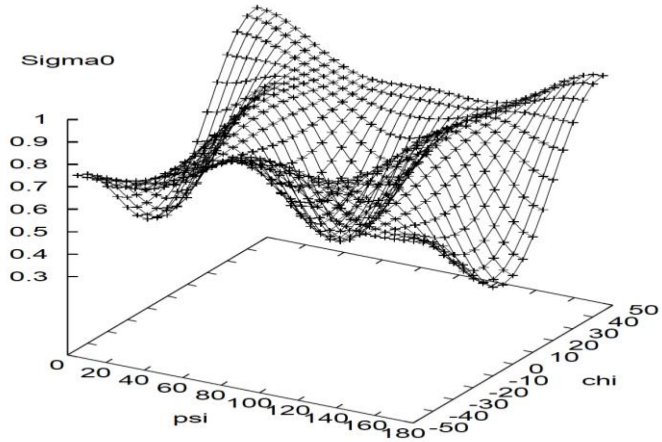
(a) Urban (Co-pol)



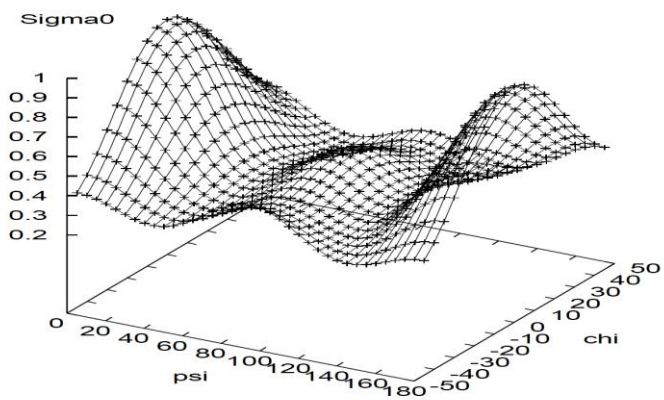
(b)Urban (Cross-pol)



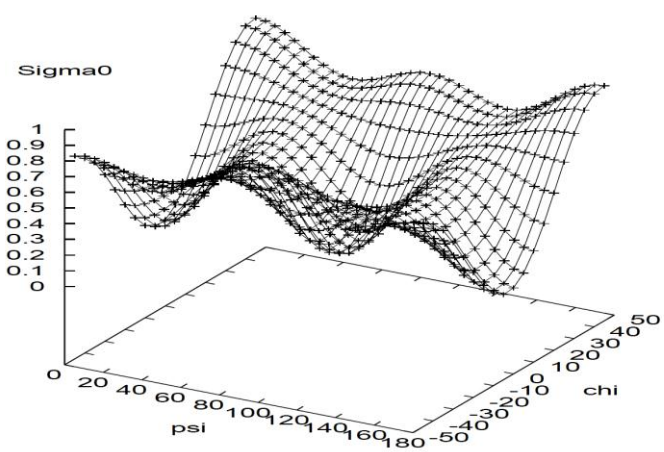
(c)Vegetation (Co-pol)



(d)Vegetation (Cross-pol)



(e)Paddy (Co-pol)



(f)Paddy (Cross-pol)

Figure 3. Examples of the calculated polarization signatures of the classes, Urban, Vegetation, and Paddy extracted from the 16 by 16 pixels of square areas which are corresponding to the designated classes.

C. Class Features

Because the PI-SAR frequency is 1.27 GHz, the contribution of leaves of vegetation is very week (penetrated) while the contribution from the ground surface and the relatively large trees is major. Therefore, polarization signature of vegetated areas is similar to the paddy fields after harvest except surface flatness. On the other hand, HH return echo signal power and polarization signature for both urban areas and vegetated areas are similar. As the results from the aforementioned reasons, all these three classes are difficult to classify. It is confirmed that three components of return echo signal power, HH, HV, and VV utilized classification makes about 80 % of classification performance.

In order to improve classification performance, the following feature is proposed by the author. That is maximum curvature of the trajectory in the eigen vector space which is converted from the polarization signature with arbitrary polarization angle response with 5 degrees intervals. The proposed feature does work for classification for the classes with different surface roughness, in particular. Although it is effective to improve classification performance, it is complicated and requires computation resources.

The proposed classification method is based on wavelet frequency components of the calculated polarization signature. Therefore not significant computation resources are not required and it is quite simple.

Fig.4 (a) and (b) shows examples of the high frequency component (HH4) of the polarization signature of image which is shown in Fig.3, and the maximum curvature of the trajectory in eigen vector space which is converted from the polarization signature image, respectively.

It is difficult to discriminate between urban and vegetation classes because the previously proposed maximum curvature of urban and vegetation is similar as shown in Fig.4 (b). On the other hand, the proposed feature of high frequency component HH4 of image shows clear difference between urban and vegetation classes. Therefore, HH4 frequency component which is derived from the level 4 of MRA output, HH4, HL4, LH4, and LL4 of the image which consists of 32 by 32 pixels with 5.6 degree step of the polarization angle of the each dimension of polarization signature is used for classification.

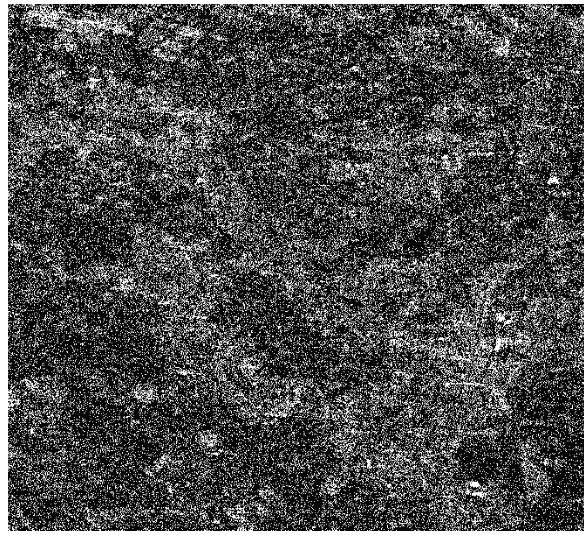
D. Classification Performance

Using training samples extracted from the training areas which are indicated in Fig.2, training performance is evaluated. A comparison is made for the following three Maximum Likelihood based classification method with (1) received three echo signal power, HH, HV, VV, with (2) three power and the previously proposed maximum curvature, and with (3) three power and the HH4 of each dimension of polarization signature.

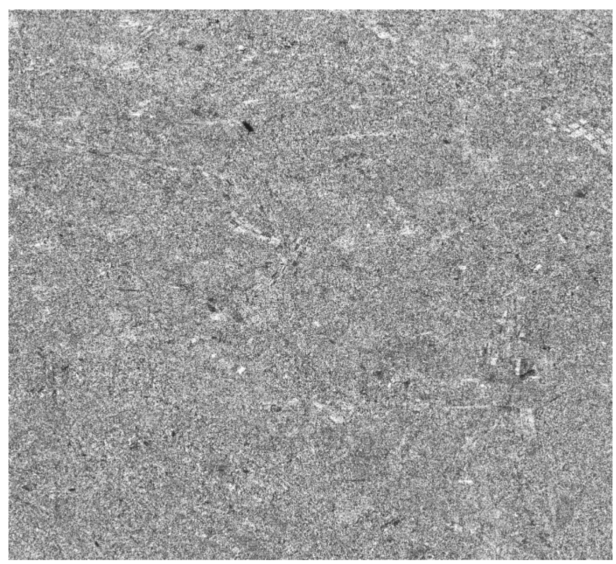
Classification performance is evaluated with confusion matrix and the k statistics which is shown in the following equation (8),

$$k = \frac{P(A) - P(E)}{1 - P(E)} \tag{8}$$

where P(A), P(E) denotes correct classification probability and classification error probability. Table 2, 3 ,4 shows the results.



(a) Examples of the high frequency component of HH4 of the polarization signature of image which is shown in Fig.3



(b) The maximum curvature image of the trajectory in eigen vector space which is converted from the polarization signature image

Figure.4. The image of the features of the previously proposed maximum curvature of the trajectory in eigen space which is converted from the polarization signature and the high wavelet frequency component image of the polarization signature.

TABLE 2. THREE RECEIVED ECHO SIGNAL POWER ONLY (K=89.05)

	Urban	Vegetation	Paddy
Urban	75.2	24.8	0.0
Vegetation	2.4	97.6	0.0
Paddy	0.8	1.6	97.6

TABLE 3. THREE POWER AND THE MAXIMUM CURVATURE (K=95.62)

	Urban	Vegetation	Paddy
Urban	89.8	10.2	0.0
Vegetation	1.2	98.8	0.0
Paddy	0.4	0.8	98.8

TABLE 4. THREE POWER AND HH4 (K=98.0)

	Urban	Vegetation	Paddy
Urban	94.9	5.1	0.0
Vegetation	0.0	100.0	0.0
Paddy	0.0	0.8	99.2

As is shown in the previous section, it is difficult to discriminate between urban and vegetation classes if only three power of features are used. It is improved by adding the features of maximum curvature and HH4, remarkably. On the other hand, k-statistics shows that 7.38% of improvement is achieved by adding the maximum curvature and 10.05% improvement is achieved by adding HH4.

IV. CONCLUSION

It is found that the proposed features of three received echo signal power, HH, HV, and VV as well as HH4 of high frequency component derived from MRA analysis with polarization signature are effective to improve classification performance for PI-SAR types of polarimetric SAR images. More than 10% of improvement of the classification performance is confirmed by adding HH4. The improvement depends on the level of MRA, because there is the most appropriate level depending on the frequency component of the polarization signature of the ground cover target.

The computational resource requirement for the previously proposed maximum curvature of the trajectory in eigen vector space which is converted from the polarization signature is twice much larger than that of the proposed MRA based feature utilized classification. Thus it may said that the proposed high frequency component of polarization signature based classification method is superior to the conventional classification method with three echo signal powers, and to the previously proposed maximum curvature based classification method in terms of classification performance and the required computation resources.

REFERENCES

[1] Zebker, H.A., vanZyl, J.J. Imaging radar polarimetry: A review. Proc. IEEE 79, 1583–1606, 1991.

[2] Ulaby, F.T., Elachi, C. Radar Polarimetry for Geoscience Applications.Artech House Inc., Dedham, 1990.

[3] Mott, M. Antennas for Radar Communications – A Polarimetric Approach. John Wiley, New York, 1992.

[4] Henderson, F.M., Lewis, A.J. Principles and applications of imaging radar, in: Manual of Remote Sensing, third ed. John Wiley, New York, 1998.

[5] Krogager, E. Analysis of the absolute and relative phase conditions of transforming matrices for the sinclair and covariance matrices in radar polarimetry. J. Opt. Soc. Am. E72B, 1A, 1993.

[6] Dong, Y. Imaging radar polarimeter from wave synthesis. J. Geophys. Res. 92 (81), 638–701, 1987.

[7] Zebker, H.A., et al., Imaging radar polarimeter from wave synthesis, J. Geophysical Research, Vol.92, No.81, pp.638-701, 1987.

[8] Krogager, E., Czyz, Z.H. Properties of the sphere, di-plane, helix decomposition, in: Proceedings, Third International Workshop on Radar Polarimetry, vol. 1, pp. 106–114, 1995.

[9] Cloude, S.R., Pottier, E. A review of target decomposition theorems in radar polarimetry. IEEE Trans. Geosci. Remote Sensing 34 (2), 498–518, 1995.

[10] Pottier, E., Lee, J.S. Unsupervised classification scheme of Polsar images based on the complex Wishart distribution and H/A/a polarimetric decomposition theorems, Proceedings of the Third European Conference on Synthetic Aperture Radar: EUSAR, 2000.

[11] Scheuchl, B., Hajinsek, I., Cumming, I.G., et al. Sea ice classification using multi-frequency polarimetric SAR data, in: Proceedings of the IGARSS'02, 2002.

[12] vanZyl, J.J., Zebker, H.A., Elachi, C. Imaging radar polarization signatures, theory and observation. Radio Sci. 22 (4), 529–543, 1987.

[13] Kohei Arai and J.Wang, Polarimetric SAR image classification with maximum curvature of the trajectory in eigen space domain on the polarization signature, Advances in Space Research, 39, 1, 149-154, 2007.

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

A New Reliability Model for Evaluating Trustworthiness of Intelligent Agents in Vertical Handover

Kailash Chander
Research Scholar, Maharishi Markendeshwar University, Mullana, Haryana, India.

Dimple Juneja
Professor, M.M Institute of Computer Technology and Business Management (MCA) , Maharishi Markendeshwar University, Mullana, Haryana, India

***Abstract*— Our previous works have proposed the deployment of mobile agents to assist vertical handover decisions in 4G. Adding a mobile agent in the 4G could lead to many advantages such as reduced consumption of network bandwidth, delay in network latency and reduction in the time taken to complete a particular task. However, this deployment demands that the deployed collection of agents shall be secure and trustworthy. Security of a mobile agent includes maintaining confidentiality, reality and integrity of not only the agent employed but also the system in which it is deployed. In fact, many conventional security solutions exists, however, very few of them addresses the challenge of introducing trusted computing in mobile agents, deployed in 4G, in particular. This paper proposes a new reliability model by implementing trust certificate for mobile agents in vertical handover.**

Keywords- Trust Certificate; Mobile Agents; Vertical Handover; Secure Network;

I. Introduction

All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later Mobile agent is an entity that migrates from one system to another to access remote resources or to interact with other agents. Agent technology finds its applications in wide areas such as user interfaces, mobile computing, information retrieval and filtering, smart messaging, telecommunications etc. Our previous works proposed an agent based smart solution for vertical handover in 4G [10]. Agents deployed at different level of PMIPv6 environment are responsible for managing user preferences, authentication of mobile node and buffering of user data in case of handover. Further, this vertical handover decision is mainly relies on the selection of the 'best' available network that could meet QoS requirements for the end-user and to meet the above requirement a novel approach for always best connected in future wireless networks [12] was also proposed. The work exploits intelligent agents for weight calculations after analyzing the explored parameters for various networks. Mobile agent based emigration framework for 4G (MAEF) [16] which helps to switch the network during critical situations of draining battery without interrupting the ongoing task were also proposed. However, in all of the above works, the trustworthiness of mobile agents deployed in the system is left unattended. The current works aims to propose a reliability model that would evaluate the trustworthiness of mobile agents operating in 4G.

Now, the deployed agents must be trustworthy and reliable wherein, agent should not be involved in the activities such as disclosure of information, denial of service and corruption of information with respect to interaction and cooperation [17]. In fact, trust is one of the basic parameters of evaluating a mobile agent-based system and it is usually computed through Direct Experience, Third Party References, Confidentiality, Persistence, Execution Trust etc. Literature [9] reflect that "Trust is the firm belief in the competence of an entity to act as expected such that this firm belief is not a fixed value associated with the entity but rather it is subject to the entity's behavior and applies only within a specific context at a given time". Using this definition the proposed work assigns weight to the mobile agents, which is further, used for calculating the credibility of a mobile agent.

The structure of the paper comprises of four sections. Section II presents the related work. Section III discusses the significance of evaluating trust in mobile agent-based frameworks. Section IV presents the new reliability model that evaluates the trustworthiness of participating agents by generating a trust certificate. Finally, conclusions and future scope is presented in Section V.

II. Related Work

The section presents the work of researchers in trusted computing in telecommunication with the aim to highlight the scope of further research in the similar direction.

When mobile agents travel from one system to another in a network, they transfer their code, data and execution state. Therefore, reliability is a vital issue for deploying the mobile agent system. Many researchers have proposed trust based reliable frameworks. For instance, MobileTrust [6] has a trust management layer added transparently on the top of the conventional security layer, which is responsible for presenting security related trust evidence for the purpose of making evaluations and decisions regarding trust relationships among mobile agents.

Work in [1] presents a fuzzy approach for reliability estimation of Mobile Agent Based Systems (MABS). It investigates the hardware oriented system reliability of MABS. The Trust Management Frame (TMF) [7] is another approach comprising of three main components i.e. trust dissemination, trust formation and trust evolution. AT-RFM [2] presents an agent tracking reliable forwarding mechanism that integrates the 'tracker' with context. The objective of tracker is to know the present status and location of the agent. Few other works such as Trust-Aware resource management model [9] uses trust in grid system. In this model overall Grid system is divided into Grid Domains, which are autonomous administrative entities consisting of a set of resources and clients managed by a single administrative authority. The study examines the integration of the notion of "trust" into resource management such that the allocation process is aware of the security implications. A trusted Certification Authority (CA) and Trusted Platform Module (TPM) [3] is proposed for authentication and delegation of identity of mobile agent environment so that entities in mobile agent environment can build trusted relationship with each other.

The literature presented above clearly indicates that researchers have been demanding the trust enabled agent-based frameworks and few of them have made good attempts to incorporate trust and reliability as two separate parameters depending upon the domain of implementation. The main intent of this work is to propose a hybrid model that would ensure both reliability and trustworthiness by implementing trust certificate for mobile agents deployed in vertical handover process in 4G.

Next section presents the background of formalizing trust and reliability in our works.

III. FORMALIZING TRUST IN MOBILE AGENT-BASED FRAMEWORKS

Trust in mobile agents is usually established through identity token, provided by an X.509 public key certificate [18]. An X.509 certificate contains a public key, a subject name in the form of a multi component distinguished name (DN), a validity period and is either signed either by a trusted third party, or by certification authority (CA). Figure 1 depicts an Example of standard X.509 Certificate [15]:

Although, an X.509 certificate is good enough to identify the identity of a mobile agent but it fails to confirm the reliability and credibility of the agent under consideration. Therefore, this paper proposes a new and improved "Trust Certificate" for a Mobile Agent. The approach provides a means to work independently or, if available, in conjunction with the X.509–PKI. The work aims to evaluate the reliability, credibility and trustworthiness of agents and hence a new trust certificate representing the weight in terms of five parameters namely, Direct Experience, Third Party References, Confidentiality, Persistence, Execution Trust of a mobile agent would be generated and hence would be improving the overall performance of the system.

```
Certificate:
   Data:
       Version: 3 (0x2)
       Serial Number: 1 (0x1)
       Signature Algorithm: md5WithRSAEncryption
       Issuer: C=ZA, ST=Western Cape, L=Cape Town, O=Thawte
Consulting cc,
               OU=Certification Services Division,
               CN=Thawte Server CA/Email=server-
certs@thawte.com
       Validity
           Not Before: Aug  1 00:00:00 1996 GMT
           Not After : Dec 31 23:59:59 2020 GMT
       Subject: C=ZA, ST=Western Cape, L=Cape Town, O=Thawte
Consulting cc,
                OU=Certification Services Division,
                CN=Thawte Server CA/Email=server-
certs@thawte.com
       Subject Public Key Info:
           Public Key Algorithm: rsaEncryption
           RSA Public Key: (1024 bit)
               Modulus (1024 bit):

00:d3:a4:50:6e:c8:ff:56:6b:e6:cf:5d:b6:ea:0c:

68:75:47:a2:aa:c2:da:84:25:fc:a8:f4:47:51:da:

85:b5:20:74:94:86:1e:0f:75:c9:e9:08:61:f5:06:

6d:30:6e:15:19:02:e9:52:c0:62:db:4d:99:9e:e2:

6a:0c:44:38:cd:fe:be:e3:64:09:70:c5:fe:b1:6b:

29:b6:2f:49:c8:3b:d4:27:04:25:10:97:2f:e7:90:

6d:c0:28:42:99:d7:4c:43:de:c3:f5:21:6d:54:9f:

5d:c3:58:e1:c0:e4:d9:5b:b0:b8:dc:b4:7b:df:36:
                   3a:c2:b5:66:22:12:d6:87:0d
               Exponent: 65537 (0x10001)
       X509v3 extensions:
           X509v3 Basic Constraints: critical
               CA:TRUE
   Signature Algorithm: md5WithRSAEncryption

07:fa:4c:69:5c:fb:95:cc:46:ee:85:83:4d:21:30:8e:ca:d9:

a8:6f:49:1a:e6:da:51:e3:60:70:6c:84:61:11:a1:1a:c8:48:

3e:59:43:7d:4f:95:3d:a1:8b:b7:0b:62:98:7a:75:8a:dd:88:

4e:4e:9e:40:db:a8:cc:32:74:b9:6f:0d:c6:e3:b3:44:0b:d9:

8a:6f:9a:29:9b:99:18:28:3b:d1:e3:40:28:9a:5a:3c:d5:b5:

e7:20:1b:8b:ca:a4:ab:8d:e9:51:d9:e2:4c:2c:59:a9:da:b9:

b2:75:1b:f6:42:f2:ef:c7:f2:18:f9:89:bc:a3:ff:8a:23:2e:
       70:47
```

Figure 1: A Sample of CA X.509 Certificate [15]

IV. THE NEW RELIABILITY MODEL FOR GENERATING TRUST CERTIFICATE

A critical look at the literature pointed towards few parameters on which trust in trusted computing is being evaluated and those parameters are Direct Experience, Third Party References, Confidentiality, Persistence, Execution Trust etc. This work exploits these stated parameters and a trust certificate is being generated as shown in Figure 2.

The trust certificate comprises of certificate version, issuer, validity dates, previous weight, current weight, trust weight and date of weight updation. The description of the components used in trust certificates are as follows:

- Certificate Version: This field describes the version of the Trust Certificate.

- Issuer: The issuer field identifies the entity who has signed and issued the Trust certificate. Mobile agents are usually expected to have FIPA standardized certificate.
- Validity Dates: Trust certificate validity period is represented two dates: the date on which the certificate validity period begins (notBefore) and the date on which the Trust certificate validity period ends (notAfter).
- Parameter Previous Weight: This field consists of previous weight of Trust Certificate. The initial value of this field will be 0.5 for new agent. For future correspondence current weight will become previous weight for an agent.
- Parameter Current Weight: This field can be calculated based on five parameters defined in Figure 2.
- Trust Weight: Trust weight field can be calculated by getting the average of Previous Weight and Current Weight.
- Date of Weight Update: This is the date on which Trust Weight is updated

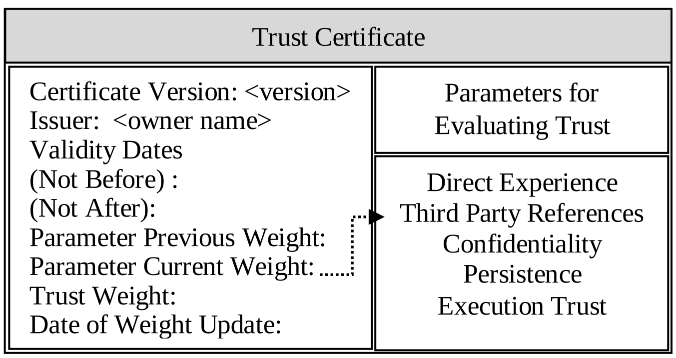


Figure 2: Proposed structure of Trust Certificate

Based on the weights i.e. previous weight (PW*i*) and current weight (CW*i*), a trust weight (TW*i*) of a mobile agent is calculated. To calculate the Current Weight (CW*i*) every parameter is assigned values in the range of [0, 1]. For instance, CW*i* =0 means distrust i.e. agent is non-trustworthy and CW*i* =1 implies agent is fully trustworthy. If the value assigned between 0 and 1 it means two entities trust each other upto an extent. In order to compute, TWi between two interacting agents, an average of PWi and CWi is calculated as given by eq.(1).

$$tw_i = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{cw_i} + \mathrm{pw_i})}{2} \tag{1}$$

Now, the challenge is how to evaluate these weights. As depicted in Figure 2 the weights are dependent on the mentioned parameters. Now, when these agents, say Agent*i*, Agent*j*, and Agent*k* interact with each other, the weight of each parameter would be computed individually as per the matrices given in Figure.3.

As shown in Figure 3(a),(c) and (e) i.e. in case of direct experience, confidentiality and execution trust, an agent*i* would award a weight '1' to agent*j* as this value have to be awarded only when the respective agents have directly experienced an interaction with each other and hence are confident and can ensure the execution trust too, else the value in the cell remains 0. This computation remains same irrespective of the facts that agents are operating in homogenous or heterogeneous environments; or belong to Intranet or Internet. In contrast, the cells in figure 3(b) and (d) have values in the range of {0-1}. For instance, when an agent*i* award values to agent*j* on the basis of third party reference, it is depending on the feedback of third party and hence may not be fully satisfied while awarding a value. In such uncertain conditions such as third party reference and agent's persistence in any environment may lead to fuzzy values which has been discritized in the range {0.0.25,0.5,0.75,1}for the sake of computational simplicity.

On the basis of above computations, the current weight for an Agent*i* would be calculated as the sum of values generated for each parameter and is given as in eq.(2).

$$cw_i = \sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{de_i} + \mathrm{tpr_i} + \mathrm{conf_i} + \mathrm{pers_i} + \mathrm{et_i}) \tag{2}$$

Where, the range of each parameter is as defined in the above matrix. The trust weight would then be computed as given in eq. (1)

Now, turning our attention to our proposed works [10, 12, 16] wherein, a trust certificate for various agents deployed at different level of PMIPv6 will now be generated. The agent of MN is responsible for managing user preferences through Compute*agent* [12]. Interface*agent* [12] provides information already collected from different MAGs and handover the same to the Compute*agent* for populating the preference list. In the similar way MAG*agent* [10] also interact with another MAG*agent* [10] to transfer the authentication and buffered data during handover. Moreover LMA*agent* [10] is responsible for maintaining the user profile and policy data which is updated to MN via MAG.

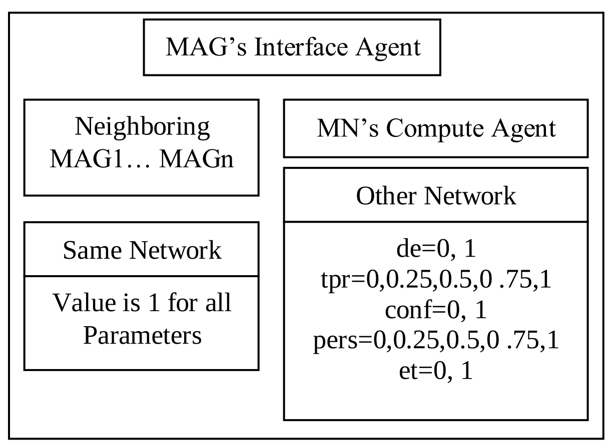


Figure 4: Interaction of Interface Agent with Compute Agent

When these agents interact, the parameters are initialized with values as given in Figure 3. As shown in Figure 4, when

Direct Experience (de_i)		
No. of Agents	$Agent_j$	$Agent_j$
$Agent_i$	1	0,1
$Agent_j$	0,1	1

(a)

Third Party References (tpr_i)		
No. of Agents	$Agent_j$	$Agent_j$
$Agent_i$	1	0,0.25,0.5, 0.75,1
$Agent_j$	0,0.25,0.5, 0.75,1	1

(b)

Confidentiality ($conf_i$)		
No. of Agents	$Agent_j$	$Agent_j$
$Agent_i$	1	0,1
$Agent_j$	0,1	1

(c)

Persistence($pers_i$)		
No. of Agents	$Agent_j$	$Agent_j$
$Agent_i$	1	0,0.25,0.5,0.75,1
$Agent_j$	0,0.25,0.5,0.75,1	1

(d)

Execution Trust (et_i)		
No. of Agents	$Agent_j$	$Agent_j$
$Agent_i$	1	0,1
$Agent_j$	0,1	1

(e)

Figure 3: Computation of Weights of Various Parameters

MAG's Interface*agent* is interacting with other MAG's agents to collect the information for best available network. The same information is then handover to Compute*agent* of MN to set the preference list for Always Best Connected. In both interactions

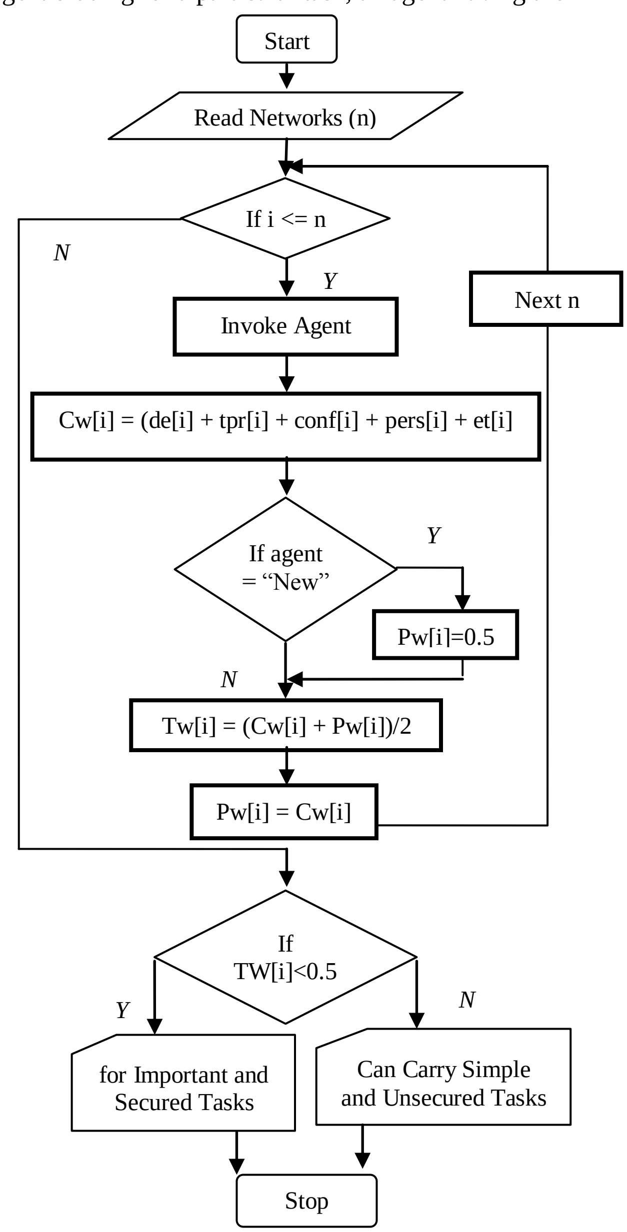
MAG vs MAG and MAG vs MN the network type may be same or it may be of different type. The values for all the parameter is set accordingly to calculate the current weight.

Algorithm: Agent TC (Trust Certificate)

----

begin
read Networks (n)
while (has more n)
invoke agent[i]
compute Cw[i]=sum(de[i], tpr[i], conf[i], pers[i], et[i])
    if (agent is new)
        Pw[i]=0.5
    else
        Pw[i]=Cw[i]
    Trust Weight Tw[i] = (Pw[i]+Cw[i]) / 2
next n
    if (Tw[i] < 0.5)
        for Simple and Unsecured Tasks
    else
        for Important and Secured Tasks
    end if
$LMA_{agent}|MAG_{agent}|Compute_{agent}|Interface_{agent} \Leftarrow$ TC
end

Now, in order to calculate the TW_i of any mobile agent for a particular network, the value of previous weights is also required. However, if an agent is interacting for the first time then default value of previous weight will be 0.5. However, current weight will become the previous weight for future interaction as shown in flow chart Figure 5. Now, as is clear from the computations, a trust certificate will be generated for each agent willing to participate and hence carry out a particular task. An agent will be assigned an important task such as secured transactions if, it scores a minimum threshold value of TW_i i.e. 0.5. An agent scoring less than this would may be allowed to carry simple and unsecured tasks such as internet surfing and voice data. However, in case of more than one agent bidding for a particular task, an agent having the



maximum TW_i value and the oldest Trust certificate will be treated as the most experienced and hence learned agent for the task to be performed successfully. The Trust certificate needs to be embedded as add-ons to the mobile agent.

V. Conclusion And Future Work

This work proposed a new reliability model for generating trust certificate for all agents participating in vertical handover procedure happening in 4G. Trusts amongst agents have always been an important and unaddressed challenge. The trust certificate generated by the proposed model not only improved the reliability and trustworthiness of agents but also added to improving robustness of the whole agent-based framework. Now, the agents can be categorized and be given a task according to their credibility. For instance, an agent with low trust weights will be assigned a task which does not demand high security whereas an agent with high trust weight may be assigned a task of carrying out financial transactions. The only limitation of this work is that it demands embedding of trust certificate in the data section of mobile agents, thus adding an overhead and increasing the complexity of mobile agent.

References

[1] Mosaab Daoud and Qusay H. Mahmoud "A Fuzzy Approach to Reliability Estimation of Mobile Agent-Based Systems", Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on Oct. 2007 p.p. 2854 – 2859.

[2] U.P.Kulkarni, Prof A R Yardi, "Agent Tracking: A Reliable Agent Forwarding Mechanism", Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)

[3] Zhidong Shen, XiaopingWu, "The Authentication and Identity Delegation about Mobile Agent System based on Trusted Computing Platform", Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on July, 2010 p.p. 672 - 676

[4] Zhidong Shen et.al., "Trust Management for Mobile Agent System Based on Trusted Computing Platforms", Wireless Communications, Networking and Information Security (WCNIS), 2010 IEEE International Conference on June, 2010 p.p 717 - 721

[5] Zhen Ye et.al., "Dynamic Trust Model Applied In Mobile Agent" Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on July 2008 p.p. 536 - 540

[6] Ching Lin and Vijay Varadharajan , "Trust Enhanced Security - A New Philosophy for Secure Collaboration of Mobile Agents", International Conference on Collaborative Computing: Networking, Applications and Worksharing, ISBN: 1-4244-0428-2, IEEE 2006

[7] Suzhen Wang et. al. "A Trust Evaluation Method of Mobile Agent System", Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on Sep. 2007 p. p. 6317 – 6320.

[8] Ching Lin et. al. "Trust Enhanced Security for Mobile Agents", Proceedings of the Seventh IEEE International Conference on E-Commerce Technology (CEC'05), 2005

[9] Farag Azzedin and Muthucumaru Maheswaran "Towards Trust-Aware Resource Management in Grid Computing Systems" Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on May 2002 p. p. 452-452

[10] Kailash Chander, Dr. Dimple Juneja et. al. "An Agent Based Smart Solution for Vertical Handover in 4G", International Journal of Engineering Science and Technology, Vol. 2(8), 2010, p.p. 3381-3390

[11] Internet Engineering Task Force (IETF). www.ietf.org

[12] Kailash Chander, Dr. Dimple Juneja "A Novel Approach for Always Best Connected in Future Wireless Networks", Global Journal of computer Science and Technology, Issue 11, Vol 05, 2011.

[13] R.Koodli, "Mobile IPv6 Fast Handovers," RFC 5268, June 2008

[14] D. Premec, NetExt Working Group, Internet Draft "Inter-technology handover in PMIPv6 domain" at www.ietf.org, March 9, 2009.

[15] http://www.pentaware.com/pw/x509.htm

[16] Kailash Chander, Dr. Dimple Juneja "Mobile Agent based Emigration Framework for 4G: MAEF" communicated to International Journal of Information and Computing Science, UK

[17] Zhidong Shen, Xiaoping Wu "A Trusted Computing Approach to Building Trusted Connect between Entities in Mobile Agents Environment"2nd International Conference on Education Technology and Computer (ICETC)- 2010.

[18] R. Housley et.al "Internet X.509 Public Key Infrastructure Certificate and CRL Profile" http://www.ietf.org/rfc/rfc2459.txt

Authors Profile

Dr. Dimple Juneja is the Principal and Professor of Computer Science at Maharishi Markandeshwar University, Mullana-Ambala, India. Her publications include more than 42 research papers in various National and International Journals and conferences. Her research interests include Agent Technologies, Semantic Web, Sensor Networks and Intelligent Systems. She is a member of Computer Society of India and International Association of Engineers, Hongkong. She is a recipient of best paper awards and the Distinguished Faculty award. Her research has been funded by Ministry of Czech Republic and Haryana State Council of Science and Technology, India. She has served on various selection and review panels. Dr. Juneja has also worked as postdoctoral researcher in computer science at Louisiana State University. She is on the editorial board of various international journals of repute and has chaired sessions at various International conferences.

Kailash Chander is the Resrach Scholar at Maharishi Markandeshwar University, Mullana-Ambala, India. His publications include more than 10 research papers in various National and International Journals and conferences. His research interests include Mobile Communications and Agent Technologies.

Some New Results about The Period of Recurring Decimal

Bo Zhang
Linyi University at Feixian
Feixian, Shandong, P.R.China

***Abstract*—This study mainly discusses period problem of recurring decimals.According to Euler theorem, this paper gives the computation formula of period of recurring decimal, relation of the period and least positive period, and the necessary and sufficient condition that the period is eaqual to least positive period.**

Keywords-Recurring decimal;Period;Prime.

I. INTRODUCTION

We know, irreducible proper fraction can be transformed recurring decimal (pure recurring decimal or mixed recurring decimal), the repetend digit of the recurring decimal is called the period of the recurring decimal, the least repetend digit is called least positive period.

Period problem of the recurring decimal is always a very interesting and difficult problem in number theory; many scholars studied it and derived some beautiful properties of the recurring decimals [1, 2, 5-10]. However, for the period and least positive period of the recurring decimal, there are not computation formulas so far.

Let us now consider 50 fractions between 1/3 and 1/62(besides the fractions which can be denoted as limited decimals), and further convert them into recurring decimals, then their least positive period are as follows.

TABLE 1. FRACTIONS AND LEAST POSITIVE PERIODS

Fraction	Least positive period	Fraction	Least positive period
1/3	1	1/36=1/4*1/9	1
1/6=1/2*1/3	1	1/37	3
1/7	6	1/38=1/2*1/19	18
1/9	1	1/39=1/3*1/13	6
1/11	2	1/41	5
1/12=1/3*1/4	1	1/42=1/6*1/7	6
1/13	6	1/43	21
1/14=1/2*1/7	6	1/44=1/4*1/11	2
1/15=1/5*1/3	1	1/45=1/5*1/9	1
1/17	16	1/46=1/2*1/23	22
1/18=1/2*1/9	1	1/47	46
1/19	18	1/48=1/16*1/3	1
1/21=1/3*1/7	6	1/49=1/7*1/7	42
1/22=1/2*1/11	2	1/51=1/3*1/17	16
1/23	22	1/52=1/4*1/13	6
1/24=1/8*1/3	1	1/53	13
1/26=1/2*1/13	6	1/54=1/6*1/9	3
1/27	3	1/55=1/5*1/11	2
1/28=1/4*1/7	6	1/56=1/8*1/7	6
1/29	28	1/57=1/3*1/19	18
1/30=1/10*1/3	1	1/58=1/2*1/29	28
1/31	15	1/59	58
1/33=1/3*1/11	2	1/60=1/20*1/3	1
1/34=1/2*1/17	16	1/61	60
1/35=1/5*1/7	6	1/62=1/2*1/31	15

By observing Table 1, we find that their least positive periods have not distinct regularity. According to Euler theorem, we will present the computation formula of period of recurring decimal, relation of the period and least positive period, and the necessary and sufficient condition that the period is eaqual to least positive period.

II. PRELIMINARIES

Definition 1 Let a/b be an irreducible proper fraction, when it can be denoted as a recurring decimal, the repetend digit of the recurring decimal(including pure recurring decimal and mixed recurring decimal) is called the period of a/b, and it is denoted by $T(a/b)$;the least repetend digit is called least positive period of a/b, and it is denoted by $\underline{T}(a/b)$.

Lemma 1[3] Let a/b be an irreducible proper fraction, where

$$b = 2^s 5^t, s \geq 0, t \geq 0, s+t > 0,$$

then a/b can be denoted as a limited decimal.

Lemma 2[3] Suppose that a/b is an irreducible proper fraction, and $(b,10)=1$, then

(i) a/b can be denoted as a pure recurring decimal.

(ii) If n_0 is least positive integer such that

$$10^n \equiv 1 (\bmod b),$$

then $\underline{T}(a/b) = n_0$.

Lemma 3[3] Suppose that a/b is an irreducible proper fraction,

where

$$b=2^s5^tb_1,(b_1,10)=1,s\ge 0,t\ge 0,s+t>0,$$

$h=\max\{s,t\}$, then

(i) a/b can be denoted as a mixed recurring decimal.

(ii) The digit of noncyclic part is h in the decimal part of mixed recurring decimal.

(iii) If n_0 is least positive integer such that

$$10^n\equiv 1(\bmod b_1),$$

then $\underline{T}(a/b)=n_0$.

III. THE MAIN RESULTS

Theorem 1 Assume that a/b is an irreducible proper fraction, $b=\prod_{i=1}^{n}p_i^{k_i}$, $k_i\ge 1$, $(p_i,10)=1$, and p_i is prime, then we have that

(i) $T(a/b)=\varphi(b)=\prod_{i=1}^{n}p_i^{k_i-1}(p_i-1)$ and

$\underline{T}(a/b)\,|\,T(a/b)$;

(ii) $\underline{T}(a/b)=T(a/b)$ if and only if

$10^i\not\equiv 10^j(\bmod b)$, for any $i,j(0\le i<j\le\varphi(b)-1)$.

Proof. (i) Since $k_i\ge 1$, $(p_i,10)=1$, and p_i is prime, thus

$$(p_i^{k_i},10)=1,\text{ i.e. }(b,10)=1.$$

From Euler theorem[4], we have that

$$10^{\varphi(b)}\equiv 1(\bmod b)\text{, i.e. } b\,|\,(10^{\varphi(b)}-1).$$

Let $10^{\varphi(b)}-1=bk$ (k denotes integer), we know from $a<b$ that $ak<bk$, hence we can set

$$ak=\overline{a_1a_2\cdots a_{\varphi(b)}}(0\le a_i\le 9),$$

where $\overline{a_1a_2\cdots a_{\varphi(b)}}$ denotes the positive integer composed of $a_1,a_2,\cdots,a_{\varphi(b)}$.

Note that

$$(10^{\varphi(b)}-1)\overline{0.\dot a_1a_2\cdots\dot a_{\varphi(b)}}$$

$$=\overline{a_1a_2\cdots a_{\varphi(b)}.\dot a_1a_2\cdots\dot a_{\varphi(b)}}-\overline{0.\dot a_1a_2\cdots\dot a_{\varphi(b)}}$$

$$=\overline{a_1a_2\cdots a_{\varphi(b)}},$$

then we can obtain

$$\frac{a}{b}=\frac{ak}{bk}=\frac{\overline{a_1a_2\cdots a_{\varphi(b)}}}{10^{\varphi(b)}-1}=\overline{0.\dot a_1a_2\cdots\dot a_{\varphi(b)}}.$$

Therefore $T(a/b)=\varphi(b)$.

In addition, since

$$\varphi(b)=\varphi\Big(\prod_{i=1}^{n}p_i^{k_i}\Big)=\prod_{i=1}^{n}p_i^{k_i-1}(p_i-1)\ [4],$$

so we have that

$$T(a/b)=\varphi(b)=\prod_{i=1}^{n}p_i^{k_i-1}(p_i-1).$$

Let n_0 be minimum positive integer such that

$$10^n\equiv 1(\bmod b),$$

then from Euler theorem, we further know that $n_0\,|\,\varphi(b)$.

By combining known conditions and Lemma 2(ii), we derived

$$\underline{T}(a/b)=n_0.$$

Consequently, we conclude that $\underline{T}(a/b)\,|\,T(a/b)$.

(ii) If $\underline{T}(a/b)=T(a/b)$, by Theorem 1 (i), we have $\underline{T}(a/b)=\varphi(b)$.

Supposing that there are $i,j(0\le i<j\le\varphi(b)-1)$ such that

$$10^i\equiv 10^j(\bmod b),$$

we then have

$$b\,|\,10^i(10^{j-i}-1).$$

From $(10,b)=1$, we obtain $b\,|\,(10^{j-i}-1)$, that is, $10^{j-i}\equiv 1(\bmod b)$.

Note that Lemma 2(ii) and $1\le j-i<\varphi(b)$, obviously, this is in conflict with $\underline{T}(a/b)=\varphi(b)$. Therefore, we have $10^i\not\equiv 10^j(\bmod b)$, for any $i,j(0\le i<j\le\varphi(b)-1)$.

Conversely, if $10^i\not\equiv 10^j(\bmod b)$, for any $i,j(0\le i<j\le\varphi(b)-1)$. Specially, by setting $i=0$, we have for any $j(1\le j\le\varphi(b)-1)$ that

$$10^{j} \not\equiv 1 (\mathrm{mod}\, b),$$

Moreover, from Euler theorem, we know that

$$10^{\varphi(b)} \equiv 1 (\mathrm{mod}\, b),$$

consequently, we have $\underline{T}(a/b) = \varphi(b).$

Combining Theorm 1(i), we can obtain

$$\underline{T}(a/b) = T(a/b).$$

The proof is completed.

Theorem 2 Let a/b be an irreducible proper fraction, where

$$b = 2^s 5^t b_1,\ s \ge 0, t \ge 0, \quad s+t>0,\ b_1 = \prod_{i=1}^{n} p_i^{k_i},$$

$$k_i \ge 1, (p_i, 10) = 1,$$

and p_i is prime, then we have that

(i) $T(a/b) = \varphi(b_1) = \prod_{i=1}^{n} p_i^{k_i - 1}(p_i - 1)$ and

$\underline{T}(a/b) \mid T(a/b);$

(ii) $\underline{T}(a/b) = T(a/b)$ if and only if

$10^{i} \not\equiv 10^{j} (\mathrm{mod}\, b_1)$, for any $i, j (0 \le i < j \le \varphi(b_1) - 1).$

Proof. From known conditions and Lemma 3, we know that $\underline{T}(a/b)$ is only related to b_1, therefore, Theorem 2 can be proved by the proof method of Theorem 1. Here, we no longer prove the proposition.

Theorem 3 If $(b_1, b_2) = 1,$ then

$$T(1/b_1)T(1/b_2) = T(1/b_1 b_2).$$

Proof. From $(b_1, b_2) = 1,$ we obtain

$$\varphi(b_1)\varphi(b_2) = \varphi(b_1 b_2) \text{ [4]}.$$

And by Thorem 1(i), we have that $T(1/b_1) = \varphi(b_1),$ $T(1/b_2) = \varphi(b_2)$ and $T(1/b_1 b_2) = \varphi(b_1 b_2).$

Consequently, we have

$$T(1/b_1)T(1/b_2) = T(1/b_1 b_2).$$

The desired result follows.

According to the above conclusion, we can also obtain the following result.

Corollary If $(b_i, b_j) = 1,$ then

$$\prod_{i=1}^{n} T(1/b_i) = T(1/\prod_{i=1}^{n} b_i).$$

Meanwhile, combine with the relevant data in table 1 and the above conclusions, we make the following guess

Guess If $(b_i, b_j) = 1,$ then

$$\underline{T}(1/\prod_{i=1}^{n} b_i) \mid \prod_{i=1}^{n} \underline{T}(1/b_i).$$

IV. Conclusions And Prospect

Through the above research, we have proposed the computation formulas of period of pure recurring decimal and mixed recurring decimal respectively, relation of the period and least positive period, and the necessary and sufficient condition that the period is eaqual to least positive period.But we think that the above conclusions are not perfect, in future studies, we will further weaken the sufficient and necessary conditions, and discuss the calculation formula of least positive period of recurring decimal and other beautiful properties.

References

[1] Liying Sun and Xiaokui Ge, An application of the division algorithm and congruence to the periodic decimal, Journal Of Guangdong Education Institute,Vol 25(3),pp.31-33,2005.

[2] Limin Xue, Law of repetend of pure repeating decimal. Journal Of Weinan Teachers College, Vol 16(4),pp.92-94,2001.

[3] Dexin Zhang, Integer theory, Beijing:Science press,1958.

[4] Chengdong Pan and Chengbiao Pan, Elementary number theory ,Beijing: Peking University Press,1992.

[5] Yuejin Wang and Shanqi Pang, The new results on the fascinating properties of recurring decimals,Journal Of Henan Normal University(Natural Science),Vol 33(4),pp.25-26,2005.

[6] Shide Zhang,The fascinating properties of the recurring decimals,Journal Of Henan Normal University(Natural Science), Vol 30(3),pp.11-14,2002.

[7] Xiaosong Chen,The period and numerals sum of repeating decimal, Journal Of Yunnan University(Natural Science),Vol 29(3),pp.217-222,2007.

[8] S.S.Wagner,Fun with repeating decimais, Mathematics Teacher, 3,pp.209-212,1979.

[9] K. Seydei, Cyciic fractions, Mathematics Teacher,11,pp.567-568,1979.

[10] M.W.Ec-Ker and S.Yates, Primes and cyciic fractions. Mathematics Teacher, 5,pp.327-390,1980.

Hierarchical Cellular Structures in High-Capacity Cellular Communication Systems

Prof. R. K. Jain, Sumit Katiyar
Research Scholar, Singhania University,
Jhunjhunu, Rajasthan, India

Prof. N. K. Agrawal Sr M IEEE
Inderprastha Engineering College, Sahibabad
Ghaziabad, India

Abstract—**In the prevailing cellular environment, it is important to provide the resources for the fluctuating traffic demand exactly in the place and at the time where and when they are needed. In this paper, we explored the ability of hierarchical cellular structures with inter layer reuse to increase the capacity of mobile communication network by applying total frequency hopping (T-FH) and adaptive frequency allocation (AFA) as a strategy to reuse the macro and micro cell resources without frequency planning in indoor pico cells [11]. The practical aspects for designing macro- micro cellular overlays in the existing big urban areas are also explained [4]. Femto cells are inducted in macro / micro / pico cells hierarchical structure to achieve the required QoS cost effectively.**

Keywords- Hierarchical cellular structures; Total frequency hopping; Adaptive frequency allocation; Communication system traffic; Code division multiple access (CDMA).

I. INTRODUCTION

The provision of capacity for the increasing traffic demand in mobile radio networks comes along with the reduction of the cell size and, hence, stronger traffic fluctuations between the cells. Moreover, improved indoor coverage is required. Hierarchical cellular structures can serve indoor users and hot spots by pico- and micro cell layers, respectively, while providing coverage in the area by the macro cell layer. Moreover, hierarchical cellular structures can compensate traffic fluctuations e.g. by shifting overflow traffic from lower to higher layers. In order to avoid interference between the layers, their frequency allocations have to be coordinated. This can be achieved by incorporating smart antenna / Intelligent antenna in hierarchical structure with adaptive-SDMA approach.

Moreover, hierarchical cellular structures become a regular feature of future mobile radio networks. Although different multiple access techniques may apply, some experiences from GSM can also be useful for the design of other hierarchical cellular networks, where several layers share the same resources [1]. The results of [1], comparative simulation study has been discussed, which aims at network configurations in a dense urban environment where a high additional traffic capacity in the pico cell layer shall be achieved solely by reusing the micro- and/or macro cell frequencies. The results of number of general conclusions for the design of hierarchical cellular networks are also discussed [1].

In order to obtain a useful knowledge about the deployment and operation of macro- micro CDMA cellular overlays, we must deal with the existing conditions of today's big urban areas, which include spatial and temporal traffic distributions, geographical characteristics, user mobility characteristics, and so on. In [5] a novel algorithmic approach for the joint deployment of macro cells and micro cells over big urban areas having spatially non-uniform traffic distributions has been discussed. Based on a discrete area representation, the proposed algorithm determines the locations, radii and required capacities of macro cells and micro cells, which guarantee the required quality of service (QoS) cost effectively. For the practical design of macro-micro CDMA cellular overlays we have to take care of the following issues in depth-

i) Effect of user's mobility.
ii) In presence of many high mobility users, the undesirable micro cellular coverage holes in hot spot areas incur a critical problem, assuming a significant volume of limited macro- cellular code division multiple access (CDMA) capacity.
iii) Even though a macro- micro cellular overlay well accommodates the traffic loads of working hours, it may loose much of its significance if it can not manage a temporal traffic variation during a day.
iv) Since only a few wideband CDMA carriers exist in the third generation wireless personal communications, it is quite difficult to fully exploit all the potentials of macro-micro cellular overlays.
v) If some operating functions, such as cell layer selections or interlayer handovers, are improperly configured, they can degrade the performance of a macro-micro cellular overlay considerably.

Since the number of mobile users is continuously growing, we shrink cell size to increase system capacity. By shrinking cell size, handoff rate is increased. To overcome these problems, hierarchical cell structure is proposed. As efficient use of radio resources is very important, utilization of all resources have to be optimized. Thus, in hierarchical cell structure, how to assign available radio resources to each user is critical question [6]. In order to adapt the changes of traffic, adaptive radio resource management can be considered in CDMA based hierarchical cell structure. The proposed scheme improves call blocking, call dropping and optimal utilization of radio resources.

II. TECHNOLOGIES FOR ENHANCEMENT OF SPECTRAL DENSITY IN HIERARCHICAL SYSTEM

A tractable, flexible and accurate model for downlink heterogeneous cellular networks (fig 1) was developed successfully. The model consist of K tiers of randomly located base stations where each tier may differ in terms of average transmit power, the supported data rate and BS density. This allows elements spanning traditional, micro, pico, and femtocell BSs to be simultaneously considered. Assuming a mobile user connects to its strongest BS, we derive its Signal-to-Interference-Ratio (SIR) distribution and use to find the coverage (equivalently outage) probability over the entire network. The accuracy of these analytical results through empirical comparisons with an actual 4G macro cell network verified [7]. Cellular networks are becoming increasingly heterogeneous due to the co-deployment of many disparate infrastructure elements, including micro, pico and femtocells, and distributed antennas. A flexible, accurate and tractable model for a general downlink HCN consisting of K tiers of randomly located BSs, where each tier may differ in terms of average transmit power, supported data rate, and BS density. Assuming 1) a mobile connects to the strongest BS, 2) the target Signal to- Interference-Ratio (SIR) is greater than 0 dB, and 3) received power is subject to Rayleigh fading and path loss. Expressions for the average rate achievable by different mobile users are derived. This model reinforces the usefulness of random spatial models in the analysis and research of cellular networks. This is a baseline tractable HCN model with possible future extensions being the inclusion of antenna sectoring, frequency reuse, power control and interference avoidance/ cancellation [8].

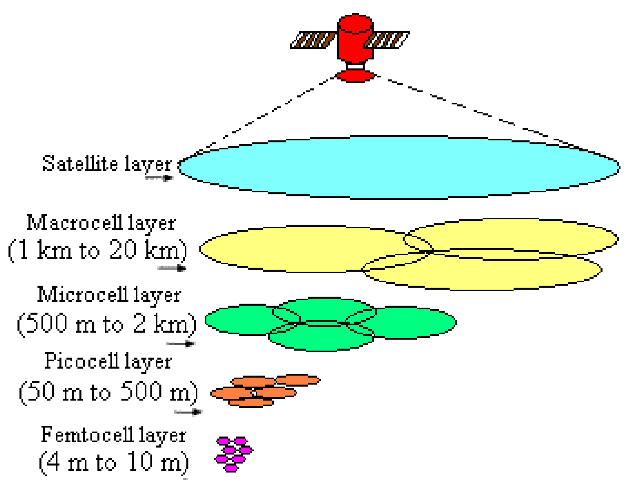


Figure 1. Hierarchical Cellular Structure

To overcome handoff problem in hierarchical cell structure, efficient use of radio resources is very important. All resources have to be optimally utilized. However, in order to adapt to changes of traffic, it is necessary to consider adaptive radio resource management. An adaptive radio resource management in CDMA based hierarchical cell structure was proposed. In this scheme, the resource shortage in micro cell is solved by increasing the number of resource and the resource shortage in macro cell is solved by decreasing traffic in macro cell. One aspect of this proposed scheme is to increase threshold velocity rather than to decrease threshold velocity when most channels in micro cell are busy. Thus, abrupt increase of macro cell load caused by decreasing threshold velocity when micro cell is overloaded can be solved. Though the number of handoff is increased in this case, the number of call loss will decrease. Since this problem occurs only during rush hours. It will prevent call from dropping or blocking although handoff rate is increased slightly. The result of proposed scheme demonstrates improvement in call dropping, call blocking and utilization of resource [6].

1) Practical designs of macro-micro CDMA cellular overlays in the existing big urban areas (having especially non-uniform traffic distributions) were proposed. The numerical results by extensive event-driven simulations show that the resulting macro-micro cellular overlays successfully cope with the existing conditions of today's big urban areas, such as spatial and temporal traffic distributions and user mobility characteristics [9].

2) The ability of hierarchical cellular structures with inter-layer reuse to increase the capacity of a GSM (Global System for Mobile Communications) radio network by applying Total Frequency Hopping (T-FH) and Adaptive Frequency Allocation (AFA) as a strategy to reuse the macro- and micro cell resources without frequency planning in indoor picocells have been discussed. The presented interference analysis indicates a considerable interference reduction gain by T-FH in conjunction with AFA, which can be used for carrying an additional indoor traffic of more than 300 Erlang/km2, i.e. increasing the spectral efficiency by over 50 %, namely 33 Erlang/km2/MHz. From these results, it can be concluded that hierarchical structures required reuse strategies that not only adapt to the current local interference situation, but additionally distribute the remaining interference to as many resources as possible. For a hierarchical GSM network this requirement is fulfilled by the T-FH/AFA technique very well [11].

3) The design of a cellular network is a complex process that encompasses the selection and configuration of cell sites and the supporting network infrastructure. This investigation presents a net revenue maximizing model that can assist network designers in the design and configuration of a cellular system. The integer programming model takes as given a set of candidate cell locations with corresponding costs, the amount of available bandwidth, the maximum demand for service in each geographical area and the revenue potential in each customer area. Based on these data, the model determines the size and location of cells, and the specific channels to be allocated to each cell. To solve problem instances, a maximal clique cut procedure is developed in order to efficiently generate tight upper bounds. A lower bound is constructed by solving the discrete optimization model with some of the discrete variables fixed. Computational experiments on seventy-two problem instances demonstrate the computational viability of our new procedure [10].

4) This paper proposes a combined channel assignment (CCA) mechanism for hierarchical cellular systems with overlaying macro cells and overlaid micro cells. The proposed CCA mechanism combines overflow, underflow, and reversible schemes, where new or handoff calls having no available

channel to use in the overlaid micro cell can overflow to use free channels in the overlaying macro cell, handoff calls from a neighboring macro cell can underflow to use free channels in the overlaid micro cell, and handoff attempts from a macro cell only region to an overlaid micro cell can be reversed to use free channels in the micro cell. We apply the CCA mechanism in two different hierarchical cellular systems of Strip type and Manhattan type and compare the CCA with the overflow channel assignment (OCA) scheme. Simulation results show that the CCA mechanism outperforms the OCA scheme by once in forced termination probability, by several times in new call blocking probability, and by 4.7% in system utilization for a hierarchical cellular system, and the CCA mechanism is more suitable for the Manhattan type than for the Strip type [1].

5) In this paper, we develop a vision for the future of wireless communications beyond the third generation, which consists of a combination of several optimized access systems on a common IP-based medium access and core network platform. Different access systems inter-work via horizontal (intra-system) and vertical (inter-system) handover, service negotiation and global roaming. These complementary access systems are optimized for different applications and environments. They are allocated to different cell layers in the sense of hierarchical cells with respect to cell size, coverage and mobility to provide globally optimized seamless services for all users. New air interfaces can also be incorporated to satisfy demands for higher data rates, increased mobility and reduced cost per bit. This vision requires international research and standardization activities to solve many technical challenges. Key issues include the global interworking of different access systems on a common platform, advanced antenna concepts and the implementation of multi-mode and multi-band terminals as well as base stations via software-defined radio concepts [12].

6) The surest way to increase the system capacity of a wireless link is by getting the transmitter and receiver closer to each other, which creates the dual benefits of higher quality links and more spatial reuse. In a network with nomadic users, this inevitably involves deploying more infrastructures, typically in the form of micro cells, hotspots, distributed antennas, or relays. A less expensive alternative is the recent concept of femtocells—also called home base-stations—which are data access points installed by home users to get better indoor voice and data coverage. In this article, author overviewed the technical and business arguments for femtocells, and described the state-of-the-art on each front. Authors also described the technical challenges facing femtocell networks, and gave some preliminary ideas for how to overcome them [13].

The demand for cellular radio services growing rapidly, and in heavy populated areas the need arises to shrink the cell sizes and scale the coverage pattern. The extension of the service into the PCN domain, railway stations, malls, pedestrian areas, markets and other hotspots further enhances this trend. The vision of future cellular systems incorporates macro, micro, pico and femto cells in hierarchical structure (fig 2).

Architectures proposed introduce the concept of remoting the antennas from the cell site, thus allowing for many micro cells to be served by a single attended center. One scheme advocates a linear transformation of the RF to optical frequencies and relaying the signals via fibers to a center. Others propose a down conversion to IF (70 MHz) which is then relayed by microwaves or by fibers.

A. Macro Cell

A conventional base station with 20W power and range is about 1 km to 20 km. Macro cell in hierarchical structure takes care of roaming mobiles.

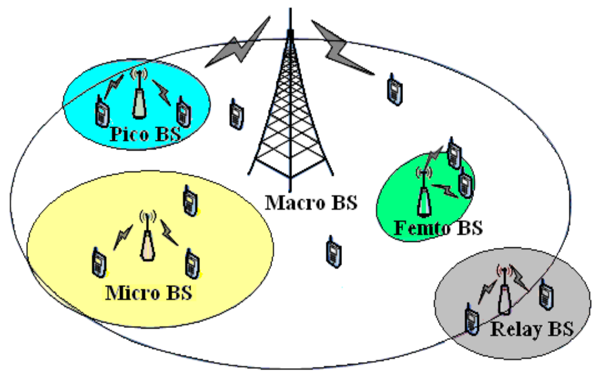


Figure 2. An example of single RAT hierarchical (Multi-tier) architecture framework

B. Micro Cell

A conventional base station with 1W to 5W power and range is about 500 m to 2 km. Micro cells and pico cells takes care of slow traffic (pedestrian and in-building subscribers). Micro cells can be classified as the following:

1) Hot Spots: These are service areas with a higher tele-traffic density or areas that are poorly covered. A hot spot is typically isolated and embedded in a cluster of larger cells.

2) Downtown Clustered Micro cells: These occur in a dense, contiguous area that serves pedestrians and mobiles. They are typically found in an "urban maze of" street canyons," with antennas located far below building height.

3) In-Building. 3-D Cells: These serve office buildings and pedestrians (fig 3). This environment is highly clutter dominated, with an extremely high density and relatively slow user motion and a strong concern for the power consumption of the portable units.

C. Pico Cell

The picocells are small versions of base stations, ranging in size from a laptop computer to a suitcase.

Besides plugging coverage holes, picocells are frequently used to add voice and data capacity, something that repeater and distributed antenna cannot do.

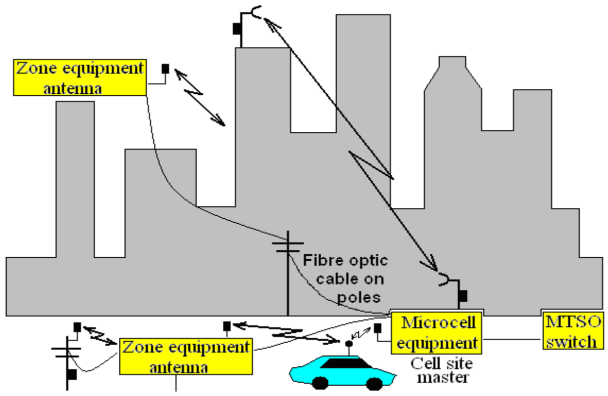


Figure 3. Microcell Installation Concept

Adding capacity in dense area, splitting cells are expensive, time consuming and occasionally impossible in dense urban environment where room for a full size base station often is expensive or unviable. Compact size picocells makes them a good fit for the places needing enhanced capacity, they can get.

Picocells are designed to serve very small area such as part of a building, a street corner, malls, railway station etc. These are used to extend coverage to indoor area where outdoor signals do not reach well or to add network capacity in areas with very dense uses.

D. Femto Cell

A femtocell is a smaller base station, typically designed for use in home or small business**.** In telecommunications, a femtocell is a small cellular base station, typically designed for use in a home or small business. It connects to the service provider's network via broadband (such as DSL or cable); current designs typically support 2 to 4 active mobile phones in a residential setting, and 8 to 16 active mobile phones in enterprise settings. A femtocell allows service providers to extend service coverage indoors, especially where access would otherwise be limited or unavailable (fig 4). Although much attention is focused on WCDMA, the concept is applicable to all standards, including GSM, CDMA2000, TD-SCDMA, Wi-MAX and LTE solutions. Femtocells are also called home base stations which are data access points installed by home users to get better indoor voice and data coverage. Data networks require much higher signal quality in order to provide the multi-Mbps data rates to individual user in the prevailing scenario (fig 5) and this can be met by using femtocells in hierarchical structure (macro, micro, pico cells along with femto cell for home user). This hierarchical structure will also result in better coverage, improved capacity, improved macro / micro cell reliability, reduced cost, reduced power consumption and RF pollution.

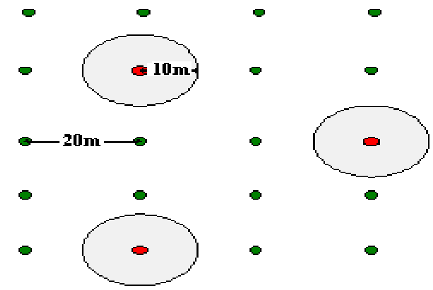


Figure 4. Illustration of FAP deployment

The capacity benefits of femtocells are:

1. Reduced distance between the femtocell and the user, which leads to a higher received signal strength. This will result in improvements in capacity through increase signal strength and reduced interference.

2. Lowered transmit power, and mitigation of interference from neighboring macro cell and femto cell users due to outdoor propagation and penetration losses. This will result in improvements in capacity through increased signal strength and reduced interference.

3. As femtocells serve only around 1-4 users, they can devote a larger portion of their resources (transmit power & bandwidth) to each subscriber. A macro / micro cell, on the other hand, has a larger coverage area (500m-20 km radius), and a larger number of users; providing Quality of Service (QoS) for data users is more difficult. Deploying femtocell will enable more efficient uses of precious power and frequency resources.

III. Smart / Adaptive Antenna

The adoption of smart antenna techniques in future wireless systems is expected to have a significant impact on the efficient use of the spectrum, the minimization of the cost of establishing new wireless networks, the optimization of service quality, and realization of transparent operation across multi technology wireless networks. Smart antennas have emerged as potentially a leading technology for achieving highly efficient networks which maximize capacity and improve quality and coverage.

Smart antenna can provide greater capacity and performance benefits than standard antennas because they can be used to customize and fine-tune antenna coverage patterns that match the traffic conditions in a wireless network or that are better suited to complex radio frequency (RF) environments. Furthermore, smart antennas provide maximum flexibility by enabling wireless network operators to change antenna patterns to adjust to the changing traffic or RF conditions in the network [16].

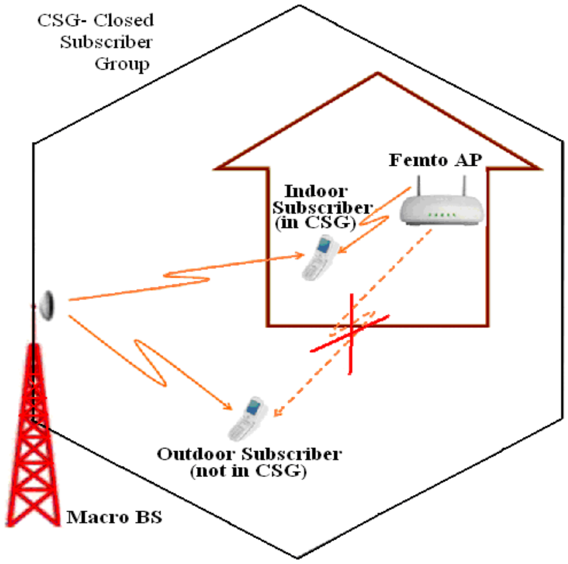


Figure 5. CSG Cell Association

Smart antennas at base stations can be used to enhance mobile communication systems in several ways:

- Increased BS range
- Less interference within the cell
- Less interference in neighboring cells
- Increased capacity by means of SFIR or SDMA

'Smart' antenna transmitters emit less interference by only sending RF power in the desired directions. Furthermore, 'smart' antenna receivers can reject interference by looking only in the direction of the desired source. Consequently 'smart' antennas are capable of decreasing CCI. A significantly reduced CCI can be taken as advantage of Spatial Division Multiple Access (SDMA) [9]. The same frequency band can be re-used in more cells, i.e. the so called frequency re-use distance can be decreased. This technique is called Channel Re-use via Spatial Separation. *In essence, the scheme can adapt the frequency allocations to where the most users are located [19].* With the inclusion of nanotechnology devices, the capability of adaptive antenna will increase manifold.

IV. APPLICATION OF NANOTECHNOLOGY

Nano- technology could provide solutions for sensing, actuation, radio, embedded intelligence into the environment, power efficient computing, memory, energy sources, man-machine interaction, materials, mechanics, manufacturing & environmental issues.

Nanotechnology will rapidly boost all these disciplines and their application areas. Economic impact is foreseen to be comparable to information technology and telecom industries. One of the central visions of wireless industry aims at ambient intelligence: computation and communication always available and ready to serve the user in an intelligent way, so that we may optimally utilize the scarcest resource- frequency spectrum and achieve faster speeds (i.e., enhanced data rate).

Mobile devices together with the intelligence that will be embedded in human environments- home, office, public places- will create a new platform that enables ubiquitous sensing, computing, storage and communication. Core requirements for this kind of ubiquitous ambient intelligence are that the devices are autonomous and robust. They can be deployed easily and require little maintenance. As data rates require more memory and computing shown in fig 6, mobile devices will be the gateways to personally access ambient intelligence and needed information. Mobile also implies limited size and restrictions on the power consumption. Seamless connectivity with other devices and fixed networks is a crucial enabler for ambient intelligence system- this leads to requirements for increased power, which together with the size limitations leads to severe challenges in thermal management. All above requirements can be addressed satisfactorily with the application of nanotechnology.

Nanotechnology may augment the sensory skills of a human based on wearable or embedded sensors and the capabilities to aggregate this immense global sensory data into meaningful information for our everyday life.

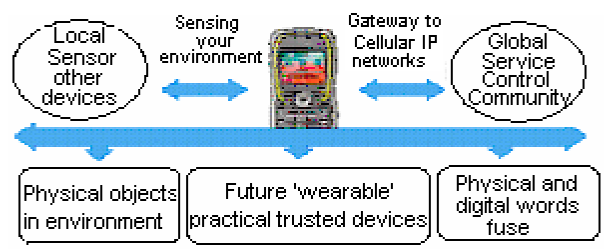


Figure 6. Mobile devices become gateway to ambient intelligence and needed information

Nanotechnology can help to develop novel kind of intelligent devices where learning is one of the key characteristic properties of the system, similarly to biological systems which grow and adapt to the environment autonomously [17].

V. HANDOFF IN CELLULAR SYSTEM

Handoff is an essential element of cellular communications. Efficient handoff algorithms are a cost-effective way of enhancing the capacity and QoS of cellular systems. Macro cell radii are in several kilometers. Due to the low cell crossing rate, centralized handoff is possible despite the large numbers of MSs are to be managed by MSC. The use of micro cells is considered the single most effective means of increasing the capacity of cellular systems. Micro cells are more sensitive to the traffic and interference than macro cells due to short term variations (e.g., traffic and interference variations), medium / long term variations (e.g., new buildings), and incremental growth of the radio network (e.g., new BSs) [18]. The number of handoffs per cell is increased by an order of magnitude, and the time available to make a handoff is decreased [19]. Using an umbrella cell is one way to reduce the handoff rate. Due to the increase in the micro cell boundary crossings and expected high traffic loads, a higher degree of decentralization of the handoff process becomes necessary [20]. Lot of handoff schemes has been developed. The schemes packages show that these schemes can significantly decrease both the number of dropped handoff calls and the number of blocked calls without degrading the quality of communication service and the soft handoff process [18], [19], [20], [21], [22], [23], [24], [25], [26].

Handoff is an integral component of cellular communications. Efficient handoff algorithms can enhance system capacity and service quality cost effectively.

VI. Proposed Network

Proposed network (Fig 7 & 8) is based on simple handoff algorithm discussed in [25] and hierarchical cellular structures with inter-layer reuse in an enhance GSM radio network is suggested [1]. A design of Macro-Micro CDMA Cellular Overlays in the Existing Big Urban Areas is suggested [4]. It is also assumed that the MS is equipped with a Rake receiver capable of performing "maximal ratio combining" of the signals it receives from the transmitting BSs [26]. The following cellular structure will be used for dense urban areas [27]:

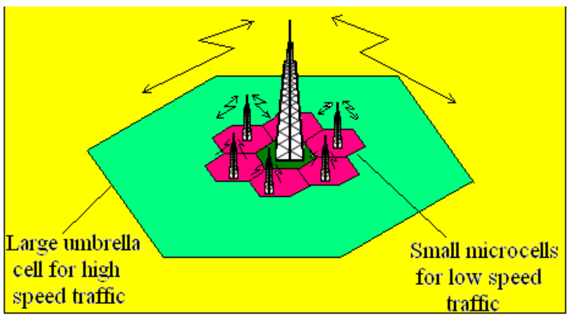


Figure 7. The Umbrella Cell Approach

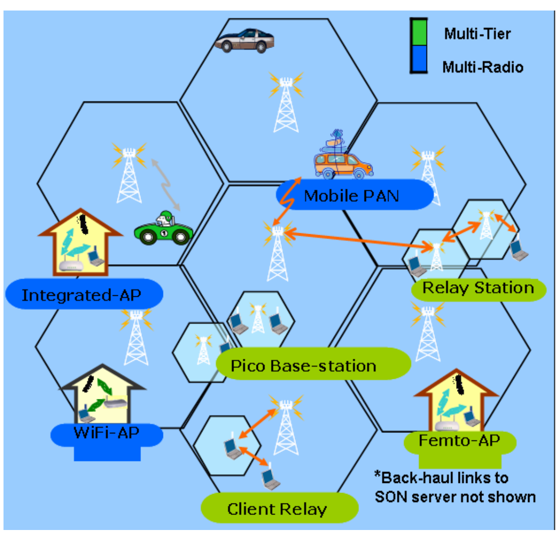


Figure 8. Hierarchical Structure [29]

1) Macro cell will be marked for fast traffic and micro cell will be marked for slow traffic in hierarchical structure. The RF resources will be dynamically allocated between macro and micro cells on the basis of velocity estimation using adaptive array antennas [24].

2) Pico cell will be marked for hotspots. An adaptive frequency allocation will be applied as strategy to reuse the macro and micro cell resources without frequency planning in indoor picocells [1].

3) Femto cell will be marked as home base stations- which are data access points installed by home users to get better indoor voice and data coverage. Femtocells enable a reduced transmit power, while maintaining good indoor coverage. Penetration losses insulate the femtocell from surrounding femtocell transmissions. As femtocells serve only around 1-4 users, they can devote a larger portion of their resources (transmit power & bandwidth) to each subscriber. A macro / micro cell, on the other hand, has a larger coverage area (500m-20 km radius), and a larger number of users; providing Quality of Service (QoS) for data users is more difficult. Deploying femtocell will enable more efficient uses of precious power and frequency resources [13].

The above proposed structure will undoubtedly enhance the spectral density with the help of diversity and adaptive approach through rake receiver and adaptive antenna respectively. The induction of pico and femto cell will reuse the RF resources of overlaid macro / micro structures which will enhance spectral density manifold. The simple technologies suggested by William C Y Lee for deployment along city streets, deployment along binding roads, deployment under the ground (subway coverage) and in-building designs etc. will also be considered in proposed hierarchical structure [28].

Conclusion

By means of adaptive frequency allocation using adaptive antenna patterns smart / adaptive antennas allow steering the transmit / receive power into certain directions by suppressing the desired power, i.e interference. Additionally the sectorization scheme offers various benefits being combined with SDMA. This has been proved beyond doubt that spectral density will be enhanced substantially. In addition to it, overlaid structures will reduce the handoff using simple handoff algorithms. Moreover RF resources of macro / micro will be reused in pico cell without frequency planning. This will further enhance the capacity of the system beyond any doubt. The use of femto cell will further increase the capacity of the system. In essence proposed network will be a highly efficient cost-effective way of enhancing the capacity and QoS of cellular systems. The above proposed system will be highly effective energy saving cellular network which will reduce power consumption and RF pollution significantly.

References

[1]. Kuen-Rong Lo, Chung-Ju Chang, Cooper Chang and C.Bernard Shung, "A combined channel assignment mechanism for hierarchical cellular systems" computer communications, pp 1943-1952, vol. 21, 1998

[2]. M. Frullone, G. Riva, P. Grazioso, and M. Barbiroli, "Advanced planning criteria for cellular systems," IEEE Personal Comm., pp. 10-15, December 1996.

[3]. M. Madfors, K. Wallstedt, S. Magnusson, H. Oloffson, P.-O. Backman, and S. Engström, "High capacity with limited spectrum in cellular systems," *IEEE Comm. Mag.*, pp. 38-44, August 1997.

[4]. Byungchan Ahn, Hyunsoo Yoon and Jung Wan Cho, "A design of Macro-Micro CDMA Cellular Overlays in the Existing Big Urban

Areas" IEEE journal on selected areas in communications, vol. 19, no. 10, October 2001.

[5]. B. Ahn, H. Yoon and J. W. Cho, "Joint deployment of macrocells and microcells over urban areas with spatially nonuniform traffic distributions" in proc. IEEE 52nd Veh. Technol. Conf., Boston, MA, Sept. 2000, pp. 2634-2641.

[6]. Taesoo Kwon and Dong-Ho Cho, "Adaptive Radio Resource Management Based on Cell Load in CDMA Based Hierarchical Cell Structure" 0-7803-7468, IEEE, 2002. http://brahms.kaist.ac.kr/down/file_F8E3W_.pdf

[7]. Harpreet S. Dhillon, Radha Krishna Ganti and Jeffrey G. Andrews, "A Tractable Framework for Coverage and Outage in. Heterogeneous Cellular Networks."- http://ita.ucsd.edu/workshop/11/files/paper/paper_323.pdf.

[8]. Harpreet S. Dhillon, Radha Krishna Ganti, Franc¸ois Baccelli and Jeffrey G. Andrews, "Average Rate Achievable in K-Tier Downlink Heterogeneous Cellular Networks"- wncg.org/publications/dl.php?file=Andrews1308428051.pdf.

[9]. Byungchan Ahn, Hyunsoo Yoon, and Jung Wan Cho, "A design of Macro-Micro CDMA Cellular Overlays in the Existing Big Urban Areas" IEEE journal on selected areas in communications, vol. 19, no. 10, October 2001.

[10]. Joakim Kalvenes, Jeffery Kennington, Eli Olinick, "Hierarchical Cellular Network Design with Channel Allocation" European journal of operational research, pp. 3-18, 2003.

[11]. Jurgen Deissner and Gerhard P. Fettweis, "A Study on Hierarchical Cellular Structures with Inter-Layer Reuse in an Enhanced GSM Radio Networks" Dresden University of Technology, Mannesmann Mobilfunk Chair for Mobile Communications Systems, pp 1-9, 1999. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.3770&rep=rep1&type=pdf.

[12]. Kurt Aretz, Martin Haardt, Walter Konhauser and Werner Mohr, "The future of wireless communications beyond the third generation" Computer Networks, pp 83-92, Volume 37 Issue 1, September 2001.

[13]. Vikram Chandrasekhar and Jeffrey G. Andrews, "Femtocell Networks: A Survey" June 28, 2008-arxiv.org/pdf/0803.0952.

[14]. http://arxiv.org/ftp/arxiv/papers/0803/0803.0952.pdf.

[15]. www.mhprofessional.com/.../boccuzzi_0071633596.pdf.

[16]. Dr. Mir Mohammad Azad, Abu Hasnat Shohel Ahmed, "Development of Smart Antenna for Future Generation Wireless Internet Connection" IJCSNS International Journal of Computer Science and Network Security, Vol. 10 No. 10, Oct 2010.

[17]. Ermolov V., Heino M., Karkkainen A., Lehtiniemi R., Nefedov N., Pasanen P., Radivojevic Z., Rouvala M., Ryhanen T., Seppala E., Uusitalo M. A., "Significance of Nanotechnology for Future Wireless Devices and Communications" 18th IEEE International Symposium PIMRC'07.

[18]. R. Beck and H. Panzer, "Strategies for handover and dynamic channel allocation in micro-cellular mobile radio systems" Proc. 39th IEEE VTC, 1989, pp. 178-85.

[19]. S. Chia and R. J. Warburton, "Handover criteria for a city microcellular radio systems" Proc. 40th IEEE VTC, 1990, pp. 276-81.

[20]. G. Liodakis and P. Stavroulakis, "A novel approach in handover initiation for microcellular systems" Proc. 44th IEEE VTC, 1994, pp. 1820-23.

[21]. Xiaomin Ma, Yun Liu, Kishor S. Trivedi, "Design and Performance Analysis of a New Soft Handoff Scheme for CDMA Cellular Systems" IEEE Transactions on Vehicular Technology, Vol. 55, No. 5, September 2006.

[22]. Xiaomin Ma, Yonghuan Cao, Yun Liu, Kishor S. Trivedi, "Modeling and Performance Analysis for Soft Handoff Schemes in CDMA Cellular Systems" IEEE Transactions on Vehicular Technology, Vol. 55, No. 2, March 2006.

[23]. Rosli Salleh and Xichun Li, "Handoff Techniques for 4G Mobile Wireless Internet" 3rd International Conference, SETIT, March 27-31, 2005, Tunisia.

[24]. Young-uk Chung and Dong-Ho Cho, "Velocity Estimation using Adaptive Array Antennas" 0-7803-6728-6/01, IEEE, 2001.

[25]. Nishith D. Tripathi, Nortel Jeffrey H. Reed and Hugh F. Van Landingham, "Handoff in Cellular Systems" IEEE Personal Communications, December 1998.

[26]. Dan Avidor, Sayandev Mukherjee and Nidhi Hegde, "On the Impact of Soft Handoff Threshold and Maximum Size of the Active Group on BS Transmit Power in the UMTS System" IEEE Transactions on Vehicular Technology, Vol.55, No.2 March 2006.

[27]. Sumit Katiyar, Prof R. K. Jain and Prof N. K. Agarwal, "Proposed Cellular Network for Indian Conditions for Enhancement of Spectral Density and Reduction of Power Consumption & RF Pollution" IEEE 2nd International Conference, ICCCT-2011, 15 - 17 September 2011.

[28]. A book on "Mobile Cellular Telecommunications Analog and Digital Systems" 2nd edition, chapter 16, pp. 563-616 by William C. Y. Lee.

[29]. IEEE C802.16ppc-10_0008r3, "Study Report on Hierarchical Networks," Inuk Jung, Nageen Himayat, January, 2011.

AUTHORS PROFILE

Prof. R. K. Jain has received his BE and ME in 1972 & 1974 from IIT Roorkie, respectively. He has 32 years industrial experience from M/S ITI Ltd. (Top Management) and more than 5 years teaching experience as professor. He is currently working as Prof. and Head of Electronics & Communication Engineering at HIET, Ghaziabad, UP, India. He is currently pursuing PhD degree course from Singhania University as research scholar. His area of research is cellular communication. He has published 9 papers in National / International Conferences and journals.

Sumit Katiyar has received his B.Sc. and M.Sc. in 2001 & 2004 respectively from CSJMU, Kanpur, UP, India. He has received his M.Tech. from JIITU, Noida, UP, India. He is currently pursuing PhD degree course from Singhania University as research scholar. He is currently working as Asst. Prof. in department of Electronics & Communication Engineering at HIET, Ghaziabad, UP, India.

Prof. N. K. Agrawal has received M.Sc. (Electronic) degree from Agra University, Agra, M.Sc (Tech) degree from Birla Institute of Technology, Pilani, Rajsthan and Ph.D degree in Microwave and Radar from Indian Institute of Technology, Roorkee in 1978. He is a senior member of IEEE (USA) and life member of ISTE and ISCEE. He was awarded Young Scientist research award by exchange program between India and France. He has visited many countries. He has worked as visiting professor at University of Shalahddin, Erbil, Iraq. His current Interest includes Antenna, Electromagnetic, Microwave Engineering, and Microwave Absorbers. He has published / presented large number of research papers in these areas.

Hybrid Query by Humming and Metadata Search System (HQMS) Analysis over Diverse Features

Nauman Ali Khan
Department of Computer Science, Quaid-I-Azam University Islamabad, Pakistan.

Mubashar Mushtaq
Department of Computer Science, Quaid-I-Azam University Islamabad, Pakistan.

***Abstract*— Retrieval of music content over web is one of the toughest job tasks and found some of the significant challenge. Song retrieval over web is the emerging problem from the category of Music Information Retrieval. Several searching techniques related to metadata and content of song are developed and implemented. In this paper we are going to propose and evaluate a new hybrid technique for song retrieval named as "hybrid query by humming and metadata search system" (HQMS).HQMS is a hybrid model that is based on metadata and query by humming. HQMS narrow downs the relevant results thus improve overall accuracy of the system. Two filters, metadata and query by humming work together and enhances overall accuracy. The system is evaluated song wise, metadata wise and Age wise. Issues related to query reformulation are also highlighted as a part of the paper. More over the evaluation validates the hypothesis and the proposed Architecture.**

Keywords- Query by humming; Metadata; MIR; Hybrid system; Pipe and Filter; Evaluation.

I. Introduction

Validation of any software system is the most crucial phase of software testing to identify the software quality, as it evaluates the validity of the system on the basis of different attributes [1]. In this paper, we have discussed the experimental setup, dataset, development and evaluation of Hybrid Query by Humming and Metadata Search System (HQMS) developed by us [2]. Query by humming is the well known Content Based Searching Technique used for song retrieval. Query by humming takes humming of any song as a query and by applying different pattern matching algorithms, evaluation of relevant retrieve results is performed [3]. HQMS is an improved version of Query by humming techniques; in which we have use metadata as a collaborative part to ordinary QbH technique. Improving the overall accuracy through result filtration is one of the core inspiration of HQMS.

One of the common mechanisms used for the song retrieval is the keyword query based searching mechanism and that totally relies on the metadata of songs. Content inside file is generally considered when either the metadata information stored with is less explained or user is not familiar to the metadata file. Content Based Searching basically searches in all parts of the file and tries to find the best match. When we are considering songs, contents can be electronic waves stored in some specific format. In the field of songs retrieval, Content Based Searching is commonly applied when a user has very less knowledge about a song that is to be searched [4].

If we observe closely, when a user is interested in searching any particular song, it mostly tries to remind the lyrics of the song, if that is not possible, it tries to remind the name of the movie or album, if that is also not possible, he tries to hum the tune of that song. QbH was introduced to provide the facility to search by Humming. Our system (HQMS) of handling some critical issues which ordinary QbH (Query By Humming) cannot resolve. HQMS is a hybrid system that searches the most relevant list of songs on the basis of metadata and humming. HQMS provides better results in the term of accuracy as compare to ordinary QbH.

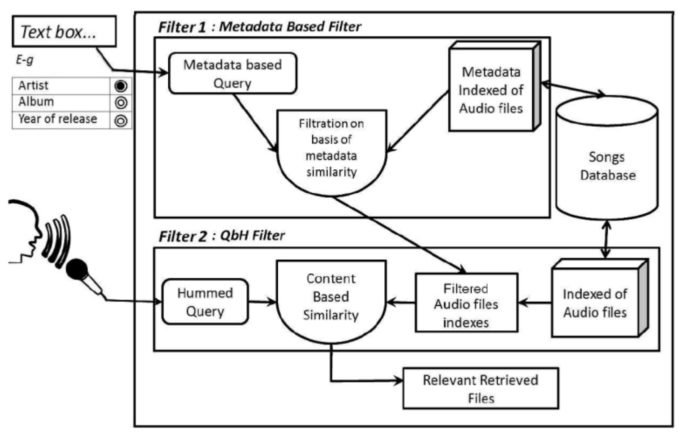


Figure. 1. Architecture of HQMS [2]

In HQMS System ordinary query by humming system was modified by an extra component of metadata. Fig. 1 explains the Architecture of HQMS. Two filters, metadata and query by humming are the core components of HQMS. Reason of introducing two filters in series is to decrease the size of dataset and to increase the overall accuracy of relevant retrieved results. The main objective of First Metadata Filter is to reduce the resultant dataset so that Second Query by Humming Filter can efficiently and accurately find the relevant results. Metadata based filter sort out the list of songs by Artist Name, Album Name, Year, Genre and Title whereas second filter only matches those files which are being filtered by first filter. In the HQMS it was described that two filters can provide better relevant files and the evaluations of HQMS given in this paper.

Basic motivation of the HQMS was that when a user is interested in finding a song when it do not know the exact lyrics of song but remember the tune and some portion of the metadata; in such case HQMS come across with better results than the ordinary QbH. Both Filters work with collaborated effort.

The rest of the paper is arranged as follows: Section II describes the literature review on query by humming, metadata and hybrid systems. Section III elaborates the HQMS in detail. Section IV discusses the experimental setup for the HQMS. Section V demonstrates develop-ment of HQMS. Section VI explains the extensive evaluation of HQMS in different directions. In Last section VII concludes with future directions.

II. LITRATURE REVIEW

Searching activity is always considered as the common activity of the end user. Various searching techniques are invented, developed and implemented. Music searching activity inside computer or over web is the sub activity of searching category. Field of searching music from large repository of internet is called as Music Information Retrieval (MIR). MIR is a vast sub field of Information Retrieval (IR) especially dedicated to resolve the issues during retrieval of music files. The process of MIR generically depends on either metadata or the content of song [5].

Metadata of song works as a main core for MIR. Several metadata based searching techniques are made like vector space model, Boolean model, indexing, invert index file, cosine measure etc [6] [7] [8] [9]. Currently storing of exact metadata is being more emphasized as per to the standards of Web 2.0. According to the standard that sort of metadata is preferred which have some meaning that somehow defines the proper context of song. As metadata of song is less as compared to ordinary document files so different string mining techniques are also applied [5].

For quick retrieval existing IR indexing and clustering techniques are also applied just to handle such critical issues associated to the Music Retrieval [9] [10]. Searches get enormous when we are dealing with huge databases of songs

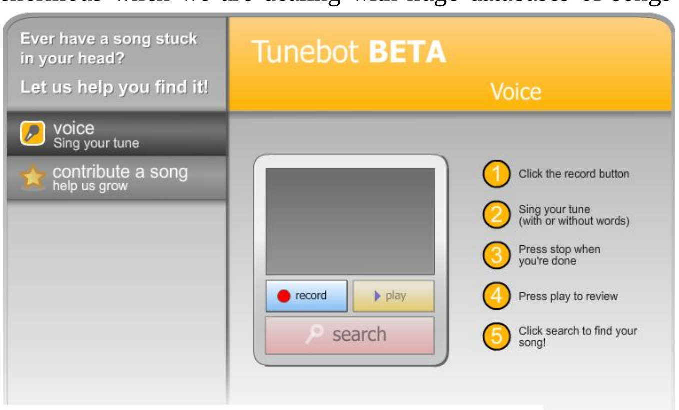


Figure. 2. Tunebot System [11]

[5]. Some of the existing Query by humming systems are explained below.

TUNE BOT is National Science Foundation funded project made by a Northwestern University Interactive Audio Lab [11]. TUNE BOT is basically a search engine that works on the basis of query by humming. System contain total of 4990 songs. Each song in the repository is linked to Amazon online shopping store. Genetic Algorithm is used to improve the overall accuracy. System improves performance by adopting explicit feedback from user. Whole system works in three steps, firstly takes the query from user as humming, then shows a list of result to user and ask whether it satisfies the need or not. By taking feedback training of Genetic Algorithm is performed and performance is improved accordingly [11] [13].

Midomi is the commercial website developed by Me-lodis

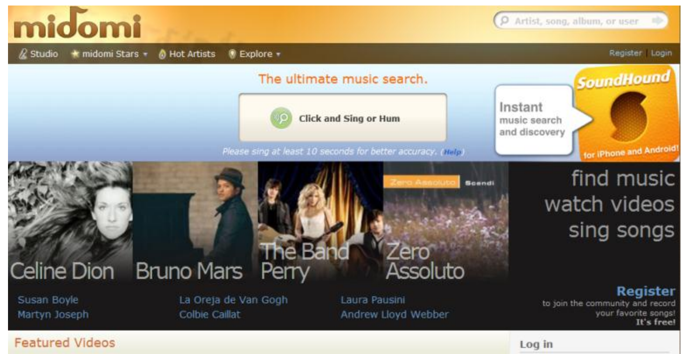


Figure. 3. Midomi System [14]

Corporation in August 15, 2006 [14]. Midomi works on query by humming, singing and whistle also provide an advance search on the basis of genre and language. Midomi got the huge repository of more than two million music files.

Figure. 4. Soundhound System [15]

Sound hound is another project of Melodis Corporation specially designed for mobile and iPods [15]. Sound Hound is also commercial software works same as mi-domi on basis of hum, sing and whistle. Currently Sound Hound is designed for iphones, ipod touch, iPods and android mobiles [15].

Musipedia is an open source project specially focusing on retrieving songs in very diverse manner [16]. For the ease of end user, system provides a facility to search song in three different categories of Melodic transcription, Contour transcription, Rhythm based. System supports query by humming, whistling, singing, contours and tapping.

Music and Audio Retrieval Tools (MaART) is open source project [17]. MaART is basically a combination of different software components specially made to evaluate different systems. In majority, system is written in C++ language. Few of the main features of MaART is extracting text from midi file, converting of any midi audio file into piano notes and extracting part from any wav file.

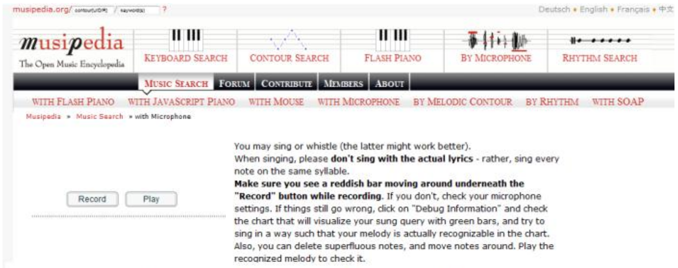


Figure. 2. Musipedia System [16]

Sloud is an ActiveX technology based project [18]. In-dexing of audio files and query is the core theme of this project. Sloud uses client and server architecture. Client side user query is transformed into segments and then matched to the existing indexes of stored repository. Pitch recognition measure is used for matching.

Music contents over web are multiplying day by day [5]. Currently the greatest challenge is to reduce the rele-vant retrieve results for the end user. Metadata based searching technique gets flop when a user is unable to provide any metadata query. To overcome this limitation Content based searching techniques plays a vital role. Content based technique look into the audio file by extracting features. Content and metadata based searching techniques works side by side. A content based searching technique enhances the overall performance of metadata search system rather than a full replacement.

III. Proposed Architecture

Results of information Retrieval System can be more accurate if detailed query is provided to a search system. This section describes our proposed HQMS. In our system, filtration feature acts as a core to the pipe and filter software architecture style thus resulting in a “Hybrid QbH and Metadata Search System” shown in Fig.1. Pipe and filter software architecture style is mostly used when output of one filter is considered as the input of second filter. In the commonly adopted approaches, there is a single filter having single input and output unit [12].

In proposed architecture, we present two filters in series, such that search query has to go from the pipe that has to pass from both filters one by one. The first filter is metadata wise search and the second follows a Query by Humming process as shown in Fig [2].

Architecture strength lies in the base concepts for building as per needs of user. Behind the adoption of a series pipe and filter architecture, is the vital goal to de-crease the number of files to be checked by a QbH system.

A. Metadata Filter

Metadata filter is actually the combination of multiple sub-filters which work altogether to get precise results. User has the privilege to provide any sort of information that has the probability of presence in metadata. Metadata information can be related to any of the attribute of the song's metadata. The provided information would be matched to provide a set of possible outcomes. The filter will be operated against the inputted query in the metadata query section. But if a user does not input any metadata than Hybrid system will work as ordinary QbH system, as this filter would not provide an output. This system is capable of handling multiple metadata queries simultaneously.

According to the survey it was examined that most of the users search song on the basis of artist name and title of the song [24], by this we conclude that there is equal possibility that a user wants to give more than one inputs as a query. To handle such situation first filter i.e., Metadata based filter will work as a multiple sub-filters. It will further reduce the number of files for the next filter.

B. Query by humming Filter

As soon as the results are being produced by the first filter (metadata), they are stored in a temporary repository. This repository acts as an input unit for the second filter (QbH) along with the hummed query from the end user. Hummed query is then transformed into time series graph and then that graph is matched accordingly to the pre-indexed files. The filtering process of this unit is to provide further filtration to the temporary repository on the basis of hummed query matched. The match that is found in temporary repository is to be further purified one by one by humming comparison.

Thus, looking at the architecture on a broader perspective, first filter concise the scope of search and second filter increases the relevance of outcome with the desired query from user.

The outcome shown after the processing of both the filters indicates the relevant retrieved files against a query. Other than simple retrieval of relevant audio files, the system can also work as a strong copyright violation detector.

The system would work for both metadata and hum, thus will provide a more authentic result. There have been many issues where a copyrighted song is sung again by a person which violates the copyrights of the owner. Our system can be used in detecting such violations and privacy; copyrights checking authorities and system like [22], will have a prominent advantage of using it.

IV. Experimental Setup

We Gathered a Dataset of total 150 Pakistani songs in six different languages, such that 25 songs for each language and 25 English of songs. Seven selected languages were Urdu, Sindhi, Punjabi, Pashto, Balochi, Kashmiri and English. After applying Stratified Random Selection five songs from each category were selected as a dataset for the Evaluation.

TABLE I
DETAILS OF DATASET OF SONGS

Language	No. of Songs
Urdu	5 Songs
Punjabi	5 Songs
Pashtu	5 Songs
Sindhi	5 Songs
Kashmiri	5 Songs
English	5 Songs
Balochi	5 Songs

Description of Dataset is described in the Table I.

Three different age users were selected for six categories of Pakistani (Cultural / Regional) languages i-e Urdu, Pashto, Punjabi, Balochi, Kashmiri, and Sindhi. Among three different categories, age of first category was in between 10year to 15year, age of Second category was in between 16year to 20year and age of Third category was in between 21year to

TABLE II
DETAILS OF USERS

Language	Age limit $\geq$11yr and $\leq$15yr		Age limit $\geq$16yr and $\leq$20yr		Age limit $\geq$21yr and $\leq$25yr	
Urdu	1 Male	1Female	1 Male	1Female	1 Male	1Female
Punjabi	1 Male	1Female	1 Male	1Female	1 Male	1Female
Pashtu	1 Male	1Female	1 Male	1Female	1 Male	1Female
Sindhi	1 Male	1Female	1 Male	1Female	1 Male	1Female
Kashmiri	1 Male	1Female	1 Male	1Female	1 Male	1Female
Balochi	1 Male	1Female	1 Male	1Female	1 Male	1Female

25year. Details of users are de-scribed in the Table II.

V. DEVELOPMENT OF HQMS

We developed HQMS architecture [2] as shown in Fig.1. Five categories were taken in the consideration while developing metadata filter that includes Album Name, Artist Name, Genre, Title of the Song and Year of Release. We used a Levenshtein string matching function for the string comparison of query and stored metadata of song [19]. By observing the human natural behavior, for searching any song we introduced query weighting as well. Motivation scenario for the giving weight to a query was

"When a user is interested in searching a song, such that he does not remember the lyrics of the song but remembers the tune of the song and little portion of Metadata the song, and in that little metadata he can further classifies the priority to each category"

For example a user is pretty much sure that the song is of *Michael Jackson* Artist but he is less sure about the genre of the song whether it is *Rock* or *Pop* to handle such situation we provided a priority based advance search. User can easily adjust weight of the different parameter of query by using a

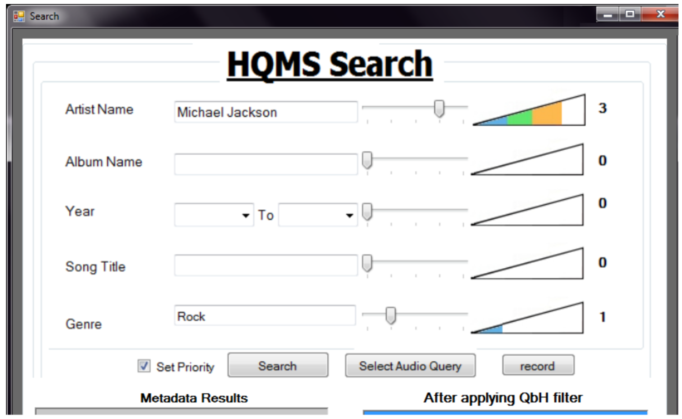
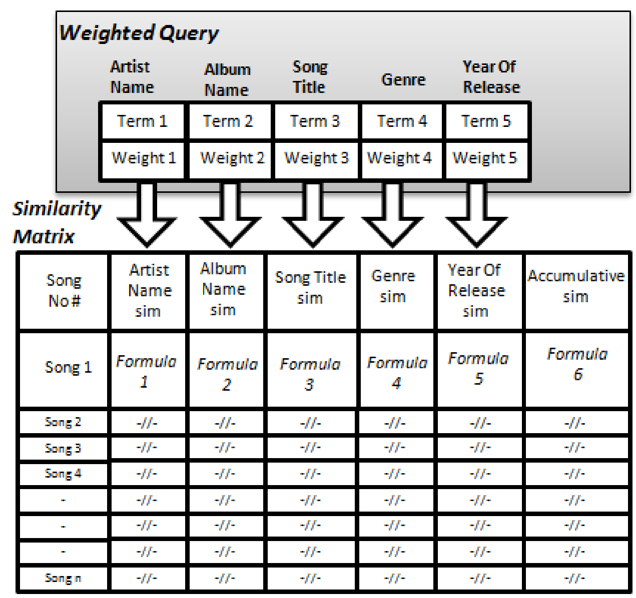


Figure. 6. HQMS System interface for metadata search

slide bar. Between the values of 0 to 5 i-e (0, 1, 2, 3, and 4) were kept. Fig. 6 explains the interface of the priority Based Advance Search.

MySQL was used as repository for storing of metadata of song [20] and, Similarity Matrix values. When a user gives a weighted query, in the backend each category of query compares to the corresponding metadata of each song by using a Levenshtein similarity function [19]. After comparison each values is multiplied by the user adjusted priority value given through the interface. At the end accumulative similarity is calculated. Fig .7 explains the whole mechanism of the metadata filter.



Formula1 =SIM(song_artist,Term1) * Weight1
Formula2 =SIM(song_Album,Term2) * Weight2
Formula3 =SIM(song_Title,Term3) * Weight3
Formula4 =SIM(song_Genre,Term4) * Weight4
Formula5 =SIM(song_YearOfRelease,Term5) * Weight5
Formula6 =(Formula1+ Formula2+ Formula3 +Formula4+ Formula5)/5

Figure. 7. Mechanism of metadata filter

Usually user finds its relevant result in the top 40% of the total searched results. We shifted the indexes of total of 40% of the total searched results to the second filter of Query By

Humming. 40% of results was actually written in the text file as the input to the second filter.

For the second filter of Query by Humming we used open source project of CompariSong developed in Java [21]. CompariSong firstly converts audio file into seg-ments. 10 segments per second were extracted and on the basis of Fundamental Frequency and Audio Power, numerical time series was converted to letters. Same procedure of extracting and assigning was done on the hummed query as well. For finding the matching be-tween hummed query and pre-indexed file Levenshtein distance measure was used [19]. Relevant retrieved re-sults were presented in a sorted list, such that most relevant at the top.

System used for evaluation was HP ProBook 4520s, Core i5 480M, 4GB Ram, 1GB Graphic card and 500GB Hard Disk. All the user recordings were taken in an empty room almost a noise free room. Earlier discussed dataset was given to each user. They heard each song once and recorded the hum for all the songs.

VI. EVALUATION OF HQMS

One male and one female as a group were selected for each category. Information related to each user was

- Name
- Age
- Gender
- Mother Language
- Region
- Qualification

The main reason to choose a diverse dataset of songs of different languages and diverse user of different ages, regions, languages and genders was just to evaluate whether each category have some effects on the other category or not. We evaluate our system in a very extensive manner and also checked the impact. We made evaluation in the following categories.

1. User Related
   i. Gender Wise
   ii. Age Wise
   iii. Language Wise
2. Metadata Related
   i. String similarity 0% and records 100%
   ii. String similarity 20% and records 80%
   iii. String similarity 40% and records 60%
   iv. String similarity 60% and records 40%
   v. String similarity 80% and records 20%
   vi. Threshold string similarity 60% and records 40%
3. Song Related
   i. Song Wise Average Accuracy
   ii. User Wise Average Accuracy

We used Mean and Standard Deviation statistical meas-ure to find the average and difference between the results. For the evaluation of query by humming system we also introduced a new measure of accuracy percentage that is explained in equation .1.1. N is the number of total songs in repository where P is the position of relevant song in the list of results.

By using this formula we can find the accuracy of Query by humming system on the basis of relevant song posi-tion such

$$Accuracy\,\% = \frac{(N-P)+1}{N} \times 100 \qquad (1)$$

that after searching if the relevant song is at the top of the result list, it means accuracy is 100%.

A. Gender Wise

Females have high pitch as compared to male so initially, we divided whole dataset into two groups of male and female for gender based evaluation and we selected only those solo songs which were sung by either male or female. For each song initially, male group recorded the query then female group recorded the query. Averages of both groups were compared for each song.

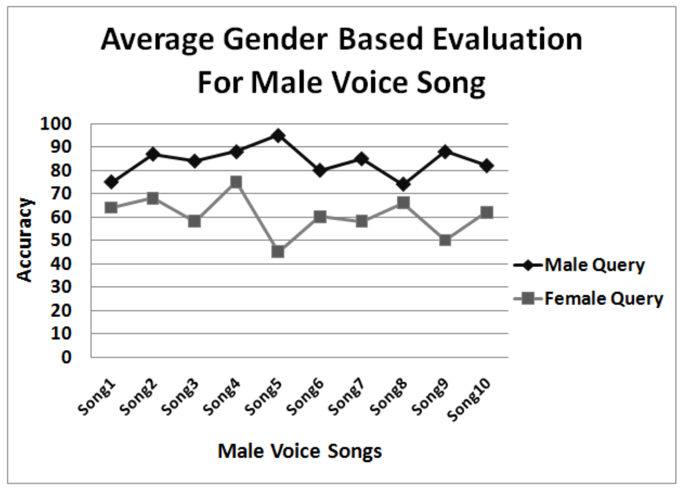


Figure. 8. Average gender based evaluation for male voice song

Fig. 8 shows the head to head comparison of male and female queries for each song. Likewise both set of groups were then given female sung songs and same procedure was adopted for comparison.

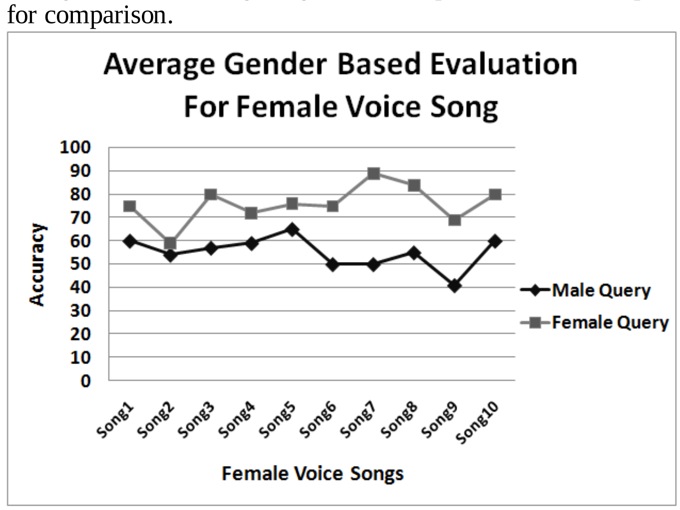


Figure. 9. Average gender based evaluation for female voice song

Fig. 8 shows Average Gender Based Evaluation of the male and female query for Male voice Songs and it was evaluated that male's queries more accurate results as compared to female's queries gave.

Respectively Fig. 9 shows evaluation of both groups for female sung song and it was examined that females can more accurately search female sung song

In graph (Fig. 10) y-axis show the accuracy in percentage and x-axis shows the songs. For each song two values are shown one the average female value and other male average accurate value.

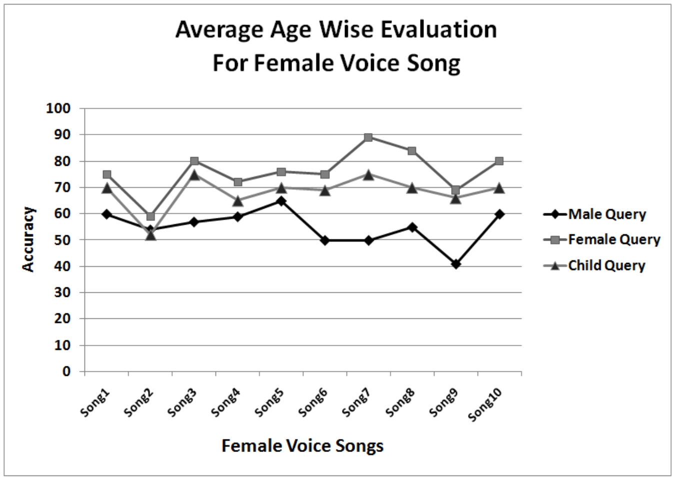


Figure. 10. Age group wise accuracy for female songs

B. Age Wise

As we described earlier in the paper that we divided our total population into three main groups age wise. The ranges of three groups are (11yr to 15yr), (16yr to 20yr) and (21yr to 25yr). The main reason to divide age wise is to evaluate whether accuracy differs with respect to age or not. We got some surprising results for the age of 11yr to 15 yr that user from this age got good accuracy for the female voice songs. The reason which we concluded was that commonly the pitch of small kids is high. High pitch is the common factor to female voice. Fig. 10. specifically shows the comparison of female, male and children group of user.

Fig. 13 show the comparison for each age groups accuracy. Average of each group is considered as a single value for every song in dataset. Three values for each song depict three average values of each age group.

C. Language Wise

Songs are recorded in different languages. Similarly song listeners also have different mother language and it is the common observation that single language users use to listen multilingual songs. By keeping these facts in mind we evaluated our system diversely. We took songs of six cultural languages of Pakistan and english language songs as dataset. Similarly we selected six different cultural language speaking users. We gave each set of users' full dataset, and directed them to listen and record there. After evaluation we concluded that if single language user sing's its own language song, the accuracy of matching is higher. Fig. 11 shows the overall average values of all language groups.

Language		Songs						
		Urdu	Pashto	Punjabi	Sindhi	Kashimiri	Balochi	English
Users	Urdu	81.2381	54.7619	68.7619	56.95238	60.28571	44.09524	63.42857
	Pashto	64.38095	79.61905	55.80952	52.09524	51.2381	68.19048	56.38095
	Punjabi	72.95238	49.61905	83.52381	59.42857	61.90476	39.2381	52.09524
	Sindhi	66.38095	54.28571	56.38095	80.85714	59.14286	56.19048	40.28571
	kashimiri	68.09524	61.14286	64.47619	46.47619	75.71429	39.52381	38.19048
	Balochi	63.42857	64.38095	44.66667	44.7619	47.42857	82.57143	45.61905

Figure. 11. Language to language comparison

In Fig. 11. the shaded diagonal values show maximum accuracy. The shaded diagonal values represent the similarity between same language songs and humming. This concludes that if single language users tries to hum or sing a song of its own language then maximum accuracy can be achieved. In this regard certain audio query reformulation or alignment techniques are to be performed so that maximum accuracy can be obtained.

This can be explain by a scenario if language "A" user tries to hum a song recorded in language "B" so in such case accuracy will be less. But if somehow language "A" audio query get transformed into language "B" accent then more relevant results can be obtained. During evaluation of HQMS we also analyzed if any two languages having similarity in regard of accent then query reformulation is not necessary. In our case two cultural languages of Pakistan Pashto and Balochi have similarity in accent as compare to other cultural languages of Pakistan. But audio query reformulation can play an

String similarity 0% and records 100%								
		Songs						
		Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
Users	Urdu	81.24	54.76	68.76	56.95	60.29	44.1	63.43
	Pashto	64.38	79.62	55.81	52.1	51.24	68.19	56.38
	Punjabi	72.95	49.62	83.52	59.43	61.9	39.24	52.1
	Sindhi	66.38	54.29	56.38	80.86	59.14	56.19	40.29
	Kashmiri	68.1	61.14	64.48	46.48	75.71	39.52	38.19
	Balochi	63.43	64.38	44.67	44.76	47.43	82.57	44.57

String similarity 20% and records 80%						
Songs						
Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
81.66	56.82	72.15	59.87	63.39	47.39	66.63
67.13	81.76	59.58	56.06	53.2	70.82	60.34
76.75	52.66	84.9	62.92	64.82	41.96	55.39
70.08	57.7	60.46	82.87	63.49	58.44	47.68
70.46	63.51	66.94	48.56	77.71	42.92	41.3
66.75	68.85	46.85	48.08	49.7	83.13	49.68

String similarity 40% and records 60%						
Songs						
Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
83.05	58.1	73.43	61.14	64.67	48.67	67.91
68.38	83.14	60.86	57.33	54.48	72.1	61.62
78	53.91	86.1	64.19	66.1	43.24	56.67
71.33	58.95	61.71	84.48	64.76	59.71	48.95
71.71	64.76	68.19	49.81	79.14	44.19	42.57
68	70.1	48.1	49.33	50.95	84.86	50.95

String similarity 60% and records 40%								
		Songs						
		Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
Users	Urdu	89.05	61.21	76.54	64.25	67.78	51.78	71.02
	Pashto	71.66	88.19	63.97	60.45	57.59	75.21	64.73
	Punjabi	81.27	57.18	90.1	67.3	69.21	46.35	59.78
	Sindhi	74.61	62.23	64.99	87.95	67.87	62.83	52.06
	Kashmiri	74.99	68.04	71.47	53.08	86.77	47.3	45.68
	Balochi	71.27	73.37	51.37	52.61	54.23	89.98	54.06

String similarity 80% and records 20%						
Songs						
Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
95.15	64.32	79.65	67.37	70.89	54.89	74.13
74.47	95.3	67.08	63.56	60.7	78.32	67.84
84.09	60	96.2	70.41	72.32	49.46	62.89
77.43	65.04	67.81	94.05	70.98	65.94	55.18
77.81	70.85	74.28	55.9	91.88	50.41	48.79
74.09	76.19	54.19	55.43	57.04	95.08	57.18

Threshold String similarity 60% and records 40%						
Songs						
Urdu	Pashto	Punjabi	Sindhi	Kashmiri	Balochi	English
89.05	61.21	76.54	64.25	67.78	51.78	71.02
71.66	88.19	63.97	60.45	57.59	75.21	64.73
81.27	57.18	90.1	67.3	69.21	46.35	59.78
74.61	62.23	64.99	87.95	67.87	62.83	52.06
74.99	68.04	71.47	53.08	86.77	47.3	45.68
71.27	73.37	51.37	52.61	54.23	89.98	54.06

Figure. 12. Metadata based evaluation

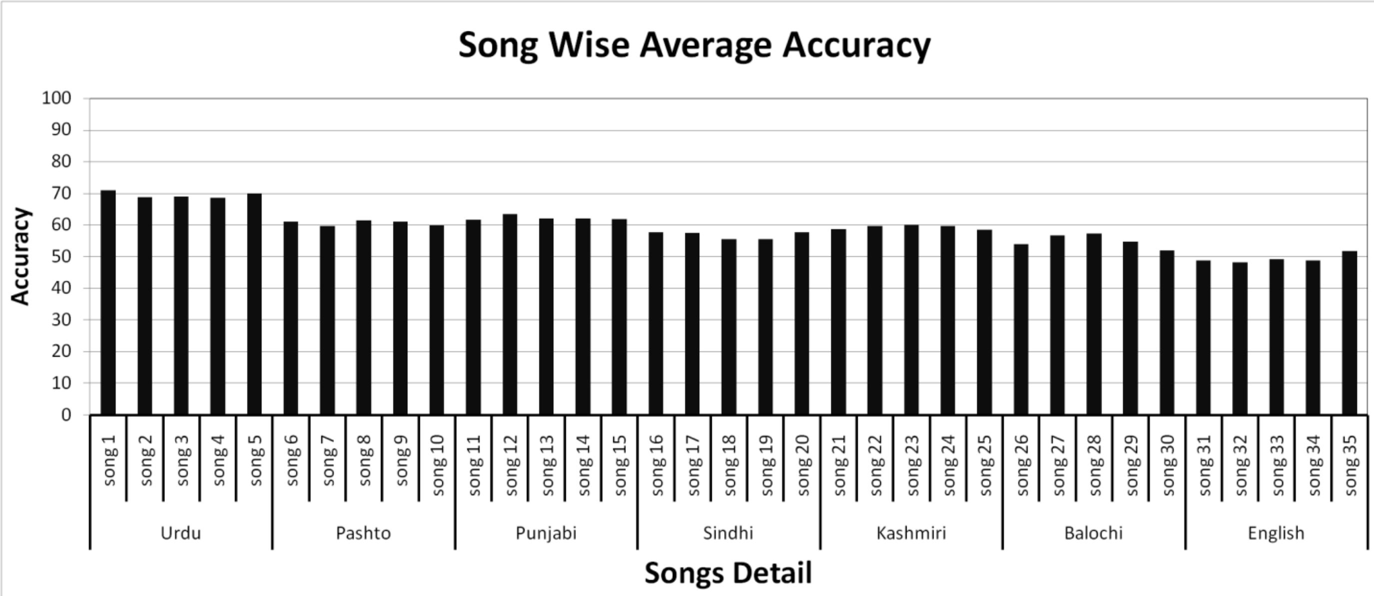


Figure. 13. Song wise average accuracy

effective role when two languages have different accents.

D. Metadata Evaluation Results

For the evaluation of metadata we tested values on different intervals and suggested threshold. We made 5 intervals for evaluation as we have limited dataset. Removal of biasness from results was the main objective for creating intervals. We choose 5 intervals on the basis of query similarity and filtration of records. Initially we gave dataset to all users to listen and then hum. If in this case we could have also asked them to give metadata of songs as well then there was a chance of biasness because by listening song one can understand lyrics and get closer to exact title. In HQMS case, two parameters are associated to metadata filter i.e. (similarity value between query and relevant song metadata) and (percentage of filtration required accordingly).

We distributed above mentioned parameter into 5 groups

- String similarity 0% and records 100%
- String similarity 20% and records 80%
- String similarity 40% and records 60%
- String similarity 60% and records 40%
- String similarity 80% and records 20%
- Threshold string similarity 60% and records 40%

Fig. 12. shows results for each category. We also suggested threshold i.e. String Similarity 60% and record 40%. Whenever a user is interested in searching a song then the query which he gives for searching is almost 60% similar to the relevant song. Secondly 40% of record shows the percentage of filtration that means we selected top 40% of total record after applying string similarity.

Fig. 12. shows 5 intervals values and 1 threshold value. As discussed earlier shaded diagonal values shows higher accuracy because of same language. The overall accuracy improves when string similarity value gets increased and filtration gets decreased. The ratio of similarity and filtration selection can be dynamic by taking explicit values from user.

E. Song Wise Average Accuracy

Songs are recorded with a variety of compositions and in different genres too. Some songs are easy to sing or hum but some songs are not memorized very quickly. To observe such variation we evaluate accuracy for each song and found that in user population group mostly user can hum more accurately for Urdu language songs. Two reasons which we concluded were, Pakistan National language is Urdu and in our population group majority of users were literate which mean they were quite familiar with Urdu language, and most of the users were having Urdu language as the medium of communication.

Fig. 13 shows the overall comparison of each song with respect to all users. Individual accuracy of song is measure just to check the song is easy to sing or hum and vice versa.

F. User Wise Average Accuracy

It is a common observation that songs are recorded in different languages and commonly one language user use to listen different language songs. To check this idea we took a small survey.

We distributed whole population into groups of language then each group was further classified into age group and gender as shown in Fig. 14. In this ``F'' represent Female and ``M'' represent Male.

According to the evaluation, age group between 20yr to 25yr has shown more accurate results as compared to rest of the group. The reason which we concluded was that users belonging to this age group are mature, and they can hum or sing more accurately as compared to other age groups.

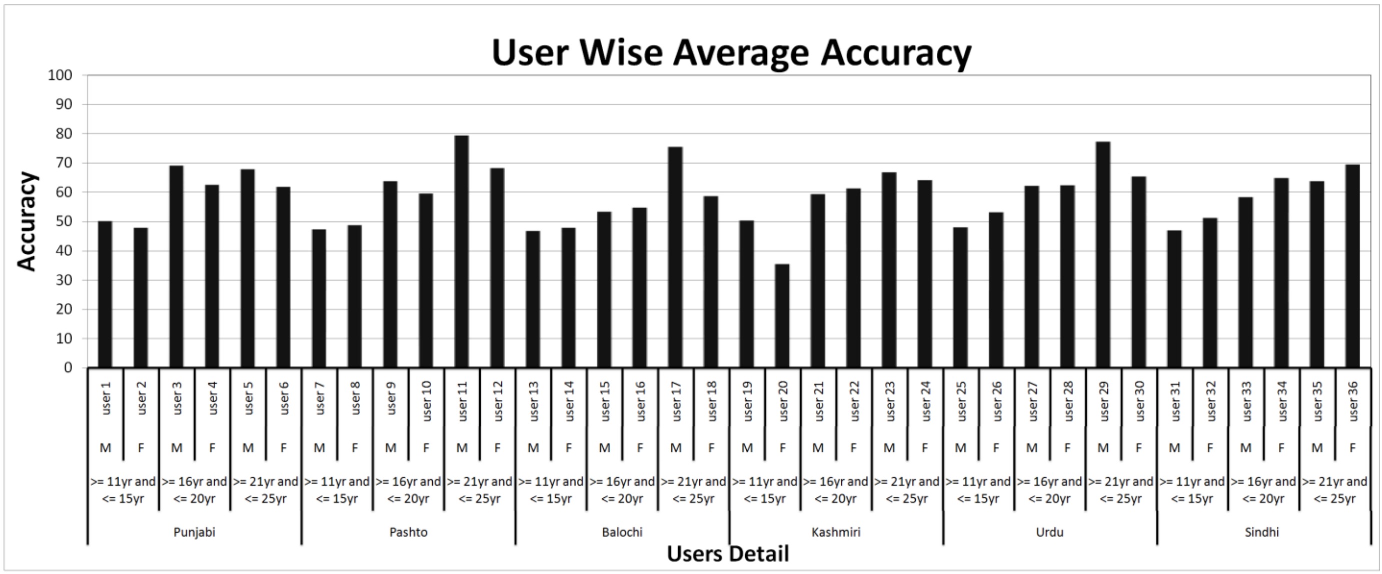


Figure. 14. User wise average accuracy

VII. Conclusion

Hybrid Query by Humming and Metadata Search System (HQMS) is the proposed model of search engine for songs [2]. This hybrid model works on the two parameters of songs i.e. metadata and content. Pipe and Filter software architecture style is used to express the model. In this paper, we implemented the proposed HQMS model and evaluated results by considering different parameters. HQMS includes two main filters of Metadata and Query by Humming. For evaluation we took sample of 36 users and 35 songs from different Cultural Languages of Pakistan. Stratified random sampling method was adopted. Evaluation of HQMS was done in three generic categories of User Related, Metadata Related and Songs Related. In future, we aim to enhance HQMS by considering different parameters and matching algorithms related to both metadata and contents of song.

Acknowledgment

The authors would like to appreciate the work and time given by G. Dzhambazov, developer of CompariSong, we are specially thankful to Danish Saif Talpur and Mudasser Ali Khan for their valuable time they served for reviewing the paper.

References

[1] G.J. Myers. The art of software testing. Wiley-India, 2008.

[2] N.A. Khan, M.A. Khan, and M. Mushtaq. Hybrid query by humming and metadata search system (HQMS). Proceedings of the 8th International Conference on Frontiers of Information Technology, page 9, 2010.

[3] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith. Query by humming: musical information retrieval in an audio database. In Proceedings of the third ACM international conference on Multimedia, pages 231–236. ACM, 1995.

[4] Y.H. Tseng. Content-based retrieval for music collections. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pages 176–182. ACM, 1999.

[5] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges. Proceedings of the IEEE, 96(4):668–696, 2008.

[6] N.C. Maddage, H. Li, and M.S. Kankanhalli. Music structure based vector space retrieval. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 67–74. ACM, 2006.

[7] N. Kosugi, Y. Nishihara, S. Kon'ya, M. Yamamuro, and K. Kushima. Music retrieval by humming-using similarity retrieval over high dimensional feature vector space. In Communications, Computers and Signal Processing, 1999 IEEE Pacific Rim Conference on, pages 404–407. IEEE, 1999.

[8] L. Chen and B.G. Hu. An implementation of web based query by humming system. In Multimedia and Expo, 2007 IEEE International Conference on, pages 1467–1470. IEEE, 2007.

[9] C. Francu and C.G. Nevill-Manning. Distance metrics and indexing strategies for a digital library of popular music. In Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, volume 2, pages 889–892. IEEE, 2000.

[10] L. Lu, H. You, and H.J. Zhang. A new approach to query by humming in music retrieval. In Proceedings of the IEEE International Conference on Multimedia and Expo. Citeseer, 2001.

[11] D. Little, D. Raffensperger, and B. Pardo. A query by humming system that learns from experience. 2007.

[12] D. Garlan and M. Shaw. An introduction to software architecture. (CMU-CS-94-166), January 1994

[13] D. Little, D. Raffensperger, and B. Pardo. Online training of a music search engine. 2007.

[14] midomi. midomi. URL: http://http://www.midomi.com/.

[15] soundhound. soundhound. URL: http://www.soundhound.com/.

[16] musipedia. musipedia. URL: http://www.musipedia.org/.

[17] MaART. Maart. URL: http://maart.sourceforge.net/.

[18] V. Lobaryev, G. Sokolov, and A. Gordeyev. Sloud query-by-humming search music engine. 2006.

[19] E.S. Ristad and P.N. Yianilos. Learning string-edit distance. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(5):522–532, 1998.

[20] AB MySQL. MySQL: the world's most popular open source database. MySQL AB, 2005.

[21] G. Dzhambazov. Comparisong: Audio comparison engine. International Book Series Number 11, page 33, 2009.

[22] K.-H. Kim, M. Lim, and J.-H. Kim. Music copyright protection system using fuzzy similarity measure for music phoneme segmentation. pages 159{164, 2009.

[23] Y. Song, Y. Wang, Y. Ju, M. Seltzer, I. Tashev, and A. Acero. Voice search of structured media data. pages 3941{3944, 2009.

AUTHORS PROFILE

Nauman Ali Khan is M.Phil candidate in the Faculty of Computer Science, Quaid-I-Azam University Islamabad, Pakistan. He received his BS(CS) honors Degree from N-W.F.P University of Engineering & Technology Peshawar. His research interests include Music Content retrieval, Query by Humming, and multimedia Searching.

Mubashar Mushtaq is an Assistant Professor of P2P multimedia Video Streaming at the Faculty of Computer Science, Quaid-I-Azam University Islamabad, Pakistan. Mubashar Mushtaq did PhD from CNRS LaBRI Lab—University of Bordeaux 1. He has received his MSc and MS degrees in Computer Science from Quaid-I-Azam University, Islamabad, Pakistan and University of Bordeaux 1, France in 2002 and 2004, respectively. His research interests include Multimedia Streaming over next generation networks, QoS, mobile IP-TV, and multimedia caching.

Semantic Retrieval Approach for Web Documents

Hany M. Harb
Computers and Systems Engineering Dept., Faculty of Eng., Al-Azhar Univ., Egypt.

Khaled M. Fouad
Computer Science Dep., Community College , Taif Univ., Kingdom of Saudi Arabia (KSA).

Nagdy M. Nagdy
Engineering Applications and Computer Systems,
Al-Baha Private College of Science,
Kingdom of Saudi Arabia (KSA).

***Abstract*— Because of explosive growth of resources in the internet, the information retrieval technology has become particularly important. However the current retrieval methods are essentially based on the full text matching of keywords approach lacking of semantic information and can't understand the user's query intent very well. These methods return a large number of irrelevant information, and are unable to meet the user's request. Systems have been established so far failed to overcome fully the limitations of search based on keywords. Such systems are built from variations of classic models that represent information by keywords. Using Semantic Web is a way to increase the precision of information retrieval systems.**

In this paper, we propose the semantic information retrieval approach to extract the information from the web documents in certain domain (jaundice diseases) by collecting the domain relevant documents using focused crawler based on domain ontology, and using similar semantic content that is matched with a given user's query. Semantic retrieval approach aims to discover semantically similar terms in documents and query terms using WordNet.

Keywords- Semantic Web; information retrieval; Semantic Retrieval; Semantic Similarity; WordNet.

I. INTRODUCTION

In the current information retrieval models and systems, only when query words appear in document, the document may be retrievable. As a result, it probably appears that the related documents may be omitted because of expression difference. This kind of problem is one of the important reasons that influences the accuracy of information retrieval. Traditional text-based information retrieval systems and search engines are mostly based on keywords matching and statistic techniques [57]. Although they are widely used nowadays, users usually suffer from the so called "too many or nothing" problem for various reasons. One common reason is that the users may not have complete domain knowledge and often can not specify appropriate and exact keywords for a valid query. The other reason is that the target documents are expressed in terms of plain-text format that is hard for the search engine to parse, thus it is difficult to understand the semantic of the documents during the retrieval process.

Aiming to solve the limitations of keyword-based models, the idea of semantic search, understood as searching by meanings rather than literal strings, has been the focus of a wide body of research in the information retrieval and the Semantic Web communities. However, these two fields have had a different understanding of the problem. The Semantic Web vision [54] was brought about the aim of helping automate tasks that require a certain level of conceptual understanding of the objects involved or the task itself, and enabling software programs to automatically find and combine information and resources in consistent ways.

At the core of these new technologies, ontologies were envisioned as key elements to represent knowledge that could be understood, used and shared among distributed applications and agents. Their potential to overcome the limitations of keyword-based search in the information retrieval context was soon envisaged, and was explored by several researchers in the Semantic Web area. A potential source of documents that is useful for information retrieval comes from the World Wide Web, it is important to develop document discovery mechanisms based on intelligent techniques such as focused crawling [55] to make this process easier for a certain domain.

In our work, we have used focused crawling to collect documents and information in a healthcare domain (jaundice diseases). Due to the huge number of retrieved documents, we require an automatic mechanism rather than domain experts in order to separate out the documents that are truly relevant to our domain of interest. The focused crawler in a domain specific search engine must crawl through the domain specific Web pages in the World Wide Web. For a crawler it is not an easy task to download the domain specific Web pages.

Ontology can play a vital role in this context. Our focus will be to identify Web pages for our domain in WWW. We present a critical semantic similarity approach for computing the semantic similarity between the terms using WordNet. We also propose the semantic retrieval approach to discover semantically similar terms in documents and query terms using WordNet by associating such terms using semantic similarity methods.

II. Related Works

In [1], authors introduced the technique of focused crawling of country based financial data. The focused crawlers yield good recall as well as good precision by restricting themselves to a limited and selected domain. The focused crawlers try to predict whether or not a target URL is pointing to a relevant and high-quality web page before actually fetching the page. Their efficient focused crawler is made for collecting the financial data for a specific country.

The approach [2] was proposed to calculate the link score. First authors calculated the unvisited URL score based on its Anchor text relevancy, its description in Google search engine and calculated the similarity score of description with topic keywords, cohesive text similarity with topic keywords and Relevancy score of its parent pages. Relevancy score is calculated based on vector space model.

In the paper [3], authors explored four kinds of semantic models and semantic information to improve focused crawling, including thesauruses, categories, ontologies, and folksonomies. Main contributions of this work are : First, A statistical semantic association model to integrate different semantic models and support semantic interoperability. Second , Include added semantic information to improve focused crawling, especially semantic markups in the Semantic Web and social annotations in Web 2.0. Third, the Semantic Association Model(SAM) that is based focused crawler which adopts heterogeneous semantic information to make predictions and decisions about relevant URLs and web pages.

In the study [4], authors analyzed four focused crawling for retrieving chemical information. These focused crawlers were formed by combining two feature representations (Latent Semantic Indexing (LSI) and Mutual Information (MI)) and two classification algorithms (Support Vector Machines (SVM) and Naive Bayes (NB)) from machine learning. The study shows that the four focused crawling can keep a high precision to collect chemistry relevant pages. It was also found that the combination of SVM and LSI provided the best performance in gathering web pages on the topic of chemistry.

Authors in [5] proposed an algorithm for crawling Web pages with limited resources. First, existing Web pages are divided into 100 clusters using both static and dynamic features. Static features are extracted from the content of a Web page, the Web page’s URL and hyperlinks. Dynamic features are extracted from changes to content, hyperlinks, page rank, and so on. The crawler fetches a sample of Web pages from a cluster to check if the pages have changed since their last download. If a significant number of pages in a cluster have changed, the other Web pages in the cluster are also downloaded. Clusters of Web pages have different change frequencies. Based on their change histories, different clusters is crawled at different frequencies. They demonstrated the superiority of their algorithm over various existing sampling-based Web page update detection algorithms.

In the paper [6] authors introduced document vector compression, which significantly reduces the size of the vectors and increases the total F-measure (cluster quality). Document vector compression involves the use of the Discrete Cosine Transform(DCT) on the document vectors to obtain a spectral representation of the document vectors. Due to the energy compaction property of the DCT, the majority of the energy is concentrated at the low frequency subbands. The high frequency subbands can be deleted without significantly degrading the content of the original vector. Any standard clustering algorithm, such as K-means , was used to cluster the compressed document vectors.

In this approach [7], vector space is taken as an example to describe the construction process of the document representation model based on query and content information. The basic idea is as following: At the initial stage of information retrieval, the traditional vector space model is adopted to represent documents. Then, the information of the query space can be introduced into the document representation model gradually, thus the document-representing vector space becomes the integration of query space and document space. This model can improve the fitness, reliability and accuracy of the feature terms of documents.

The approach [8] is for representing text data. The method translated the text clustering problem into query processing. The intuition behind this approach is if a set of documents belongs to the same cluster, authors expected that they respond similarly to the same queries, which can be any combination of terms from the vocabulary. While in information retrieval, the target is to retrieve relevant document(s) to a query, in text clustering, the goal is finding relevant queries which generates high quality clusters (lowest inter-cluster and highest intra-cluster similarities). In this paper, authors proposed approach to generate relevant and non-redundant queries from the domain taxonomy which is extracted from document collection. Using this new model, the terms in BOW model are transformed to the similarity scores of Bag-Of-Queries (BOQ) model. The effectiveness of the proposed approach is evaluated by extensive numerical experiments using benchmark document data set.

It was proposed [9] to use WordNet for document expansion, proposing a new method: given a full document, a random walk algorithm over the WordNet graph ranks concepts closely related to the words in the document. This is in contrast to previous WordNet-based work which focused on WSD to replace or supplement words with their senses. The method discovered important concepts, even if they are not explicitly mentioned in the document.

The goal in the work [10] was to study the use of the WordNet expansion technique over a collection with minimal textual information. The integration of knowledge through the use of ontologies has been very successful in many systems. Specifically, WordNet has been used with success in many works related to information retrieval, image retrieval, disambiguation and text categorization.

In the study [11], authors developed a new information retrieval system integrating Semantic Web with Multi-agent that handles the processing, recognition, extraction, extensions and matching of content semantics to achieve the following objectives: (1) Using Resource Description Framework (RDF) to analyze and determine the semantic features of users’ queries, to present a new algorithm to extract semantics in the content and build up semantic database; (2) to present a new matching algorithm using semantics extracted from content which can feedback useful and accurate information meeting users' requirements; (3) a new Information Retrieval based on Multi-agent is put forward, the Agents in this model can adapt users' own interests and hobbies, collect information based on users' behavior, dig up semantics in internet and feedback and share information between different users, so the search results will be more in line with users' needs and help users to complete complex tasks.

Authors in [12] introduced ontology into query expansion and makes good use of semantic relations of concepts in ontology to expand query keywords and to make the retrieval results more accuracy and comprehensive. Experimental results showed that this method can improve the precision and recall ratios of information retrieval.

The method proposed in [13] focused on semantic based expansion. There are three important improvements in the query expansion. First of all, this method categorizes the query terms based on their semantic similarities, and expands each category on words which show the relationship between words in the same group, as a result in this method selected words are not related to only an individual query term. Therefore it avoids outweighting problem in query expansion. Secondly, it avoids selecting vague and noise words to expand the query. Thus it avoids making the query noisy. Thirdly, it uses spreading activation algorithm to select candidate expansion words. Using spreading activation algorithm eases the selection of appropriate depth for hierarchical relations.

Authors in [14] proposed a new semantic similarity based model (SSBM) and they used this model in document text clustering. The model analyzed a document to get the semantic content. The SSBM assigns new weights to reflect the semantic similarities between terms. Higher weights are assigned to terms that are semantically close. In this model, each document was analyzed to extract terms considering stemming and pruning issues. They used the adapted Lesk algorithm to get the semantic relatedness for each pair of terms. SSBM solved the ambiguity and synonym problems that may lead to erroneous grouping and unnoticed similarities between text documents.

III. THE PROPOSED FRAMEWORK

Ontologies play an important role in providing a controlled vocabulary of concepts, each with an explicitly defined and machine understandable semantics. They are largely used in the next generation of the Semantic Web which focuses on supporting a better cooperation between humans and machines.

Due to the tremendous size of information on the Web, it is increasingly difficult to search for useful information for certain domain. For this reason, it is important to develop document discovery mechanisms based on intelligent techniques such as focused crawling. In a classical sense, crawling is one of the basic techniques for building data storages. Focused crawling goes a step further than the classical approach. It was proposed to selectively seek out pages relevant to a predefined set of topics called crawling topics.

In order to leave a lot of irrelevant noisy pages out, we propose an ontology-based focused crawling framework for Web. Crawling topics are based on our domain ontology. We focus on building an effective web-based documents discovery crawler that can autonomously discover and download pages from the web relevant to our domain, that is jaundice diseases. This is considered as semantic-based focused crawling, it makes use of an ontology to improve decision accuracy. The figure 1 shows the architecture of the proposed system.

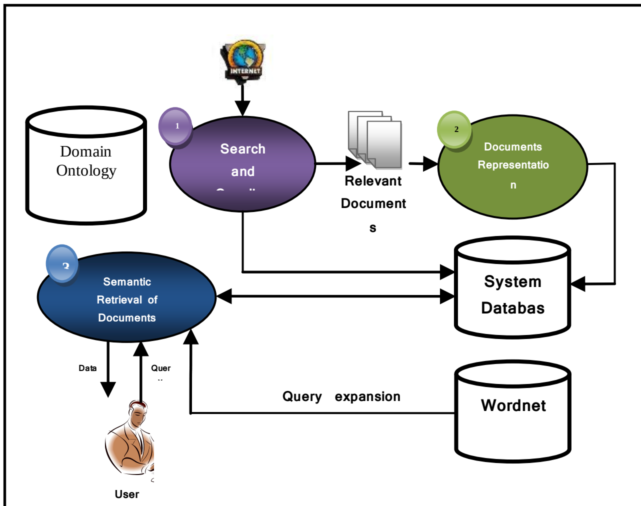


Figure 1 The architecture of the proposed system

A. The Domain Ontology

The term "Ontology" [21] is becoming frequently used in many contexts of database and artificial intelligence researches. However, there is not a unique definition of what an ontology is [22,23]. An initial definition was given by Tom Gruber: "an ontology is an explicit specification of a conceptualization" [22]. However, this definition is general and remains still unsatisfied for many researchers. In [24] Nicola Guarino argues that the notion of "conceptualization" is badly used in the definition. We note that many real-world ontologies already combine data instances and concepts [25]. The definition in [21] differs from this point of view as we show later . Informally, an ontology is defined as an intentional description of what is known about the essence of the entities in a particular domain of interest using abstractions, also called concepts and the relationships among them.

Ontologies [26] are designed for being used in applications that need to process the content of information, as well as, to reason about it, instead of just presenting information to humans. They permit greater machine interpretability of content than that supported by XML, and OWL, by providing additional vocabulary along with a formal semantics. Because of the intrinsic complexity of the concepts involved, the medical domain is one of the most active ones in defining and using ontologies.

The ontology in our system is focused in the medical domain that is "Jaundice diseases". Jaundice [27], is a yellowing of the skin, conjunctiva (clear covering over the sclera, or whites of the eyes) and mucous membranes caused by increased levels of bilirubin in the human body. When red blood cells die, the heme in their hemoglobin is converted to bilirubin in the spleen and in the hepatocytes in the liver. The bilirubin is processed by the liver, enters bile and is eventually excreted through feces. Consequently, there are three different classes of causes for jaundice. Pre-hepatic or hemolytic causes, where too many red blood cells are broken down, hepatic causes where the processing of bilirubin in the liver does not function correctly, and post-hepatic or extrahepatic causes, where the removal of bile is disturbed. Figure 2 shows part of our ontology for "Jaundice diseases". Jaundice diseases [28] are divided into three types as shown, as it follows: Pre-hepatic jaundice is caused by anything which causes an increased rate of hemolysis (breakdown of red blood cells). Hepatic (in hepatocellular jaundice there is invariably cholestasis) jaundice causes include acute hepatitis, hepatotoxicity, Gilbert's syndrome. Post-hepatic jaundice, also called obstructive jaundice, is caused by an interruption to the drainage of bile in the biliary system. The most common causes are gallstones in the common bile duct, and pancreatic cancer in the head of the pancreas.

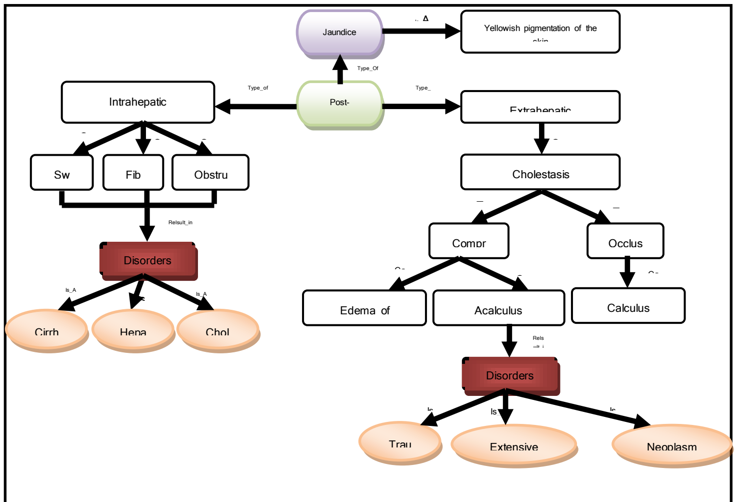


Figure 2: Part of the domain ontology

Our domain ontology is represented in Web Ontology Language (OWL) [29]. The Web Ontology Language (OWL) [30] describes classes, properties, and relations among these conceptual objects in a way that facilitates machine interpretability of Web content. OWL is the result of the Web Ontology Working Group (now closed) and descends from DAML+Oil, which is in turn an amalgamation of DAML and OIL.

B. *Search and Crawling Web*

A web crawler [15] is a program that collects web content from the World Wide Web automatically and stores this content into the storage. It starts with a list of URLs (seeds) to be visited. When it visits these pages, it parses all links from these pages. After collecting all these links, the web crawler inserts them in the URL queue. Web crawler continuously visits the unseen links and also scans them for discovering more links and put only the unseen links in the URL queue. The basic steps in the crawler [15] are:

1. Fetch a page
2. Parse it to extract all linked URLs
3. For all the URLs not seen before, repeat (1)–(3).

One of the features that characterizes a focused crawler [16] is the way it exploits hyper-textual information. Traditional crawlers convert a web page into plain text extracting the contained links, which will be used to crawl other pages. Focused crawlers exploit additional information from Web pages, such as anchors or text surrounding the links. Semantic focused crawlers [17, 18, 19] are considered as a subset of focused crawlers enhanced by various semantic web technologies. Table 1 shows different category of semantic focused crawlers and its definitions.

TABLE I. DIFFERENT CATEGORY OF SEMANTIC FOCUSED CRAWLERS AND ITS DEFINITION

Crawler category	**Definition**
Ontology-based focused crawlers	The focused crawlers that utilize ontologies to link a crawled web document with the ontological concepts (topics), with the purpose of organizing and categorizing web documents, or filtering irrelevant web pages with regards to the topics.
Metadata abstraction focused crawlers	The focused crawlers that can abstract and annotate metadata from the fetched web documents, in addition to fetching relevant documents.
Other semantic focused crawlers	The focused crawlers that employ other semantic web technologies than ontology-based filtering and metadata abstraction.

1) *Search and Crawling Web approach*

In our approach, we crawl through the Web and add Web pages contents to the database, which are related to a specific domain and discard Web pages which are not related to the domain [20]. The block diagram of the search and web crewing is shown in figure 3.

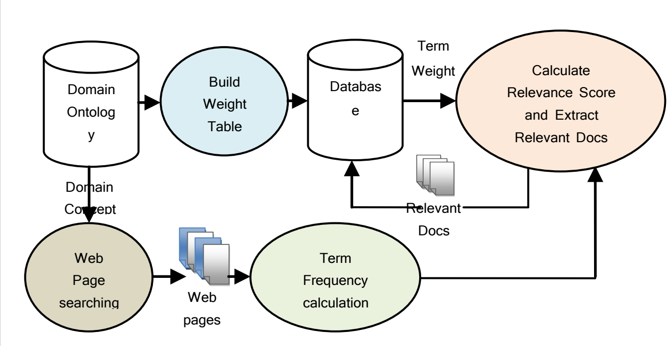


Figure 3: block diagram of the search and web crewing

2) *Building the weight table*

In order to build the weight table, we must determine some weights to each term in our ontology The strategy of assigning weights is that, the more specific term will have more weight on it. The terms which are common to more than one domain have less weight. The sample weight table for some terms of a given ontology is shown in table 2. The weight table is built with the assistance of the knowledge experts.

3) Relevance Score Calculation

Although documents are retrieved selectively through restricted queries and by focused crawling, we still need a mechanism to evaluate and verify the relevance of these documents to the predefined domain of Jaundice domain. To remove unexpected documents, first we automatically remove those that are blank, too short, duplicated documents, or those that are in a format that is not suitable for text processing. We then perform the relevance calculation to extract the relevant documents and discard the irrelevant document to our domain.

In the relevance calculation, the relevancy of a Web page to a specific domain is calculated. Relevance calculation algorithm [20], which calculates the relevance score of a Web page, is shown in figure 4.

TABLE II. WEIGHT TABLE FOR THE PART OF ONTOLOGY

Concepts	Weight
Jaundice	1
Biliruin	1
Post-hepatic	0.9
Prehepatic	0.9
Hepatic	0.9
Hepatitis	0.8
Cholestasis	0.8
Thalassemia	0.7
Damaged Hepatocytes	0.6
Tumours	0.4
Trauma	0.2
Compression	0.1
Disorder	0.1

Input: A Web page (P), a weight table.
Output: The relevance score of the Web page (P).
Procedure:
Get a Web page (P).
Read Ontology (O) (C: Concept in the ontology, and C_c: Concept count in ontology).
Do While $j <= C_c$ //Calculate and save weight table
Open Weight Table (WT) of the domain concepts that contains concepts and its weights.
Calculate Concept Weight (CW_j).
Save CW_j in DB table.
EndDo
Read WT.
Get (T_C : Concept Count in weight table).
Do While $i <= T_C$ //Calculate Relevance Score for the page
Reset Relevance Score (RS_P) of the Web page (P) (RS_P=0).
Get (T_i) and its (W_i).
Calculate (TF_P :Term Frequency in P).
Calculate $TW_P = TF_P \times W_i$. (TW_P: Term Weight in P).
Calculate $RS_P = RS_P + TW_P$.
Output RS_P (for the Web page).
Save RS_P and its Content in DB table.
EndDo

Figure 4 : Algorithm of calculation of relevance score for the Web pages

In our approach we go along the links that are found in domain specific pages to crawl the result web page. The figure 5 shows the crawling algorithm that is based on the relevance of domain ontology.

Input: Crawling Level (C_L), Start URL list (U_{List}) from search engine, Relevance threshold (R_T).
Output: The Relevant Web pages (P).
Procedure:
Do While $i <= CU_{List}$ // CU_{List} : the count of URL in the U_{List}.
Get URL.
Do While $j <= C_L$ //Crawling until the Crawling Level
Do Crawling Function.
Get Web Page.
Calculate relevance score of the Web page RS_P.
If ($RS_P > RT$) // if the page is relevant to the domain ontology and take high relevance score.
Get the Page content.
Save the content in the database.
Else // if the page is not relevant to the domain ontology and take high relevance score.
Discard the Page
Endif
EndDo
EndDo

Figure 5 : Algorithm of crawler based on domain relevance

C. Documents Representation

In information retrieval, the most widely accepted document representation model in text classification is probably vector space model [31]. The Vector Space Model is adapted in our proposed system to achieve effective representations of documents. Each document is identified by n-dimensional feature vector where each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight. The weight is a function of the *term frequency*, *collection frequency* and *normalization factors*. Different weighting approaches may be applied by varying this function. Hence, a document j is represented by the document vector d_j:

$d_j = (w_{1j}, w_{2j}, \ldots, w_{nj})$ Where, w_{kj} is the weight of the k_{th} term in the document j. Figure 5 shows the approach for representation and classification of the output documents of web pages.

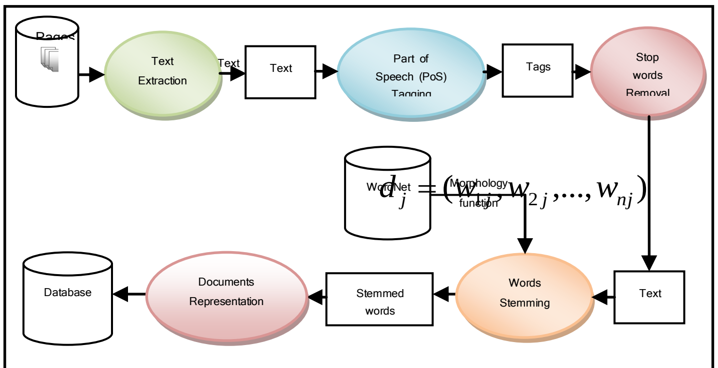


Figure 5 : Approach of documents

The first step of the documents classification process is to extract textual data from the web pages. Then convert each page into individual text document to apply text preprocessing techniques on it. This step is applied on input Web documents dataset by scanning the web pages and categorizing the HTML tags in each page. Then exclude the tags that contain no textual information like formatting tags and imaging tags (i.e. <HTML>, <BODY>, <IMG>, etc.).

Also exclude all the scripts and codes that are found in the page like JavaScript and VBScript. Then extract the textual data from other tags (like paragraphs, hyperlinks, and metadata tags) and store it into individual text documents as input for next steps. To extract the text from Web documents, we used the open source high-performance .NET C# module that was created to parse HTML [32] for links, indexing and other purposes.

2) Part of Speech (PoS) Tagging

The PoS tagger [33] relies on the text structure and morphological differences to determine the appropriate part-of-speech. This requires the words to be in their original order. This process is to be done before any other modifications on the corpora. For this reason, if it is required, PoS tagging is the first step to be carried out. After this, stop word removal is performed, followed by stemming. This order is chosen to reduce the amount of words to be stemmed. We used Stanford POS Tagger to tag the tokens [34].

The output of this step is the tags of each tokens. For example: "*Jaundice is yellowish pigmentation of the skin*", when this sentence is passed to the POS Tagger. The output is the tagged text as follows: "*Jaundice/NNP is/VBZ yellowish/JJ pigmentation/NN of/IN the/DT skin/NN*". The table 3 shows the different symbols that is found in PoS tagging.

TABLE III. SYMBOLS OF POS TAGGING

CC	Coordinating conjunction	**IN**	Preposition or subordinating conjunction	**MD**	Modal
CD	Cardinal number	**JJ**	Adjective	**NN**	Noun, singular or mass
DT	Determiner	**JJR**	Adjective, comparative	**NP**	Proper noun singular
EX	Existential there	**JJS**	Adjective, superlative	**NPS**	Proper noun plural
FW	Foreign word	**LS**	List item marker	**PDT**	Predeterminer
POS	Possessive ending	**RBR**	Adverb, comparative	**TO**	To
PP	Personal pronoun	**RBS**	Adverb, superlative	**UH**	Interjection
PP$	Possessive pronoun	**RP**	Particle	**VB**	Verb, base form
RB	Adverb	**SYM**	Symbol	**VBD**	Verb, past tense
VBG	Verb, gerund or present participle	**VBP**	Verb, noun-3rd person singular present	**WDT**	Wh-determiner
VBN	Verb, past participle	**VBZ**	Verb, 3rd person singular present	**WP**	Wh-pronoun
WP$	Possessive wh-pronoun	**WRB**	Wh-adverb		

3) Stop Words Removal

Stop words, i.e. words thought not to convey any meaning, are removed from the text. In this work, the proposed approach uses a static list of stop words with PoS information about all tokens. This process removes all words that are not nouns, verbs or adjectives. For example, stop words removal process will remove all the words like: he, all, his, from, is, an, of, your, and so on. Removing these words will save spaces for storing document contents and reduce time taken during the search process.

4) Words Stemming

The stem is the common root-form of the words with the same meaning appear in various morphological forms (e.g. player, played, plays from stem play). In the proposed approach, we used the morphology function [35] provided with WordNet [36, 37] that is used for stemming process. Stemming will find the stems of the output terms to enhance term frequency counting process because terms like “diseases” and “diagnosing” come down from the same stem “disease” and "diagnose". This process will output all the stems of extracted terms [38, 39].

5) Documents Representation

As part of the key vocabulary extraction process from documents, tf × idf (term frequency times inverted document frequency), takes place. Terms (T_k) in the documents is represented as the document-term frequency matrix ($D_j \times TF_{jk}$) as shown in figure 6.

D_j/T_k	T_1	T_2	....	T_k
D_1	TF_{11}	TF_{12}	....	TF_{1m}
D_2	TF_{21}	TF_{22}	....	TF_{2m}
D_3	TF_{31}	TF_{32}	....	TF_{3m}

Figure 6 : The document-term frequency matrix

D_j is referring to each document that exists in the system database where j=1,….,n. Term frequency TF_{jk} is the number of how many times the distinct term T_k occurs in document D_j where k=1,…,m.

The calculation of the terms weight W_{jk} of each term T_k is done by [40, 41, 42]:

$$W_{jk} = TF_{jk} \times idf_k \qquad (1)$$

where the document frequency df_k is the total number of documents in the database that contains the term T_k. The inverse document frequency is :

$$idf_k = \log_2 n - \log_2 df_k + 1 \qquad (2)$$

where (n) is the total number of documents in the database.

$tf \times idf$ is a mathematical algorithm [40, 41, 42], which is used to efficiently find key vocabulary that best represents the texts by applying the term frequency and the inverted document frequency together. $tf(\mathrm{T_k}, \mathrm{D_j})$ is the term frequency of term $\mathrm{T_k}$ that appears in Document $\mathrm{D_j}$, and (n) is the total number of documents of the corpus. $df(\mathrm{T_k})$ is the number of the documents in which the Term $\mathrm{T_k}$ appears at least once and represents how often Term $\mathrm{T_k}$ appears in other documents. $tf \times idf$ for Term $\mathrm{T_k}$ is defined as:

$$tf \times idf(\mathrm{T_k}, \mathrm{D_j}) = tf(\mathrm{T_k}, \mathrm{D_j}) \times \log(\frac{n}{df(\mathrm{T_k})}) \qquad \text{.........(3)}$$

For vocabulary with a low or rare appearance frequency, the value of $tf \times idf$ is low, compared to that with a high appearance frequency, thus resulting words successfully classifying the documents. In the term selection process, a list of all terms contained in one document from the text collection is made. Then, the document selection process chooses term $\mathrm{T_k}$ that maximizes *W(k),* which is expressed as a vector for document D_j as follows.

Document D_j includes $tf \times idf(\mathrm{T_k}, \mathrm{D_j})$, which is $tf \times idf$ for the most appropriate term.

$$W(k) = \sum_{j=1}^{n} tf \times idf(\mathrm{T_k}, \mathrm{D_j}) \qquad \text{.......................(4)}$$

D. Semantic Retrieval of Documents

Semantic retrieval [43] plays an increasingly important role in information retrieval. It overthrew the shackles of traditional idea of information retrieval. Semantic matched on information considerably improves the information recall and precision ratio. Given a query, if we can get enough semantic knowledge, acquire semantic similarity of the known query and optional data, then will get a result set which is sorted according to semantic similarity.

Nowadays semantic retrieval mainly implements concept retrieval [44, 45] by interaction terms, which does not take the concept's attributes and other information in to consideration. This semantic retrieval method based on concepts often cannot meet practical requirements.

So, we organize concepts with ontology, calculating the semantic similarity between concepts, whose basis is that there are some semantic correlations between two concepts. There are several semantic similarity methods were used which have certain limitations despite the advantages. No one method replaces all the semantic similarity methods. When a new information retrieval system is going to be build, several questions arises related to the semantic similarity matching function to be used. In [46] authors discussed the survey of different similarity measuring methods used to compare and find very similar concepts of an ontology.

In our approach, we depend on the semantic similarity based on Wordnet [47]. Five commonly used semantic similarity measures based on WordNet are discussed in [48, 49]. In [50] the authors conducted a comparative study on how different term semantic similarity measures including path-based, information content-based and feature-based similarity measure affect document clustering. WordNet is a controlled vocabulary and thesaurus offering a taxonomic hierarchy of natural language terms developed at Princeton University [56].

Figure 7 shows the block diagram of information retreieval of documents using semantic similarity between query and documents data.

Query expansion refers to the process of adding new necessary terms to a user's initial query. The purpose is to improve retrieval performance Query expansion reformulates the original query that enables users' desired information to be retrieved.

The major process of query expansion is the modification of the original query with new relevant and meaningful terms. With query expansion [51], the user is guided to formulate queries which enable useful results to be obtained. The main aim of query expansion [52, 53] (also known as query augmentation) is to add new meaningful terms to the initial query. Our approach uses query expansion that computes good weights for the new terms introduced into the query by using semantic similarity based on wordnet. Queries is first syntactically analyzed and reduced into term vectors as performed in documents in section 3.3.5. Very infrequent or very frequent terms are eliminated. Each term in this vector is represented by its weight.

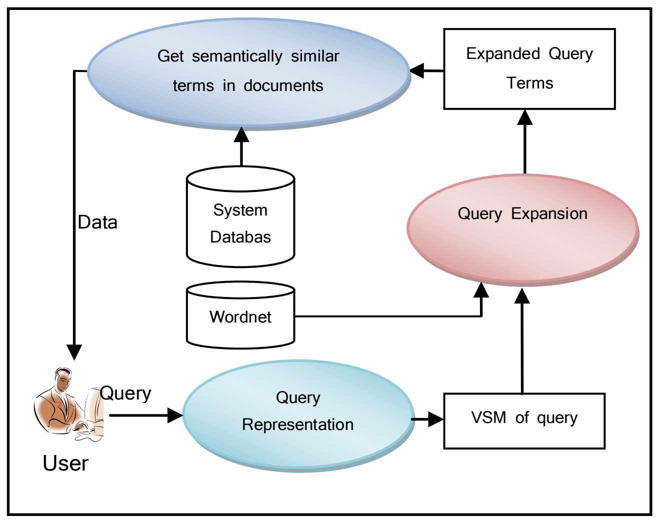


Figure 7 Semantic Retrieval of Documents

The query is augmented by synonym terms in wordnet, using the most common sense of each query term. Then, the query is augmented by terms higher or lower in the tree hierarchy in XML file which are semantically similar to terms already in the query.

The neighborhood of the term is examined and all terms with similarity greater than threshold t are also included in the query vector. This expansion may include terms more than one level higher or lower than the original term.

Each query term i is assigned a weight as in formula 5 in which the summation is taken over all terms j that is introducing terms to the query. It is possible for a term to introduce terms that already existed in the query. It is also possible that the same term is introduced by more than one other terms. After expansion and reweighting, the query vector is normalized by document length, like each document vector.

$$q_i' = \begin{cases} q_i + \sum_{sim(i,j)\geq t}^{j\neq i} \frac{1}{n}\, q_j\ sim(i,j)\,, & i \text{ had weight } q_i \\ \sum_{sim(i,j)\geq t}^{j\neq i} \frac{1}{n}\, q_j\ sim(i,j)\,, & i \text{ is a new term} \end{cases} \quad \text{..............(5)}$$

where t is a user defined threshold, n is the number of hyponyms of each expanded term j (for hypernyms n will be equal 1). The algorithm of the semantic retrieval of documents is shown in figure 8.

IV. Conclusion

We have proposed the approach of semantic retrieving for web documents in certain domain, extracting relevant information based on the semantic web. We have studied and implemented a focused crawler enabling us to retrieve web documents in the domain of jaundice diseases from the Web.

Semantic information retrieval method have exploited the advantages of the semantic web to retrieve the relevant data. It outperforms VSM, the classic information retrieval method and demonstrates promising performance improvements over other semantic information retrieval methods in retrieval.

Input: Query Vector $q = (q_1, q_2, .., q_i)$, Document Vector $d = (d_1, d_2, \ldots, d_K)$, Thresholds T.
Output: Document similarity value *Sim(d, q)*.
Procedure:

// --- Get the semantic similarity of each term with another in the query of the same vector.
Do While $i <= q_{count}$ // q_{count} : the count of terms in the query vector q.
 Get q_i.
 Compute $q_i' = q_i + \sum_{sim(i,j)\geq t}^{j\neq i} q_j\ sim(i,j)$
EndDo

// --- Get Expanded terms based on wordnet.
Do While $i <= q_{count}$
 Get q_i.
 Open Wordnet.
 If *sim(i,j) >= T* // *sim(i,j) : similarity between the term i in the query and term j in the wordnet.*
 Retrieve the term *j* from wordnet.
 Add term *j* to the query vector *q*.
 Endif
EndDo

// --- Re-weighting the terms in the query.
Do While $i <= q_{count}$

$$q_i' = \begin{cases} q_i + \sum_{sim(i,j)\geq t}^{j\neq i} \frac{1}{n}\, q_j\ sim(i,j)\,, & i \text{ had weight } q_i \\ \sum_{sim(i,j)\geq t}^{j\neq i} \frac{1}{n}\, q_j\ sim(i,j)\,, & i \text{ is a new term} \end{cases}$$

EndDo

// --- Compute the document similarity (The similarity between an expanded and re-weighted query *q* and a document *d*).
Do While $i <= q_{count}$

$$sim(q,d) = \frac{\sum_i \sum_j q_i q_j\ sim(i,j)}{\sum_i \sum_j q_i q_j}$$

EndDo

Figure 8 : Algorithm of the semantic retrieval of documents

References

[1] D. Manas, C. Hasan, S. Debakar, A. Khandakar. (2010). Focused Web Crawling: A Framework for Crawling of Country Based Financial Data. 978-1-4244-6928-4/10, IEEE.

[2] H. Debashis, S. Biswajit, K. Amritesh. (2010). Adaptive Focused Crawling Based on Link Analysis. 201O 2nd International Conference on Education Technology and Computer (ICETC). 978-1-4244-6370-1, IEEE.

[3] H. Rui, L. Fen, S. Zhongzhi. (2008). Focused Crawling with Heterogeneous Semantic Information. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 978-0-7695-3496-1/08, IEEE.

[4] X. Zhaojie, G. Li, L. Chunyang, L. Xiaoxia, and Y. Zhangyuan. (2007). Focused Crawling for Retrieving Chemical Information. Innovations in Hybrid Intelligent Systems, ASC 44, pp. 433–438, Springer-Verlag Berlin Heidelberg.

[5] T. QINGZHAO and M. PRASENJIT. (2010). Clustering-Based Incremental Web Crawling. 2010 ACM 1046-8188/2010/11-ART17.

[6] T.W. Fox. (2005). Document Vector Compression and Its Application in Document Clustering. 0-7803-8886-0/05, IEEE.

[7] Z. ZHOU, H. JIANG, J. MA, X. YANG. (2010). Study on Application of Document Representation Model Based on Query and Content Information in Website Search Engine. 2010 International Conference on Web Information Systems and Mining, 978-0-7695-4224-9/10, IEEE.

[8] M. Masoud. (2010). Query-Relevant Document Representation for Text Clustering. 978-1-4244-7571-1/10, IEEE.

[9] A. Eneko, A. Xabier, O. Arantxa. (2010). Document Expansion Based on WordNet for Robust IR. COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics, Volume, pages 9–17, Beijing. ACM.

[10] D. Manuel, M. Maria, U. L. Alfonso, and P. Jose. (2010). Using WordNet in Multimedia Information Retrieval. CLEF 2009 Workshop, Part II, LNCS 6242, pp. 185–188, Springer-Verlag Berlin Heidelberg.

[11] J. LUO, X. XUE. (2010). Research on Information Retrieval System Based on Semantic Web and Multi-Agent. 2010 International Conference on Intelligent Computing and Cognitive Informatics. 978-0-7695-4014-6/10, IEEE.

[12] W. Hongsheng, Q. Jiuying, S. Hong. (2009). Expansion Model of Semantic Query Based on Ontology. Web Mining and Web-based Application. WMWA '09. IEEE.

[13] M. Shabanzadeh, M.A. Nematbakhsh, and N. Nematbakhsh. (2010). International Conference on Intelligent Control and Information Processing August 13-15, 2010 - Dalian, China. 978-1-4244-7050-1/10,IEEE.

[14] G. Walaa and K. Mohamed. (2009). Enhancing Text Clustering Performance Using Semantic Similarity. LNBIP 24, pp. 325–335. Springer-Verlag Berlin Heidelberg.

[15] [D. Manas, C. Hasan, S. Debakar, A. Khandakar. (2010). Focused Web Crawling: A Framework for Crawling of Country Based Financial Data, 978-1-4244-6928-4/10, IEEE.

[16] M. Alessandro and G. Fabio. (2007.). Adaptive Focused Crawling. The Adaptive Web, LNCS 4321, pp. 231–262, 2007, Springer-Verlag Berlin Heidelberg .

[17] D. Hai, K. Farookh, and C. Elizabeth. (2009). State of the Art in Semantic Focused Crawlers. O. Gervasi et al. (Eds.): ICCSA 2009, Part II, LNCS 5593, pp. 910–924, Springer-Verlag Berlin Heidelberg.

[18] D. Hai, H. Farookh, C. Elizabeth. (2008). A Survey in Semantic Web Technologies-Inspired Focused Crawlers. 978-1-4244-2917-2/08, IEEE.

[19] D. Hai Dong, K. Farookh. (2010). Focused Crawling for Automatic Service Discovery. Annotation and Classification in Industrial Digital Ecosystems, Industrial Electronics, IEEE Transactions on Volume : PP , Issue:99 .

[20] M. Debajyoti, B. Arup, S. Sukanta. (2007). A New Approach to Design Domain Specific Ontology Based Web Crawler. 0-7695-3068-0/07, IEEE.

[21] S. Wei, L. Da-Xin. (2006), Using Ontologies for Semantic Query Optimization of XML Database, R. Nayak and M.J. Zaki (Eds.): KDXD

2006, LNCS 3915, pp. 64 – 73, 2006, Springer-Verlag Berlin Heidelberg.

[22] Gruber, T. (1993), A translation approach to portable ontology specifications. In: Knowledge Acquisition, 5(2) ,199-220.

[23] Chandrasekaran, B., Josephson, J., Benjamins, V. (1999), What are ontologies, and why do we need them? In: IEEE Intelligent Systems, 20-26.

[24] Guarino, N., Giaretta, P. (1995), Ontologies and knowledge bases: towards a terminological clarification. In: Knowledge Building Knowledge Sharing, ION Press. 25-32.

[25] Noy, N., Hafner, C.D. (1997), The state of the art in ontology design. AI Magazine 3, 53-74.

[26] V. Aida, G. Karina, S. David, B. Montserrat, (2010), Using ontologies for structuring organizational knowledge in Home Care assistance, international journal of medical informatics 79 (2010) 370–387, Elsevier Ireland Ltd.

[27] http://www.medic8.com/healthguide/articles/jaundice.html.

[28] http://en.wikipedia.org/wiki/Jaundice.

[29] http://www.w3.org/TR/owl-ref/

[30] K.K. Breitman, M.A. Casanova and W. Truszkowski. (2007), OWL, In: " Semantic Web: Concepts, Technologies and Applications ", Springer-Verlag London Limited.

[31] L. Ying. (2009). On Document Representation and Term Weights in Text Classification. IGI Global.

[32] Majestic-12 : Projects : C# HTML parser (.NET). http://www.majestic12.co.uk/projects/html_parser.php.

[33] S. Julian, K. Dimitar. (2004). WordNet-based Text Document Clustering. ROMAND '04 Proceedings of the 3rd Workshop on RObust Methods in Analysis of Natural Language Data. ACM.

[34] The Stanford Natural Language Processing Group. http://nlp.stanford.edu/software/tagger.shtml.

[35] wordnetdotnet - Revision 262. http://wordnetdotnet.googlecode.com/svn/trunk/Projects/Thanh/ .

[36] B. Rujiang, W. Xiaoyue, L. Junhua. (2010). Extract Semantic Information from WordNet to Improve Text Classification Performance. AST/UCMA/ISA/ACN 2010, LNCS 6059, pp. 409–420. Springer-Verlag Berlin Heidelberg.

[37] F. Christiane. (2010). WordNet. In: R. Poli et al. Theory and Applications of Ontology: Computer Applications, (pp. 231-243). 231-243, DOI: 10.1007/978-90-481-8847-5_10, Springer Science+Business Media B.V.

[38] G.Tarek, F. Mohammed, A. Mostafa. (2008). Web Document Clustering Approach using WordNet Lexical Categories and Fuzzy Clustering. Proceedings of International Workshop on Data Mining and Artificial Intelligence (DMAI' 08), 24 December, 2008, Khulna, Bangladesh. 1-4244-2136-7/08, IEEE.

[39] G.Tarek, F. Mohammed, A. Mostafa. (2010). Fuzzy Document Clustering Approach using WordNet Lexical Categories. K. Elleithy (ed.), Advanced Techniques in Computing Sciences and Software Engineering, DOI 10.1007/978-90-481-3660-5_31, Springer Science+Business Media.

[40] S. Mu-Hee, L. Soo-Yeon, K. Dong-Jin, L. Sang-Jo. (2005). Automatic Classification of Web Pages based on the Concept of Domain Ontology. Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC'05). 0-7695-2465-6/05, IEEE.

[41] S. Mu-Hee, L. Soo-Yeon, K. Dong-Jin, L. Sang-Jo. (2006). Ontology-Based Automatic Classification of Web Documents. D.-S. Huang, K. Li, and G.W. Irwin (Eds.): ICIC 2006, LNAI 4114, pp. 690 – 700, Springer-Verlag Berlin Heidelberg.

[42] S. Mu-Hee, L. Soo-Yeon, K. Dong-Jin, L. Sang-Jo. (2006). Ontology-Based Automatic Classification of Web Pages. Advances in Intelligent and Soft Computing, Volume 34/2006, 483-493, DOI: 10.1007/3-540-31662-0_37. Springer-Verlag Berlin Heidelberg.

[43] W. Hongsheng, H. Xiaoguang. (2010). Research on Similarity of Semantic Web. International Conference on Computer Application and System Modeling (ICCASM 2010). 978-1-4244-7237-6/10, IEEE.

[44] Q. Sheng, G. Ying. (2008). Measuring Semantic Similarity in Ontology and Its Application in Information Retrieval. Congress on Image and Signal Processing. 978-0-7695-3119-9/08, IEEE.

[45] L. Gang, Z. Cheng, Z. Li. (2009). Text Information Retrieval Based on Concept Semantic Similarity. Fifth International Conference on Semantics, Knowledge and Grid. 978-0-7695-3810-5/09, IEEE.

[46] K.Saruladha, G.Aghila, R. Sajina. (2010). A Survey of Semantic Similarity Methods for Ontology based Information Retrieval. Second International Conference on Machine Learning and Computing. 978-0-7695-3977-5/10, IEEE.

[47] S. Michel, M. Radja, D. Gayo, S. Ana. (2010). Ontologies in the Health Field. In: Data Mining and Medical Knowledge Management: Cases and Applications. Pages: 37-56 pp. 10.4018/978-1-60566-218-3.ch002. IGI Global.

[48] L. Haisheng, T. Yun, Y. Ben, C. Qiang. (2010). Comparison of Current Semantic Similarity Methods in WordNet. International Conference on Computer Application and System Modeling (ICCASM 2010). 978-1 -4244-7237-6/10, IEEE.

[49] H. Tran and S. Dan. (2008). Word Similarity In WordNet. Modeling, Simulation and Optimization of Complex Processes, 293-302, DOI: 10.1007/978-3-540-79409-7_19. Springer-Verlag Berlin Heidelberg.

[50] Z. Xiaodan, J. Liping, H. Xiaohua, N. Michael, X. Jiali, Z. Xiaohua. (2010). Medical Document Clustering Using Ontology-Based Term Similarity Measures. In: International Journal of Data Warehousing and Mining (IJDWM). Pages: 62-73 pp. IGI Global.

[51] J. Bhogal, A. Macfarlane, P. Smith. (2006). A review of ontology based query expansion. Information Processing and Management 43 (2007) 866–886. Elsevier Ltd.

[52] V. Giannis, V. Epimenidis, R. Paraskevi. (2005). Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web. WIDM'05, November 5, Bremen, Germany. ACM.

[53] H. Angelos, V. Giannis, V. Epimenidis, P. Euripides, M. Evangelos. (2009). Information Retrieval by Semantic Similarity. In: Medical Informatics: Concepts, Methodologies, Tools, and Applications. Page: 647-665 pp. IGI Global.

[54] F. Miriam, C. Iván, L. Vanesa, V. David, C. Pablo, M. Enrico. (2010). Semantically enhanced Information Retrieval: An ontology-based approach. Web Semantics: Science, Services and Agents on the World Wide Web, doi:10.1016/j.websem.2010.11.003. Elsevier B.V.

[55] L. Hiep, G. Susan, W. Qiang. (2009). Ontology Learning Through Focused Crawling and Information Extraction. 2009 International Conference on Knowledge and Systems Engineering. 978-0-7695-3846-4, IEEE.

[56] F. Christiane. (2010) . WordNet. Theory and Applications of Ontology: Computer Applications, DOI 10.1007/978-90-481-8847-5_10, C_ Springer Science+Business Media B.V.

[57] R. Manjunath. (2010). Information Retrieval. In: Web-Based Supply Chain Management and Digital Signal Processing: Methods for Effective Information Administration and Transmission. 182-194 pp. DOI: 10.4018/978-1-60566-888-8.ch014. IGI Global.

AUTHORS PROFILE

Hany M. Harb is professor of Computers and Systems Engineering Department - Faculty of Engineering AlAzhar university. Doctor of philosophy (Ph.D.), Computer Science, Illinois Institute of Technology (IIT) , Chicago , Illinois, USA, 1986 He is Chairman of Computers and Systems Engineering department, Chairman of Systems and Networks Unit in Al-Azhar university, and manager of WEB-Based Tansik program. He has supervision of many master's and doctoral degrees in the department of Systems and Computers Engineering.

Khaled M. Fouad received his Master degree of AI and expert systems. He is currently a PhD candidate in the faculty of engineering AlAzhar University in Egypt. He is working now as lecturer in Taif University in Kingdom of Saudi Arabia (KSA) and is assistant researcher in Central Laboratory of Agriculture Expert Systems (CLAES) in Egypt. His current research interests focus on Semantic Web and

Expert Systems.

Nagdy M. Nagdy is professor of engineering applications and computer systems, Department of Systems Engineering and Computer Engineering - Faculty of Engineering Al-Azhar University. He is working now in Al-Baha Private College of Science, Kingdom of Saudi Arabia (KSA) .He received his Ph.D in 1986. He has supervision of some master's and doctoral degrees in the department of Systems Engineering and Computer and Electrical Enginee

Prediction of Users Behavior through Correlation Rules

Navin Kumar Tyagi
Marathwada Institute of Technology
Bulandshahr, UP, India

A. K. Solanki
Meerut Institute of Engineering and Technology
Meerut, UP, India

***Abstract*— Web usage mining is an application of Web mining which focus on the extraction of useful information from usage data of severs logs. In order to improve the usability of a Web site so that users can more easily find and retrieve information they are looking for, we proposed a recommendation methodology based on correlation rules. A correlation rule is measured not only by its support and confidence but also by the correlation between itemsets. Proposed methodology recommends interesting Web pages to the users on the basis of their behavior discovered from web log data. Association rules are generated using FP growth approach and we used two criteria for selecting interesting rules: Confidence and Cosine measure. We also proposed an algorithm for the recommendation process.**

Keywords- Web usage mining; FPgrowth; Cosine measure; Usability; Association rules.

I. INTRODUCTION

The ease and speed with which information exchange and business transactions can be carried out over the Web has been a key driving force in the rapid growth of the Web. Recommendation systems have become popular among users in World Wide Web environment. Web sites generates huge amount of usage data which consists useful information about the users behaviour.Automatic discovery of user access patterns from server log is known as web usage mining . The term web usage mining was introduced by Cooley in 1997.Data mining techniques such as association rules, sequential patterns, clustering and classification can be used to analyze the web site usage data. Association rules mining is one of the important and widely used data mining technique. It is highly successful technique for extracting useful information from very large databases [1, 2, 3, and 4]. In web environment, HTTP server log contains historical user sessions. Web sessions reflect user behavior while navigating throughout a web site and considered as an important source of information about users. Association rules shows similarities between web pages derived from user behavior, can be utilized in Recommender systems. The main objective of such recommendation is to suggest web pages which are useful for the user. Proposed system generates association rules from web log data and then correlation analysis is performed to obtain interesting rules. Pages visited by a user are matched with the antecedent of the rules and consequents of matching rules become the recommendations. In this way proposed system can enhance the usability of the site.

This paper is organized as follows. In section II association rule mining and correlation analysis are presented. In section III we proposed a Methodology and algorithm to predict web pages for the users. An example is presented in section IV. We evaluated the performance of proposed system through example in section V. Section VI presented some related work and conclusion is given in section VII.

II. ASSOCIATION RULES MINING

Association rules [5] are used to show the relationship between data items. These uncovered relationships are not inherent in the data. Association rules are frequently used by retail stores to assist in marketing, advertising, floor management, and inventory control. An association rule A→B represents a relationship between itemsets A and B and it is characterized by two measures, support and confidence. The support of the rule is the percentage of transactions in the database that contain AUB and confidence or strength of the rule is the ratio of the number of transactions that contain AUB to the number of transactions that contain A.

Association rule mining can be viewed as two-step process. In first step frequent itemsets that satisfies a minimum support are generated from the transactional database and in second step strong association rules that satisfies minimum confidence are generated.Apriori[6] can be used to generate frequent itemsets, but it can suffer from two nontrivial costs[3] . It may need to generate a huge number of candidate sets and it may also need to repeatedly scan the database and check a large set of candidates by pattern matching.

An interesting method FP-growth can be used to generate frequent itemsets without candidate generations. This method works on divide and conquers strategy. It compresses the database representing frequent itemsets into a frequent pattern tree or FP tree, which retains the itemset information. It then divides the compressed database into a set of conditional databases; each associated with one frequent item and mines each such database separately. FP growth algorithm is efficient and scalable for mining long and short frequent patterns and is about an order of magnitude faster than the apriori algorithm. It is also faster than Tree-Projection algorithm, which recursively projects a database into a tree of projected databases. To generate association rules from frequent patterns, following steps are to be performed.

For each frequent itemset l, generate all nonempty subsets of l.

- For every nonempty subset s of l, generate the rule s→ (l-s) if support (l)/support(s) ≥ Min_conf, where Min_conf is the minimum confidence threshold.

A. Correlation analysis

An association rule is interesting or not can be assessed either subjectively or objectively. The user can judge if a given rule is interesting, and this judgment, being subjective, may differ from one user to another. However objective interestingness measures based on the statistics behind the data can be used to extract uninteresting rules. Support and confidence measures are insufficient to filter out uninteresting rules as confidence of rule A→B is only an estimate of the conditional probability of itemset B given itemset A. It does not measure the real strength of

The correlation and implication between A and B. In order to overcome this weakness, a correlation measure can be used to augment the support-confidence framework for association rules. This leads to correlation rules of the following form

A→B [support, confidence, correlation]

A correlation rule is measured not only by its support and confidence but also by the correlation between itemsets A and B.Many different correlation measures [3] such as lift, chi square, cosine and all _confidence can be used to perform correlation analysis. Lift between two itemsets A and B can be given by the following equation.

$$\text{confidence}(A\rightarrow B)/\text{support}\ (B) \qquad (1)$$

If the resulting value of equation (1) is less than 1, then occurrence of A is negatively correlated with occurrence of B(Fig.3b).If the resulting value is greater than 1,then A and B are positively correlated(Fig.3a).If the resulting value is equal to 1, then A and B are independent (Fig 3c).For two itemsets A and B, the Cosine Measure can be defined by the following equation.

$$support\ (A \cup B)/\sqrt{support(A) \times support(B)} \qquad (2)$$

The Cosine Measure can be viewed as a harmonized lift measure. Cosine value is only affected by the support of A,B and AUB not by the total number of Transactions.Moreover,Cosine Measure is null invariant as it is not affected by the number of null trasactions.This property is important for measuring correlations in large transaction Databases. Support-confidence framework can be augmented with a correlation measure to mine correlation rules. It can reduce the number of rules generated and leads to the discovery of more meaningful rules. It is better to augment Cosine measure with left when the result is not conclusive.

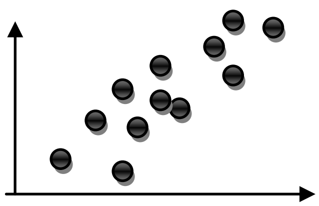


Figure 1a: Positive correlation

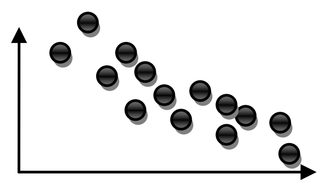


Figure 1b: Negative correlation

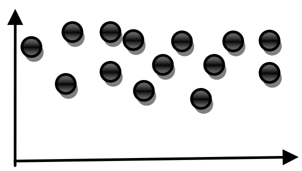


Figure 1c: No correlation

III. METHODOLOGY

Web server stores large volume of data as a result of access to a website. Data may includes date and time of request, URL requested, amount of data, IP address of client, browser and operating system information etc .In proposed methodology (Fig.3) server logs are preprocessed to get sequential list of pages that were visited in the sessions[16]. In Web environment, sessions and pages can be treated as transactions and items respectively. FP growth method [3] is used to generate frequent itemsets and then association rules are generated from frequent itemsets. Cosine measure is used to filter out uninteresting rules. In order to produce better results cosine measure may be augmented with lift measure. We consider dependencies only between 1-page set i.e. single Web pages. Interesting rules are stored in knowledgebase. When a user request for a page, then it is matched with the antecedent part of rules in the knowledgebase and a recommendation list of pages with highest confidence presented to the user [13]. We proposed an algorithm in pseudo codes for overall process of recommendation.

Algorithm

Inputs: Database of Sessions (D)

Minimum Confidence (α)

Minimum support(s)

Threshold cosine value (d)

Output: Recommended web pages

Begin

K_b=ϕ // K_b; knowledgebase

Determine the set of frequent page set L using FP Growth method.

For each l ϵ L // l; frequent page set

Generate association rules $p_i \rightarrow p_j$ and determine Cosine (p_i, p_j). // where i≠ j

If (Confidence ($p_i \rightarrow p_j$) ≥ α && Cosine (p_i, p_j) >d) Then

K_b =K_b U ($p_i \rightarrow p_j$)

Else

Remove the Rule

End If

End For

For each visited page p_i of user u_i

P_i is matched with the antecedent of rules in knowledge base.

End For

Return (consequents from matching rules)

End

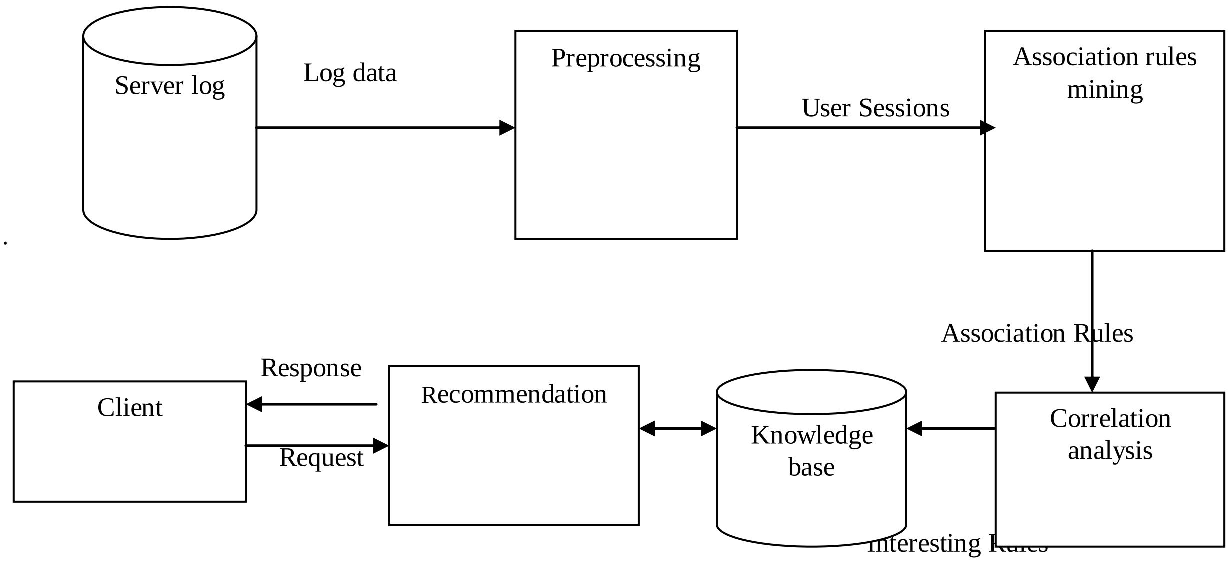


Figure 2.Proposed Methodology

IV. EXAMPLE

Let us consider an example set of nine user sessions within a website which contains five pages (table I), D {A, B, C, D, E}.We used data of Table I to construct FP tree (Fig.3) and then tree is mined [3] to get frequent patterns (Table.II).Association rules generated from frequent patterns are shown in figure 4 and recommendation list for each page is shown in Table III.

TABLE I. DATABASE OF TRANSACTIONS

Session id	Pages
1	A,B,E
2	B,D
3	B,C
4	A,B,D
5	A,C
6	B,C
7	A,C
8	A,B,C,E
9	A,B,C

TABLEII.FREQUENT PATTERNS (SUPPORT COUNT=2)

Page	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns
E	{{B,A:1},{B,A,C:1 }}	<B:2,A:2>	<B,A,E:2> <B,E:2> <A,E:2>
D	{{B,A:1},{B:1}}	<B:2>	<B,D:2>
C	{{B,A:2},{B:2},{A: 2}}	<B:4,A:2>, <A:2>	<B,C:4> <A,C:4> <B,A,C:2>
A	{B:4}	<B:4>	<B,A:4>

C→B	E→A	B→C
A→C	D→B	A→E
C→A	A→B	B→E
E→B	B→A	B→D

Figure 4.Interesting Association rules (α=25%, d=0.5)

V. PERFORMANCE EVALUATION

In this study, we used FP Growth method to find frequent item set, which is faster than Apriori method. The execution time for the two algorithms [15] for different support values on a data set is shown in Fig. 5. Cosine measure is used to prune the generated association rules (only positive correlation between page set has been taken in to account).

Performance of a Recommender system can be evaluated on the basis of three measures: Recall, Precision and F1. Precision measures the degree to which the system produces accurate recommendations. It is the number of relevant web pages retrieved divide by the total number of web pages in the recommendation set. On the other hand Recall measures the ability of the system to produce all of the page views which are likely to be visited by the user and it is the number of relevant web pages retrieved divide by the total number of web pages that actually belong to the user sessions.F1 measure attains its maximum value when both precision and recall are maximized.

Recall= Relevant and Retrieved/ Relevant (3)

Precision= Relevant and Retrieved/ Retrieved (4)

F1= 2(Precision. Recall)/ (Precision+ Recall) (5)

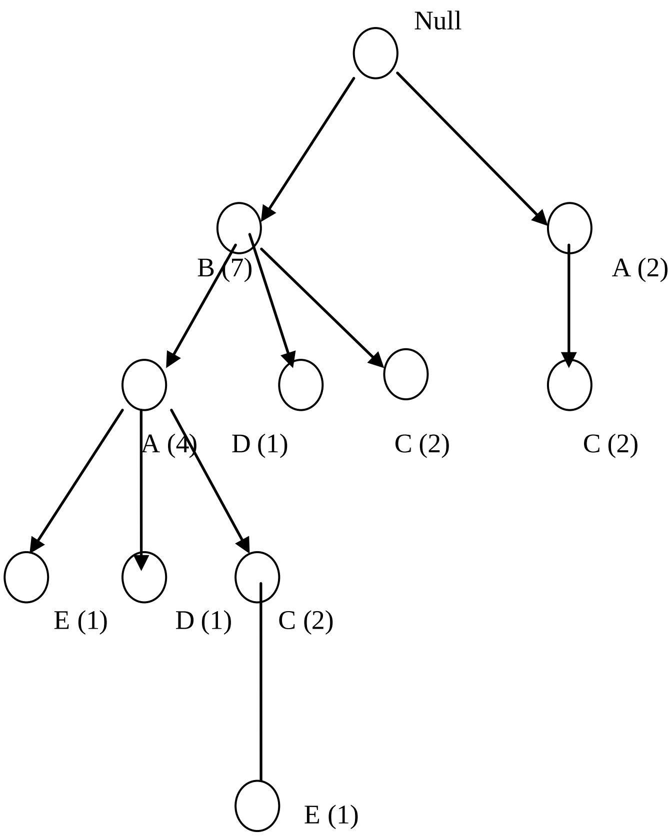


Figure 3.FP Tree

TABLE III.RECOMMENDATION LIST

Page	**Recommended Pages**
A	{C,B},E
B	{C,A}, {E,D}
C	{A,B}
D	B
E	{B,A}

We obtained the values of precision, recall and F1 using (3), (4) and (5) for above mentioned example as shown in Table IV.

TABLE IV.PERFORMANCE METRICS

Precision	**Recall**	**F1**
0.53	0.93	0.67

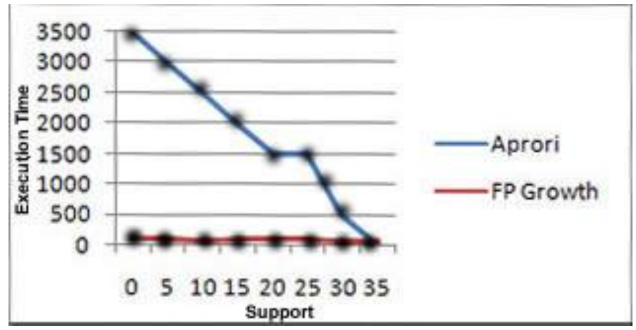


Figure 5.A comparison between FP Growth and Apriori

VI. Related Work

Association Rules Mining is one of the important Data Mining Technique. Association Rules can be used for the recommendation of Web Pages. In [7] Complex association rules have been used for the recommendation of Web pages. [8] Discovered the association rules by using data cube structure and applying OLAP operations. In [9] coordination is achieved between caching and prefetching.Collaborating filtering technique can be used to recommend Web pages within a Web site[10].

This approach uses Association rules mining to form a set of predictive rules, which are further pruned by using minimum reaching distance(MRD) information. Two Rule learning algorithms, Set covering and CN2 to analyze sequences of WWW Pages visits in click stream data are presented in [11].A simplified WWW data model[12] can be used to represent data in the cache of Web browser to mine association rules .These rules are stored in Knowledgebase and prefetched the pages according to user interest. [13] Presented a Recommendation model by generating association rules. An integrated system (Web Tool) for applying Data mining Techniques such as association rules or sequential patterns on access log files is presented in [14].

VII. Conclusion

In this paper we proposed a recommendation methodology based on correlation rules. Association rules are generated from log data by using FP Growth algorithm and then Cosine measure is used for generating correlation rules. We considered only positive correlated rules in our recommendation process and other types of rules (negative and independent) have been pruned. Proposed methodology can recommend web pages to the users which are interesting to them. Moreover negative Correlation may be used to remove the links which are uninteresting to the users.

References

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," *Proc.of ACM SIGMOD Conf. on Management of Data*, 1993, pp. 207-216.

[2] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. of the Eleventh Int. Conf. on Data Engineering*, 1995,pp. 3-14.

[3] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.

M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "A Generalized Affinity-Based Association Rule Mining for Multimedia Database Queries," *Knowledge and Information Systems (KAIS): An International Journal*, vol. 3, no. 3, August 2001, pp. 319-337.

[4] Margaret H.Dunham and S.Sridhar,"Data Mining: Introdutory and Advanced Topics", Pearson Education Publisher, 2008.

[5] Agrawal, R., R. Srikant,"Fast Algorithms for Mining Association Rules, In Proc. of international conference on very large databases, 1994, pp.487-499.

[6] Przemyslaw Kazienko,"Mining Indirect Association Rules for Web Recommendation", International journal of applied math and Computer Science, 2009, vol.19, No.1, pp.165-186.

[7] Galina Bogdanova and Tsvetanka Georgieva,"Discovering the association rules in OLAP data cube with daily downloads of Folklore Materials",International conference on Computer Systems and Technologies 2005.

[8] A.Nanopoulos,.Katsaros and Y. Manolopoulos,"Exploiting Web Log Mining for Web Cache Enhancement", WEBKDD 2001.

[9] M. Shyu, C.Haruechaiyasak and N Zhao," Collaborative Filtering by Mining Association Rules from User Access Sequences"

[10] Peter Berka,"Click stream data analysis using rule based approach".

[11] W.Xu, W.Song, and H.Yang,"Pre-Fetching Web Pages Through data mining based prediction".

[12] A.Jorge, M.Alves, and P.Azevedo,"Recommendation with association rules: A Web mining Application".

[13] F.Masseglia, P.Poncelet, and M.Teisseire,"Using data mining techniques on Web access logs to dynamically improve hypertext structure".

[14] Cornelia Győrödi, Robert Győrödi, prof. dr. ing. Stefan Holban, "A Comparative Study of Association Rules Mining Algorithms".

[15] Murat Ali Bayır, Ismail H. Toroslu, Ahmet SAR"A Performance Comparison of Pattern Discovery Methods on Web Log Data".

AUTHORS PROFILE

Navin Kumar Tyagi received M.Tech in Computer science and Engineering from Kurukshetra university, Kurukshetra (India) in 1998 and currently pursuing Ph.D in Computer Science and Engineering from Bhagwant University, Ajmer (India).Presently, he is working as Assistant Professor and Head of Computer Science and Information Technology Department in Marathwada Institute of Technology, Bulandshahr (India).He has more than 10 years of teaching experience. His areas of interest include Web usage mining, Software Engineering, Operating Systems and Data Structures.

A.K Solanki received M.E. in Computer Science and Engineering from NIT Allahabad (India) in 1996 and Ph.D in Computer Science and Engineering from Bundelkhand University Jhansi (India) in 2005. Presently, he is working as Professor and Director in Meerut Institute of Engineering and Technology, Meerut (India). He has more than 22 years of teaching experience. Professor Solanki is appointed as an executive committee member of national executive council of Indian society of technical education (ISTE) for three years 2009-2012 for Uttar Pradesh and Uttarakhand Region. He is also the member of selection and inspection committee of AICTE, UPTU and other Universities. Professor Solanki research contribution is in the field of Data warehousing and Web Mining.

EGovernment Stage Model: Evaluating the Rate of Web Development Progress of Government Websites in Saudi Arabia

Osama Alfarraj, Steve Drew
ICT School, Griffith University,
Brisbane, Australia

Rayed Abdullah AlGhamdi
King Abdulaziz University,
Jeddah, Saudi Arabia

***Abstract* – This paper contributes to the issue of eGovernment implementation in Saudi Arabia by discussing the current situation of ministry websites. It evaluates the rate of web development progress of vital government websites in Saudi Arabia using the eGovernment stage model. In 2010, Saudi Arabia ranked 58th in the world and 4th in the Gulf region in eGovernment readiness according to United Nations reports. In particular, Saudi Arabia has ranked 75th worldwide for its online service index and its components compared to the neighbouring Gulf country of Bahrain, which was ranked 8th for the same index. While this is still modest in relation to the Saudi government's expectation concerning its vision for eGovernment implementation for 2010, and the results achieved by the neighbouring Gulf countries such as Bahrain and the United Arab Emirates on the eGovernment index, the Saudi government has endeavoured to meet the public needs concerning eGovernment and carry out the implementation of eGovernment properly. Governments may heed the importance of actively launching official government websites – the focus of this study – as the main portals for delivering their online services to all the different categories of eGovernment (including G2C, G2B, and G2G). However, certain Saudi ministries have not given due attention to this vital issue. This is evidenced by the fact that some of their websites are not fully developed or do not yet exist, which clearly impedes that particular ministry from appropriately delivering eServices.**

Keywords- eGovernment; Saudi Arabia government websites; web development progress; eGovernment stage model.

I. INTRODUCTION

In applying the concept of eGovernment in Saudi Arabia, the supreme Royal Decree number 7/B/33181 of 7 September 2003 was established to transform Saudi society into an information society by initiating and supporting new strategies and efforts to facilitate the electronic delivery of government services [1, 2]. However, this decree was not launched until 2005 [2].

The vital initiatives for implementing eGovernment should be followed and given due attention in order to enhance the delivery and use of eGovernment services. One of these initiatives is the active launch of official government websites as the main portals for delivering online services to all of the different categories of eGovernment, including G2C, G2B, and G2G. However, some Saudi ministries do not seem to be paying adequate attention to this issue, as some of their websites are not well developed or do not yet exist, which absolutely impedes that particular ministry from appropriately delivering eServices.

In the study, 'E-Government in Saudi Arabia: Can it overcome its challenges?' conducted by Sahraoui et al. [2], the researchers indicated that only 13 out of 22 Saudi ministries have an online presence. This represents 60% of the ministries. Conducted in 2006, this study was based on a survey as well as online browsing and the accessing of over 25 government websites to evaluate these websites [2]. Moreover, the researchers indicated that the number of Saudi ministries with an online presence has remained the same since the study that had been done by Abanumy et al in 2003 [2], who also found that 'only 13 ministries had online presence, while 8 did not' (as illustrated in Table I below), and that none of these websites were accessible to disabled people as cited in [2, 3].

TABLE I. ONLINE SURVEY FOR SAUDI GOVERNMENT WEBSITES CONDUCTED BY ABANUMY AND MAYHEW IN 2003

Stage reached	Assessment elements	Number	%
No presence	No official website available	8	38%
Emerging presence	**e.g.** agency name, agency phone number, address, operating hours, general frequently asked question	0	0%
Enhanced presence	**e.g.** organisational news, publication, online policy (security, privacy)	3	14%
Interactive presence	**e.g.** officials' e-mail, post comment online, simple two-way communication, download organization's forms	10	48%
Transactional presence	**e.g.** e-form, e-payment	0	0%
Seamless	Full integration across organisation	0	0%

Source: Adapted from Abanumy et al. [3].

According to [4], 'six out of every ten government departments with an Internet connection have their own website. Less than 10% of the websites are hosted only in English. The majority of websites are hosted in both English and Arabic (52%) a shift from Arabic only in 2007' (p. 24).

Although developing eGovernment web portals requires tremendous effort and resources, including human resources, software and hardware, there seems to be a delay in building and launching the government web portals, even though each government agency is in charge of developing its own website and has full responsibility for doing so [5, 6]. For example, one of the main government websites which still lacks an online presence is the Ministry of Hajj (pilgrimage). Approximately two million people from all over the world come at a certain time every year to perform Hajj in Makkah, Saudi Arabia. Such an influx of visitors requires comprehensive information and services to facilitate their travel, such as applying for a Hajj Visa or other related services. The availability a website that would provide online services and facilitate information sharing would benefit all Muslims worldwide, as well as the people of Saudi Arabia. While another website, www.hajinformation.com, does provide information on all issues related to Hajj, it is not an official website for the ministry and lacks online services. The failure to develop such essential websites for all citizens, residents and businesses is delaying the implementation of eGovernment in Saudi Arabia as a part of the eGovernment project.

II. AIM AND SIGNIFICANCE OF THIS STUDY

Despite the emphasis on the concept of eGovernment in the literature, there is still a lack of research which evaluates the progress of government websites, specifically in Saudi Arabia, to show where these websites stand in terms of their readiness to deliver eServices. According to [7], Saudi Arabia was ranked as number 58 worldwide for eGovernment readiness index in 2010 and 4th among the Gulf countries. This position is far from the expectation for 2010, as the Saudi government had predetermined that 'by the end of 2010, everyone in the Kingdom will be able to enjoy from anywhere and at any time – world class government services offered in a seamless, user friendly and secure way by utilizing a variety of electronic means'. As it is now 2011, the timetable for the eGovernment program set by the Saudi government is not being achieved as expected in light of what has been done so far and as indicated in the literature. Thus, this study was conducted to show the level of web readiness of Saudi's government websites, which has played a significant role in delaying the delivery of eServices.

III. RESEARCH METHODOLOGY

The aim of this study is to evaluate the web development of government websites in Saudi Arabia and then repeat the same evaluation for Bahraini government websites to reveal the level of readiness of Saudi government websites in comparison to Bahrain. In addition to helping determine the current status of the Saudi government websites, this study can also determine whether the websites are ready to enter the transactional stage as online service providers. As the study by Sahraoui et al. [2] to rate Saudi websites was conducted in 2006 and no current evaluation study was available in the literature at the time of writing, Saudi ministry websites need to be re-evaluated to note any differences in the subsequent four years (from 2006 to 2010). This will primarily be done by reviewing the relevant published literature and using the eGovernment stage model adopted by [2, 8, 9].

The researchers browsed and visited the same websites of Saudi ministries that Sahraoui et al. [2] evaluated as well as some others. Moreover, the same government websites were also selected in the Bahrain context for the evaluation. The researchers used a checklist (assessment elements) that helps to determine the proper stage of each government website selected, which has been attached as an appendix to this study.

IV. EGOVERNMENT STAGE MODEL

As shown in Figures 1 and 2 below, the eGovernment stage model has five main stages: (I) Emerging presence; (II) Enhanced presence; (III) Interactive presence; (IV) Transactional presence; and (V) Seamless or connected [10, 11].

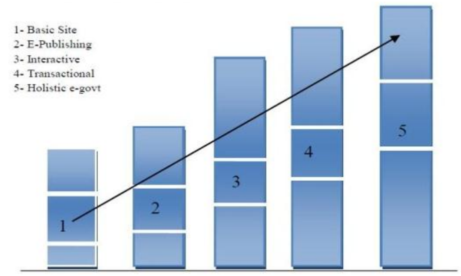


Figure 1. Stages of the eGovernment model (Adopted from [12, 2]).

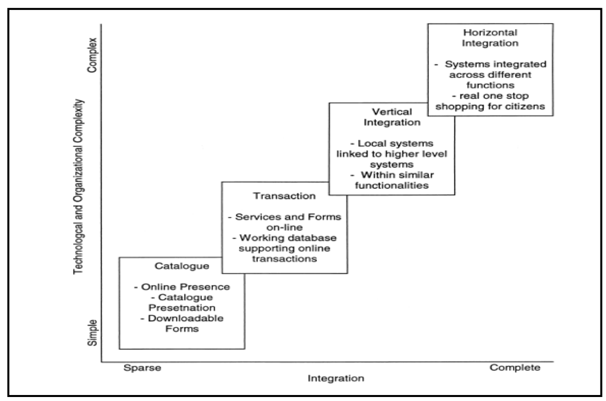


Figure 2. Dimensions and Stages of the eGovernment model (Adopted from [11]).

The stages of eGovernment are further clarified below.

1) Stage I is referred to as emerging. To fall into this stage of eGovernment online presence, there should be an official website for the country containing information about it and at there must be links for the country's ministries and departments, such as health, education and so on [13].

2) Stage II is the enhanced stage wherein the government provides more information to citizens on public policy and the government as well as other information such as reports and regulations, all of which is easily and continuously accessible through archives [10].

3) Stage III is the interactive stage. In this stage the government provides downloadable forms for other services in order to enhance the ease and convenience of the service requester. Simple two-way communication with the ability to post comments online is also offered [13].

4) Stage IV is the transactional stage, which is when the government has started to provide online services and allows citizens to access these services 24/7 in order to represent G2C interactions. Examples of these services are applications for ID cards and online license renewals [10].

5) In Stage V, the connected stage, governments activate back offices; that is, they have transformed themselves into an online entity that meets their citizens' needs and can respond to their citizens in easy and modern ways. Thus, it represents the most developed level of online government initiatives and has the following characteristics:

a) Horizontal connections (among government agencies)

b) Vertical connections (central and local government agencies)

c) Infrastructure connections (interoperability issues)

d) Connections between governments and citizens

e) Connections among stakeholders (government, private sector, academic institutions, NGOs and civil society) [10].

V. Evaluation of Saudi and Bahraini Government Websites Using the eGovernment Stage Model

A. Saudi Context

Only a few studies have been conducted to evaluate the rate of developmental progress of government websites in Saudi Arabia using the eGovernment stage model. The study done by [2] evaluated the Saudi ministries' web portals in accordance with the eGovernment stage model. The researchers of this study browsed and visited some of the main Saudi government websites to determine at which stage each one of these government websites was for the purpose of showing their readiness. As the researchers state, 'they are mostly situated between stage II and III of the e-government stage model, hence not yet fully transactional' [2]. As only 13 out 22 Saudi ministries had an online presence, with the majority of these websites being placed between stages II and III, none were providing online services. Table II depicts the majority of Saudi government websites that were placed between stages II and III. These websites are considered to be only information providers rather than service providers, with the exception of Saudi Telecom, a government organization which was placed at stage IV as a service provider.

As the study by [2] was done in 2006 and no current evaluation study was available in the literature at the time of writing, the evaluation of the Saudi ministries needs to be done again to note the differences in the next four years (from 2006 to 2010) and to show the current level of these web portals. Thus, the same assessment elements were employed by Sahraoui and his colleagues in 2006 were used in this study. The researchers browsed and visited the same Saudi ministries' websites as well as some others (see Table III) and were evaluated according to the eGovernment stage model offered by United Nations to determine whether or not these official web portals have started delivering their eServices to their citizens, residents, and businesses. This links with the above discussion on whether the Saudi eGovernment has implemented and delivered its eServices well in the specified time set in 2005 by the 'Yesser' program, which stated: 'By the end of 2010, everyone in the Kingdom will be able to enjoy – from anywhere and at any time – world class government services offered in a seamless, user friendly and secure way by utilizing a variety of electronic means' [14, 2, 15, 16]. As seen in Table III below, 28 official Saudi government websites of various ministries were browsed and visited via the online survey.

The online survey examined the following criteria: whether the ministry has an online presence, information, downloadable applications, online applications, transaction inquiry, transaction online (services online), and whether an English version is available for the same website and other checklist criteria (see appendix), as carried out by [2] and previously by [9, 8].

TABLE II. Online survey of Saudi government websites conducted by Sahraoui et al. on 6 April 2006

	Authority	Application Download	Application Online	Transaction Inquiry	Transaction Online	English Version	Stage	URL
1	The E-government Program	No	Yes	Yes	No	No	?	www.gov.sa
2	Communication and Information Technology Commission	Yes	Yes	Yes	No	No	II	www.citc.gov.sa
3	Saudi Telecom	Yes	Yes	Yes	Yes	Yes	IV	www.stc.com.sa
4	Ministry of Agriculture	Yes	No	Yes	No	No	III	www.agrwat.gov.sa
5	Ministry of Civil Service	Yes	Yes	Yes	No	No	III	www.mcs.gov.sa
6	Ministry of Commerce and Industry	Yes	Yes	Yes	No	No	III	www.commerce.gov.sa
7	Council of Saudi Chambers of Commerce and Industry	Yes	Yes	Yes	*	Yes	III	www.saudichambers.org.sa
8	Ministry of Defense and Aviation	Yes	Yes	No	No	Yes	III	www.gaca.gov.sa
9	Ministry of Water and Electricity	Yes	Yes	No	No	No	III	www.mow.gov.sa
10	Saudi Ports Authority	No	No	No	No	Yes	III	www.ports.gov.sa
11	Ministry of Interior Passport Authority	Yes	No	Yes	No	No	III	www.passport.gov.sa
12	Ministry of Health	No	No	Yes	No	No	III	www.moh.org.sa
13	Ministry of Education	No	No	No	No	Yes	III	www.moe.gov.sa
14	Royal Embassy of Saudi Arabia - London	Yes	Yes	Yes	No	Yes	III	www.mofa.gov.sa
16	Ministry of Hajj	No	No	No	No	No	III	www.hajinformation.com
16	Ministry of Higher Education	Yes	No	Yes	No	No	III	www.mohe.gov.sa
17	Ministry of Communications and Information Technology (IT)	No	No	No	No	Yes	III	www.mcit.gov.sa

This online survey revealed that only two of the selected ministries (Ministry of Hajj and General Presidency of Youth Welfare) still have no online presence, while 26 have an online presence, which shows improvement compared to the online survey done by [2].

One of these two ministries lacking an online presence is the Ministry of Hajj (pilgrimages), which is a very important ministry that benefits not only the residents, citizens, and businesses of Saudi Arabia, but all Muslims worldwide as well.

TABLE III. ONLINE SURVEY FOR SAUDI GOVERNMENT WEBSITES CONDUCTED BY THE RESEARCHERS OF THIS STUDY IN SEPTEMBER 2010

No.	Authority	Has presence	Information	Application Download	Application Online	Transaction Inquiry	Transaction Online	English Version	Stage	URL
1	Saudi eGovernment national portal	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.saudi.gov.sa
2	Communication and Information Technology Commission	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.citc.gov.sa
3	Saudi Telecom	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.stc.com.sa
4	Ministry of Agriculture	Yes	Yes	Yes	No	Yes	No	No	III	http://www.moa.gov.sa
5	Ministry of Civil Service	Yes	Yes	Yes	Yes	Yes	Yes	No	IV	http://www.mcs.gov.sa
6	Ministry of Commerce and Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.mci.gov.sa
7	Council of Saudi Chambers of Commerce and Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.saudichambers.org.sa
8	Ministry of Defense and Aviation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	ttp://www.gaca.gov.sa
9	Ministry of Water and Electricity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.mowe.gov.sa/
10	Saudi Ports Authority	Yes	Yes	No	No	No	No	Yes	II	http://www.ports.gov.sa
11	Ministry of Interior (Passport Authority)	Yes	Yes	Yes	Yes	Yes	No	No	III	http://www.gdp.gov.sa
12	Ministry of Health	Yes	Yes	No	Yes	Yes	No	Yes	IV	http://www.moh.gov.sa
13	Ministry of Education	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.moe.gov.sa
14	Ministry of Hajj	No	Yes	No	No	No	No	Yes	III	http://www.hajinformation.com/
15	Ministry of Foreign Affairs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.mofa.gov.sa
16	Ministry of Higher Education	Yes	Yes	Yes	Yes	Yes	Yes	Yes	V	www.mohe.gov.sa
17	Ministry of Communications and Information Technology (IT)	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.mcit.gov.sa
18	Ministry of Finance	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.mof.gov.sa
19	Ministry of Justice	Yes	Yes	Yes	No	No	No	No	III	http://www.moj.gov.sa
20	Ministry of Labor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.mol.gov.sa
21	Ministry of Municipalities and Rural Affairs	Yes	Yes	Yes	Yes	Yes	Yes	No	IV	http://www.momra.gov.sa/
22	Ministry of Petroleum and Mineral Resources	Yes	Yes	Yes	No	No	No	Yes	II	http://www.mopm.gov.sa
23	Ministry of Economy and Planning	Yes	Yes	Yes	Yes	No	No	Yes	III	http://www.mep.gov.sa
24	Ministry of Islamic Affairs, Endowment, Dawa and Guidance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.moia.gov.sa
25	Ministry of Social Affairs	Yes	Yes	Yes	Yes	No	No	No	III	http://mosa.gov.sa
26	Ministry of Transport	Yes	Yes	Yes	Yes	Yes	No	Yes	III	http://www.mot.gov.sa
27	Ministry of Interior	Yes	Yes	Yes	Yes	Yes	Yes	Yes	IV	http://www.moi.gov.sa
28	General Presidency of Youth Welfare	Yes	*	*	*	*	*	*	*	http://www.gpyw.gov.sa

The website of Ministry of Hajj, which was provided and examined in the table above, is an unofficial website and does not officially belong to the Ministry of Hajj.

Thus, specific attention should be given to the development of this Ministry's web portal. Moreover, the online survey revealed that only two government websites, namely, the Saudi Ports Authority and the Ministry of Petroleum and Mineral Resources, are designated as falling under Stage II (enhanced presence) of the eGovernment stage model, which means that they are only providing basic items that are not even in downloadable forms. Therefore, these two government websites are at a very low level and need to be developed as both are essential ministries. Table IV below summarises the results of the online survey according the number of government organizations in each stage.

As seen in Table IV, the online survey indicated that 11 government ministry websites are currently at stage III (Interactive presence), which means that these ministries do not yet provide online services.

TABLE IV. THE NUMBER OF SAUDI GOVERNMENT WEBSITES IN EACH STAGE, AS SURVEYED IN SEPTEMBER 2010

Stage No.	Stage reached	Assessment elements	Number of Saudi government ministries
	No presence	No official websites available	2
I	Emerging presence	**e.g.** agency name, agency phone number, address, operating hours, general frequently asked questions	-
II	Enhanced presence	**e.g.** organisational news, publication, online policy (security, privacy)	2
III	Interactive presence	**e.g.** officials' e-mail addresses, ability to post comments online, simple two-way communication, can download the organisation's forms	11
IV	Transactional presence	**e.g.** e-form, e-payment and some query services	12
V	Seamless	Full integration across the organisation	1

On the other hand, the online survey found that 12 Saudi ministry websites are currently at stage IV (transactional presence), demonstrating that these ministry websites are considered to be online services providers. The majority of these websites have basic online services such as online query services and eForms. Additionally, the online survey found that only one ministry is currently at stage at stage V (seamless), namely, the Ministry of Higher Education. Consequently, none of the government websites, except for the Ministry of Higher Education, achieved full integration or even a high performance at the level of transactional presence by the end of 2010 as recommended by the Yesser program. In comparison with informational websites, transactional websites usually receive high scores for the United Nations web index for eGovernment readiness [10]. This is why the 'survey ranked the UAE in 5th position in terms of transactional services, just behind developed countries like Sweden, Denmark, Norway and the US', as shown in Figure 3 below [17, 18]. The transactional stage was defined by [19] as one 'in which citizens will be able to conduct business online with governments'.

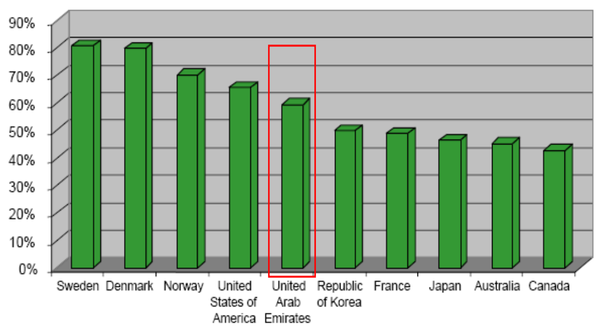


Figure 3. Transactional services: Top 10 countries in 2008 [10].

B. Bahraini Context

Bahrain is a neighbouring country to Saudi Arabia and was used in this study as a country of comparison. Table V provides a comparison of attributes between Saudi Arabia and Bahrain in relation to the readiness of their government websites to deliver eServices.

The same online evaluation survey was conducted evaluate and show the rate of web development progress for Bahraini government websites compared to those evaluated in the Saudi Arabian context (as show in table VI). This online evaluation survey has revealed that all 23 selected ministries have an online presence and that none of the selected ministries are at Stage II (enhanced presence) as mere information providers. Furthermore, the online evaluation found that only six government websites are designated as being in Stage III (Interactive presence) of the eGovernment stage model, indicating that none of these ministries have provided online services yet.

On the other hand, the online survey found that eight Bahraini government websites are currently at Stage IV (transactional presence), demonstrating that these ministry websites are considered to be online service providers with e-forms and payment options. Surprisingly, the online survey revealed that nine ministries are currently at Stage V (seamless). Consequently, the majority of the selected ministries for Bahrain (which are similar to those selected for Saudi Arabia) are currently placed in the highest stage of the eGovernment model and have indeed transformed themselves into an online entity that meets its citizens' needs and responds to its citizens in easy and developed ways. Thus, this represents the most developed level of online government initiatives. Table VII summarises the results of the online evaluation survey according to the number of government organizations in each stage.

TABLE V. THE ATTRIBUTES OF SELECTING BAHRAIN AS A COUNTRY OF COMPARISON

Country / Attribute	Saudi Arabia	Bahrain
Culture	▪ Islam is practised by the majority of people and is part of many aspects of life ▪ The family/tribal structure is one of the cultural aspects used to identify social structure and individual identity ▪ Conservative dress for both men and women is compulsory	▪ Islam is practised by the all people and it is part of many aspects of life ▪ The family/tribal structure is one of the cultural aspects used to identify social position and individual identity ▪ Conservative dress for both men and women is not compulsory, but this does not mean that it should be too free. It should be acceptable and not offend public decency
Government and the country	▪ Monarchy ▪ Constitution governed in accordance with Islamic Law ▪ Legal system based on Sharia law (Islamic law) ▪ Unification of the Kingdom as a public holiday alongside with *Eid ul-Fitr* (festivity of conclusion of the fast) and *Eid al-Adha* (Festival of Sacrifice)	▪ Monarchy ▪ Legal system based on Islamic law and English common law ▪ There are several public holidays which include New Year's Day. Mouloud (Birth of the Prophet), Eid al-Fitr (End of Ramadan), Eid al-Adha (Feast of the Sacrifice), National Day (two days), Labour day, Accession Day, Arafat day, Al-Hijrah (Islamic New Year), Ashoura which is the tenth day of Muharram in the Islamic calendar
People	▪ Arabs	▪ Arabs
Religion	▪ Islam	▪ Islam
Language used	▪ Arabic is the official language ▪ English is a commercial language	▪ Arabic is the official language ▪ English is a commercial language
Geography	▪ Middle East, bordering the Persian Gulf and the Red Sea	▪ Middle East, archipelago in the Persian Gulf, east of Saudi Arabia
Economy	▪ Oil-based economy as a major economy ▪ Main exports: oil, gas, cereals	▪ It is well diversified and a home for multinational firms in the Gulf region ▪ Planning depends heavily on oil ▪ Main exports: petroleum and petroleum products and aluminum
UN eGovernment readiness as 2010	▪ Fourth ranking amongst the Gulf countries	▪ First ranking amongst the Gulf countries
Gulf Cooperation Council (GCC)	▪ Saudi Arabia is a member of the Gulf Co-operation Council	▪ Bahrain is a member of the Gulf Co-operation Council

Source: [20], [21], [22], [23], [7], [24].

TABLE VI. SURVEY FOR BAHRAINI GOVERNMENT WEBSITES CONDUCTED BY RESEARCHERS IN SEPTEMBER 2010

No.	Authority	Application Download	Application Online	Transaction Inquiry	Transaction Online	English Version	Stage	URL
1	Bahraini eGovernment national portal	Yes	Yes	Yes	Yes	Yes	V	http://www.e.gov.bh/pubportal
2	Telecommunication regulatory authority	Yes	Yes	Yes	Yes	Yes	IV	http://www.tra.org.bh/
3	Bahraini Telecom	Yes	Yes	Yes	Yes	Yes	IV	http://www.batelco.com/portal/
4	Ministry of Municipalities and Agriculture Affairs	Yes	Yes	Yes	Yes	Yes	IV	http://websrv.municipality.gov.bh/mun/index_en.jsp
5	Ministry of Civil Service	Yes	Yes	Yes	No	No	III	http://www.csb.gov.bh/csb/wcms/ar/home/
6	Ministry of Industry and Commerce	Yes	Yes	Yes	Yes	Yes	V	http://www.moic.gov.bh/moic/en
7	Bahrain chamber of commerce and industry	Yes	Yes	Yes	No	Yes	III	http://www.bahrainchamber.org.bh
8	Bahrain defence force	Yes	Yes	Yes	No	Yes	III	http://www.bdf.gov.bh/ar/default.asp
9	Electricity and Water Authority	Yes	Yes	Yes	Yes	Yes	V	http://www.mew.gov.bh
10	General organisation of sea ports, Bahrain	Yes	Yes	Yes	Yes	Yes	IV	http://www.gop.bh/index.asp
11	Ministry of Interior Passport Authority	Yes	Yes	Yes	Yes	Yes	IV	http://www.gdnpr.gov.bh/
12	Ministry of Health	Yes	Yes	Yes	Yes	Yes	V	http://www.moh.gov.bh
13	Ministry of Education	Yes	Yes	Yes	Yes	Yes	V	http://www.moe.gov.bh
14	Ministry of Foreign Affairs	Yes	Yes	Yes	No	Yes	III	http://www.mofa.gov.bh/
15	Ministry of Finance	Yes	Yes	Yes	Yes	Yes	V	http://www.mof.gov.bh
16	Ministry of Justice & Islamic affairs	Yes	Yes	Yes	Yes	Yes	V	http://www.moj.gov.bh
17	Ministry of Labor	Yes	Yes	Yes	Yes	Yes	IV	http://eservices.mol.gov.bh
18	National oil & gas authority	Yes	Yes	Yes	Yes	Yes	IV	http://www.noga.gov.bh
19	Ministry of Works	Yes	Yes	Yes	Yes	Yes	IV	http://www.works.gov.bh
20	Ministry of Social development	Yes	Yes	Yes	Yes	Yes	V	http://www.social.gov.bh
21	Ministry of Transport	Yes	Yes	Yes	No	Yes	III	http://www.transportation.gov.bh/
22	Ministry of Interior	Yes	Yes	Yes	Yes	Yes	V	http://www.interior.gov.bh
23	General organization for youth and sports	Yes	Yes	Yes	No	Yes	III	http://www.goys.gov.bh/

TABLE VII. THE NUMBER OF BAHRAINI GOVERNMENT WEBSITES IN EACH STAGE, AS SURVEYED IN SEPTEMBER 2010

Stage No.	Stage reached	Assessment elements	The number of Bahraini government ministries
	No presence	No official websites available	0
I	Emerging presence	e.g. agency name, agency phone number, address, operating hours, general frequently asked questions	-
II	Enhanced presence	e.g. organisational news, publication, online policy (security, privacy)	0
III	Interactive presence	e.g. officials' e-mail, ability to post comments online, simple two-way communication, can download the organisation's forms	6
IV	Transactional presence	e.g. e-form, e-payment and some query services	8
V	Seamless	Full integration across the organisation	9

According to [7], 'what is noteworthy is that some developing countries have begun to catch up with higher-income countries despite these challenges. Bahrain (0.7363), for example, has made significant strides in the two years since the previous survey, moving up in the rankings to 13th place in 2010 from 42nd place in 2008'. Therefore, Bahrain has already attained excellent results in respect to eGovernment readiness, having already ranked as number 13 worldwide (as illustrated in table VIII), where it ranked first amongst the Gulf countries as well as the region and third in Asia.

TABLE VIII. TOP 20 COUNTRIES IN EGOVERNMENT DEVELOPMENT

Rank	Country	E-government development index value	Rank	Country	E-government development index value
1	Republic of Korea	0.8785	11	Singapore	0.7476
2	United States	0.8510	12	Sweden	0.7474
3	Canada	0.8448	13	Bahrain	0.7363
4	United Kingdom	0.8147	14	New Zealand	0.7311
5	Netherlands	0.8097	15	Germany	0.7309
6	Norway	0.8020	16	Belgium	0.7225
7	Denmark	0.7872	17	Japan	0.7152
8	Australia	0.7863	18	Switzerland	0.7136
9	Spain	0.7516	19	Finland	0.6967
10	France	0.7510	20	Estonia	0.6965

Source: Adapted from [30].

In fact, Bahrain ranked 8th place in eGovernment development in 2010 as demonstrated in Table IX below, while in the same context, Saudi Arabia ranked 75th [7].

According to [7], 'a country's strength in online service provision correlates positively with its use of new technology such as the emerging tools for social networking' (p. 76).

This reflects the growth and the high level of online services provided by the country with advanced Web 2.0 tools such as online discussion forums, live chat and online polls on government portals and websites. Such facilities can assist in getting citizens involved in government decision making.

TABLE IX. TOP 20 COUNTRIES IN ONLINE SERVICE DEVELOPMENT

Rank	Country	Online service index value	Rank	Country	Online service index value
1	Republic of Korea	1.0000	11	France	0.6825
2	United States	0.9365	12	Netherlands	0.6794
3	Canada	0.8825	13	Denmark	0.6730
4	United Kingdom	0.7746	14	Japan	0.6730
5	Australia	0.7651	15	New Zealand	0.6381
6	Spain	0.7651	16	Malaysia	0.6317
7	Norway	0.7365	17	Belgium	0.6254
8	Bahrain	0.7302	18	Chile	0.6095
9	Colombia	0.7111	19	Israel	0.5841
10	Singapore	0.6857	20	Mongolia	0.5556

Online presence for any government website has several stages which should be gone through to maintain a high level of services provided, as mentioned previously in the eGovernment stage model section.

While attaining transactional presence is a bit challenging, it is not impossible. However, 'only a few countries are able to offer many transactional services online at this time' [7]. Bahrain is one country that has already reached this stage (as illustrated in Figure 4) and offers a wide range of integrated transactional e-services by having comprehensive back office integration systems and advanced networks. As these systems are extremely secure, they allow citizens to operate e-services with confidence [7]. This is missing in the Saudi context as the online services provided by government websites are poor and lack quality.

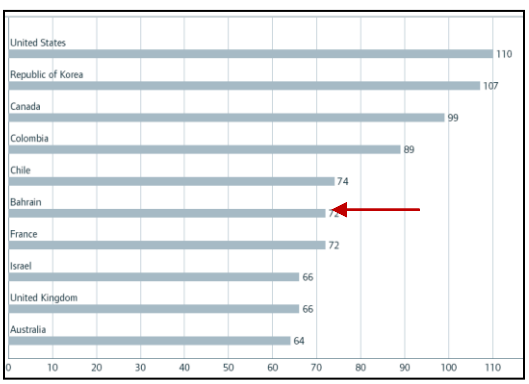


Figure 4. Selected countries with high transactional presence scores (Adapted from [7]).

Moreover, Bahrain has achieved a high ranking with respect to the e-participation index, which is also a part of the eGovernment readiness index for the whole country (this involves: 1. the web measure index; 2. the telecommunication infrastructure index; 3. the human capital index 4; and e-participation index) as this index can determine the level of eGovernment readiness for each country compared to others in the world. The goal of e-participation is to 'improve the citizen's access to information and public services; and participation in public decision-making' [13].

Thus, it has three parts, which are e-information, e-consultation and e-decision making, which have to be examined to determine the level of e-participation. This index easily reflects the level of engagement on the part of citizens/residents and businesses in the decision making of a particular government. Bahrain stands out among the Gulf countries as it is ranked in 11th place (as shown in Table X below) internationally and stands at 4th place in respect to the quality of e-participation websites worldwide (see Table XI). Conversely, Saudi Arabia is ranked in 102nd place with an e-participation index score (0.1000) that is considered to be very low [7]. Furthermore, Saudi government websites lack in the deployment and utilization of new technologies of Web 2.0, such as forums and web-based collaborative technologies that can effectively connect the public with the government in easy and efficient ways as is being done in Bahrain.

TABLE X. TOP 20 COUNTRIES IN THE E-PARTICIPATION INDEX FOR 2010

Rank	Country	2010 e-participation index value	2010 rank	2008 rank	Change +/(-)
1	Republic of Korea	1.0000	1	2	1
2	Australia	0.9143	2	5	3
3	Spain	0.8286	3	34	31
4	New Zealand	0.7714	4	6	2
4	United Kingdom	0.7714	4	25	21
6	Japan	0.7571	6	11	5
6	United States	0.7571	6	1	(5)
8	Canada	0.7286	8	11	3
9	Estonia	0.6857	9	8	(1)
9	Singapore	0.6857	9	10	1
11	Bahrain	0.6714	11	36	25
12	Malaysia	0.6571	12	41	29
13	Denmark	0.6429	13	3	(10)
14	Germany	0.6143	14	74	60
15	France	0.6000	15	3	(12)
16	Netherlands	0.6000	15	16	1
17	Belgium	0.5857	17	28	11
18	Kazakhstan	0.5571	18	98	80
19	Lithuania	0.5286	19	20	1
20	Slovenia	0.5143	20	55	35

TABLE XI. QUALITY OF E-PARTICIPATION WEBSITES OF SELECTED COUNTRIES FOR 2010

Range	Country	Score (%) E-information	E-consultation	E-decision making	Total
Over 60%	Republic of Korea	87.50	78.79	75.00	78.95
	Australia	100.00	60.61	68.75	68.42
	Kazakhstan	87.50	66.67	62.50	68.42
	Bahrain	75.00	66.67	56.25	64.91
30-60%	Spain	75.00	63.64	37.50	57.89
	Kyrgyzstan	50.00	63.64	31.25	52.63
	Mongolia	62.50	54.55	43.75	52.63
	Israel	50.00	51.52	50.00	50.88
	New Zealand	50.00	54.55	43.75	50.88
	United Kingdom of Great Britain	50.00	60.61	31.25	50.88
	Japan	87.50	39.39	50.00	49.12
	United States of America	50.00	54.55	37.50	49.12
	Canada	75.00	36.36	50.00	45.61
	China	37.50	39.39	62.50	45.61
	Colombia	75.00	39.39	43.75	45.61
	Mexico	87.50	51.52	12.50	45.61
	Slovenia	50.00	45.45	43.75	45.61
	Chile	75.00	39.39	31.25	42.11
	Cyprus	50.00	30.30	62.50	42.11
	Estonia	50.00	42.42	31.25	40.35
	Singapore	50.00	48.48	18.75	40.35
Under 30%	Belarus	37.50	33.33	18.75	29.82
	France	37.50	36.36	12.50	29.82
	Netherlands	75.00	18.18	31.25	29.82
	Belgium	62.50	12.12	43.75	28.07
	Kenya	37.50	33.33	12.50	28.07
	Kuwait	75.00	21.21	18.75	28.07
	Turkey	37.50	18.18	37.50	26.32

Source: Adapted from [7].

The brief discussion above reveals that the Bahrain government websites rank at a high level as e-service providers alongside with advanced and developed countries websites in the world. So far, Bahrain is the only country in the Gulf region that has managed to attain this result and serves as a very good example that can be followed by Saudi Arabia. Furthermore, the United Nations results showed a wide divergence between Bahrain and Saudi Arabia in terms of online service development, as Bahrain was ranked in 8th place worldwide compared to Saudi Arabia, which ranked 75th place.

The Saudi Arabian government should heed the web development progress of its ministries' websites. Moreover, it should treat this as a serious issue that impedes the current implementation of eGovernment, particularly in delivering eServices within the specified timeframe. Therefore, the Saudi Arabian government should speedily respond in handling this delay in the portal development of its ministries' websites and follow this up gradually with the top management or high authorities in the country in order to maintain a steady rate of web development for these government websites. Furthermore, engaging the public (citizens, residents and businesses) in decision-making is crucial and must be done to increase the level of transparency between the government and the public as well as to meet the public needs in easy, modern and effective ways.

References [25, 26] showed inactive role of the government supporting the growth of online activities in the country. They indicated that the government support is a critical key to promote online activities in the country as people and businesses in Saudi Arabia have tendency to feel more confident and secure with the online activities comes through the government or under its supervision.

VI. CONCLUSION

Literature in the field of eGovernment implementation and specifically in the development of web portals for government agencies in developing countries, especially in Saudi Arabia, is still lacking. This motivates researchers to further explore this with the aim of bringing about the expected benefits of eGovernment implementation by pointing out the main barriers to eGovernment by analysing the web development progress of government websites. It is clear that some Saudi ministries have made progress in developing their websites to enhance the implementation of eGovernment applications.

However, these websites still need to provide more comprehensive online services that can adequately serve the residents/citizens. Providing eServices for car registration renewal via the traffic department website or for registering people in continuous educational programs through the education ministry website are examples of needed services. However, as the majority of Saudi ministry websites lack such services that would reinforce the concept of eGovernment among society, the Saudi government should pay close attention to the slow development of its government websites and take measures to remedy this immediately where applicable. This is particularly crucial as the majority of the selected websites in this study fall between Stages III and IV of the eGovernment stage model and some are providing poor online services as compared to Bahrain, which has achieved excellent results in this regard.

REFERENCES & BIBLIOGRAPHY

[1] Abu Mgayd, Y 2004, 'Electronic Government in General Organizations in Saudi Arabia', Master thesis, King Saud University.

[2] Sahraoui, S, Gharaibeh, G & Al-Jboori, A 2006, 'E-Government in Saudi Arabia: Can it Overcome its Challenges', paper presented to eGovernment Workshop '06 (eGOV06), Brunel University, West London UB8 3PH, UK.

[3] Abanumy, A, Al-Badi, A & Mayhew, P 2005, 'e-Government Website Accessibility: In-Depth Evaluation of Saudi Arabia and Oman', *The Electronic Journal of e-Government*, vol. 3, no. 3, pp. 99-106.

[4] Communications and Information Technology Commission 2008, *Internet Usage in the Kingdom of Saudi Arabia: Government Sector*, Riyadh, Kingdome of Saudi Arabia.

[5] Yesser e-Government Team 2006, *The National e-Government Strategy and Action Plan (Yesser report)*, Ministry of Information and Communication Technology, Riyadh.

[6] Almarabeh, T & AbuAli, A 2010, 'A General Framework for E-Government: Definition Maturity Challenges, Opportunities, and Success', *European Journal of Scientific Research*, vol. 39, no. 1 (2010), pp. 29-42.

[7] United Nations 2010, *E-Government Survey 2010: Leveraging e-government at a time of financial and economic crisis*, Department of Economic and Social Affairs, Division for Public Administration and Development Management, New York.

[8] Abanumy, A 2003, 'The Way to e-Ministries in Saudi Arabia', *The economist*, p. 11.

[9] Abanumy, A, Mayhew, P & Al-Badi, A 2003, 'An Exploratory Study of EGovernment in two GCC Countries', in *International Business Information Management Conference*,, Cairo, Egypt, pp. 387-95.

[10] United Nations 2008, *United Nations e-Government Survey 2008: From E-Government to Connected Governance*, Department of Economic and Social Affairs, Division for Public Administration and Development Management, New York.

[11] Al-Khouri, AM & Bal, J 2007, 'Electronic Government in the GCC Countries', *International Journal of Social Sciences*, vol. 1, no. 2, pp. 83-98.

[12] Sahraoui, S 2005, 'E-Government in the Arabian Gulf: Government Transformation Vs. Government Automation', paper presented to eGovernment Workshop '05 (eGOV05), Brunel University, West London UB8 3PH, UK

[13] United Nations 2008, *E-government Report for Government of Saudi Arabia*, Department of Economic and Social Affairs, Division for Public Administration and Development Management, New York.

[14] Yesser Program Team 2010, *Kingdom of Saudi Arabia e-Government Program (Yesser)*, The Ministry of Communication and Information Technology, Riyadh, KSA, <http://www.yesser.gov.sa>.

[15] AL-Shehry, A, Rogerson, S, Fairweather, NB & Prior, M 2006, 'The Motivations for Change Towards E-Government Adoption: Case Studies from Saudi Arabia', paper presented to eGovernment Workshop '06 (eGOV06), Brunel University, West London, UB8 3PH.

[16] Al-Shehry, AM 2008, 'Transformation Towards E-Government in the Kingdom of Saudi Arabia: Technological and Organisational Perspectives', Doctoral thesis, De Montfort University, Leicester, UK.

[17] National Media Council 2009, *UAE Yearbook 2009*, Trident Press Ltd, London.

[18] UAE Interact 2010, *News & Information on the United Arab Emirates*, viewed 4 October 2010, <http://www.uaeinteract.com>.

[19] Coursey, D & Norris, DF 2008, 'Models of E-Government: Are They Correct? An Empirical Assessment', in *Public Administration Review*, vol. 68(3), pp. 523-36.

[20] Central Intelligence Agency 2010, *The Country profiles for Saudi Arabia and Bahrain* viewed 3 September 2010, <https://www.cia.gov/>.

[21] AME Info FZ LLC 1996 - 2011, *Country Profile of Bahrain* viewed 4 September 2010, <http://www.ameinfo.com/>.

[22] Foriegn & Commonwealth Office 2011, *The Country Profiles for Saudi Arabia & Bahrain* viewed 4 September 2010, <http://www.fco.gov.uk/en/>.

[23] British Broadcasting Corporation 2011, *Bahrain and Saudi Arabia Country Profiles* viewed 6 September 2010, <http://www.bbc.co.uk>.

[24] Ministry of Foreign Affairs 2005-2011, *All About the Kingdom of Saudi Arabia*, viewed 7 September 2010, <http://www.mofa.gov.sa>.

[25] AlGhamdi, R, Drew, S & Alshehri, M 2011, 'Strategic Government Initiatives to Promote Diffusion of Online Retailing in Saudi Arabia', in *The Sixth International Conference on Digital Information Management*, Melbourne, Australia.

[26] AlGhamdi, R & Drew, S 2011, 'Seven Key Drivers to Online Retailing in KSA', in P Kommers & P Isaías (eds), *Proceedings of the IADIS International Conference on e-Society 2011*, Avila, Spain, pp. 237-44.

APPENDIX

No.	Authority	Emerging presence					Enhanced presence			Interactive presence				Transactional presence			Seamless	URL
		agency name	Agency contacts	Agency address	Operation hour	General frequently asked questions	organizational news	publications	online policy	officials' e-mail	post comment online	two-way communication	Download forms	e-form	e-payment	e-services	Full integration across organization	Website link
1	Saudi eGovernment national portal	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓	×	×	×	×	http://www.saudi.gov.sa
2	Communication and Information Technology Commission	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓	✓	×	×	×	http://www.citc.gov.sa
3	Saudi Telecom	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓	✓	✓	×	✓	×	http://www.stc.com.sa
4	Ministry of Agriculture	✓	✓	✓	×	×	✓	✓	×	✓	×	✓	✓	✓	×	✓	×	http://www.moa.gov.sa
5	Ministry of Civil Service	✓	✓	✓	×	×	✓	✓	✓	×	×	✓	✓	✓	×	✓	×	http://www.mcs.gov.sa
6	Ministry of Commerce and Industry	✓	✓	✓	×	×	✓	✓	×	×	×	✓	✓	✓	×	✓	×	http://www.mci.gov.sa
7	Council of Saudi Chambers of Commerce and Industry	✓	✓	✓	×	×	✓	✓	✓	×	×	✓	✓	✓	×	✓	×	http://www.saudichambers.org.sa
8	Ministry of Defense and Aviation	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓	×	http://www.gaca.gov.sa
9	Ministry of Water and Electricity	✓	✓	✓	×	×	✓	✓	×	×	×	✓	✓	✓	×	✓	×	http://www.mowe.gov.sa/
10	Saudi Ports Authority	✓	✓	✓	×	×	✓	✓	×	×	×	✓	×	×	×	×	×	http://www.ports.gov.sa
11	Ministry of Interior Passport Authority	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓	✓	×	×	×	×	http://www.gdp.gov.sa
12	Ministry of Health	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	×	✓	×	http://www.moh.gov.sa
13	Ministry of Education	✓	✓	✓	×	×	✓	✓	×	×	×	✓	✓	×	×	×	×	http://www.moe.gov.sa
14	Ministry of Hajj	✓	×	✓	×	✓	✓	×	×	×	✓	✓	✓	×	×	×	×	No official website
15	Ministry of Foreign Affairs	✓	✓	✓	×	✓	✓	✓	✓	×	×	✓	✓	✓	×	✓	×	http://www.mofa.gov.sa
16	Ministry of Higher Education	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	http://www.mohe.gov.sa
17	Ministry of Communications and Information Technology (IT)	✓	✓	✓	×	×	✓	✓	×	×	✓	✓	✓	×	×	×	×	http://www.mcit.gov.sa
18	Ministry of Finance	✓	✓	✓	×	✓	✓	✓	×	×	×	✓	✓	×	×	×	×	http://www.mof.gov.sa
19	Ministry of Justice	✓	×	×	×	×	✓	✓	✓	×	✓	✓	✓	✓	×	×	×	http://www.moj.gov.sa
20	Ministry of Labor	✓	✓	✓	×	✓	✓	✓	✓	×	×	✓	×	✓	×	✓	×	http://www.mol.gov.sa
21	Ministry of Municipalities and Rural Affairs	✓	✓	✓	×	✓	✓	✓	×	×	×	✓	✓	✓	×	✓	×	http://www.momra.gov.sa/
22	Ministry of Petroleum and Mineral Resources	✓	×	×	×	×	✓	✓	×	×	×	×	✓	×	×	×	×	http://www.mopm.gov.sa
23	Ministry of Economy and Planning	✓	✓	×	×	✓	✓	✓	×	✓	✓	✓	×	×	×	×	×	http://www.mep.gov.sa
24	Ministry of Islamic Affairs, Endowment, Dawa and Guidance	✓	✓	×	×	×	✓	✓	✓	×	✓	✓	✓	✓	×	✓	×	http://www.moia.gov.sa
25	Ministry of Social Affairs	✓	✓	✓	×	✓	✓	✓	×	×	✓	✓	✓	×	×	×	×	http://www.mosa.gov.sa
26	Ministry of Transport	✓	✓	✓	×	×	✓	✓	×	×	✓	✓	✓	×	×	×	×	http://www.mot.gov.sa
27	Ministry of Interior	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓	×	http://www.moi.gov.sa
28	General Presidency of Youth Welfare	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	http://www.gpyw.gov.sa No access

Analysis of *k*-Coverage in Wireless Sensor Networks

Rasmi Ranjan Patra
Dept. of Mathematics,
Utkal University,Vani Vihar
Bhubaneswar, Odisha, India

Prashanta Kumar Patra
College of Engineering and Technology,
Bhubaneswar, Biju Patnaik University of
Technology, Odisha, India

Abstract— **Recently, a concept of wireless sensor networks has attracted much attention due to its wide-range of potential applications. Wireless sensor networks also pose a number of challenging optimization problems. One of the fundamental problems in sensor networks is the coverage problem, which reflects the quality of service that can be provided by a particular sensor network. The coverage concept is depending from several points of view due to a variety of sensors and a wide-range of their applications. One fundamental issue in sensor networks is the *coverage* problem, which reflects how well a sensor network is monitored or tracked by sensors. In this paper, we formulate this problem as a decision problem, whose goal is to determine the degree of coverage of a sensor network, which is covered by at least *k* sensors, where *k* is a predefined value. The sensing ranges of sensors can be same or different. Performance evaluation of our protocol indicates that degree of coverage of wireless sensor networks can be determined within small period of time. Therefore energy consumption of the sensor networks can be minimized.**

Keywords- Wireless sensor networks; coverage; k-coverage; connectivity.

I. Introduction

In computer networking there is a great value of wireless networking because it has no difficult installation, no more expenditure and has lot of way to save money band time. In the field of wireless networking there is another form of networking which is called as wireless sensor network. A type of wireless networking which is comprised on number of numerous sensors and they are interlinked or connected with each other for performing the same function collectively or cooperatively for the sake of checking and balancing the environmental factors. This type of networking is called as Wireless sensor networking. Basically wireless sensor networking is used for monitoring the physical conditions such as weather conditions, regularity of temperature, different kinds of vibrations and also deals in the field of technology related to sound.

Total working of wireless sensor networking is based on its construction. Sensor network initially consists of small or large nodes called as sensor nodes. These nodes are varying in size and totally depend on the size because different sizes of sensor nodes work efficiently in different fields. Wireless sensor networking have such sensor nodes which are specially designed in such a typical way that they have a microcontroller which controls the monitoring, a radio transceiver for generating radio waves, different type of wireless communicating devices and also equipped with an energy source such as battery. The entire network worked simultaneously by using different dimensions of sensors and worked on the phenomenon of multi routing algorithm which is also termed as wireless ad hoc networking.

There are mainly three types of coverage problem like Area Coverage, Point Coverage, Barrier Coverage .In Area of coverage, the main objective of the sensor network is to cover (monitor) an area (also referred sometimes as region). In the point coverage problem, the objective is to cover a set of points. There are two types of coverage approach (a) random point coverage (b) deterministic point Coverage. In barrier Coverage minimize the probability of undetected penetration through the barrier (sensor network).

The main goal of this idea is to determine the degree of coverage of an area covered by two or more sensors & then find out the lowest degree of coverage.

II. Related Work

Wireless sensor networks (WSNs) have attracted a great deal of research attention due to their wide-range of potential applications. A WSN provides a new class of computer systems and expands people's ability to remotely interact with the physical world. In a broad sense, WSNs will transform the way we manage our homes, factories, and environment. Applications of WSNs [1] include battlefield surveillance, biological detection, home appliance, smart spaces, and inventory tracking.

Sensors in a network can cooperatively gather information from an interest region of observation and transmit this collected information to a base station. There are two types of data sent to the base station: 1. event-driven and 2. On-demand. In the former case, the data is sent to the base station when one or more sensors detect an event in the vicinity. In the latter case, the data is sent from the sensors to the base station based on an explicit request.

An important problem addressed in literature is the sensor coverage problem. This problem is centered around a fundamental question: "*How well do the sensors observe the physical space ?*" As pointed out in [2], the coverage concept is a measure of the quality of service (QoS) of the sensing function and is subject to a wide range of interpretations due to a large variety of sensors and applications. The goal is to have each location in the physical space of interest within the sensing range of at least one sensor.

Since sensors may be spread in an arbitrary manner, one of the fundamental issues in a wireless sensor network is the *coverage problem*. In general, this reflects how well an area is monitored or tracked by sensors. In the literature, this problem has been formulated in various ways. For example, the *Art Gallery Problem* is to determine the number of observers necessary to cover an art gallery (i.e., the service area of the sensor network) such that every point in the art gallery is monitored by at least one observer. This problem can be solved optimally in a 2D plane, but is shown to be NP-hard when extended to a 3D space [3].

Reference [4] defines a sensor coverage metric called *surveillance* that can be used as a measurement of quality of service provided by a particular sensor network, and centralized optimum algorithms that take polynomial time are proposed to evaluate paths that are best and least monitored in the sensor network. The work further investigates the problem of how well a target can be monitored over a time period. While it moves along an arbitrary path with an arbitrary velocity in a sensor network. Localized exposure-based coverage and location discovery algorithms are proposed in [5].

On the other hand, some works are targeted at particular applications, but the central idea is still related to the coverage issue. For example, sensors' on-duty time should be properly scheduled to conserve energy. Since sensors are arbitrarily distributed, if some nodes share the common sensing region and task, then we can turn off some of them to conserve energy and thus extend the lifetime of the network. This is feasible if turning off some nodes still provide the same "coverage" (i.e., the provided coverage is not affected).Author in [6] proposes a heuristic to select mutually exclusive sets of sensor nodes such that each set of sensors can provide a complete coverage the monitored area. Author in [7] proposes a probe-based density control algorithm to put some nodes in a sensor-dense area to a doze mode to ensure a long-lived, robust sensing coverage. A coverage preserving node scheduling scheme is presented in [8] to determine when a node can be turned off and when it should be rescheduled to become active again.

In this work, we consider a more general sensor coverage problem. Given a set of sensors deployed in a target area, we want to determine if the area is sufficiently *k-covered,* in the sense that every point in the target area is covered by at least k sensors, where k is a predefined constant. As a result, the aforementioned works [6,7] can be regarded as a special case of this problem with $k = 1$. Applications requiring $k > 1$ may occur in situations where the stronger environmental monitoring is necessary, such as military applications. It also happens when multiple sensors are required to detect an event.

For example, the triangulation-based positioning protocols [3,4, 7] require at least three sensors (i.e., $k \geq 3$) at any moment to monitor a moving object. Enforcing $k \geq 2$ is also necessary for fault-tolerant purpose. In paper [9], a novel solution is proposed to determine whether a sensor network is k-covered. The sensing range of each sensor can be a unit disk or a non-unit disk. The solution can be easily translated to a distributed protocol where each sensor only needs to collect local information to make its decision. Instead of determining the coverage of each location, our approach tries to look at how the perimeter of each sensor's sensing range is covered, thus leading to an efficient polynomial time algorithm. As long as the perimeters of sensors are sufficiently covered, the whole area is sufficiently covered. In this paper, we propose a simple solution to determine the degree of coverage of a sensor network. In paper [9], authors consider the perimeter only but not determine it in a mathematical way. So, in this paper we consider the intersection area and calculate in set theory method and also calculate the area of intersection geometrically. We consider for same and different sensing range of a sensor. After finding the degree of coverage, it is easy to find out which node always be active

III. SYSTEM MODEL

A. Definitions & Notations

1) Sensing Range :

Within which range a sensor can sense a particular area.

Suppose an area A is covered by a sensor S, when it is covered within the sensing range of s.

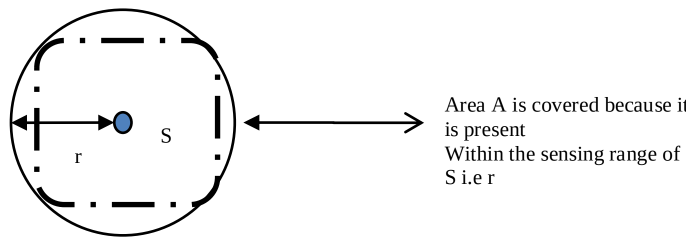


Figure 1. Area within sensing range.

2) Communication Range :

Within which range a sensor can communicate with another.

3) Degree of Coverage

When an area is covered by a sensor s, then the degree of coverage of particular that area is one because it is covered within the sensing range of only one sensor**.**

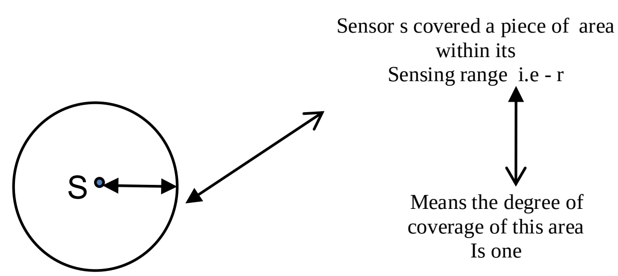
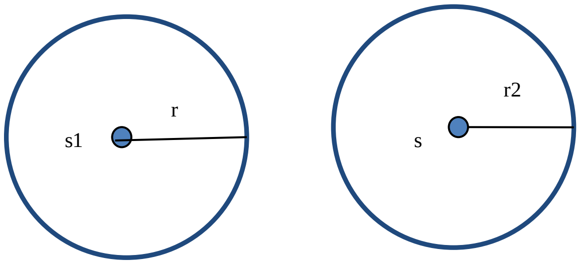


Figure 2. Sensor S within it's sensing range.

B. Condition for Intersection

1) Consider two sensors S1 and S2. Both two sensors are intersect with each other when sum of radii is less than and equal to distance between centers.



Both are intersect when Sum of radii <= Distance between centers

Figure 3. Condition for intersection of two sensors

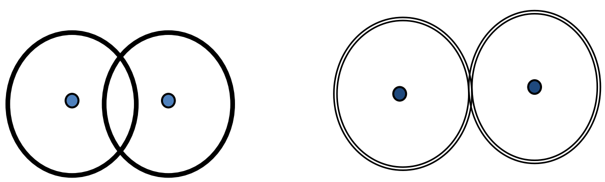
$$r1+r2 <= \sqrt{((x2-x1)^2+(y2-y1)^2)} \quad \textbf{(1)}$$

This idea also states that how to know one sensor should be always active or go to sleep mode.

2) Based on condition for intersection some special case arises:-

$$r1+r2 <= \sqrt{((x2-x1)^2+(y2-y1)^2)}$$

$\downarrow$ => D(s1,s2)



Case 1 Case 2

Figure 4. Two sitution for two sensors for intersection

a) Some Special Cases for Intersection

- When two sensor sensing range separate with each other

=> d(s1,s2) > r1+r2 means A∩B =Ø

Figure 5. Two separated sensors

b) When one sensor present within another sensor sensing range

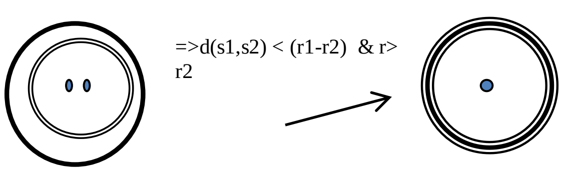
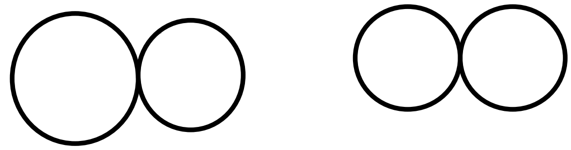


Figure 6. Overlapped sensors.

c) When two sensor only touched means intersect with each other without creating any area:



=> d(s1,s2) = r1+r2

Figure 7. Touched sensors.

IV. Determination of *K*-Coverage

A. Determination of Area of Intersection

1) When two sensors sensing range intersect with each other creating any area.

- **Case1**

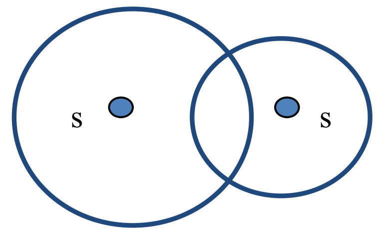
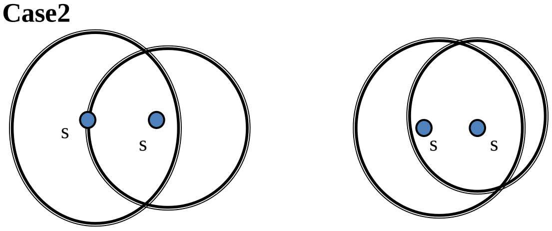


Figure 8. Two intesect sensors

=> r1-r2 < d(s1,s2) < r1+r2 Means A∩B ≠ Ø

(2)

- **Case2**



=> r1-r2 < d(s1,s2) < r1+r2

Figure 9. Two intesect sensors

2)*Find the area of intersection between two circles*

$$x^2 + (y-1)^2 = 1 \quad (0,1)$$

$$(x-1)^2 + y^2 = 1 \quad (1,0) \quad \textbf{(3)}$$

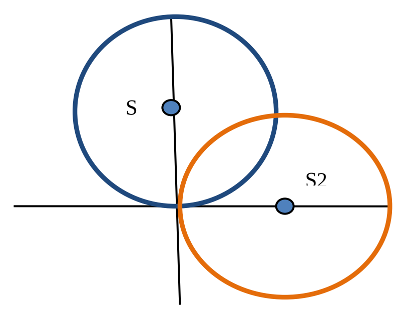


Figure 10. Two intesect sensors

3) *We can see that if we draw the line y=x on the graph, we will split the intersection area into two equal pieces.*

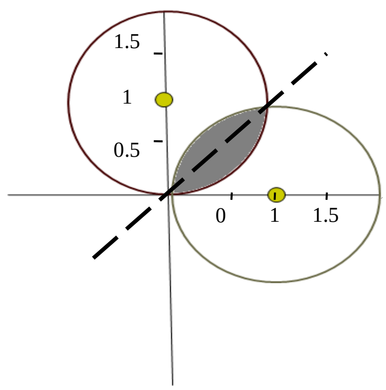


Figure 11. Two intesect sensors

Now, notice that we can form a triangle in the second circle, from the dashed line to the center at (1,1). It will be a 45º-45º-90º right triangle.

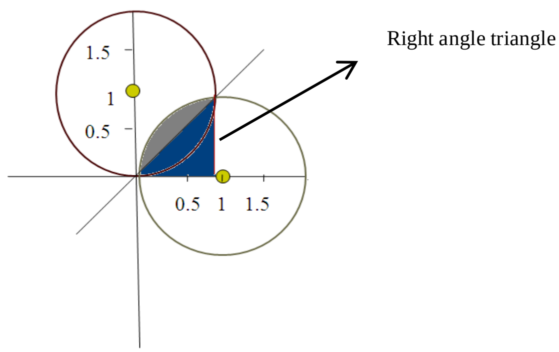


Figure 12. Two intesect sensors

We know the area of the sector of the circle is π/4 since we are dealing with a quarter of the circle with radius 1 and we know that the area of the triangle is 1/2.

(Area of the triangle is 1/2 b* h

=> ½*1*1 = ½)

The area of the gray shaded region is

=> π/4-1/2= π-2/4 => It is the half of the intersection area only

The area of entire intersection will be

=> 2* π-2/4

=> π-2/2

We found the area of intersection in the case the radii for both circles was equal to 1.

In the more general case in which both circles have radius *r (*and have their centers shifted accordingly), then $(\pi\text{-}2/2)r^2$, this formula is compute the intersection area

B. Determination Degree of Coverage

1) When two sensors overlap each other then Degree of coverage of A and B is one by the defination of coverage.

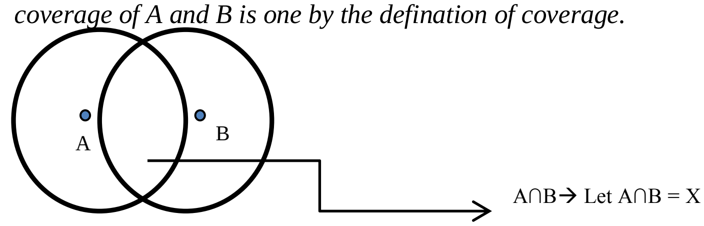


Figure 13. Two Intesect Sensors

Equation

(Let A∩B = X)

AUB= |A-B|+ |B-A|+|A∩B|

=1-X+1-X+X

=>2-2X+X=|AUB|

=>2-X=AUB

=>2-AUB= X

=>2-AUB=A∩B

2) Suppose If one sensor covered a area then degree of coverage is one then, if two sensor covered a area then degree of coverage is of this area is 2.

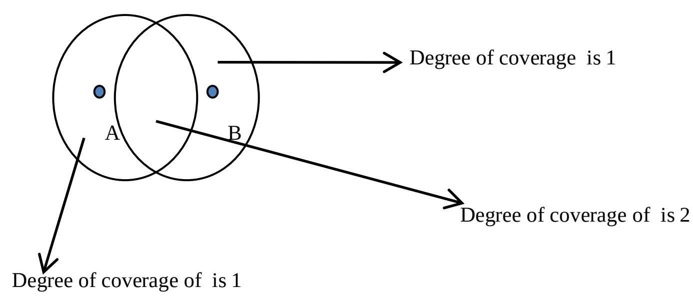


Figure 14. Two Intesect Sensors

Then |AUB|=|A|+|B|-|A∩B|

=1+1-2

=0

|AUB |= 0

Then put |AUB|=0 In equation 1 then we get→ |A∩B|=2

So, according to above contradiction method it shows that degree of coverage of intersection area of two sensor sensing range is 2.

3)*For degree of Coverage*

Lemma-1

If two sensors intersect with each other, then degree of coverage of intersection area is two according to the 4(i) equation. Likewise if 3 sensor intersect then, degree of coverage of that intersection area is three. Same way how many sensors intersect with each other is equal to the degree of coverage of that intersects area.

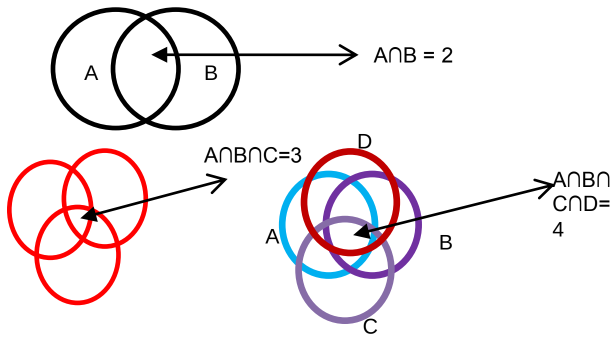


Figure 15. Multiple Intesect Sensors

V. PERFORMANCE EVALUATION

The four graphs devloped by MATLAB 7.4.0 after putting the equation used in the algorithms.As shown in figure 16, the result of average degree of coverage with different sensing ranges.As shown in figure 17,energy consumption for different nodes when degree of coverage is different.As shown in figure 18. Coverage Detection Time in millisecond for Different Nodes (100 to 1000) for different Sensing Range. As shown by figure 19, Find out the coverage detection time with different nodes with different sensing ranges. Degree of coverage for different nodes when Sensing Range equal to Communication Range, Communication Range twice of Sensing range and Communication Range greater than double of sensing range.

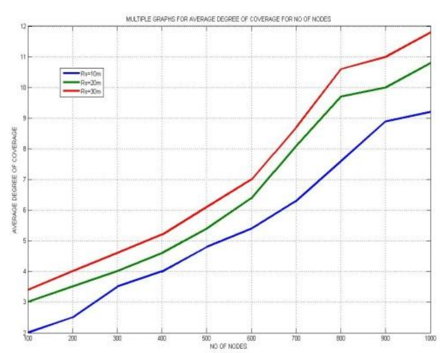


Figure 16. Average Degree of Coverage with different nodes(100 to 1000) with Various Sensing Ranges(10M,20M,30M).

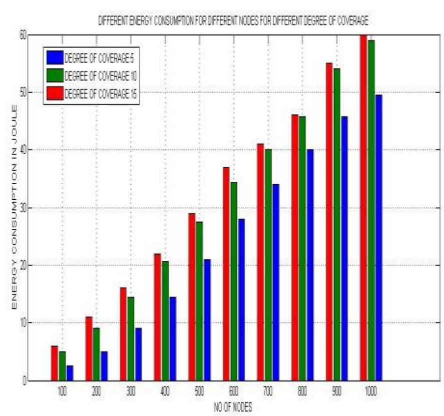


Figure 17. Energy Consumption for different nodes (100 to 1000) for different Degree of Coverage.

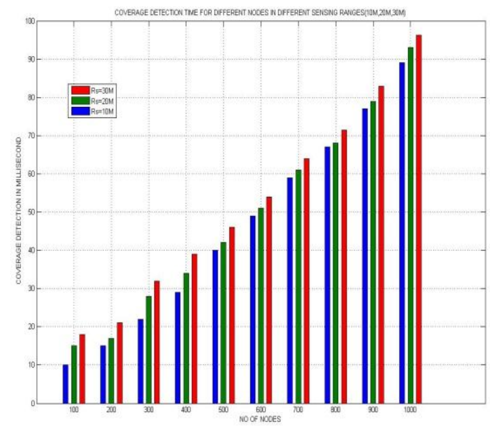


Figure 18.Coverage Detection Time for Different Nodes (100 to 1000) for different Sensing Range.

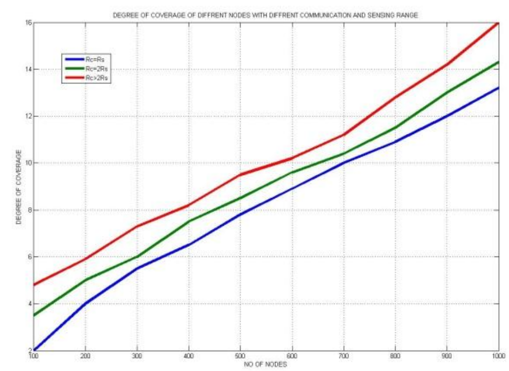


Figure 19.Degree of coverage for different Nodes when Communication Range equal to Sensing range, Communication Range equal to Double of sensing range and Communication Range Greater than double of sensing range.

VI. CONCLUSION AND FUTURE SCOPE

In this paper, we have proposed a solution to find out the degree of coverage in a sensor network with irrespective of same and different sensing range. We consider the intersection area & try to find out in a mathematical way using set theory method. With the proposed techniques, we also discuss some applications like whether a node goes to sleep or active state.

Our proposed model is very simple and efficient. This paper is proposed for easily finding degree of coverage the However, time complexity calculation & simulation are yet to be done in order to prove the efficiency of the protocol.

REFERENCES

[1] G. T. Huang, “Casting the Wireless Sensor Net,” Technology Review, pp 50-56, July/August 2003.

[2] D. Tian and N. D. Georganas, A coverage-preserving Node scheduling scheme for large wireless sensor networks. In ACM Int'l Workshop on Wireless Sensor Networks and Applications (WSNA), 2002.

[3] J. O'Rourke, Computational geometry column 15. Int'l Journal of Computational Geometry and Applications, Pp 215–217, 1992.

[4] S. Meguerdichian,F. Koushanfar, M. Potkonjak, and M.B.Srivastava. “ Coverage problems in wireless ad-hoc sensor Networks”. In IEEE INFOCOM, pp 1380–1387, 2001.

[5] S. Meguerdichian, S. Slijepcevic, V. Karayan, and M. Potkonjak. "Localized algorithms in wireless ad-hoc networks: location discovery and sensor exposure." In ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing (MobiHOC), pp 106–116, 2001.

[6] S.Slijepcevic and M.Potkonjak. Power efficient organization of wireless sensor networks. In IEEE Int'l Conf. on Communications (ICC), pp 472–476, 2001.

[7] F. Ye, G. Zhong, S. Lu, and L. Zhang, "PEAS: A robust energy conserving protocol for long-lived sensor networks". In Int'l Conf. on Distributed Computing Systems (ICDCS), 2003.

[8] D. Tian and N. D. Georganas. "A coverage-preserving node scheduling scheme for large wireless sensor networks". In ACM Int'l Workshop on Wireless Sensor Networks and Applications (WSNA), 2002.

[9] Chi-Fu Huang and Yu-Chee Tseng . "The coverage problem in a wireless sensor network". Department of Computer Science and Information Engineering National Chiao-Tung University Hsin-Chu, 30050, Taiwan

AUTHORS PROFILE

Rasmi Ranjan Patra received Master In Computer Application With 1st Class With distinction from O.U.A.T, Odisha, India in 2001, M.Tech in Computer Science and Technology from C.E.T , Bhubaneswar ,India in 2010 and currently pursuing Ph.D Degree in Utkal University, India .He is working as Assistant professor in Department of Computer Science and Application under Orissa University of Agriculture and Technology(O.U.A.T).He has Published many papers at national /international Journals and Conferences in the areas of Sensor Network. Mr.Patra has authorize one book in Computer Science area.

Prashant Kumar Patra received Bachelor Degree in Electronics Engineering with 1st Class Distinction from SVRCET (NIT), Surat, India, M.Tech. Degree in Computer Engineering from I.I.T., Kharagpur and Ph. D. in Computer Science from Utkal University, India in the year 1986, 1993 and 2003 respectively. He is presently working as Professor in the Department of Computer Science & Engineering, College of Engineering & Technology, Bhubaneswar, a constituent college of Biju Patnaik University of Technology, Orissa, India. He has published many papers at National/International journals and Conferences in the areas of Soft computing, Image processing and Pattern recognition. Dr. Patra is a recipient of J.C. Bose award by IETE, India.

PSO Based Short-Term Hydrothermal Scheduling with Prohibited Discharge Zones

G.Sreenivasan
Associate Professor,
Dept. of Electrical Engineering,
Intell Engineering College,
Anantapur, AP, India.

Dr. C.H.Saibabu,
Professor,
Dept. of Electrical Engineering,
JNTU Kakinada, Kakinada,
AP, India

Dr.S.Sivanagaraju
Associate Professor,
Dept. of Electrical Engineering,
JNTU Kakinada, Kakinada,
AP, India.

Abstract— **This paper presents a new approach to determine the optimal hourly schedule of power generation in a hydrothermal power system using PSO technique.. The simulation results reveal that the proposed PSO approach appears to be the powerful in terms of convergence speed, computational time and minimum fuel cost.**

Keywords- Hydrothermal scheduling; Particle Swarm Optimization; Valve - point loading effect; Prohibited Discharge Zones.

I. INTRODUCTION

The optimal scheduling of generation in a hydrothermal system involves the allocation of generation among the hydroelectric and thermal plants so as to minimize the total operation costs of thermal plants while satisfying the various constraints on the hydraulic and power system network. In Short-term scheduling it is normally assumed that the largest dam levels at the end of the scheduling period have been set a medium term scheduling process that takes into account longer term river inflow modeling and load predictions. The short term scheduler than allocates this water (Power) to the various time intervals in an effort to minimize thermal generation costs while attempting to satisfy the various unit and reservoir constraints.

The main constraints include the time coupling effect of the hydro sub problem, where the water flow in an earlier time intervals affects the discharge capability at a later period of time, the time varying system long demand, the cascade nature of the hydraulic network, the varying hourly reservoir inflows, the physical limitations on the reservoir storage and turbine flow rate and loading limits of both thermal and hydro plants. Further constraints could be depending on the particular requirements of a given power system, such as the need to satisfy activities including, flood control, irrigation, fishing, water supply etc., The hydrothermal scheduling problem has been the subject of intensive investigation for several decades now.

Most of the methods that have been used to solve the hydrothermal co-ordination problem make a number of simplifying assumptions in order to make the optimization problem more tractable.

The performances of different stochastic techniques have been studied in the literature [6-14]. Though stochastic techniques have been proved to be very efficient and having faster performances than the conventional methods, there are some limitations in the goodness of the solutions to the problem that are obtained in [13]. From the literature it is found that particle swarm optimization technique has the fastest convergence rate to the global solution amongst all algorithms and has highest potential of finding more nearly global solutions to hydrothermal co-ordination problems [13]. Early works on PSO have shown the rich promise of emergence of a relatively simple optimization technique this is easier to understand compared to other evolutionary computation techniques presently available eg. Genetic algorithm and evolutionary programming.

Another advantage of PSO can be the possibility of tuning smaller number of free, tunable parameters to arrive at the desired goal. The PSO technique has been applied to various fields of power system optimization. Yu et al applied PSO technique to solve short-term hydrothermal scheduling [16] with an equivalent thermal unit having smooth cost functions connected to hydel systems. Here the constraints were handled by penalty function method [16]. But the performance of PSO to Short-term hydrothermal scheduling for interconnected individual thermal units with non-smooth cost function has not been tested yet.

In this paper PSO method is proposed for short-term optimal scheduling of generation in a hydrothermal system which involves the allocation of generation among the multi-reservoirs cascaded hydro plants and thermal plants with prohibited discharge zones and valve point loading effects so as to minimize the fuel cost of equivalent thermal plant while satisfying the various constraints on the hydraulic and power system network.

To validate the PSO based hydrothermal scheduling algorithm, the developed algorithm has been illustrated for a test system [11]. The same problem has been solved by GA and the results are compared. The performance of the proposed method is found to be quite encouraging as compared with other methods.

II. PROBLEM STATEMENT

NOMENCLATURE :

C	Composite Cost function

C_i fuel cost of i^{th} thermal unit
P_{GTjM} Output power of i^{th} thermal unit at time 'm'
P_{GHjM} Output power of j^{th} hydro unit at time 'm'
P_{GTi}^{min}, P_{GTi}^{max} Lower and Upper generation limits for i^{th} thermal units
a_i, b_i, c_i and d_i, e_i –cost curve co - efficients of i^{th} thermal unit
P_{Dm} Load demand at time 'm'
P_{GHj}^{min}, P_{GHj}^{max} Lower and Upper generation limits for i^{th} thermal unit
Q_{Hjm} Water discharge rate of i^{th} reservoir at time 'm'
V_{HJm}, Storage volume of j^{th} reservoir at time 'm'
Q_{HJ}^{min}, Q_{HJ}^{max} Minimum and Maximum water discharge rate of j^{th} reservoir
V_{HJ}^{min}, V_{HJ}^{max} Minimum and Maximum Storage volume of j^{th} reservoir
P_{Lm} Total transmission line losses at time 'm'
C_{1j}, C_{2j}, C_{3j}, C_{4j}, C_{5j}, C_{6j}, Power generation co - efficients of j^{th} hydro unit
I_{Hjm} Inflow rate of j^{th} reservoir at time 'm'
S_{Hjm} Spillage of j^{th} reservoir at time 'm'
T_{lj} Water transport delay from reservoir l to j
R_{uj} Set of upstream units directly above j^{th} hydro plant
N_{GT} Number of thermal generating units
N_{GH} Number of hydro generating units
m, M Time index, scheduling period
V_{Hj}^{begin} Initial storage volume of j^{th} Reservoir
V_{Hj}^{end} Final storage volume of j^{th} Reservoir

III. MATHEMATICAL FORMULATION

Hydrothermal Scheduling involves the optimization of a problem with a non-linear objective function, with a mixture of linear, non-linear and dynamic network flow constraints. The problem difficulty is compounded by a number of practical considerations and unless several simplifying assumption are made, this problem is difficult to solve for practical power systems as shown in fig 1.

Due to Zero incremental cost of hydro generating units, the prime objective of the short-term hydrothermal scheduling problem becomes to minimize the fuel cost of thermal plants, while making use of the availability of hydropower as much as possible, such that the load demands P_D supplied from hydro plants and a thermal plant in the intervals of the generation scheduling horizon can be met and simultaneously, all the equality and inequality operation constraints are satisfied.

The objective function and associated constraints of the Hydrothermal scheduling problem are formulated as follows.

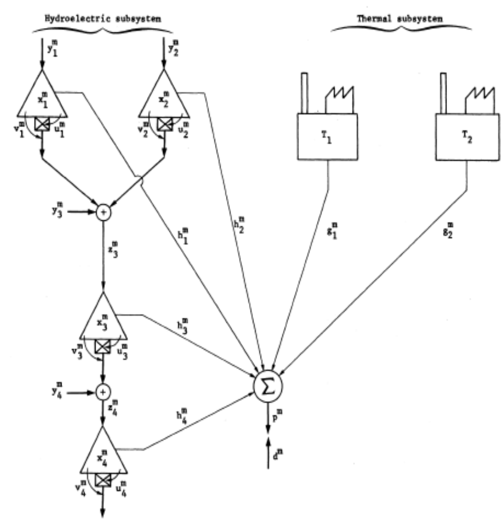


Fig.1 Practical Hydrothermal Power system Network

A. Objective Function

The total fuel cost for running the thermal system to meet the load demand in scheduling horizon is given by C. The objective function is expressed mathematically, as

$$\text{Minimize} \quad C = \sum_{i=1}^{N_{GT}} C_i\,(P_{GTi}) \tag{1}$$

When considering valve-point effects, the fuel cost function of each thermal generating unit is expressed as the sum of a quadratic and a sinusoidal function. The total fuel cost in-terms of real power output can be expressed as :

$$C = \sum_{m=1}^{M}\sum_{i=1}^{N_{GT}}\Big[a_i + bi\,P_{GTim} + c_i P_{GTim}^{\ 2} + \left|d_i \sin\left\{e_i\left(P_{GTi}^{\min} - P_{GTim}\right)\right\}\right|\Big] \tag{2}$$

subject to a number of unit and power system network constraints.

B. constraints

This non-linear constrained hydrothermal scheduling optimization problem is subjected to a variety of constraints depending upon practical implications like the varying system load demand, the time coupling effect of hydro subsystem, the cascading nature of the hydraulic network, the time varying hourly reservoir inflows, thermal plant and hydro plant operating limits, system losses, reservoir storage limits, water discharge rate limits, hydraulic continuity constraints and initial and final reservoir storage limits. These constraints are discussed below.

1) Power balance constraints(Demand Constraints)

This constraint is based on the principle of equilibrium between the total active power generation from the hydro and thermal plants and the total system demand plus the system losses in each time interval of scheduling 'm'

$$\sum_{i=1}^{N_{GT}} P_{GTim} P_{im} + \sum_{j=1}^{N_{GH}} P_{GHjm} = P_{Dm} + P_{lossm} \quad m \in M \qquad (3)$$

2) Thermal Generator Constraints

The operating limit of equivalent thermal generator has a lower and upper bound so that it lies in between these bounds.

$$P_{GT}^{min} \le P_{GTim} \le P_{Gi}^{max}, \quad m \in M \qquad (4)$$

3) Hydro Generator Constraints

The operating limit of hydro plant must lie in between its upper and lower bounds.

$$P_{GHj}^{min} \le P_{GHjm} \le P_{GHj}^{max}, \quad j \in N_{GH}, m \in M \qquad (5)$$

HYDRAULIC NETWORK CONSTRAINTS

The hydraulic operational constraints comprise the water balance (Continuity) equations for each hydro unit (System) as well as the bounds on reservoir storage and release targets.

These bounds are determined by the physical reservoir and plant limitations as well as the multipurpose requirements of the hydro system. These constraints include :

1) Reservoir Capacity Constraints

The operating volume of reservoir storage limit must lie in between the minimum and maximum capacity limits.

$$V_{Hj}^{min} \le V_{Hjm} \le V_{Hj}^{max}, \quad j \in N_H, m \in M \qquad (6)$$

2) The Water Discharge Constraints

The variable net head operation is considered and the physical limitation of water discharge of turbine, Q_{Hjm}, Must lie in between maximum and minimum operating limits, as given by

$$Q_{Hj}^{min} \le Q_{Hjm} \le Q_{Hj}^{max}, j \in N_H, \quad m \in M \qquad (7)$$

3) Reservoir end conditions

The desired volume of water to be discharged by each reservoir over the scheduling period,

$$V_{Hjm}\Big|_{m=0} = V_{Hj}^{begin}$$

$$V_{Hjm}\Big|_{m=m} = V_{Hj}^{end} \qquad j \in N_H \qquad (8)$$

4) Hydraulic Continuity Equation Constraint

The storage reservoir volume limits are expressed with given initial and final volumes as

$$V_{HjmH} = V_{Hjm} + \sum_{u=1}^{Ru} \left[Q_{Hu}(m-\tau_{lj}) + S_u(m-\tau_{lj}) \right] - Q_{Hj\,(m+1)} - S_{j\,(m+1)} + \gamma_{j(m+1)} \text{ for } j \in N_H, m \in M \qquad (9)$$

Where $\tau_{lj}$is the water delay time between reservoir *l* and its upstream u at interval' m'.

$R_u$is the set of upstream units directly above the hydro plant 'j'.

5) Power Generation Characteristics

The Power generated from a hydro plant is related to the reservoir characteristics as well as the water discharge rate. A number of models have been used to represent this relationship. In general, the hydro generator power output is a function of the net hydraulic head, H, reservoir volume, V_H, and the rate of water discharge, Q_H,

$$P_{GHjm} = f(Q_{Hjm}, V_{Hjm}) \text{ and } V_{Hjm} = f(H_{jm}) \qquad (10)$$

The model can also be written in-terms of reservoir volume instead of the reservoir net head, and a frequently used functional is

$$P_{GHjm} = C_{1j}V^2_{Hjm} + C_{2j}Q^2_{Hjm} + C_{3j}V_{Hjm}Q_{Hjm} + C_{4j}V_{Hjm} + C_{5j}Q_{Hjm} + C_{6j}, j \in N_H, m \in M \quad (11)$$

Net head variation can only be ignored for relatively large reservoirs, in which case power generation is solely dependent on the water discharge. In setting the generation levels of the thermal plants, a quadratic cost function is used to model the fuel input power output characteristic of thermal units.

IV. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization is one of the most recent developments in the category of combinatorial metaheuristic optimizations. This method has been developed under the scope of artificial life where PSO is inspired by the natural phenomenon of fish schooling or bird flocking. PSO is basically based on the fact that in quest of reaching the optimum solution in a multi-dimensional space, a population of particles is created whose present coordinate determines the cost function to be minimized. After each iteration the new velocity and hence the new position of each particle is updated on the basis of a summated influence of each particle's present velocity, distance of the particle from its own best performance, achieve so far during the search process and the distance of the particle from the leading particle, i.e. the particle which at present is globally the best particle producing till now the best performance i.e. minimum of the cost function achieved so far.

Let x and v denote a particle position and its corresponding velocity in a search space, respectively. Therefore, the i^{th} particle is represented as $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$ in the 'd' dimensional space. The best previous position of the i^{th} particles recorded and represented as $pbest_i = (pbest_{i1}, pbest_{i2}, \ldots, pbest_{id})$. The index of the best particle among all the particles in the group is represented by the $gbest_d$. The rate of the velocity for the particle i is represented as $v_i=(v_{i1}, v_{i2}, \ldots, v_{id})$.

The modified velocity and position of each particle can be calculated using the current velocity and the distance from $pbest_{id}$ to $gbest_d$ as shown in the following formulas:

$$v_{id}^{k+1}=w \times v_{id}^{k}+c_1 \times rand(\ ) \times (pbest- x_{id}^{k})+c_2 \times rand(\ ) \times (gbest_d-x_{id}^{k}) \quad (12)$$

$$x_{id}^{k+1}=x_{id}^{k}+ v_{id}^{k+1}$$
$$i=1,2,\ldots,N_p, d=1,2,\ldots,N_g \quad (13)$$

where, N_P is the number of particles in a group, Ng the number of members in a particle, k the pointer of iterations, w the inertia weight factor, C_1, C_2 the acceleration constant, rand()the uniform random value in the range [0,1], v_i^k the velocity ofa particle i at iteration k, $v_d^{min} \le v_{id}^k \le v_d^{max}$ and $x_i^k$is the current position of a particle i at iteration k.In the above procedures, the parameter v^{max} determined the resolution, with which regions are to be searched between the present position and the target position.

If v^{max} is too high, articles might fly past good solutions. If v^{max} is too small, particles may not explore sufficiently beyond local solutions. The constants C_1 and C_2 represent the weighting of the stochastic acceleration terms that pull each particle toward the pbest and gbest positions. Low values allow particle to roam far from the target regions before being tugged back. On the other hand, high values result in abrupt movement toward or past, target regions. Hence, the acceleration constants C_1 and C_2were often set to be 2.0 according to past experiences. Suitable selection of inertia weight 'w' provides a balance between global and local explorations, thus requiring less iteration on average to find a sufficiently optimal solution.

As originally developed,' w 'often decreases linearly from about 0.3to -0.2 during a run. In general, the inertia weight w is set according to the following equation:

$$w = w_{max}-\frac{w_{max} - w_{min}}{iter_{max}} \times iter \quad (14)$$

where $iter_{max}$ is the maximum number of iterations and 'iter' is the current number of iterations.

V. PSO Based Hydrothermal Scheduling

Taking the number of particles to be N, the no. of Scheduling intervals as m and the number of hydro unit, as N_H, each initial trial vector Q (j, m, p) denoting the particles of population to be evolved for P = 1, 2, …. N is selected. The discharge of j^{th} hydro plant at m^{th} interval is randomly generated as $Q_{GHjm} \sim \upsilon\ (Q_{GHj}^{min}, Q_{GHj}^{max})$

Let $P_K = [P_{GT1}, P_{GT2}, \ldots P_{GTi}, \ldots PG_{TNT}, Q_{GH}, Q_{GH1}, Q_{GH2}, \ldots Q_{GHj} \ldots, Q_{GHNH}]^T$ be a trail matrix designating the K^{th} individual of population to be evolved and

$$P_{GTi} = [P_{GTi1}, P_{GTi2}, \ldots P_{GTim}, \ldots P_{GTiM}],$$
$$Q_{GHj} = [Q_{GHj1}, Q_{GHj2}, \ldots Q_{GHjm}, \ldots Q_{GHjM}]$$

The elements P_{GTim} and Q_{GHjm} are the power output of the i^{th} thermal unit and the discharge rate of the j^{th} hydro plant at time interval m. The range of elements P_{GTim} and Q_{GHjm} should satisfy the thermal generating capacity and the water discharge rate constraints in equations (3) and (7) respectively.

Assuming the spillage in Eq (9) to be zero for simplicity the hydraulic continuity constraints are

$$V_{Hjo} - V_{HjM} = \sum_{m=1}^{M} Q_{GHjm} - \sum_{m=1}^{M} \sum_{l=1}^{R_{uj}} Q_{GHl}(m-\tau ij) - \sum_{m=1}^{M} I_{Hjm}, j \in N_H \quad (15)$$

To meet exactly the restrictions on the initial and final reservoir storage in eq.(9), the water discharge rate of j^{th} hydro plant in the dependent interval 'd' is then calculated by

$$Q_{GHjd}=V_{Hjo}-V_{HjM}+ \sum_{m=1}^{M} I_{Hjm} + \sum_{\substack{m=1 \\ m\pm d}}^{M} \sum_{l=1}^{R_{uj}} Q_{GHl}(m-\tau ij) - \sum_{m=1}^{M} Q_{GHjm}, j \in N_H \quad (16)$$

The dependent water discharge rate must satisfy the constraints is Eq (7).After knowing the water discharges, the reservoir volumes of different intervals are determined. Then, the hydro generations are calculated from Eq (11). Knowing the calculated hydro generations, P_{GHjm} and the given load demand P_{Djm} for m =1, 2 …..m, thermal generations P_{GTi} can be calculated as

$$P_{GTim}= P_{Dm} + P_{Lossm} - \sum_{j=1}^{N_H} P_{GHjm} \quad (17)$$

Also to meet exactly the power balance constraints in Eq (3), the thermal power generation P_{GTdm} of the dependent thermal generating unit can then be calculated using the following equation.

$$P_{GTdm}= P_{Dm} - \sum_{\substack{i=1 \\ i\neq d}}^{N_{GT}} P_{GTim} - \sum_{j=1}^{N_H} P_{GHjm} \quad (18)$$

The dependent thermal generation must satisfy the constraints in Eq. (4).All the generation levels, discharges, reservoir water volumes and initial and final reservoir storage volumes must be checked against their limiting values as per eq's.(4)–(11).

Stopping Rule :

The iterative procedure of generating new solutions with minimum function value is terminated when a predefined maximum number of iterations (generations) is reacted.

VI. PSO ALGORITHM

The computational process of PSO technique can be described in the following steps.

Step 1 Input parameters of the system and specify the upper and lower boundaries of each variable.

Step 2 Initialize randomly the particles of the population according to the limit of each unit including individual dimensions, searching points and velocities. There initial particles must be feasible candidate solutions that satisfy the practical operating constraints.

Step 3 Let, Qp = [q_{11}, q_{12}, ……, q_{1m}, q_{21}, q_{22}, …. q_{2m},…q_{n1}, q_{n2}, …., q_{nm}], be the trait vector denoting the particles of population to be evolved. The elements of q_{jm} are the discharges of turbines of reservoirs at various intervals subjected to their capacity constraints in (7). q_{id}, be the dependent discharge of i_{th} hydro plant at d_{th} interval is randomly selected from among the committed 'm' intervals. Then, knowing the hydro discharges, storage volumes of reservoirs V_{jm} are calculated by (9). Then P_{GHjm} is calculated from (11) for all the intervals.

Step 4 Compare each particle (4 x 24) evaluation value with its P_{best} the best evaluations value among P_{best} is denoted as g_{best}.

Step 5 Update the iteration as K = K+1 ; inertia weight, velocity& position by (12-14).

Step 6 Each particle is evaluated according to its updated position, only when satisfied by all constraints. If the evaluation value of each particle is better than the previous P_{best}. The current value is set to be P_{best}.
If the best P_{best} is better than g_{best}, the value is set to be g_{best}.

Step 7 If the stopping criterion is reacted, then go to Step-8, otherwise go to Step-2.

Step 8 The individual that generates the latest g_{best} is the solution of the problem and then print the result and stop.

VII. NUMERICAL RESULTS

A. Test System

To verify the applicability and to evaluate the performance of the proposed PSO algorithm, a test system has been adapted from [22], [23]. It consists of a multi chain cascade of four hydro units, and a number of thermal units represented by an equivalent thermal plant. The schedule horizon is one day with 24 intervals of 1 hour each.

The cost of thermal generation can be obtained in two ways:

a) By using a standard economic dispatch technique to find the optimal operation cost of the on-line thermal generators.

b) By assuming the thermal generation is represented by an equivalent single plant, where characteristic can be determined as described in [1].

The hydraulic Sub-system is characterized by the following:

c) A multi chain cascade flow network, with all of the plants on one stream;

d) Reservoir transport delay between successive reservoirs;

e) Variable head hydro plants;

f) Variable natural inflow rates into each reservoir;

g) Variable load demand over scheduling period.

The data of the test system considered here are the same as in [10] and the additional data with valve point loading effect are also same as in Reference[11].

The hydro Sub-system configuration is shown in fig 1.

The hydraulic test network models most of the complexities encountered in practical hydro networks. The load demand, hydro units power generation Coefficients, river inflows, reservoir limits are given in reference [11].

The fuel cost function of the equivalent thermal plant unit with valve point loading is

$$C_i(P_{GTi}) = 5000 + 19.2\,P_{GTi} + 0.002\,P_{G2Ti} + \left|700 \,\mathrm{Sin}\,(0.085\,(P_{GTi}^{min} - P_{GTi})\right|$$

And the inequality constraint limit of this unit is

$$500_{\,(MW)} \leq P_{GTi} \leq 2500_{(MW)}$$

The Spillage rate for the hydraulic system is not taken into account for simplicity and further the electric loss from the hydro plant to the load is taken to be negligibly small.

To demonstrate the effectiveness of the proposed PSO method, the system is considered with prohibited discharge zones and with valve point loading effects.

B. Simulation Results

In short term hydrothermal scheduling problem, the two important parameters, that can be allowed to vary, are the satisfaction of the final reservoir levels and the cost of thermal generation. The present work has been implemented in command line of Matlab-7.0 for the solution of hydrothermal scheduling. The program was run on a 2.70 GHz, Pentium-® Dual core, with 1GB RAM PC. After a number of trails of run with different values of PSO parameters tuning , such as inertia weight, number of particles, maximum allowable velocity, the details of key parameters selected are: w_{max}=0.9,w_{min}=0.4,N=20,c_1=c_2=2.0,$iter_{max}$=100.

The optimal hydro generations , optimal hydro discharges, hydro reservoir levels with minimum cost obtained by the proposed PSO methods are reported in tables 6-8 respectively.

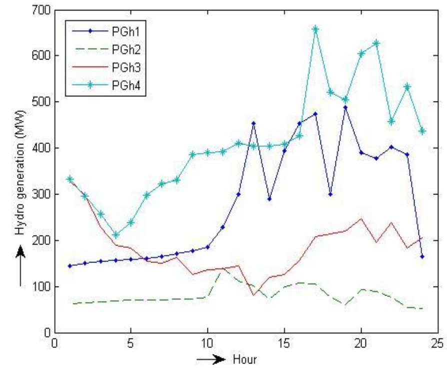


Fig.2 Hourly Hydro plant Power Generations

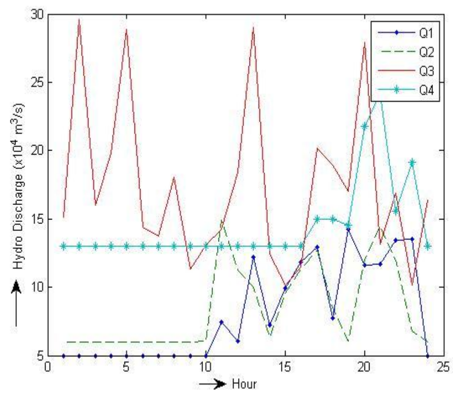


Fig.3 Hydro plant Discharge

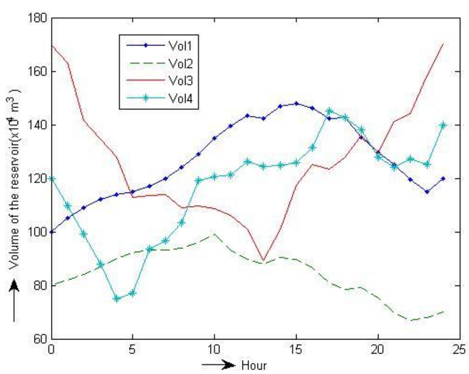


Fig.4 Hydro Reservoir Storage Volumes.

TABLE: 2 HOURLY PLANT DISCHARGES (X 10^4 M^3)

Hour	Q1	Q2	Q3	Q4
1	5.0000	6.0000	20.0855	13.0000
2	5.0000	6.0000	14.8462	13.0000
3	5.0000	6.0000	14.5574	13.0000
4	5.0000	6.0000	29.9338	13.0000
5	5.0000	6.0000	19.7629	13.0000
6	5.0000	6.0000	19.5213	13.0000
7	5.0000	11.1817	13.9664	13.0000
8	5.0000	8.5639	21.9058	13.0000
9	5.0000	11.6833	10.6531	13.0000
10	11.6845	11.1089	14.8067	13.0000
11	9.7279	10.0058	19.9333	13.0000
12	9.9741	10.3444	15.4261	13.0000
13	11.7367	8.4407	14.2368	13.0000
14	6.7460	11.6675	15.6500	13.0000
15	14.0606	6.1333	14.0058	13.0000
16	6.4655	11.7517	19.5437	13.0000
17	10.3722	8.9621	11.0288	13.0000
18	11.8711	8.9506	11.4574	13.0000
19	14.7748	10.8394	28.1152	15.7367
20	13.4112	6.6569	28.3459	13.0041
21	6.3168	8.4911	15.4668	22.2915
22	10.5628	6.5219	18.7661	23.0842
23	7.2958	8.6968	12.7706	23.4654
24	5.0000	6.0000	21.0000	13.0000

TABLE:1 HYDRO THERMAL POWER GENERATIONS AND OPERATING COST OF EQUIVALENT THERMAL UNIT.

Hour	Hydro1	Hydro2	Hydro3	Hydro4	Equivalent Thermal	Operating Cost	Loss	Demand
1	143.8144	61.8151	321.2825	332.7922	510.2957	14804.9367	0.0000	1370.0000
2	149.4909	63.5051	246.4740	296.0982	634.4318	18665.8623	0.0000	1390.0000
3	153.7395	66.0341	231.7937	256.5596	651.8732	18649.0470	0.0000	1360.0000
4	156.5676	68.5559	270.9775	211.3765	582.5224	16643.8241	0.0000	1290.0000
5	157.9804	70.2331	222.8005	236.0156	602.9704	16655.0631	0.0000	1290.0000
6	160.8036	71.0705	209.2636	242.4304	726.4318	20472.3135	0.0000	1410.0000
7	165.0319	108.7630	172.6008	247.8405	955.7637	25839.2282	0.0000	1650.0000
8	170.6580	87.6200	192.4158	306.5693	1242.7369	31761.7471	0.0000	2000.0000
9	177.6717	103.0794	130.8265	329.9760	1498.4464	38941.6308	0.0000	2240.0000
10	387.8581	96.4377	168.7674	352.5204	1314.4164	33180.3073	0.0000	2320.0000
11	336.3422	88.5497	182.2171	355.8591	1267.0319	33234.1502	0.0000	2230.0000
12	344.0508	87.0328	182.7510	386.6004	1309.5650	32919.0708	0.0000	2310.0000
13	394.6999	75.0250	194.3870	378.5038	1187.3842	31317.4820	0.0000	2230.0000
14	248.5863	88.4380	223.0302	384.7371	1255.2084	32716.0318	0.0000	2200.0000
15	469.3308	58.5306	227.5507	408.6391	965.9489	25699.8086	0.0000	2130.0000
16	239.6118	87.2711	279.7513	416.9959	1046.3699	27468.6170	0.0000	2070.0000
17	367.1793	71.1986	209.9862	421.2548	1060.3812	27083.0721	0.0000	2130.0000
18	400.2948	67.3058	221.6939	430.3770	1020.3285	27167.5092	0.0000	2140.0000
19	447.9425	69.5130	389.0751	501.1761	832.2933	21670.4962	0.0000	2240.0000
20	384.6560	51.9737	364.7770	447.0443	1031.5490	27629.5083	0.0000	2280.0000
21	198.8359	62.5183	279.2299	638.0259	1061.3901	27071.1105	0.0000	2240.0000
22	308.8412	53.9597	340.2419	583.4025	833.5546	21694.0029	0.0000	2120.0000
23	225.5577	65.7688	255.4101	619.4911	683.7723	18472.1612	0.0000	1850.0000
24	165.0319	51.6079	364.9660	436.9659	571.4283	17001.5918	0.0000	1590.0000

TABLE: 3 HOURLY STORAGE VOLUME OF HYDRO RESERVOIRS ($\times 10^4$ M^3)

Hour	Vol1	Vol2	Vol3	Vol4
0	100.0000	80.0000	170.0000	120.0000
1	105.0000	82.0000	158.0145	109.8000
2	109.0000	84.0000	151.3683	99.2000
3	112.0000	87.0000	145.8109	87.8000
4	114.0000	90.0000	128.8771	74.8000
5	115.0000	92.0000	123.1142	81.8855
6	117.0000	93.0000	118.5929	83.7317
7	120.0000	87.8183	118.6265	85.2891
8	124.0000	86.2544	109.7208	102.2229
9	129.0000	82.5711	111.0677	108.9858
10	128.3155	80.4622	113.4427	115.5071
11	130.5876	79.4564	108.0733	116.4735
12	130.6135	77.1120	118.0151	125.3792
13	129.8768	76.6713	128.6151	123.0323
14	135.1308	74.0038	135.9449	124.8390
15	132.0702	76.8705	147.0202	131.7723
16	135.6047	73.1188	144.6632	134.1984
17	134.2325	71.1567	161.3625	135.4352
18	130.3614	68.2061	164.5039	138.0852
19	122.5866	64.3667	159.5126	136.3543
20	115.1754	65.7098	152.9999	142.8939
21	115.8586	66.2187	163.2585	131.6312
22	113.2958	68.6968	170.7429	120.0043
23	115.0000	68.0000	171.9461	124.6541
24	120.0000	70.0000	170.0000	140.0000

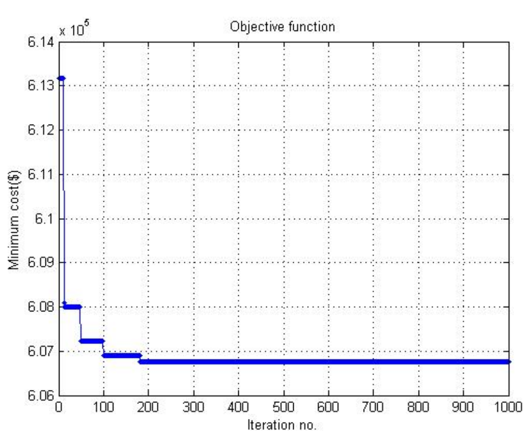


Fig.5 Convergence characteristic of PSO Algorithm for the test case.

TABLE: 4 SUMMARY OF TEST RESULTS

Method	Operating cost of equivalent thermal unit (in $)		Execution time(in sec)	
	Without Prohibited Discharge Zones	With Prohibited Discharge Zones	Without Prohibited Discharge Zones	With Prohibited Discharge Zones
PSO	602255.525216	606758.572623	111.916082	122.182410
GA	606958.862515	607154.561663	398.333785	422.213541

VIII. CONCLUSION

In this paper an approach of particle swarm optimization has been proposed and demonstrated to solve short – term hydrothermal scheduling problem. In the algorithm, the thermal generator units are represented by and equivalent unit. The generator load power balance equations and total water discharge equation have been subsumed into system model .constraints on the operational limits of the thermal and hydro units on the reservoir volume limits are also included in the algorithm. the numerical results show that the proposed approach is better than generic algorithm in terms of having better solution quality and good convergence characteristics. The PSO approach can easily be extended to other complex optimization problems faced by the utilities.

REFERENCES

[1]. A.J.Wood, B.F.Wollenberg, “Power Generation Operation and Control”, John Wileyand Sons, New York, 1984

[2]. M.F. Carvalloh, S.Soares, “An efficient hydrothermal scheduling algorithm”, IEEE Trans. PWRS 4 (1987) 537-542.

[3]. L.Engles, R.E. Larson, J.Peschon, K.N. Stanon, “Dynaming programming applied to hydro and thermal generation scheduling”, in: IEEE tutorial course text, 76CH1107-2-PWR, IEEE, New York, 1976.

[4]. S.chang, C.Cen, L.fung, P.B.Luh, “Hydroelectric Generation Scheduling With An Effective Differential Dynamic Programming”, IEEE Trans. Power syst. 5(1990) 737-743

[5]. Q. Xia, N.Xiang, S.Wang, B.Zang, M.Huang, “Optimal Daily Scheduling Of Cascaded Plants Using A New Algorithm Of Nonlinear Minimum Cost Network Flow”, IEEE Trans. Power syst. 3(3) (1988) 929-935

[6]. K.P. Wong, Y.W. Wong, “Short Term Hydrothermal Scheduling”: part 1. Simulated annealing approach, IEE Proc. C 141 (1994) 497-501

[7]. Y.-G. Wu, C,=Y. Ho, D.-Y. Wang, “A diploid genetic approach to short-term scheduling of hydrothermal systems”, IEEE Trans. Power syst. 15(November (5)) (2000) 1268-1274

[8]. S.O. Orero, M.R. Irving, “A genetic algorithm modeling framework and solution technique for short-term optimal hydrothermal scheduling”, IEEE Trans. Power Syst. 13(2) (1998) 501-518

[9]. P.-H Chen, H.-C. Chang, “Genetic added scheduling of hydraulically coupled plants in hydro-thermal coordination”, IEEE Trans. Power syst. 11(may(2)) (1996) 975-981.

[10]. P.K. Hota, R. Chakrabarti, P.K. Chattopadhyay, “Short-term hydrothermal scheduling through evolutionary technique”, EPSR 52 (November (2)) (1999) 189-196

[11]. N. Sinha, R.Chakrabarti, P.K. Chattopadhyay, “Fast evolutionary technique for short-term hydrothermal scheduling”, IEEE Trans. Power syst. 18(1) (2003) 214-220.

[12]. K.K. Mandal*, M.Basu, N.Chakraborty, "Particle Swarm Optimization technique based short-term hydrothermal Scheduling", Appled soft computing 8(2008) 1392-1399

[13]. Chandrasekar Samudi, GauthamP.Das, Piyush C. Ojha,Sreeni. T.S, 978-1-4244-1904-3/08/ 2008 IEEE

[14]. Wang, C., Shahidehpour, S.M., "Power generation Scheduling for multi- area hydrothermal power systems with tie-line constraints", cascaded reservoirs and uncertain data.", IEEE Trans. PWRS, Vol. 8, No. 3, 1993, pp. 1333-1340

[15]. Soares, S., Lyra, C., Tavayes, WO., "Optimal generation scheduling of hydrothermal power system.", IEEE Trans. PAS Vol. 99, No 3, 1980 1106-1114.

[16]. B.Yu, X. Yuan, J.Wang, "Short-term hydro-thermal scheduling using Particle swarm optimization method", Energy convers. Manage.48 (2007) 1902-1908.

[17]. Z.L., Gaing, "Particle swarm optimization to solving the economic dispatch considering the generator constraints", IEEE Trans. 18 (August(3)) (2003)1187-1195.

[18]. J.B. Park, K.S. Lee, J.R. Shin, K.Y. Lee, "A particle swarm optimization for economic dispatch with non-smooth cost functions", IEEE Trans. Power Syst. 20 (1) (2005) 34-42

[19]. K.Price, R,Storn, "Differential evolution: numerical optimization made easy". Dr. Dobbs J.22(4) (1997) 18-24

[20]. K.K. Mandal*, M.Basu, N.Charaborty, "Particle Swarm Optimization technique based short-term hydrothermal Scheduling", Appled soft computing 8(2008) 1392-1399

[21]. Chandrasekar Samudi, Gautham P.Das, Piyush C. Ojha,Sreeni. T.S "Hydrothermal Scheduling Using Particle Swarm Optimization, 978-1-4244-1904-3/08/ 2008 IEEE.

[22]. Wang, C., Shahidehpour, S.M., "Power generation Scheduling for multi- area hydrothermal power systems with tie-line constraints, cascaded reservoirs and uncertain data.", IEEE Trans. PWRS, Vol. 8, No. 3, 1993, pp. 1333-1340

[23]. Soares, S., Lyra, C., Tavayes, WO., "Optimal generation scheduling of hydrothermal power system.", IEEE Trans. PAS Vol. 99, No 3, 1980 1106-1114.

[24]. B.Yu, X. Yuan, J.Wang, "Short-term hydro-thermal scheduling using Particle swarm optimization method", Energy convers. Manage.48 (2007) 1902-1908.

AUTHORS PROFILE

G.Sreenivasan received his graduation in 1996, Masters in 2001 from JNT University college of engineering, Anantapur and pursuing Ph.D at JNT University, Kakinada. Presently he is working as Associate Professor in the Department of Electrical and Electronics Engineering, Intell Engineering College, Anantapur since 2001. His area of interest is applications of GA and PSO to Power System Operation and Control. He has published research papers in 1 international and 2 national journals and has publications in national and international conferences. He has participated several workshops conducted by various engineering colleges at south India. Recently he co-authored the text book, "Power System Operation and Control" under the publication of Pearson Education, New Delhi. He is the life member of ISTE and IE.

Dr. CH. SaiBabu received his graduation from Andhra university and Masters from REC, Warangal and Ph.D from JNTU,Hyd. He is the Professor in the Department of Electrical and Electronics Engineering and working as Director of Energy Management and Empowerment at JNT University, Kakinada. He has about 10 publications in international journals and 44 national and international conferences. He authored a text book, "Elements of Power Electronics" published by S.Chand and Co. He is the member of AICTE team and Member of governing council for various affiliated engineering colleges of JNTU, Kakinada.

Dr. S.Sivanagaraju received his graduation in 1998, Masters in 2000 from IIT, Kharagpur and his Ph.D from JNT University, Hyderabad in 2004. He is working as Professor in the Department of Electrical and Electronics Engineering JNT University, Kakinada, Andhra Pradesh since 1992. He has received 2 National awards (Panditmadan Mohan malaviya Memorial prize award and Best paper prize award) form the Institute of Engineers (India) for the year 2003 – 2004. He is also a referee for the IEEE proceedings – generation, transmission and distribution and international journal of emerging electric power systems. He has about 50 publications in national and international journals and conferences to his credit. His areas of interest are in Distribution Automation, Genetic Algorithm Applications to Distribution Systems and Power Systems.

mSCTP Based Decentralized Mobility Framework

[1]Waqas A. Imtiaz, [2]M. Afaq, [3]Mohammad A.U. Babar
[1,2,3]Department of Electrical Engineering
[1,3]IQRA National University, Peshawar, Pakistan
[2]City University of Science and Information Technology, Peshawar, Pakistan

***Abstract*— To conceive the full potential of wireless IP services, Mobile Nodes (MNs) must be able to roam seamlessly across different networks. Mobile Stream Control Transmission Protocol (mSCTP) is a transport layer solution, which unlike Mobile IP (MIP), provides seamless mobility with minimum delay and negligible packet loss. However, mSCTP fails to locate the current IP address of the mobile node when Correspondent Node (CN) wants to initiate a session. In this paper, we propose DHT Chord to provide the required location management. Chord is a P2P algorithm, which can efficiently provide the IP address of the called MN by using its key-value mapping. The proposed decentralized mobility framework collectively exploits the multihoming feature of mSCTP, and efficient key-value mapping of chord to provide seamless mobility. Suitability of the framework is analyzed by preliminary analysis of chord lookup efficiency, and mSCTP handover procedure using overlay weaver and NS-2. Performance analysis shows that mSCTP multihoming feature and Chord efficient key-value mapping can provide a non-delayed, reliable, and an efficient IP handover solution.**

Keywords - mSCTP; Chord; Multihoming; Seamless mobility.

I. INTRODUCTION

Basic design of Internet Protocol stack was laid on the assumption that, "*all nodes have fixed IP addresses*" [21]. This concept of fixed IP addresses worked flawlessly until communication nodes became mobile. Since then, IP mobility is a major issue, which needs to be resolved because when a MN changes its PoA, its IP address changes and results in termination of the ongoing session [1].

To exploit the full potential of wireless IP services, MNs must be able to wander seamlessly across different set of networks. Seamless mobility consists of two basic components: *handover management* and *location management*. Handover Management allows the MN to change its PoA without terminating the ongoing session. Location Management allows the MN to maintain its reachability for new connections after changing its PoA.

Layered Internet Protocol Stack enables us to provide seamless mobility at different layers [16]. MIP at the network layer provides a complete seamless mobility solution; however, its handover mechanism introduces unavoidable delay and dependence on additional network components [7]. A suitable handoff solution is required, which can provide mobility at user end without relying on additional network components, along with minimum delay and maximum security. SCTP is a transport layer protocol that encompasses revolutionary features like multihoming, multistreaming, and four-way handshake for connection establishment. SCTP multihoming feature can provide handover solutions; however, SCTP alone cannot support handoffs because it is not able to add or delete IP addresses during an active association. mSCTP, an extension of SCTP, enables SCTP to dynamically add/delete IP addresses during an active association and enables SCTP to perform handovers using its multihoming feature [1][7]. However, mSCTP fails to provide seamless mobility, when a CN wants to initiate a session with MN.

Location management in wireless IP services provides the IP address of the called MN. Traditional location management schemes like DDNS and SIP incorporates client server models, and suffers from their well-known drawbacks like congestion, centre point of failure and bottlenecking [16]. Such scheme is required, which along with mSCTP can provide an abrupt, decentralized, efficient, and reliable location management. DHT chord, a P2P algorithm, can provide the required location management by using its efficient key-value mapping as name to IP mapping.

Chord is a decentralized lookup system, which provides one and only operation: *given a key and it efficiently maps the key onto a peer* [1][7]. Chord forms a one-dimensional identifier circle, consisting of nodes and keys placed inside them, which ranges from 0 to 2^m - 1, where *m* is the number of bits in the identifier circle. Each node and key is assigned an *m-bit* identifier; in order place them efficiently in the overlay network. *0(Log N)* messages are required for key-value mapping over the identifier circle, where *N* represents the number of nodes [12]. A complete decentralized approach, unnecessary complexity and no need for advanced technological changes, led us to investigate the performance and suitability of DHT Chord as location manager.

Efficient key-value mapping of chord can provide the necessary location management. Each node uses an identifier-locator set in the overlay network such that identifier refers to key i.e. Node ID, and locator represents the value or IP address in the key-value pair. Efficiency of the framework is increased via successor pointers, which efficiently reduces the number of messages during a chord query process. This paper proposes a complete decentralized mobility framework without unnecessary complexity and no need for advanced technological changes in the current internet architecture. Proposed framework exploits the multihoming feature of mSCTP for handover management at the transport layers, and efficient key-value mapping of chord for location management at the application layer to provide an efficient, robust and scalable mobility solution.

II. RELATED WORK

In the near future, LTE and 4G technologies will not be able to support network-controlled handovers [47], therefore, an efficient user centric approach for seamless mobility is required. Kim and koh [5] evaluates the performance of MIP and mSCTP over IPv6 networks. Their analysis shows that mSCTP performs better than MIP via its multihoming feature. Zeadally and Siddiqui [10] show that mSCTP performs better than MIP and SIP in terms of handover latency and packet transmission after hadoffs. Ferrus and Brunstrom in [6], figures that transport layer is a worthwhile approach for handoffs and hence deserves more attention than the existing network and application layer solutions.

Park and kim [22] determines the performance of SCTP along with Session Initiation Protocol (SIP), in order to improve the Quality of Service (QoS) for real-time media. Their evaluation shows that mSCTP can perform better than UDP. Fu and Atiqquzama in [16] developes an analytical model to evalutae the performance of DNS as location manager. SCTP Draft proposes the use of Mobile IP for location managenet along with mSCTP, however this approach brings unavoidable delay via additional components.

P2P can perform lookups in a fraction of seconds with distributed content placement and discovery [11]. Cirani and Veltri in [14] propses an architecture for Distributed Location Service (DLS) which provides efficient location management. Sethom and Afifi in [13] presents PALMA (peer to peer architecture for location management), using tapestry for LM in mobile networks. However, PALMA cannot support handover management. Kunzmann and Hanks in [19] introduces a noval architecture for Next Generation Internet that is completely decentralized and relies on DHT algorithm.

III. DECENTRALIZED MOBILITY FRAMEWORK

A. Addressing Scheme:

An addressing scheme is required to identify and place mobile node when it enters the overlay network. Our addressing scheme consists of an *identifier-locator* set where identifier corresponds to the unique identity of the MN, and locator represents the current IP address in the network.

Each node inside the network has a *UID (Unique Identifier)* [1][14]. UID consists of three basic components, ***name: device: ID***. Name contains the owner's name; it can be selected in any desired form, i.e. surname or name initials. Device refers to the type of the device, e.g. laptop, mobile, and PDA etc. ID is a unique identity that can be user's mobile number, email address or NIC number, e.g. *xyz: laptop: 17301xxxx*. Any naming scheme can be adapted as chord provides flexible naming mechanism.

Locator is referred to as *TL (Temporary Locator)* [1], which represents the value in the key-value pair. It is identical to the current IP address of the MN. TL has the ability to update itself as soon as MN attains a new IP address.

B. Node Entry:

Mobile node must be able to join and publish its TL value as soon as it enters the network. When mobile node joins the network, it receives a new IP address and updates its UID-TL pair. After updating its pair, mobile node submits a query for mapping between its UID and base node in order to publish its TL value. Mapping process enables the mobile node to locate its corresponding base node, which will carry its TL value.

MN can update its UID-TL pair periodically by sending update messages to the base node. BN replies with an acknowledgment message. ACK messages are necessary to indicate the presence of the base node, as it is also mobile and can enter or leave the network at any instant. In case of acknowledgment failure, MN needs to locate another BN to publish its UID-TL value [1].

On reception of UID-TL value from the MN, base node publishes pointers towards its successor nodes. These pointers can shorten the query process by reducing the number of messages during a query. Successor pointers have the ability to time out if they are not periodically updated by their respective BN. Fig. 1 shows the process, where BN N8 publishes pointers towards its successors N12, N16 and N28. Base node N8 periodically updates these pointers information which helps in coping with any updated information regarding TL values [1].

C. Location Search:

When CN wants to start a session, it requires MN's IP address, i.e. TL value. To obtain the current TL value of the MN, CN simply hashes its UID value and submits a query to its successor node. Successor node of the CN maps the given UID to the responsible base node and provides it with an IP address of the querying CN. This IP address helps to transfer the requested TL values directly to the CN.

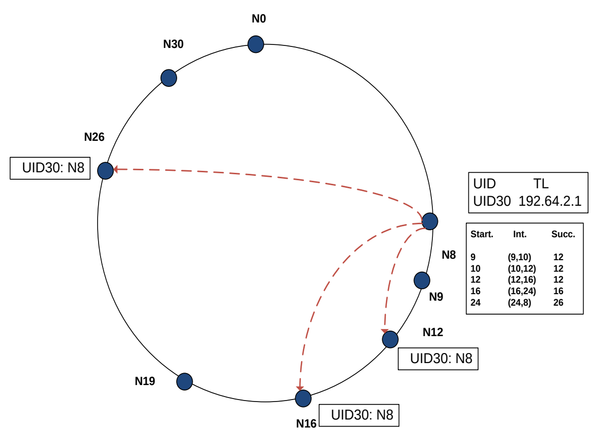


Figure 1: N8 publishing pointers towards its successor nodes

Pointers stored at every successor node shorten the number of *0(Log N)* messages required to locate an object. For instance, the successor node encounters a node with a pointer towards the base node. In this case, the conventional query process terminates, and the query instantaneously redirects towards the base node, which provides the necessary TL value. Fig. 2, shows node N22 querying its successor node N26 for node N30 TL value. N22 first submits a query to its successor node N26. N26 having a pointer for UID30, redirects the query to N8 instead of searching for the closest interval in its finger table. Node N8 takes the required IP address of the querying node and directly transfers TL value of N30 to node N22. Thus, the use of successor pointers can reduce the number of query messages and hence the time required to find the required UID-TL pair.

D. Location Update:

As mobile moves into a different network, its TL value and corresponding base node changes, this may involve the update of mobile node's location information. When MN enters the overlapping region, it attains a new IP address and updates its TL value. MN also sends an update message to its corresponding BN1 containing both TL values from network1 and network2 via interface 1. However, MN is still in AP1, and uses interface 1 as a primary path for communication.

When MN switches to network 2 in the overlapping region, it finds a new BN i.e. BN2 and updates its location with BN2 via interface 2, which is now the primary path for communication. Interface 1 is used as secondary path for redundancy purposes. MN after joining network2 informs BN2 of its previous base node, i.e. BN1. BN2 on receiving the necessary information sends a redirect query message towards BN1 [1]. This enables BN1 to redirect all the queries regarding UID value of the MN towards BN2, until its UID-TL value, and pointers time out in BN1 as shown in Fig. 3.

When MN leaves the overlapping region and completely transfers to network2, it updates its UID-TL value by sending an update message to the responsible BN2 via interface 2. TL value on interface 1 is dropped, so that the interface is free for further communication sessions.

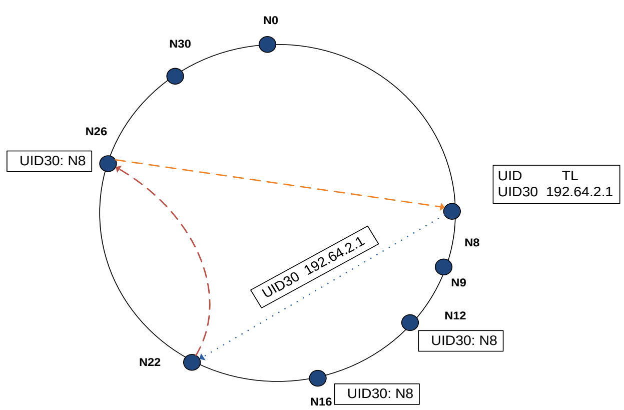


Figure 2: Query Process Made Easy with Successor Pointers

IV. HANDOVER PROCEDURE

A. Handover Procedure:

mSCTP decentralized mobility framework mainly undertakes the sessions originated from CN towards MN. DHT Chord in this scenario provides the current TL value of the MN, which initiates the session from CN towards the MN. Once the session is established, an mSCTP handover procedure supports the on-going session and handovers when required [8].

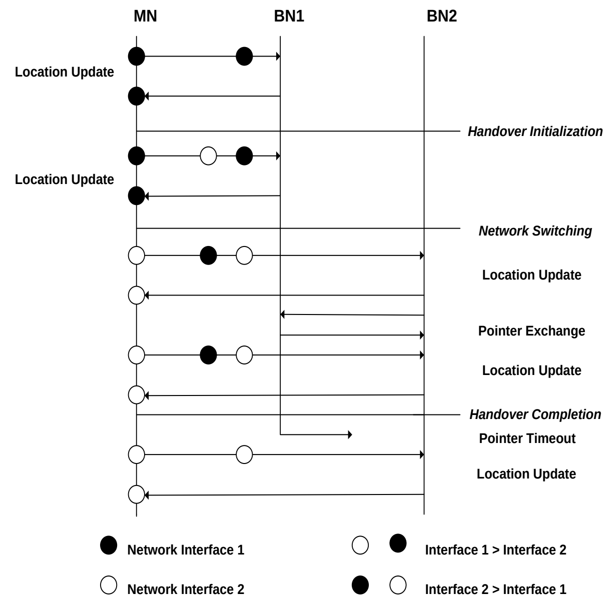


Figure 3: Location Update Process for MN [1]

1) Association Establishment:

In order to start a new session, CN simply hashes the UID value of a MN and submits a query to its closest successor. CN informs its mSCTP stack and initiates the basic SCTP association initialization process after receiving the called TL value. CN sends an INIT chunk towards the MN on the obtained TL (IP address) value. MN replies to INIT chunk with an INIT-ACK chunk, followed by connection establishment with the exchange of COOKIE-ECHO and COOKIE-ACK messages.

Fig. 4 shows an example where, MN N18 enters a new network i.e. network 1, and publishes its UID-TL pair via base node N24. On reception of the TL value, BN publishes the respective pointers towards its successor nodes. Now CN N2 wants to initiate session with the MN N18. CN first hashes UID of N18 and submits a query for its TL value. Node N24 provides the required TL value, which enables the CN to start a session with MN by the exchange of INIT, INIT-ACK, COOKIE-ECHO and COOKIE-ACK chunks between N2 and N18 between MN and CN.

2) Data Transport and Handover:

After association establishment, CN starts sending packets towards MN over the acquired TL value. BN contains the current location information of the MN, and periodically updates CN. MN periodically updates it TL value with its corresponding BN via update messages. CN also keeps itself updated by periodically querying its respective BN. Base node updates its successor nodes by periodically publishing pointers towards them, along with ACK replies to the MN.

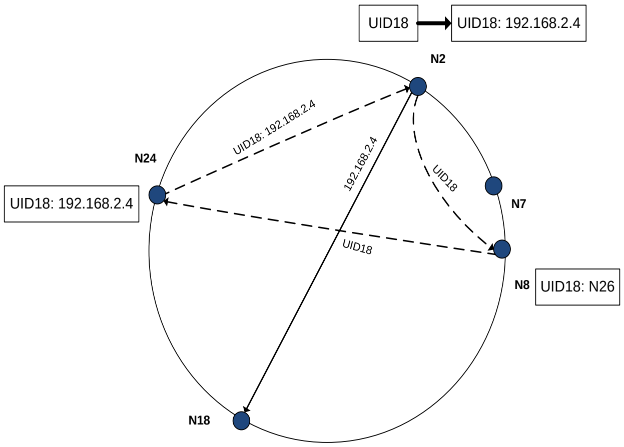


Figure 4: Association Establishment process using Chord as Location Manager

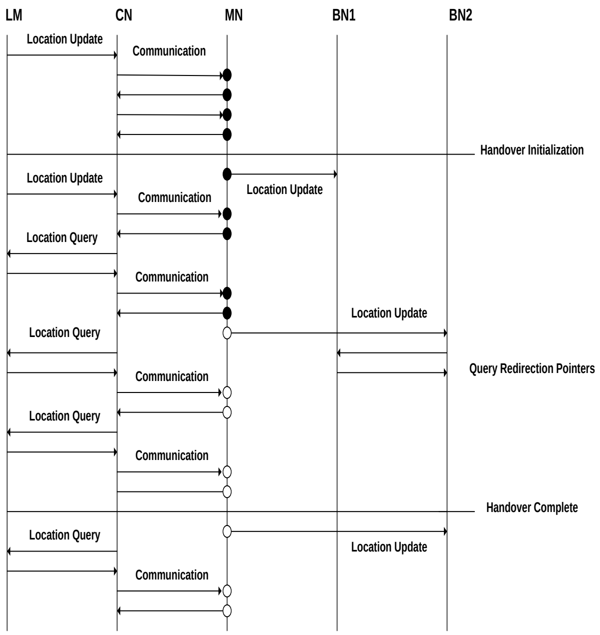


Figure 5: Data Transfer and Handover Procedure

Fig. 5 shows the data transport and handover process followed by MN and CN. LM handles any changes in MN's TL value and periodically updates CN about the BNs carrying the corresponding MN's TL value. Figure shows that CN continuously investigates location manager in order to keep itself updated of the BNs and MN TL value. When MN moves from network 1 towards network 2, it updates its respective BN i.e. BN2. BN1 after receiving query redirect message from BN2 updates its CN to locate BN2 for MN current TL value [1]. MN on the other hand performs a soft handover by adding the newly attained TL value to its ongoing SCTP association using mSCTP DAR extension.

V. Preliminary Analysis

A. mSCTP Handover Latency

1) Analytical Analysis:

"Handover latency is the time at which an existing IP address becomes unavailable for end to end data transmission by movement of a node into a new network, to the time at which the end node receives a sequence of end to end transmission using the newly obtained IP address" [7].

T_{mSCTP} = mSCTP Handover Delay
T_{md} = Movement Detection
T_{ac} = Address Configuration
T_{DAR} = Dynamic Address Configuration Time
$T_{add\text{-}IP}$ = Add-IP Address Chunk Exchange Time
$T_{pc\text{-}IP}$ = Primary-IP Address Change Chunk Exchange Time
$T_{del\text{-}IP}$ = Delete-IP address Chunk Exchange Time
T_{pc} = Primary Interface Change Time

mSCTP handover delay is accounted by MN's movement detection, Dynamic Address Reconfiguration (DAR) between MN and the CN, and configuration of the newly attained IP address. Thus the handover latency of mSCTP becomes [1][5][7] :

$$T_{mSCTP} = T_{md} + T_{ac} + T_{DAR} \qquad (1)$$

DAR includes the exchange of ADD-IP, DELETE-IP and PRIMARY CHANGE-IP messages along the processing time required to transfer the process these messages, so T_{DAR} in eq. 1 becomes [1],

$$T_{DAR} = T_{add\text{-}IP} + T_{pc\text{-}IP} + T_{del\text{-}IP} + T_{pc} \qquad (2)$$

$$T_{add\text{-}IP} + T_{pc\text{-}IP} + T_{del\text{-}IP} + T_{pc} = (T_{MN\text{-}CN} + T_{CN\text{-}MN}) + (T_{MN\text{-}CN} + T_{CN\text{-}MN}) + (T_{MN\text{-}CN} + T_{CN\text{-}MN}) + T_{pc} \qquad (3)$$

$$T_{add\text{-}IP} + T_{pc\text{-}IP} + T_{del\text{-}IP} + T_{pc} = 3(T_{MN\text{-}CN} + T_{CN\text{-}MN}) + T_{pc} \qquad (4)$$

As MN is a multihoming device, so SCTP can utilize the primary interface to transmit all data chunks. Secondary interface is used to configure the newly acquired IP address. Thus, the time taken during movement detection and address configuration (T_{md} + T_{ac}) can be neglected, because MN and CN can communicate during these processes without session termination.

Therefore, the total handover latency in Eq. 1 becomes [1]

$$T_{mSCTP} \cong \underbrace{T_{md} + T_{ac}}_{} + 3(T_{MN\text{-}CN} + T_{CN\text{-}MN}) + T_{pc} \qquad (5)$$

(≅0 due to

$$T_{mSCTP} \cong 3(T_{MN\text{-}CN} + T_{CN\text{-}MN}) + T_{pc} \qquad (6)$$

As explained in [8] [9]; that ASCONF chunks can be transmitted by bundling them with data chunks. Therefore, delay introduced by exchanging DAR control chunks between CN and MN, $3(T_{MN\text{-}CN} + T_{CN\text{-}MN})$, can be neglected as no extra time is spent in transmitting these chunks from MN towards CN. Only delay significant enough is the processing time of a MN during DAR procedure, which mainly includes switching of data transmission from one interface to another.

So the total theoretical handover latency for mSCTP in Eq. 6 becomes

$$T_{mSCTP} \cong \underbrace{T_{md} + T_{ac} + 3(T_{MN\text{-}CN} + T_{CN\text{-}MN})}_{} + T_{pc} \quad (7)$$

(≅0 bundling of ACONF chunks with Data chunks)

$$T_{mSCTP} \cong T_{pc} \quad (8)$$

Eq. 8 shows that the handover latency of mSCTP only accounts for the processing time of the node that is required to switch between interfaces. Thus, mSCTP can efficiently provide the required seamless mobility with minimum handover delay.

2) Simulation scenario:

A simulation scenario is designed to analyze the performance of mSCTP handover by measuring its handover latency. Handover latency is determined by measuring the time when the handover takes place to the time when a new packet arrives at the CN from the newly switched interface. Ns-2 uses SCTP multihoming feature to perform the required handover. Processes like that of mobile node movement detection, address configuration etc. are not taken into account, as delay introduced by them is negligible as shown in Eq. 7. Simulation scenario is run for 60 sec, and handover mechanism is introduced at 30 sec. Handover occurs when MN switches data transmission from one interface to another along with APs. Fig. 6 shows the data transmitted from MN towards the CN during 60 sec. Data transmission curve shows that there is no significant packet loss or delay during the handover process as the bytes transmitted keeps on increasing. There is a slight bent in the curve at 30 sec, which can be accounted for the processing time required to switch between the interfaces as mentioned in Eq. 8.

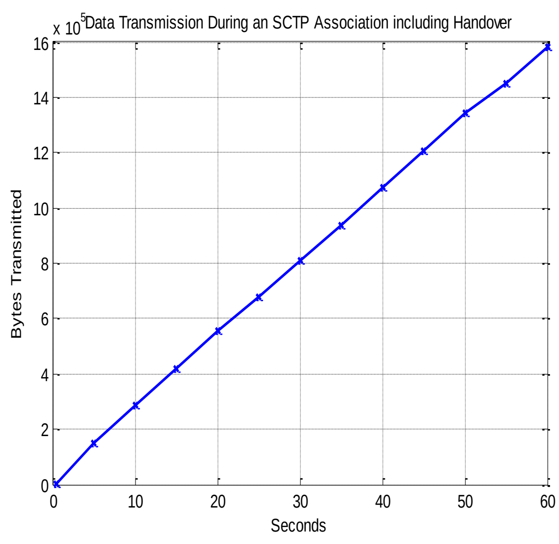


Figure 6: Data Transmitted During an SCTP Association Including Handover at 30 sec

B. CHORD Successful Value Retrieval:

Suitability and success of chord as location manager depends on how efficiently it can retrieve the required key-value pair, and how often does it replies to queries requested by CN. Moreover chord must be able to self organize itself under worst network conditions. Overlay weaver is used to analyze the performance of chord, which is an overlay construction tool kit that supports multiple p2p lookup algorithms like CAN, Chord etc. It can invoke single or multiple nodes on the structured overlay network using multiple instances of DHT Shell. DHT shell is a layered command language interpreter that is used to control DHT and its algorithms. Each instance of DHT shell acts as a node on the overlay network.

1) Test Scenario:

To create a large overlay network in a rather small environment, we invoked several instances of DHT shell, over limited machines connected to one another in Local Area Network (LAN). The LAN consists of 8 computers with the following specifications, Intel Core 2 Duo 2 GHz processor and 3 GB RAM. Cisco 3600 series router and D-Link switch using Ethernet links connect these computers. Each node is running Windows XP and is equipped with Overlay weaver and Apache Ant build tool.

Each node is assigned a different port number to differentiate it from other nodes in the overlay network running on the same computer. 50 nodes of DHT shell are invoked on every computer, to get a large overlay network up to 400 nodes. Tests are performed manually, so no stabilization time or rate of nodes entering or leaving the network is considered. The ability of chord algorithm to support multiple values for a single key is used in our tests for performance analysis.

Successful retrieval of key-value pair is determined by making 25 queries in randomly selected nodes while increasing and decreasing the number of nodes. Four tests are conducted by varying the number of values associated with a key. Nodes are randomly selected to insert the key-value pair at the start of the test, using *Put <key> <value>*. To retrieve the required value in randomly selected nodes, *Get <value>* command is used at each instant. The number of queries successfully answered determines success of value retrieval. For decreasing network size, percentage of key-value retrieval is determined by decreasing the number of nodes in the overlay network from 400 to 100 nodes manually, and queries are made at random instances to retrieve the required key-value pair. The results obtained are explained below.

2) Results and Observations:

Results are observed as the percentage of queries successfully answered for a key-value i.e. UID-TL query. For this purpose, we first increased the number of nodes by introducing multiple DHT shells in each pc, and randomly made queries at different network sizes. Number of queries successfully answered are observed using DHT shell.

Fig. 7 shows that as we increase the number of nodes, amount of queries successfully answered somewhat decreases. However, the percentage of queries successfully answered still remains higher than 95%. Wrong/incomplete finger table entries and dropped UDP packets can be responsible for this decrease in successful retrieval of the key-value pair. Increase in the number of nodes, in the overlay network does not have any considerable effect on the performance of chord, as successful gets remains more than 95% in most cases.

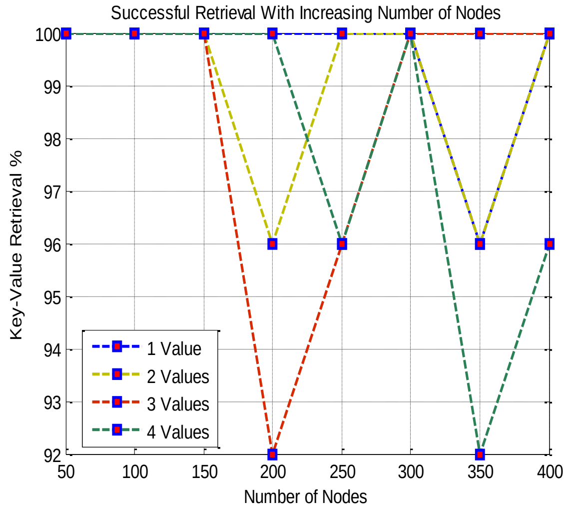


Figure 7: Successful Queries with Increasing Number of Nodes

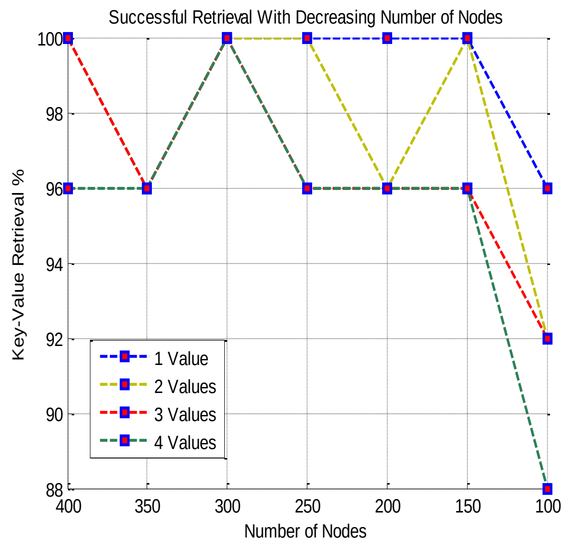


Figure 8: Successful Queries with Decreasing Network Size

When we decrease the number of nodes in the overlay network, value retrieval of chord remains efficient enough to operate as LM as shown in Fig. 8. In case of nodes, failures about 92 % of the queries are replied successfully. Even when the number of nodes is reduced from 400 to 100 nodes, chord is still able to retrieve the required key-value value. With decrease in the number nodes, the time taken to retrieve the required value increases, due to finger table updates or route failures. To account for this effect, we discarded any query that took more than 5 seconds. Performance analysis of chord value retrieval under both scenarios shows that chord can efficiently retrieve the required TL value, and hence can efficiently perform as location manager.

VI. CONCLUSION

Seamless handoff solutions are required for future IP networks to facilitate ubiquitous users. Mobile Stream Control Transmission (mSCTP) protocol is a transport layer protocol, which enables MN to roam seamlessly across different networks. Unlike MIP, it provides handover solutions at user ends, with minimum delay and negligible packet loss via its multihoming feature. Preliminary analysis shows that mSCTP handover delay is as small as the processing time of the device, and it eliminates delay via agent discovery and registration process by exploiting its multihoming feature. However due to its inherent transport layer mechanism, mSCTP cannot provide location management.

Traditional location management schemes like DDNS and SIP incorporates client server models, which makes them vulnerable to centre point of failure, congestion and bottlenecking. DHT Chord can solve the problem by providing a complete decentralized approach using its efficient key-value mapping. An identifier-locator set is created using chord's key-value pair, which contains UID-TL pair to provide the required location management. Efficiency of the location management scheme is increased with the successor pointers. These pointers efficiently reduce the number of messages during query and hence the time in finding the required TL value. Performance analysis of chord value retrieval under both scenarios shows that, chord can efficiently retrieve the required TL value, and hence can efficiently perform as location manager.

This paper proposes a decentralized mobility framework for IP based handovers that does not requires any evolutionary technology changes to the current internet architecture. Efficient lookup algorithm, scalability, flexible naming, authorization support, and no central point of failure make DHT Chord an ideal candidate for location management. Supported by mSCTP multihoming feature and DAR extension, chord can efficiently provide the required name-IP mapping and support non-delayed handover procedures. Besides technical advantages, end users will gain added functionalities and more flexibility from this mobility framework.

REFERENCES

[1] Waqas Ahmed, M. Jashim, "An SCTP Based Decentralized Mobility Framework", Masters Thesis, Department of Electrical Engineering, Blekinge Institue of Technlogy, Karlskrona, Sweden, July 2010.

[2] SCTP for Beginners, http://tdrwww.exp-math.uni-essen.de, [Online], [Accessed: Aug 5th, 2011].

[3] Muhammad Omer Chughtai, "Transport Protocol for Video Traffic over Mobile WiMAX", Master's Thesis, Department of Electrical Engineering, Comsats Institute of Information Technology, Islamabad, Pakistan, 2009.

[4] Randall Stewart, Paul D. Amer, Why is SCTP needed given TCP and UDP are widely available?, September 2007, http://www.isoc.org/briefings/017/, [Online], [Accessed: Aug 5th, 2011]

[5] Kim and S. Koh, "Analysis of handover latency for mobile IPv6 and mSCTP," *IEEE International Conference on Communication Workshops*, pp. 420-429, May 2008.

[6] L. Budzisz, R. Ferrus, A. Brunstrom, and F. Casadevall, "Towards transport-layer mobility: Evolution of SCTP multihoming," *Computer Communication*, vol. 31, pp. 980-998, Dec. 2008.

[7] J. K. Song, "Performance Evalution of Handoff between UMTS/802.11 Based on Mobile IP and Stream Control Transmission Protocol", Master's Thesis, Department of Computer Science, North Carolina State University, 2005.

[8] R. Stewart, Q. Xie, M. Tuexen, M. Maruyama, and M. Kozuka. (2007, Jun.) Internet-Draft SCTP Dynamic Address Reconfiguration. [Online]. http://tools.ietf.org/html/drafts/ietf-tsvwg-addip-sctp-22.

[9] S. J. Koh, Q. Xie, and S. D. Park. (2005, Oct.) Internet Draft mSCTP for IP Handover Support. [Online], http://www.ietf.org/proceedings/65/IDs?drafts-sjkoh-msctp-01.txt.

[10] S. Zeadally and F. Siddiqui, "An Empirical Analysis of Handoff Performance for SIP, Mobile IP, and SCTP Protocols," *Wireless personal communications*, vol. 43, pp. 589-603, 2007.

[11] I. Stoica, R. Morris, and D. Karger, "Chord: A scalable peer-to-peer lookup service for internet application," *ACM SIGCOMM Computer Communication*, vol. 31, pp. 149-160, 2001.

[12] S. Sarmady,"A survey on Peer-to-Peer and DHT", Grid Lab, School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia, 2007.

[13] K. Sethom, H. Afifi, G. pujolle, "PALMA: A P2P based Architecture for Location Management", IFIP *International Conference on Mobile and Wireless Communications Networks*, 2005.

[14] S. Cirani, L. Veltri ,"Implementation of a Framework for DHT-based Distributed Location Service", *International Conference on Software, Telecommunication and Computer Networks*, p 279-83, 2008.

[15] M. Zangrilli, D. Bryan, "A Chord-based DHT for Resource Lookup in P2PSIP", Internet Dreaft draft-zangrilli-p2psip-dsip-dhtchord-00, SIPeerior Technologies, February 25, 2007.

[16] A. A. S. R. Reaz, M. Atiquzzaman, and S. Fu, "Performance of DNS as location manager," *IEEE International Conference on Etectro Information Technology*, p. 6, May 2005.

[17] J. K. Song, "Performance Evalution of Handoff between UMTS/802.11 Based on Mobile IP and Stream Control Transmission Protocol", Master's Thesis, Department of Computer Science, North Carolina State University, 2005.

[18] S.Thomson, Y. Rekhter, Paul Vixie, "Dynamic Updates in the Domain Name System (DNS UPDATE)", RFC 2136

[19] O. Hanka, C. Spleib, G. Kunzmann, and J. Eberspächer, "A novel DHT-Based network architecture for the Next Generation Internet," *Eighth International Conference on Networks*, pp. 332-341, May 2009.

[20] F. Dabek, E. Brunskill, M. F. Kaashoek, and D. Karger, "Building peer-to-peer systems with chord, a distributed lookup service," *Proceeding of the eighth Workshop on operating System*, pp. 81-86, Aug. 2002.

[21] W. M. Eddy, "At What Layer Does Mobility Belong?," *IEEE Communication Magazine*,, vol. 42, no. 10, pp. 155-159, Oct. 2004.

[22] L. M. Campoy, " LTE-Advance Evolution to match Explosion of Mobile Data Traffic", 2nd International ICST Conference on Mobile Lightweight", Barcelone, Spain, 2010.

[23] H. Park, M. Kim, S. Lee, S. Kang, and Y. Kim, "A mobility management scheme using SCTP-SIP for real-time services across heterogeneous networks," *A mobility management scheme using SCTP-SIP for rea proceedingsof the 2009 ACM symposium on Applied Computing*, pp. 196-200, 2009.

Audio Watermarking with Error Correction

Aman Chadha, Sandeep Gangundi, Rishabh Goel, Hiren Dave, M. Mani Roja
Department of Electronics & Telecommunication
Thadomal Shahani Engineering College
Mumbai, INDIA

***Abstract*— In recent times, communication through the internet has tremendously facilitated the distribution of multimedia data. Although this is indubitably a boon, one of its repercussions is that it has also given impetus to the notorious issue of online music piracy. Unethical attempts can also be made to deliberately alter such copyrighted data and thus, misuse it. Copyright violation by means of unauthorized distribution, as well as unauthorized tampering of copyrighted audio data is an important technological and research issue. Audio watermarking has been proposed as a solution to tackle this issue. The main purpose of audio watermarking is to protect against possible threats to the audio data and in case of copyright violation or unauthorized tampering, authenticity of such data can be disputed by virtue of audio watermarking.**

Keywords- watermarking; audio watermarking; data hiding; data confidentiality.

I. INTRODUCTION

Over the years, there has been tremendous growth in computer networks and more specifically, the internet. This phenomenon, coupled with the exponential increase of computer performance, has facilitated the distribution of multimedia data such as images, audio, video etc. Data transmission has been made very simple, fast and accurate using the internet. However, one of the main problems associated with transmission of data over the internet is that it may pose a security threat, i.e., personal or confidential data can be stolen or hacked in many ways. Publishers and artists, hence, may be reluctant to distribute data over the Internet due to lack of security; copyrighted material can be easily duplicated and distributed without the owner's consent. Therefore, it becomes very important to take data security into consideration, as it is one of the essential factors that need attention during the process of data distribution. Watermarks have been proposed as a way to tackle this tough issue. This digital signature could discourage copyright violation, and may help determine the authenticity and ownership of an image.

Watermarking is "the practice of imperceptibly altering a Work to embed a message about that Work" [1]. Watermarking can be used to secretly transmit confidential messages, for e.g., military maps, without the fact of such transmission being discovered. Watermarking, being ideally imperceptible, can be essentially used to mask the very existence of the secret message [6]. In this manner, watermarking is used to create a covert channel to transmit confidential information [5]. Watermarking is an effective means of hiding data, thereby protecting the data from unauthorized or unwanted viewing. Watermarking is becoming increasingly popular, especially for insertion of undetectable identifying marks, such as author or copyright information to the host signal. Watermarking may probably be best used in conjunction with another data-hiding method such as steganography, cryptography etc. Such data-hiding schemes, when coupled with watermarking, can be a part of an extensive layered security approach.

To combat online music piracy, a digital watermark could be added to all recordings prior to release, signifying not only the author of the work, but also the user who has purchased a legitimate copy. Audio watermarking is defined as "the imperceptible, robust and secure communication of data related to the host audio signal, which includes embedding into, and extraction from, the host audio signal" [4]. Digital audio watermarking involves the concealment of data within a discrete audio file. Intellectual property protection is currently the main driving force behind research in this area. Several other applications of audio watermarking such as copyright protection, owner identification, tampering detection, fingerprinting, copy and access control, annotation, and secret communication, are in practice. Other related uses for watermarking include embedding auxiliary information which is related to a particular song, like lyrics, album information, or a hyperlink etc. Watermarking could be used in voice conferencing systems to indicate to others which party is currently speaking. A video application of this technology would consist of embedding subtitles or closed captioning information as a watermark [7].

II. IDEA OF THE PROPOSED SOLUTION

A. Process of Watermarking

The block diagram for watermarking is as shown below:

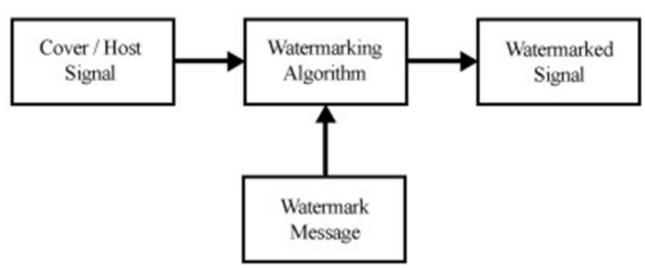


Figure 1. General watermarking block diagram

We can summarize the entire process of hiding and retrieving data as follows:

- Read the data to be hidden.
- Read the cover, i.e., host, in which data is to be hidden.
- Apply watermarking methods on the host.
- Hide the data in the host.
- Retrieve the original data at the receiver end.

B. Mean Square Error

Mean Square Error (MSE) [2], first introduced by C. F. Gauss, serves as an important parameter in gauging the performance of the watermarking system. The following factors justify the choice of MSE as a convenient and extensive standard for performance assessment of various techniques of audio watermarking:

1) Simplicity: It is parameter-free and inexpensive to compute, with a complexity of only one multiply and two additions per sample. It is also memory less, i.e., MSE can be evaluated at each sample, independent of other samples.

2) Clear physical meaning: It is the natural way to define the energy of the error signal. Such an energy measure is preserved even after any orthogonal or unitary linear transformation. The energy preserving property guarantees that the energy of a signal distortion in the transform domain is the same as that in the signal domain.

3) Excellent metric in the context of optimization: The MSE possesses the properties of convexity, symmetry, and differentiability.

4) Used as a convention: It has been extensively employed for optimizing and assessing a wide variety of signal processing applications, including filter design, signal compression, restoration, reconstruction, and classification.

MSE is essentially a signal fidelity measure [14],[15]. The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity/fidelity or, conversely, the level of error/distortion between them. Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors.

Suppose that x = {x_i | i = 1, 2, . . . , N} and y = {y_i | i = 1, 2, . . . , N} are two finite-length, discrete signals, for e.g., visual images or audio signals. The MSE between the signals is given by the following formula:

$$MSE(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \qquad (1)$$

Where,

N is the number of signal samples.
x_i is the value of the i^{th} sample in x.
y_i is the value of the i^{th} sample in y.

III. Implementation Steps

Audio Watermarking can be implemented in 3 ways:

- Audio in Audio
- Audio in Image
- Image in Audio

A. Audio in Audio

In this method, both the cover file and the watermark file are audio signals. The watermark signal must have fewer samples as compared to those of the cover audio signal. Further, this method can be implemented with the help of two techniques, namely, Interleaving and DCT.

1) Using Interleaving: It is a way to arrange data in a non-contiguous way so as to increase performance [3]. The following example illustrates the process of interleaving:

Original signal:	AAAABBBBCCCCDDDDEEEE
Interleaved signal:	ABCDEABCDEABCDEABCDE

In this technique, the samples of watermark audio are inserted in between the samples of the cover audio file [9],[10]. In terms of complexity, this is the simplest method of audio watermarking.

2) Using Discrete Cosine Transform: This technique is based on the Discrete Cosine Transform (DCT) [11]-[13]. In this technique, we take the DCT of both the cover audio and the watermark audio signals. Upon zigzag scanning, the high frequency DCT coefficients of the cover audio file are replaced with the low frequency DCT coefficients of the watermark audio file. During transmission, the Inverse Discrete Cosine Transform (IDCT) of the final watermarked DCT is taken. In this particular technique, since both the host, i.e., cover and watermark signal are in the audio format, we implement this method using a 1D DCT which is defined by the following equation:

$$F(u) = \sqrt{\frac{2}{N}}\sum_{i=0}^{N-1} A(i) * \cos\left(\frac{u(2i+1)\pi}{2N}\right) * f(i) \qquad (2)$$

Where,

$$A(i) = \begin{cases} \frac{1}{\sqrt{2}} & for\ u = 0 \\ 1 & otherwise \end{cases}$$

f(i) is the input sequence.

B. Audio in Image

This watermarking implementation uses DCT for embedding audio file in an image. Here we take DCT of both the cover image and the watermark audio files. The low frequency coefficients of both the DCT's are taken. The high frequency coefficients of the DCT of the image are replaced with the low frequency coefficients of the DCT of the watermark audio file. During transmission, the IDCT of the final watermarked DCT is taken. This technique involves both an audio signal (watermark) and an image (host). Hence, we implement this method using a 1D DCT for the audio signal and a 2D DCT for the image. Equation (2) defines a 1D DCT while the corresponding equation for a 2D DCT is defined by the following equation:

$$F(u,v)=\sqrt{\frac{2}{N}}\sqrt{\frac{2}{M}}\sum_{i=0}^{N-1}A(i)*\cos\left(\frac{u(2i+1)\pi}{2N}\right)*\sum_{j=0}^{M-1}A(j)*\cos\left(\frac{v(2j+1)\pi}{2M}\right)*f(i,j) \quad (3)$$

Where,

$$\mathrm{A(i)}=\begin{cases}\frac{1}{\sqrt{2}} & \mathit{for}\ u\ =\ 0\\ 1 & \mathit{otherwise}\end{cases}$$

$$\mathrm{A(j)}=\begin{cases}\frac{1}{\sqrt{2}} & \mathit{for}\ v\ =\ 0\\ 1 & \mathit{otherwise}\end{cases}$$

f(i, j) is the 2D input sequence.

C. Image in Audio

In this implementation, as deployed previously, DCT is used for embedding an image in an audio file. Here we take the DCT of both the cover audio and the watermark image files. This is followed by zigzag scanning so as to ascertain the low frequency and high frequency DCT coefficients. The high frequency DCT coefficients of the audio signal are replaced with the low frequency DCT coefficients of the watermark image file. While transmitting, the IDCT of the final watermarked DCT is taken.

Since this method is similar to the 'Audio in Image' technique with respect to the parameters involved, i.e., this technique also involves both an audio signal (watermark) and an image (host), hence, we may implement this method using a 1D DCT for the audio signal and a 2D DCT for the image. Equation (2) defines a 1D DCT while, (3) defines a 2D DCT.

IV. RESULTS

A. Audio in Audio

The spectra of the input, i.e., original cover and watermark audio signal are as shown below:

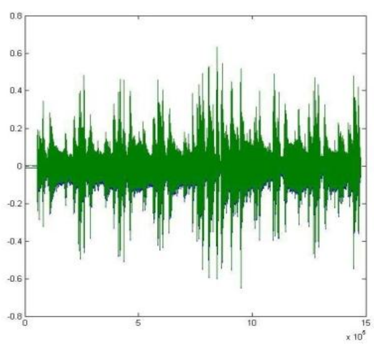


Figure 2. Original cover audio signal

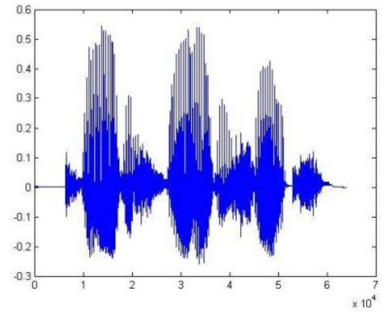


Figure 3. Original watermark audio signal

1) Using Interleaving: The spectra of the output, i.e., watermarked audio signal and the recovered audio signal, obtained by interleaving are as shown below:

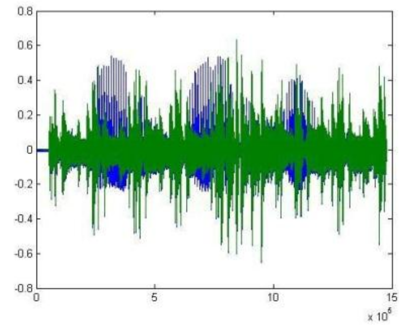


Figure 4. Watermarked audio signal

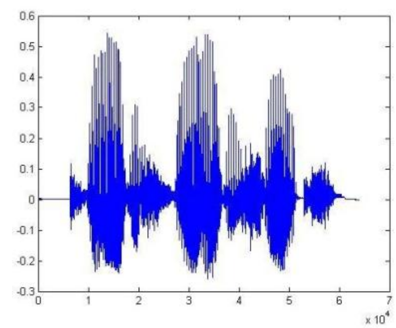


Figure 5. Recovered audio watermark signal

2) Using DCT: The spectra of the output, i.e., watermarked audio signal and the recovered audio signal, obtained by DCT based 'Audio in Audio' watermarking are as shown below:

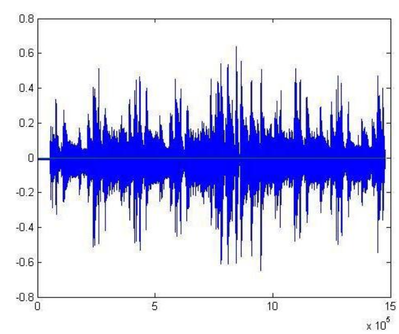


Figure 6. Watermarked audio signal

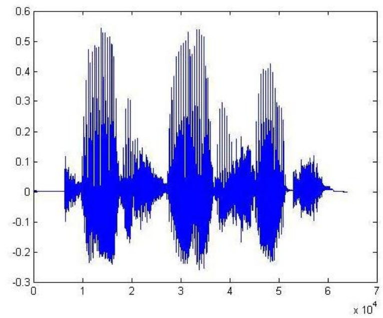


Figure 7. Recovered audio watermark signal

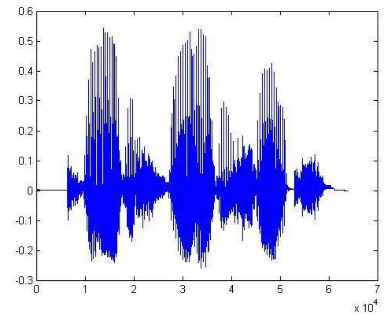


Figure 11. Recovered audio watermark signal

B. Audio in Image

The input, i.e., original cover image and spectrum of the watermark audio signal is as shown below:

Figure 8. Original cover image

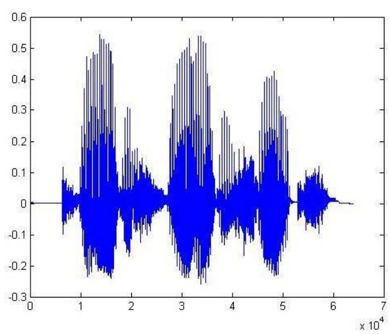


Figure 9. Original watermark audio signal

The output, i.e., watermarked image and the spectrum of the recovered audio signal, obtained by 'Audio in Image' watermarking is as shown below:

Figure 10. Watermarked image

C. Image in Audio

The input, i.e., spectrum of the original cover signal and the watermark image is as shown below:

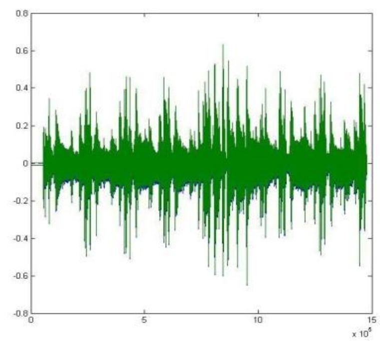


Figure 12. Original cover audio signal

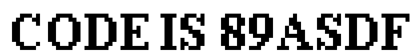


Figure 13. Original watermark image

The output, i.e., spectrum of the watermarked signal and the recovered watermark image, obtained by 'Image in Audio' watermarking is as shown below:

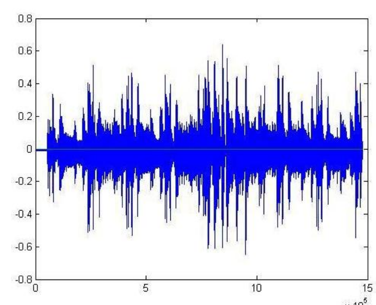


Figure 14. Watermarked image

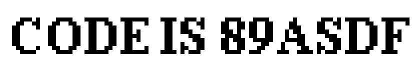


Figure 15. Recovered watermark

For all practical purposes, it is preferable to have a quantitative measurement to provide an objective judgment of the extracting fidelity. This is done by calculating the MSE in each case. Results, thus obtained, have been tabulated.

TABLE I. CALCULATION OF MSE WITHOUT NOISE

Types	MSE	
	Watermarked signal	***Watermark***
Audio in Audio (Interleaving)	3.5×10^{-4}	0
Audio in Audio (DCT)	3.4×10^{-3}	2.48×10^{-9}
Image in Audio	3.2×10^{-3}	0
Audio in Image	23.3	2.47×10^{-9}

V. ERROR CORRECTION USING HAMMING CODES

During transmission, when data travels through a wireless medium over a long distance, it is highly probable that it may get corrupted due to fluctuations in channel characteristics or other external parameters. Hence, hamming codes, which are Forward Error Correction (FEC) codes, may be used for the purpose of error correction in audio watermarking.

Hamming codes can detect up to two simultaneous bit errors and correct single bit errors; thus, reliable communication is possible when the "hamming distance" between the transmitted and received bit patterns is less than or equal to one [8]. In contrast, simple parity codes cannot correct errors and can only detect an odd number of errors. In this application, we are using the (15, 11) form of the hamming code. The improvement in the quality of received signal is clearly visible from the following waveforms:

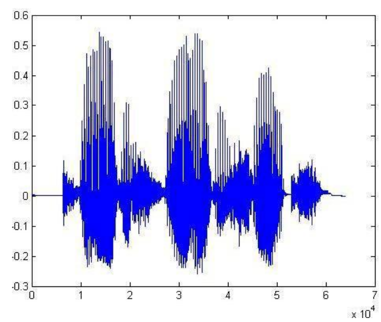


Figure 16. Desired audio signal

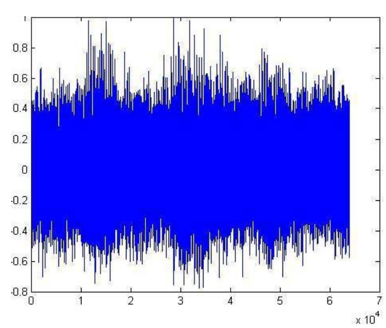


Figure 17. Recovered audio signal without Hamming Code

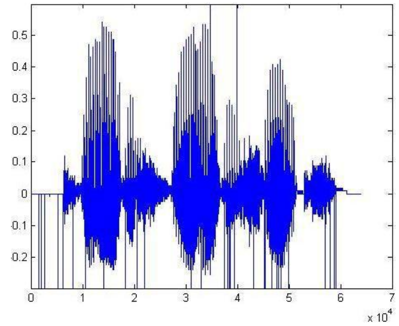


Figure 18. Recovered audio signal with Hamming Code

The corresponding results for Audio Watermarking with noise have been tabulated as follows:

TABLE II. CALCULATION OF MSE WITH NOISE

Types	MSE	
	Watermarked with noise	***Watermark with noise***
Audio in Audio (Interleaving)	0.010	0.010
Audio in Audio (DCT)	0.002	0.008
Image in Audio	0.003	1.92×10^{3}
Audio in Image	323.3	0

VI. CONCLUSION

DCT is an effective and robust algorithm for audio watermarking as the audio signal retrieved is clearly audible. Hamming Codes enable data correction in case of data corruption during transmission and help in recovering the original audio signal.

REFERENCES

[1] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography, 2nd Edition, Morgan Kaufmann, 2008, p. 31.

[2] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - A new look at signal fidelity measures," IEEE Signal Processing Magazine, Vol. 26, No. 1, pp. 98-117, January 2009.

[3] S. Shiyamala and Dr. V. Rajamani, "A Novel Area Efficient Folded Modified Convolutional Interleaving Architecture for MAP Decoder," International Journal of Computer Applications (IJCA), Vol. 9, No. 9, p. 1, November 2010.

[4] N. Cvejic, Academic Dissertation, Faculty of Technology, University of Oulu, "Algorithms for audio watermarking and steganography," p. 5.

[5] Shahreza S.S. and Shalmani M.T.M., "Adaptive wavelet domain audio steganography with high capacity and low error rate," Proceedings of the IEEE International Conference on Information and Emerging Technologies, (ICIET'07), pp. 1729-1732, 2007.

[6] Dr. H. B. Kekre and A. A. Archana, "Information hiding using LSB technique with increased capacity," International Journal of Cryptography and Security, Vol. 1, No. 2, p. 1, October 2008.

[7] M. Adya, M.Tech. Thesis, Indian Institute of Technology Kharagpur

(IIT-KGP), “Audio watermark resistant to mp3 compression,” p. 13.

[8] D. MacKay, “Information Theory, Inference and Learning Algorithms,” Cambridge University Press, September 2003 pp. 218-222.

[9] C. Zeng and Y. Zhong, "An interleaving-decomposition based digital watermarking scheme in wavelet domain," 8th International Conference on Signal Processing, Vol. 4, Nov. 2006.

[10] P. Agarwal and B. Prabhakaran, “Reliable Transmission of Audio Streams in Lossy Channels Using Application Level Data Hiding,” Journal of Multimedia, Vol. 3, No. 5, December 2008, pp. 1-8.

[11] J. Liu and Z. Lu, “A Multipurpose Audio Watermarking Algorithm Based on Vector Quantization in DCT Domain,” World Academy of Science, Engineering and Technology (WASET), Issue 55, July 2009, pp. 399-404.

[12] R. Ravula, M.S. Thesis, Department of Electrical and Computer Engineering, Louisiana State University and Agricultural and Mechanical College, “Audio Watermarking using Transformation Techniques,” pp. 8-51.

[13] Y. Yan, H. Rong, and X. Mintao, “A Novel Audio Watermarking Algorithm for Copyright Protection Based on DCT Domain,” Second International Symposium on Electronic Commerce and Security, China, May 2009, pp. 184 - 188.

[14] G. Casella and E.L. Lehmann, Theory of Point Estimation. New York: Springer-Verlag, 1999.

[15] T. N. Pappas, R. J. Safranek and J. Chen, “Perceptual criteria for image quality evaluation,” in Handbook of Image and Video Processing, 2nd ed., May 2005.

Development of a Computer Aided Transport Monitoring System (CATRAMS) for Manufacturing Organizations

Dipo Theophilus Akomolafe
Ph.D., MBCS, MNCS
Dept of Mathematical Sciences
Ondo State University of Science and Technology,
Okitipupa, Nigeria

***Abstract*—Presently, there are different types of monitoring systems and devices being used to monitor vehicles, products, processes and activities in manufacturing organizations. Each of these devices has their unique strengths and weaknesses but one problem that is common to them is that there is no knitted relationship between the devices and parameters necessary for effective and efficient monitoring system. Therefore, there is the need to develop a system that will address this shortcoming. CATRAMS is an integration of computer and communication facilities to monitor and control movement of vehicles and goods. Its objectives are to provide detail information on movement of vehicles and reduce likely operational delays associated with movement of vehicles and goods. The development of the system was carried out by studying some existing devices to know their limitations, designing of road transport database, specifying hardware requirements and integrating the hardware and software resources to make a complete system. The system was tested using data collected from some manufacturing industries and it was found out that detail information about movement of vehicles and goods could be provided by the system.**

***Keywords-* transport; manufacturing; database; hardware; communication, route; goods; passengers.**

I. Introduction

The benefits of transport monitoring systems and other systems based simply on mini and micro -computers particularly robotics are now widely recognized and used in all spheres of endeavour. In the area of transportation, computer is used to aid the initial techno-economic feasibility study, the actual design and construction of transportation means and other related activities. However, computer is yet to be fully integrated into monitoring of transportation operations in the manufacturing organization. This inadequacy is having negative impact on the productivity and capacity utilization of machinery and efficient and effective use of human resources. Manufacturing organization is an entity that is set up primarily to transform raw materials into finished goods. It is evident from the activities of this type of organization that its operations revolve around mobility.

Raw materials, processing equipment and employees are needed to be moved from one place to another so also finished goods must be appropriately delivered. It is therefore clear that transportation is crucial to the realization of their corporate goals and objectives.

It is visible that an organization cannot enjoy one hundred percent commitment from its employees generally especially those that are charged with the responsibility of transporting raw materials and finished goods from one place to another. Consequently, various measures are being put in place to address this problem by organizations but a deep study of some of these measures showed that they cannot totally curtail some of the practices of vehicle drivers. It is against this backdrop that this study is being carried out. The study is primarily intended to meet the challenges posed by moving both human and materials resources from one place to another.

Transportation system generally can be classified broadly into three types; land, air and water. The concern of this study is land transportation with emphasis on road transportation. Road transportation is the movement of goods and passengers from one place to another by a means which can be bicycle, motorcycle or vehicle. Road transportation has a multi-modal feature and because of this, it has the ability to penetrate every area and is the most commonly used mode of transportation in every part of the world and by extension the most widely used means of transportation in manufacturing organizations to move men and materials from one place to another. Road transportation can either be regulated or unregulated. It is regulated when the mode (bicycle, motor cycle and vehicle) of transportation is owned, managed and controlled by a manufacturing organization having more than five vehicles in their fleet while unregulated system refers to when organization rely on outside operator. This study is focused on regulated system because of its relative advantages over other modes, the possibility of defining definite route(s) of vehicle, ease of identification, flexibility in statistical analysis, ease of information, openness of operation, international coloration and possibility of modeling and simulation base on available statistics.

II. Objectives

Many problems are associated with road transportation and these ranges from uninformed road accident, diversion of products//commodities, faults, misappropriation of proceeds among others. These problems had wreaked havoc and will continue to have negative impact on the growth of manufacturing organization unless an effective system is put in place to prevent its occurrence where possible and reduce its effects. The primary objective of this study is to develop an effective and efficient system for monitoring of vehicles used in manufacturing organizations. Other objectives are to:

a) extend computer application to the area of monitoring of vehicles in a manufacturing organization;

b) demonstrate the possibility of using a computer program to control fleet of vehicles in a manufacturing organization;

c) use real time computing in transport monitoring and

d) provide effective and reliable means to remove delays and sharp practices of employees associated with transport operation.

III. METHODOLOGY

The development of CATRAMS involves three definite and distinct phases and these are the design of the software, hardware specification and the interconnection of hardware and software. Before going into the details of the design, it is imperative to examine some existing systems, computer network, communication facilities, monitoring and control devices in order to determine the platform for the development of CATRAMS and identify the computer and communication resources required for the system.

IV. DESCRIPTION OF DEVELOPMENTS IN COMPUTER AND COMMUNICATION

Computer Network is the interconnection of computer resources with communication system in order to achieve the full exchange of information within the network and allow sharing of resources. In recent time, computer network has gained wide acceptability in its application in business and science as a tool in the design and implementation of systems. However, the ability and capability of computer network, as a pre requisite for effective monitoring has over the years not been fully extended to monitoring of vehicles in the manufacturing sector. There had been some technological developments in computer and communication over the last few years that can be used to design and implement effective monitoring systems for road transportation. Among the recent developments are:

a) The introduction of digital transparent network, which meant that many of the functions that were hitherto carried out inside network could now be, carried out either inside or outside the network;

b) The range of services that could be offered by the administration had increased and this had come about because the switches that are now being employed in the network are controlled not by hardwired logic but by computers and this allows the exchange to perform many more functions in addition to switching simple cells.

c) The availability of wireless communication system.

Therefore the development of CATRAMS is made possible due to these advancements that simply involve among other things the integration of advances in computer networks into the operation of regulated transport system.

V. EXISTING SYSTEMS OF VEHICLE MONITORING

Monitoring in its real sense means the act of guiding a situation in order to prevent (or avoid) likely events from occurring in the course of the situation or to know that any of the likely event(s) has occurred and suggest probably way(s) to report any occurrence. Monitoring had been accepted as a scientific tool that is widely used to guide likely occurrence and to categorically state the required action to take in case the unexpected happens. Researchers and corporations had developed many systems that are used either on stand-alone or along with other device(s) for the purpose of monitoring. A careful observation of these systems generally showed that the process of monitoring basically involve communication between source and remote locations [10]. [7] identified the various procedures involved in the act of communication from source to remote locations and proposed that the effectiveness of any system developed for such purpose depends to a great extent on the communication components.

The various components of a network as identified by [4] include inter connection of computer, transmitter, transmission medium the receiver and other appropriate hardware depending on the type of network. These components are in conformity with identification proposed in [3]. The numerous existing monitoring devices were designed using different technologies and adopted diverse modes in their implementation. While some were implemented manually, some were activated electrically, robotically and mechanically. Among the existing devices are Automatic Point Location (APL) and Specific Time Alert (STA) [8] and enhancing road transportation through geospatial technology [1].

These and other monitoring devices that were not actually designed for vehicle monitoring in the manufacturing organization possess some negative effects which posed challenges for further researches into monitoring devices. Some of these effects are similar to those identified in [6] and these facilitated the design of Hazard Alert Device (HAD) by Bell Inc. The primary objective of HAD was to remove the negative effects of previous devices and provide effective monitoring device. The shortcomings of APL and STA are:

a) lack of communication between the base station and remote location

b) lack of adequate information on vehicles plying the road

c) inadequate information on goods in transit along the network

However shortly after HAD was introduced, it was discovered that the limitations of the devices it was supposed to overcome were not totally removed and in addition to this, the followings shortcomings were discovered:

(i) Limitation of coverage area
(ii) They were not a stand-alone system

(iii) They were developed with technologies that were gradually becoming obsolete
(iv) The system restrained its scope of applicability
(v) It supported only voice communication

VI. Components of the Existing System

Some of the existing systems basically involved the use of communication equipment that supports voice information within a specific area and within a pre-determined range. The major rationale for the design of most of the existing systems was to be able to communicate with vehicles within a certain range.

The concern for such design device was not to monitor vehicles throughout the course of the journey but to determine the probable arrival at a specified location. Automatic Point Location, which is a widely used system, is more of a security device than a monitoring system. Though it has the ability to indicate the particular location of vehicle at every point in time but cannot provide detailed information of the vehicle in case of accident. This feature is common to most of the existing devices; they have security capability but lacked functional database.

Walkie-talkie and radiophone are other devices for monitoring [9]. The application of these systems is simple and flexible. It involved communication-using walkie-talkie between radiophone room and drivers. However there is a limit to the coverage area of both radiophone and walkie-talkie among other deficiencies. These devices are wholly communication facilities and their actions and results can only be fed into the computer manually for storage and future references.

It is clear from the above analysis that monitoring systems generally are made up of either communication infrastructures or combination of this with computer facilities. Some of the existing devices were made using communication systems like transmitter, transmission medium and receiver integrated with computer or communication facilities alone. Therefore, the conception of CATRAMS is being directed towards addressing these limitations and to provide a stress and hitch free monitoring device that is wider in scope, applicability and coverage.

VII. Dessign

The design of CATRAMS involved three major phases; the design of software, hardware and integration of both. Succinctly, the system is basically inter-connection of computer resources with communication facilities. The computer resources required are computer hardware and software.

Computer Hardware: The computer hardware used in the design of the system were system unit with inbuilt wireless facilities and peripheral devices like input, output, printers and scanner, network cables and MODEM.

Software: The first phase was the design of the software. In designing the software, the first consideration was to examine the type(s) of information that could be exchanged on a network and these are data, sound and video. These forms had been examined in [1] and [2]. The software required for the system is the system software and application packages. The operating systems used are Windows Server 2003 for the servers and Windows XP for other client computers. The program application was developed using Visual Basic 6 (VB 6) language. Visual Basic 6 was chosen for the development and implementation of the application program because it provides fast and easy properties to create application for Microsoft Windows.

In addition it offers users with a complete set of tools to simplify rapid application development. Visual Basic is made up of two parts; Visual and Basic. The visual aspect generally refers to the method used to create the Graphical User Interface (GUI). Rather than writing numerous lines of code to describe the appearance and location of interface elements, it makes it possible to simply drag and drop rebuilt objects into a specific place on screen.

The Basic part refers to the Beginners All-purpose Symbolic Instruction Code) language, a language that is generally used by more programmers than any other language in the history of computing. Visual Basic has evolved from the original BASIC language and now contains several hundred, statements, functions and keywords, many of which relate directly to the windows GUI.

VIII. Program Description

CATRAMS is a system designed to provide up to date information on vehicles transporting goods and passengers from one location to another using any of the current systems especially those that works via satellite and radio communication as connectivity. It also established a computer aided monitoring system in which computers can be connected from different remote stations (clients) to the central computer system (server). This system runs on the network to provide the server with adequate and timely information that will trigger an action to be performed anytime there is a request and aid decision making.

CATRAMS is a window-based system designed to run on Windows 95/98, NT or ME developed using Visual Basic 6.0. The system runs with executable file (TMS.EXE) after which a welcome screen appeared and displays the system information and the author. After the welcome screen, the log in screen would be displayed. The login is a control measure to authenticate user id and password. When the user id is certified correct, the main menu will be displayed showing the following information:

A. GENERAL INFORMATION

1) *System Header*
2) *System Administration*
3) *Parameter setup*

B. TECHNICAL INFORMATION

1) *Vehicle Information*
2) *Drivers Information*
3) *Vehicle Movement Information*
4) *Goods/Passenger Information*

C. INPUT PROCESSING

1) *Data Input*

D. REPORT UTILITIES

1) *Backup Data Files*
2) *Restore Data Files*
3) *Reset Data Files*

IX. PROGRAM DESIGN

The program design utilized the visual design process, which is peculiar to VB generally, and the integrated development environment, which allowed the application program to be developed, run, tested and debugged.

VISUAL DESIGN PROCESS: DESIGNING CATRAMS GUI WITH VISUAL BASIC FOLLOWED THESE STEPS:

A. *Definition of requirements*

B. *Planning of the user interface which centered on the production of process plan/model of the interface, planning the user interface element and planning/implementing standards and mechanism*

C. *Building the user interface element which involved creating and modifying user interface element and functionality /*

D. *User feedback and monitoring. This involved prototype testing and monitoring and / application of user feedback*

The actual design of the program began with the gathering of data. These data were the components of information imperative for effective monitoring. At this stage, all the information gathered was analyzed; sorted and universal description was allotted to each of the situation to be monitored. The information included those relating to the organization, the driver(s), vehicle(s) goods/passengers and the journey. The following technical sub-headings were identified and used:

1) *Drivers*
2) *Vehicles*
3) *Journey*
4) *Goods/Passengers*
5) *Users log in*

All the above listed technical sub-headings and relevant data underneath each of them were modeled into a relation before their table structure were appropriately designed.

X. TABLE DEFINITION

The following tables formed from the identified headings were defined:

Drivers

This table contained relevant data on all the drivers engaged by the user organization. It showed the name, residential address, department in which the driver is attached in case of situation where driver may be borrowed for a particular assignment, date of birth, license number, date of issue and employment number among other relevant data included. The table is shown in Table 1

Vehicles

All data pertaining to each of the vehicles were described in the vehicle data table structure. The table contains information relating to the historical background and current information of all the vehicles and the status of their ownership. It clearly showed data, which clearly distinguished vehicles owned by the organization from those on hire. This is shown in Table 2

Journey

A brief summary of the journey being undertaken by all the vehicles at every point in time was defined in the table structure of journey. The table journey was specifically designed for the purpose of distinguishing the destination of each of the vehicles and purpose of the journey. In additionally, the journey data is the backbone of the monitoring system as it contained pertinent data relating to the journey and likely occurrences along the journey. In a nutshell, the table contains data that are being monitored by the system from the point of departure till destination. The table structure is as shown in Table 3.

Goods/Passengers

It is pertinent for an organization to maintain accurate record of goods and passengers. This is to enable the organization to have a compendium of information on goods and passengers for use, when:

a) *Unexpected occurrence like accident occurs.*

b) *Take accurate stock of goods and luggage.*

c) *Keep tracks of the goods and passengers' entrance and departure*

The table structure is shown in Table 4

The interfaces were designed using Visual Basic 6.0

The user interface consists of a set of screens. In other words, a screen is called whenever the user clicks on one of the application‘s forms. A screen can be composed of several components that constitute different parts of the screen content i.e. different record of the database. From the scenarios and function specifications discussed above, the following interrelated set of screens was developed:

Entry Login Screen: This screen is launched and accessed from the Internet, using a browser. It is a HTML page meant for controlling access to designated areas.

Vehicle Registration Screen: Here a user can view the list of all registered vehicles within the organization and other details of the vehicles.

Journey Screen: From this screen a user has access to all the fields that relate to any vehicle journey.

Drivers Screen: From the vehicle screen a user can enter the drivers screen on _Enter Driver from where he/she can search for any driver.

Beside these, other screens are goods, passengers and route

XI. IMPLEMENTATION

The required Software was installed on networked computers (servers and workstations) appropriately. The computers and other devices were connected together using appropriate cables. The table was populated by using the appropriate forms designed for that purpose. The following forms had been designed to allow easy entry of data; vehicle, driver, journey, passenger and goods. The user is merely expected to click any of the interface forms to display the appropriate form. At this level, the system makes use of Microsoft Access 2007 as the back engine. Information can be exchanged among facilities available within the level since the computer hardware used in each of the nodes had been interconnected to allow free communication.

Data Entry Forms

Data entry forms are the interfaces designed to allow users to enter, modify and delete data values of a record in the database. The data entry forms serve as a link between the user and the database files. It contains features such as text boxes, command buttons, pictures, charts, labels and other features. These features facilitate data update operations. The data entry forms designed in this research and which are for each of the tables are for the purposes of insertion of new records, modification of existing record, deletion of record, saving record, cancelling of operation and close operation .

Password Authorization System

Gaining access to the main program requires that each user must enter his or her name and password. Where invalid password is entered, access to the program is denied. The procedure of password verification involves matching the entered password with the stored password. If the name and password are valid, access to the program will automatically be granted but is denied if there is a mistake in any of them. The procedure of password creation is simple and interactive hence new users can always be added. Thereafter, any desired transaction can be carried out using the appropriate form

Hardware:

The hardware required and used for its implementation:

Digital camera: This is used to take the images of any scene that may occur during the journey

Multimedia Handset: This is another device that can be used to take images of any scene, recording and communicating with the base station

Computers: The computers used for the implementation of the system have a minimum 512MB Random Access Memory (RAM) and 45 G Hard Disk.

The system was tested using two different manufacturing companies and it was found out that:

a) Movement of vehicles was effectively monitored from the point of departure till they got to their final destination Information relating to goods in transit could be accessed at all the workstations

b) Accurate data relating to the passengers were made available immediately it was desired

c) Diversion of products by drivers was quickly detected

XII. CONCLUSION

The design of CATRAMS was compelled by the shortcomings of the present devices of monitoring. These devices served as the platform for the design of the system. A database of a manufacturing organization was proposed using relational database model. The application program was developed using Visual basic and was installed in computers in a network after the installation of the appropriate operating systems. The system was tested using real data from two manufacturing companies in different locations and the results showed significant improvement over the existing systems.

TABLE 1 DRIVERS' TABLE

No	NAME	TYPE	SIZE	KEY	DESCRIPTION
1	Drvnum	Text	15	Primary	Driver's employment number
2	Drvsname	Text	20		Driver's surname
3	Drvfname	Text	20		Driver' first name
4	Drvadd	Memo	0		Driver's residential address
5	Drvlcno	Text	25		Driver's license number
6	Drvlctype	Text	15		Driver's license type
7	Drvdept	Text	15		Driver's department

TABLE 2 TABLE DEFINITIONS FOR VEHICLE.

NO	NAME	TYPE	SIZE	KEY	DESCRIPTION
01	Vehno	Text	15	Primary	Vehicle Number
02	Typeofveh	Text	50		Type of Vehicle
03	Chasisno	Text	25		Chasis Number
04	Engine	Text	25		Engine Number
05	Yrofmk	Date/Time	8		Year of Make
06	Vehcolor	Text	12		Vehicle Colour
07	Inspol	Text	25		Insurance Policy
08	Spdometer	Long	4		Speedometer range
09	Maxspall	Long	4		Maximum speed allowed
10	Ownername	Text	30		Owner's Name
11	Date Hired	Date			Date Hired
12	H Duration	Long	2		Duration of Hire
13	HAmt	Long	10		Amount paid for hiring

TABLE 3 TABLE DEFINITION OF JOURNEY

NO	NAME	TYPE	SIZE	DESCRIPTION
01	Jnycode	Text	6	Journey code
02	Jnyppse	Memo	0	Journey purpose
03	Jnydepart	Text	25	point of departure
04	Jnydte	Date/Time	8	date of the journey
05	Jnydest	Date/Time	8	Destination
06	Jnytime	Date/Time	8	time of departure
07	Jnydist	Long	4	estimated distance of the journey
08	Ptmofarrvl	Date/Time	8	probable time of arrival
09	Tmofarrvl	Date/Time	8	actual time of arrival
10	Timediff	Date/Time	8	difference between probable time of arrival and actual time of arrival
11	Delay	Memo	0	type of delay
12	Causes	Integer	2	causes of delay

TABLE 4 TABLE DEFINITIONS FOR PASSENGERS

NO	NAME	TYPE	SIZE	KEY	DESCRIPTION
01	Vnum	Text	15	Primary	Vehicle number
02	Gd/Passallowm	Text	20		Goods and Male passengers Allowed
03	Gds/Passallf	Text	20		Goods and female passengers Allowed
04	Gd/Passdepartf	Text	20		Goods and female passengers at departure
05	Gd/Padepartm	Text	20		male passengers at departure
06	GdPassall	Text	20		Total goods and passengers Allowed
07	Drsnfordiff	Memo	0		reason for difference
08	Others	Memo	0		Others

REFERENCES

[1] Akinyokun O.C. (1994) Computers and Programming Using Fortran. Second edition. Ondo state New Era Commercial Press, Alagbaka G.R.A., Akure, Ondo State.

[2] Akomolafe, D.T. et all, (2009) "Enhancing road monitoring and safety through the use of geospatial technology," International Journal of Physical Sciences Vol.4 (5) pp 343-348

[3] Aram Falsafi, Kaveh Pahlavam, Ganing Young (1995) "Transmission Techniques for Radio LANS- A Comparative Performance Evaluation Using Ray Tracing" (www,cwins.wpi.edu)

[4] Falaki S.O. (2000). "Hardware Requirements for the Implementation of a Computer Network,". Course Material.

[5] Lacky, R. W., Salz,J.,and Weldon, E.J., Jr., (2003) "Principles of Data Communications", j McGraw-Hill Book Company, New York, Martin, James, (1998) "Telecommunications and the Computer," Prentice-Hall, Inc., Englewood Cliffs, N. J.,

[6] Miller, M.J., Vucetic B and Barry, L (1993) ―Satellite Communications Mobile and Fixed Services‖ Norwell, M A Kluiver Academic Publishers

[7] Miya, K. (ed.) (1975) "Satellite communications Engineering," Kyowa Book Company, Tokyo,

[8] Ola, J.S., (2007) ―Development of Specific Alert System for Road Transportation‖ International of Physical Sciences 207 Vol. 3(4)

[9] Taub, Herbert, and D. L. Schilling, (1971) Principles of Communication Systems, McGraw-Hill Book Company, New York.

[10] Wilkinson, R.I., (1956) "Theories for Toll Traffic Engineering in the USA," Bell System Tech.

The Performance between XEN-HVM, XEN-PV and Open-VZ during live-migration

Igli Tafa, Elinda Kajo, Ariana Bejleri, Olimpjon Shurdi, Aleksandër Xhuvani
Polytechnic University of Tirana,
Faculty of Information Technology, Computer Engineering Department,
Tirana, Albania

***Abstract*— The aim of this paper is to compare the performance between three hypervisors: XEN-PV, XEN-HVM and Open-VZ. We have simulated the migration of a virtual machine by using a warning failure approach. Based on some experiments we have compared CPU Consumption, Memory Utilization, Total Migration Time and Downtime. We have also tested the hypervisor's performance by changing the packet's size from 1500 byte to 64 byte. From these tests we have concluded that Open-VZ has a bigger CPU Consumption than XEN-PV, but the Total Migration time is smaller than in XEN-PV. XEN-HVM has a worse performance than XEN-PV, especially regarding to Downtime parameter.**

Keywords- Hypervisor; XEN-PV; XEN-HVM; Open-VZ; CPU Consumption; Memory Utilization; Downtime.

I. INTRODUCTION

One of the most interesting technologies in the field of information technology nowadays is Virtualization. This technology gives some advantages regarding cost, source and energy consumption, tolerance to failures, isolation to different attacks etc. Anyway, in this technology there are some black holes which have to do with the performance of the applications related to communication speed, sources or their energy consumption.

To realize a virtualization, it's needed to establish a hypervisor. The hypervisor is the administrator and the manager of the sources used by the virtual machines. The hypervisor can be established above the hardware and this is called full virtualization, or it can be established above the operating system and this is called OS virtualization. The full virtualization has got the advantage to integrate physical machines with different characteristics for example Intel x86 with AMD without doing any modification in Operating System's kernel. The performance offered by this type of virtualization is not high (i.e the communication with I/O devices is slow etc).

To increase the performance, para-virtualization approach is used. This approach requires the modification of the kernel of Guest Operating Systems. It is also required that the processors have the same characteristics. In this way, the communication between applications in virtual machines and I/O devices will be realized through virtual I/O drivers which rise above the hypervisor. This will give an increase of communication speed between applications and I/O devices.

One of the most important points in the technology of virtualization is live migration. This means that if a virtual machine which is running an application undergoes a discontinuity or it's CPU is loaded heavily, then the application can be migrated from this virtual machine to another. The virtual machine migration includes the transfer of page memories that are working; the transfer of the sources that are participating in this application (i.e network card, disc etc) and CPU's status. Each virtual machine has got its own CPU, its own physical memory (which is shared between different virtual machines), its own I/O etc. Memory migration is the most crucial point of virtual machine's migration. There are some methods of its migration, but the most used is the iteration method with Pre-Copy approach. In this method, at first the modified pages are transferred, which are registered from a table in memory which is managed by XEN. This table is created with bitmap method which marks with ''1'' anytime a memory page is written. The modified pages are iterated again until the end when CPU's status is transferred. In this method, the interruption time "downtime" because of the migration is not high, it is calculated in the order of milliseconds, despite of migration capacity.

In our article we have studied three types of hypervisors and we have measured their performance regarding the file's time of transfer, CPU efficiency and memory utilization from a virtual machine to another in the same physical host or between virtual machines in different physical hosts. The tested hypervisors are Xen-PV, Xen-FV and OpenVZ,. Xen hypervisor rises above the bare hardware, whereas OpenVZ is a hypervisor which rises above the Host Operating System.

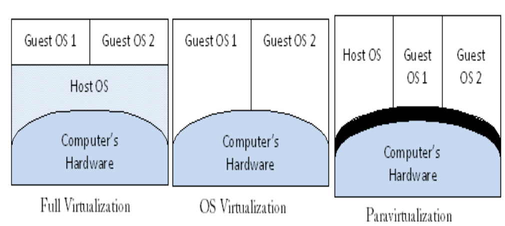


Figure 1. The three types of hypervisors

Xen Hypervisor can pass from a PV level to FV if we raise **Xen/Qemu**. This will make possible the emulation of I/O drivers in user's space. To achieve this is needed that the hardware supports this technology. Intel and AMD processors recently support virtualization (Intel VT dhe AMD-V). Using these processors we can raise the Full Virtualization technique.

This means that a GuestOS can be a Windows XP. These processors use VTX/SVM instructions. Generally Full Virtualization decreases the communication's performance with I/O disks. This technique used by Xen is often called Xen-HVM. Usually HostOS is called Dom0 and GuestOS is called DomU. Unlike Xen-HVM where an application uses two system calls to access a hardware driver, Xen-PV uses special calls who will offer the possibility to access virtual drivers who are managed by Dom0 and can be connected directly with the hardware. Referring to XenPV, the hypervisor lies in ring 0, whereas GuestOS lie in ring 1. The applications are in the third ring, the second ring is not used, like it's shown in Figure 2. In Xen-HVm the ring 0 is reserved for GuestOS and the virtualized hardware lies in ring 1.

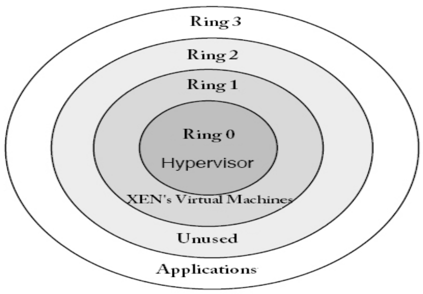


Figure 2. The rings of x86 architecture and the way they are used by XEN

OpenVZ is an OS Virtualization. GuestOS is called container or Virtual Private Server. Unlike Xen where each Guest has got its own kernel, in OpenVZ all the containers have one kernel in common with the HostOS. Anyway every GuestOS has got its own IP, its own I/O and its own memory. Since every GuestOS is a process in OpenVZ hypervisor, this method offers a better possibility than Xen in scalability but weaker in isolation. OpenVZ can modify the Linux's kernel giving to every unmodified Linux-based OS the possibility to be executed as a process in Linux.

All the above hypervisors support the SMP (Symmetric Multi-Processor) technique. This means that some GuestOS can use some host processors at the same time.

II. RELATED WORKS

In reference [15] is shown the difference between OpenVZ and XEN and is analyzed their performance. From the experiments it can be seen that OpenVZ has got a higher speed then Xen. The speed in reading is almost the same. OpenVZ has got a good performance because the hypervisor introduces a smaller complexity then XEN, expect that GuestOS in OpenVZ are treated like processes and have a common kernel with the host.

In [3] are used different measurements using different tools like **''pktgen''**, a module that incorporates in Linux's kernel and serves to generate traffic (packets with different sizes) from one host to another. Another benchmark called **'' stress tool''** is used to measure CPU consumption and memory utilization. It is noticed that when packet's size decreases from 1500B to 64B, it is not utilized the whole bandwidth offered for all the hypervisors, anyway OpenVZ has the best performance. In [3] is tested the case when are used different streams using packets with different sizes for both hypervisors from a virtual machine to another in a computer network connected with a gigabit switch. In all the tested cases, OpenVZ has the best performance.

In [2] is tested the performance between XEN and OpenVZ. In this system is built for the first time the multilayer approach where Web Server works in a layer, DB works in another layer and the PHP interface works in another layer. Using this multi-layer approach, the system's performance in scalability, isolation and speed is higher than in analogue cases [4]. The tool used to measure their performance is called **RUBIS**. When the number of the applications increases, the average response time of the packets RTT in OpenVZ is four times smaller than in the first case. This occurs because XEN has a bigger overhead than OpenVZ. Based on [2] there are a lot of miss cache for instruction in L2 cache. Anyway, from [2] is seen that OpenVZ consumes more CPU because of the common kernel between the host and guests and because of the fair CPU sharing between containers. To measure the overhead here is used the tool **''Oprofile''**. This tool generates data anytime there is a Hardware event, i.e anytime it happens a miss cache. This tool [5] is adapted to the XEN performance and is called **''Xenoprof''**. To measure the CPU consumption in XEN is used the tool **'' xentop –b''** which gives detailed information about the CPU consumption of every GuestOS. There isn't any specific tool to measure the CPU consumption of the containers in OpenVZ, anyway here the data here are measured from the generation of the time report which gives the time CPU spends in every container in /proc/vz/vstat. To measure the performance of every hypervisor is used **RUBIS** benchmark which can increase the CPU load increasing the number of threads generated by a script in C. Is measured the throughput, the response time of the packets and CPU consumption. In all of these three cases is concluded again that OpenVZ has the best performance.

In [6] is shown that the creation of a multilayered disc increases the migration performance of virtual machines. Xen has lower scalability than OpenVZ, because for OpenVZ the GuestOSs are processes, although they consume a lot of memory and processing [4]. In reference [8] is compared the performance of CPU consumption for the same applications between XEN-PV, Xen-FV. It is seen that XEN-PV consumes less CPU. To measure the CPU performance here is used the tool **''SAR''**. XEN-PV has the highest speed of writing SAS disc.

III. BACKGROUND

In this article we want to test three parameters: CPU Consumption, Memory Utilization and the Migration time of the virtual machine due to a warning failure. To create a warning failure is used a tool of CentOS 5.5 called **''Heartbeat''**. Using this tool we will get notified if a machine has ''dead'' and so the Hypervisor will migrate the applications (actually not only the applications) that were running in the ''dead'' machine to maintain the continuity. In our article we will realize a script in C which will simulate the stop of the virtual machine in the physical host, regardless the fact that in reality it hasn't stopped yet. So, the virtual machine in the first host will finish it's execution only in the moment that the virtual machine will start it's execution in the physical host where it is migrated. In this case the performance will be better than the case when ''Heartbeat'' acts normally [9]. The case of the study of an uncontrolled failure will be a study object in the future. However, we will examine the case when the virtual machine is relocated in the same physical host, without passing the network. The virtual machine's migration passes some steps:

a) The migration of memory pages that are in RAM who belong to the application that was being executed in the virtual machine.

b) The migration of the drivers of I/O devices.

c) The migration of virtual I/O discs as part of the activity of the virtual machine.

d) The migration of CPU-statuses.

This method is called pre-copy [10]. The purpose is that during the migration of the applications, to reduce the time of the interrupt down-time as much as possible. In pre-copy approach, the down-time is lower, but there is a problem with the total time of the migration as a result of the iteration of dirty pages, which are saved in a bitmap table in RAM. As we know all the virtual machines have the possibility to share the common memory, I/O discs, CPU etc and all these processes are managed by the hypervisor. The application that is going to be tested is a 180MB application played online (game). We will examine the parameters mentioned above using the hypervisors:

XEN-PV, Xen-FV, Open-VZ

We have used a computer and have exploited it in all the possible cases. The parameters of the computer we have used are:

Intel Core i7 920, Quad Core +, L2 4x256 KB, L3 = 8 MB, Asus, Three Channel DDR3 1600 Mhz, RAM 3x2GB, 64 bit processor, Hyperthread Technology, Freq 3.2 Ghz, VT Support, Turbo Boost Support.

We will start the experiment with XEN-PV and then with the two other hypervisors. The purpose is to find the hypervisor with the better performance during the migration of a virtual machine.

IV. THE EXPERIMENTAL PHASE

A. The simulation of warning failure in x0 virtual machine

Referred to figure 3, initially we will prepare a warning failure of x0 virtual machine. It means that x0 virtual machine in reality is operating, but the hypervisor and the other virtual machine built above the hypervisor are informed from heartbeat tool which is included in CentOS 5.5, that it is stopped as we explained in section 3.To simulate a warning failure we should create a script in C programming language and we call it ***heartcare***. This script is located in /proc and sends a message to heartbeat every time we want to execute it. At this moment heartbeat is informed for the virtual machine which should get interrupted (in our example it is x0 virtual machine), and at the same time heartbeat informs x1 virtual machine and Xen Hypervisor for this situation. Thus the hypervisor begins to migrate x0 virtual machine to x1 virtual machine based on pre-copy approach, which is explained in section 3.

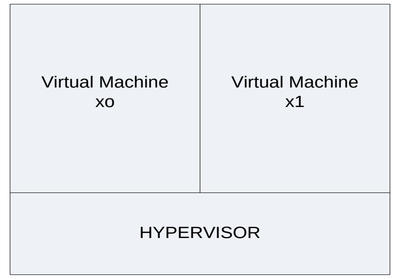


Figure 3. Two Virtual machines that Lay above the Hypervisor

B. Xen-PV

As we explained in section 3 initially we have installed Xen as hypervisor, above it is installed Dom0 with CentOS 5.5 version and 2 virtual machines DomU (GuestOS) each has Ubuntu 10.04 Server installed. In x0 virtual machine is executed a 180 MB application.

1) CPU consumption in Xen Hypervisor before and after the migration of x0

At first we will evaluate CPU consumption of Xen Hypervisor before migrating the x0 virtual machine. The migration occurs at the moment when a warning failure signal from ***heartcare*** script is sent to "Heartbeat" tool. To evaluate the CPU consumption in Xen first we have been located in /proc directory and typed the command xentop –b. The output results of this command are saved in a matrix form in a script called ***XenCProc*** which is located in /proc/xen. This script presents the CPU consumption for every 5 sec. At the moment when we push s key in the keyboard it will give us the average of CPU consumption up to this moment. The value is **2,23%**. This is because the resources, memory consumption, virtual disks I/O, virtual network etc are not being used heavily.

After x0 virtual machine is migrated, the CPU consumption at the first moment increases slightly, then it is increased up to **9,63 %**, in 1,65 sec; this is the peak of consuming, because of page faults. When the page faults increase, the CPU consumption increases too. This result depends from the iteration of dirty pages which are maintained by bitmap table in "Grant Shared Table" located in RAM and managed by the Hypervisor. As it looks in table 1 after 2,54 sec CPU consumption decreases to **3,11%**. After 3,66 sec CPU consumption is decreased to **2,11%**. This is the stabilized value. If we compare both cases before and after migration, the CPU consumption after the stabilization phase in the second case decreases up to 0,12%. The reason is the reduction of resources which were implemented to x0 virtual machine.

TAB 1. CPU CONSUMPTION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
2,23 %	0
9,63%	1,65
3,11%	2,54
2,11%	3,66

2) Memory utilization in Xen Hypervisor before and after x0 migration

To evaluate the memory utilization in Xen before migration we will use the tool named "MemAccess" located in /etc [11]. Initially the memory utilization is 10,6 %. After the migration of x0 virtual machine, memory utilization increases 10,7% for 1,55sec. This is the peak of memory utilization value. After that value the memory utilization will be stabilized at 10,5% at 2,04 sec (see table 2). If we compare the memory utilization inertia with CPU consumption, it is clear that the memory has more stability because of its native nature. During the migration in memory are just added some extra code (pages migrated from x0 virtual machine). This extra code is replaced in dirty bit map table located in Grant Shared Table.

TAB 2. MEMORY UTILIZATION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
10,6%	0
10,7%	1,55
10,5%	2,04

As it look from table 2 the stability of memory utilization happens after 2,04 sec from the migration process. If we compare table 1 and table 2 again, the peak of memory utilization happens after 1,55 sec while the peak of CPU consumption after 1,65 sec. This is because the iteration process does not affect directly to memory but it affects CPU consumtion. Also the CPU should adapt some additional parameters during the migration such as memory management, I/O disk refresh etc.

3) Average Total Time migration of x0 virtual machine to x1.

Initially we should clarify that the migration has occured in the same physical host. At the moment when ***heartcare*** script send a message to heartbeat tool to crash x0 virtual machine, a counter is programmed to start and it is implemented into that script. This counter will evaluate the total transferring time. At the end of migration another message is sent to heartcare script. This message is sent from ***XenCProc*** script because the last of phase of pre-copy migration is dedicated to CPU status of x0 virtual machine [10].

The CPU status can be identified using ***XenCProc*** script because the CPU status is the first argument saved in stack [12]. The ID of CPU status is in the end of the transfer. At the final transfer, ***XenCProc*** sends a message to ***heartcare*** script. The total time is shown in display. The average total transferring time in our test is evaluated **2,66 sec**. This is a very effective time, because the application installed is 180 MB (of course just a little size of this application is being transfered, because most of this is located in the hypervisor which is similar to a SAN device between two virtual machines, this is not part of our study).

4) Downtime during the migration of x0 virtual machine to x1 virtual machine.

This is a very critical case, because live migration phase depends from this parameter. To evaluate the downtime we will refer to XenCProc. Based on [10] downtime is evaluated as the transfer time of CPU status. Thus we should evaluate the total transferring time of Program Counter Register (also the same thing will be done with the execution instructions at the moment when the warning failure occurs in x0 virtual machine) of x0 virtual machine to x1 virtual machine. PC register is encapsulated in the shared memory of the Hypervisor. So in the same manner, with total transferring time transferring we should identify the last process.

As we know when an interrupt occurs, CPU saves its status and PC counter. So we should identify the ID of the first process. This ID is recorded in ***XenCProc*** at the moment when heartbeat stops x0 virtual machine, then it passes to /proc/xentop file.

The downtime algorithm is:

a) Heartcare sends a message to XenCProc

b) XenCProc saves the ID of the first process

c) Then we type xentop command

d) ID process is transported to xentop file

e) CPU status iss transferred, it send automatically a sys call to the hypervisor

f) Xen look the xentop file and starts the CPU status in x1 virtual machine

g) The downtime is saved at XenCProc

h) It is showed in display

The downtime is evaluated **4 ms**. It is a small value. There are some reasons:

a) We are doing a migration inside a physical host

b) CPU is very fast, see section 3.

c) There are some extra parameters such as Turbo BOOST

d) The application is not big (It can be considered small, only 180 MB)

e) There are no data dependency [13] etc.

Now we will repeat from a-d the experiments by changing the MTU (Message Transfer Unit). By changing the MTU value, the packet size will change automatically. It will affect the transferring time, downtime, memory utilization and CPU consumption too. The data packets are transferred from network virtual driver of x0 virtual machine to x1. Both virtual drivers form a team and are connected by a bridge soft which is managed by Xen. To change the packet data size we can change MTU from 1500B, which is standard of Ethernet Network Adapter, to 500 B and 64 B. For each VM we type the command:

Ifconfig eth0 mtu 500

This is a temporary value and we suppose that packet data size is 500 B. We should clarify that the results taken till now belong to the case when the packet data size is 1500 B.

MTU = 500 B

TAB 3. CPU CONSUMPTION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
2,26 %	0
11,47%	1,87
3,87%	2,96
2,24%	3,91

TAB 4. MEMORY UTILIZATION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
11,3%	0
11,9%	2,30
11,2%	2,41

MTU=64 B

TAB 5. CPU CONSUMPTION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
9,24 %	0
21,13%	2,78
16,32%	4,50
8,03%	5,21

TAB 6. MEMORY UTILIZATION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
11,9%	0
12,6%	3,42
11,6%	4,05

If we compare the tables 1-6, we see that the CPU consumption increases when packet data size decreases. The same thing happens with the memory utilization. The reason is the increasing of the overhead, because small packets have more context switch and more overhead [14].

TAB.7 THE AVERAGE TOTAL MIGRATION TIME AND DOWNTIME FOR DIFFERENT MTU SIZES

Packet data size	Average Total Migration time of x0_VM	Downtime
500 B	3,37 sek	6ms
64 B	5,12 sek	9 ms

C. Xen-FV

If we want to use Xen as a Full virtual machine we should have a hardware that supports it. As we see in section 3 the parameters of our computer match with our requirements [15]. Also we should built QEMU on Xen, thus we should emulate the hardware in user space [16],[17]. The Full virtualization in Xen has the same characteristics as VMWare which means that we can build OS with different native nature and different architecture, such as Windows in DomU. Also in Full virtualization it is not necessary to modify kernel OS Host or Guest. Nevertheless Full Virtualization has some disadvantages such as the increase of access time in I/O disks, because there are 2 trap instructions to access a disk [18]. The Full virtualization includes an additive complex layer presented by QEMU emulation software. In order to emulate network drivers in both GuestOS we should install e1000 emulator in /root directory.

MTU 1500 B

TAB 8. CPU CONSUMPTION IN XEN-HVM AFTER X0 VIRTUAL MACHINE MIGRATION IN X1.

CPU rate consuming (%)	Time (sec)
2,66	0
10,52	2,14
4,08	3,25
3,16	4,31

TAB 9. MEMORY UTILIZATION IN XEN-HVM AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
11,5	0
11,9	2,26
11,5	2,84

MTU = 500 B

TAB 10. CPU CONSUMPTION IN XEN-HVM AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
3,72	0
14,59	2,51
6,11	4,33
3,6	5,70

TAB 11. MEMORY UTILIZATION IN XEN-HVM AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
12,4	0
13,2	2,92
12,4	3,54

MTU=64 B

TAB 12. CPU CONSUMPTION IN XEN-HVM AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
11,8	0
24,5	3,91
19,1	6,84
11,7	8,16

TAB 13. MEMORY UTILIZATION IN XEN-HVM AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
13,02	0
13,7	5,7
12,9	6,16

If we compare the tables 8-13, the CPU consumption and memory utilization in Xen-HVM are bigger than in Xen-PV.

TAB.14 AVERAGE TOTAL MIGRATION TIME AND DOWNTIME FOR DIFFERENT MTU SIZE

Packet data size	Average Total time migration of x0_VM	Downtime
1500 B	4,10 sec	8 ms
500 B	5,24 sec	11 ms
64 B	6,08 sec	16 ms

As it look from the table 14, downtime and Average Total time are increase when the number of packet size is decrease.

D. OpenVZ

1) The Evaluation of CPU consumption in OpenVZ

To evaluate the CPU consumption in OpenVZ we don’t have any specific tool nevertheless we can measure the CPU wasted time in /proc/vz/vstat. To evaluate the CPU consumption we create a script in C which is called **traceproc.**

It traces the active and idle processes in hypervisor by scanning the status of each process in vstat file. Each process has a wake bit in Process Status Register, if it is 1 this process is active and if it is 0 the process is idle. In **Traceproc** script located in /proc/vz we have implemented a formula:

The availability of the process= (Time for each active process)/(Total CPU time) x100% (1)

The sum of the availability active processes = CPU Availability (2)

In reality this formula doesn't calculate the CPU availability, because when the processes are idle they still spend CPU time, consequently their consume CPU. Thus for the idle process we should build a semaphore variable [7] in order to make them sleep. In this way they will not consume CPU. Semaphore variables are built in a script in C called **semaphore**, which records the ID of all idle processes. This information is taken from **Traceproc** script. For each passive process we generate a thread which sends a signal to these processes.

In this manner, the passive processes are transformed in sleep processes. At the moment when CPU sends an interrupt message for one of the sleeping processes, the semaphore script is the first that takes this signal. This script reads the ID of calling processes, records it in a specific address into a specific register and then calls the specific thread. The thread wakes up the sleeping process. Thus the process can take the interrupt launched from CPU. This is a very dangerous approach because the script is implemented in user space, it means that after the interrupt request from CPU, the generated thread can't wake the process up. So the process is going to sleep forever. Nevertheless after this modification to **traceproc** script we will evaluate the CPU consumption by using the formula:

CPU consumption= Sum of active processes/ Total nr of process (3)

We should emphasize that this script gives us an approximately value of CPU consumption in OpenVZ hypervisor.

2) The evaluation of Memory Utilization in OpenVZ

We have to use a tool named **stream_tool** [8] to evaluate the memory utilization. There is one problem, this tool cannot evaluate the dynamic changing i.e the iteration of dirty pages while the x0 machine migrates to x1. So we should build a script that finds the number of page faults and multiples them with the page size.

Nevertheless we cannot find the appropriate number of transferred pages in a unit of time in case a page miss occurs. So we should implement another tool called **Bonnie ++**, which calculates the bandwidth transfer for 2 disks. We take RAM_VM1 as first disk and RAM_VM2 as second disk and we can calculate the total number of transferred pages for each iteration by using the formula:

The nr of transferred pages= Total size transferred (B)/ Page size (4)

The calculated from **stream** benchmark at 0 time:

Total memory utilization= (Time before a page fault occurs) + (Nr of transferred page while a page fault occurs) x (Nr of page faults) x (Page size) (5)

All these formulas are implemented in **Mem0 script**, written in C language.

3) The evaluation of transferring time and downtime

To evaluate the transferring time we can use the same script we did in previous cases, but this script is located in /proc/vz.

MTU 1500 B

TAB 15. CPU CONSUMPTION IN OPENVZ HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consumption (%)	Time (sec)
2,24 %	0
9,67%	1,52
3,20%	2,24
2,18%	3,04

TAB 16. MEMORY UTILIZATION IN OPENVZ HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
10,8%	0
10,9%	1,42
10,8%	1,57

MTU = 500 B

TAB 17. CPU CONSUMPTION IN OPENVZ HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consuming (%)	Time (sec)
2,28 %	0
11,48%	1,77
3,88%	2,91
2,26%	3,84

TAB 18. MEMORY UTILIZATION IN OPENVZ HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
11,2%	0
11,7%	1,59
11,2%	2,22

MTU=64 B

TAB 19. CPU CONSUMPTION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

CPU rate consuming (%)	Time (sec)
9,76 %	0
22,53%	2,1
17,11%	3,29
8,62%	4,27

TAB 20. MEMORY UTILIZATION IN XEN HYPERVISOR AFTER X0 VIRTUAL MACHINE IS MIGRATED TO X1.

Memory Utilization	Time
11,9%	0
12,8%	3,11
11,6%	3,58

If we compare tables 15-20 in OpenVZ, the CPU consumption and memory utilization is just a little bit more than the parameters in Xen, the reason is that all the Containers and Hosts share the same fair resources such as CPU, but the transition time in OpenVZ is smaller than that on Xen-PV, because in OpenVZ each container is considered a process.

TAB.21 AVERAGE MIGRATION TOTAL TIME AND DOWNTIME WITH DIFFERENT MTU SIZES

Packet data size	Average Total Migration time of x0_VM	Downtime
1500 B	2,06 sec	3 ms
500 B	2,57 sec	4 ms
64 B	3,72 sec	5 ms

In Tab 21 is presented the Average Total Migration time of x0_VM is smaller than that of XEN. The same thing happens with downtime. The reason is the low complexity of OpenVZ, Overhead and Context Switch; because each container is considered a process.

V. CONCLUSIONS

From the above experiments we conclude the following results:

1) CPU Consumption and Memory Utilization in XEN-PV are lower than in Open-VZ because Open-VZ uses the same kernel for Host OS and Guests OS by fair sharing the CPU. XEN has got a better sharing of the CPU between Host OS and Guests OS.

2) XEN-HVM consumes more CPU because of the emulator's complexity (QEMU).

3) All the parameters for the three hypervisors increase with the decrease of packet's size. This causes a slower performance.

4) The Total Migration Time and Downtime are smaller in Open-VZ than in XEN because in OPEN-VZ the overhead is smaller (every OS works as a process).

VI. FUTURE WORKS

As a future intention we would want to test and compare the performance for five hypervisors XEN-HVM, XEN-PV, Open-VZ, KVM-FV and KVM-PV in a LAN. Also we will test these hypervisors not using a warning failure, but simulating an unwarning failure.

REFERENCES

[1] Jianhua Che, Qinming He, Kejiang Ye, and Dawei Huang, "Performance Combinative Evaluation of Typical Virtual Machine Monitor", 2010

[2] Daniel Schlosser, Michael Duelli, and Sebastian Goll,, "Performance Comparison of Hardware Virtualization Platforms" , 2010

[3] Pradeep Padala, Xiaoyun Zhu, Zhikui Wang, Sharad Singhal, Kang G. Shin, "Performance Evaluation of Virtualization Technologies for Server Consolidation", 2007

[4] Christopher Clark, Keir Fraser, Steven Hand, Jakob Gorm Hanseny,Eric July, Christian Limpach, Ian Pratt, Andrew Warfield, "Live Migration of Virtual Machines" , 2009

[5] A. Menon, J. R. Santos, Y. Turner, G. J. Janakiraman,and W. Zwaenepoel, "Diagnosing performance, overheads in the Xen virtual machine environment.In *Proceedings of the First ACM/USENIX*

International Conference on Virtual Execution Environments(VEE), pages 13–23", 2005

[6] Katharina Haselhorst, Matthias Schmidt, Roland Schwarzkopf, Niels Fallenbeck, Bernd Freisleben, "Efficient Storage Synchronization for Live Migration in Cloud Infrastructures", 2010

[7] Andrew Tanenbaum, Modern Operating System 4-th edition, chap 2, Semaphores, 2009

[8] Lucas Nussbaum, Fabienne Anhalt, Olivier Mornard, Jean-Patrick Gelas "Linux-based virtualization for HPC clusters", 2007

[9] Espen Braastad, "Management of high availability services using virtualization", 2006

[10] Michael R. Hines, Umesh Deshpande, and Kartik Gopalan, "Live migration with post copy", 2007

[11] Jin Heo Xiaoyun, Zhu, Pradeep Padala, Ann Arbor, Zhikui Wang, "Memory Overbooking and Dynamic Control of Xen Virtual Machines in Consolidated Environments ", 2009

[12] Andrew Tanenbaum, Modern Operating System 4-th edition, chap 2 processes, 2009

[13] John Henessy, David Peterson, Computer Organization and Design 4 –th edition, chap 4, p. 330, "Pipeline Data Path and Control", 2010

[14] Andrew Tanenbaum, 2009, Modern Operating System 4-th edition, chap 3, Size of Page Memory

[15] Nathan Regola, Jean-Christophe Ducom, "Recommendations for Virtualization Technologies in High Performance Computing" , 2010

[16] Chunqiang Tang, "FVD: a High-Performance Virtual Machine Image Format for Cloud", 2008

[17] Daniel P. Berrang´e, "Taking full advantage of QEMU in the Xen user-space", 2007

[18] Andrew Tanenbaum, 2009, Modern Operating System 4-th edition, chap 1, Virtual Machines

AUTHORS PROFILE

Igli TAFA – Assistant Pedagogue in Polytechnic University of Tirana. PhD Student. His interesting field is Virtual and Cloud Technology.

Elinda KAJO – Pedagogue in Polytechnic University of Tirana. She is graduated in PhD at 2008. She is interesting in JAVA Technology and C Programming

Ariana BEJLERI - Pedagogue in Polytechnic University of Tirana. She is interesting in Programming Languages Technology

Olimpion SHURDI - Pedagogue in Polytechnic University of Tirana. He is very interesting in Hardware and Computer Network Technology

Aleksander XHUVANI - Professor in Polytechnic University of Tirana. He is lider of Software Laboratory in University. His interesting field is Software Engineering.

An Ontology Based Reuse Algorithm towards Process Planning in Software Development

Shilpa Sharma
Sanghvi Institute of Management and Science
Indore, India

Maya Ingle
Indore Institute of Computer Applications
Indore, India

Abstract— **The process planning task for specified design provisions in software development can be significantly developed by referencing the knowledge reuse scheme. Reuse is considered to be one of the most promising techniques to improve software excellence and productivity. Reuse during software development depends much on the existing design knowledge in meta-model, a "read only" repository of information. We have proposed, an ontology based reuse algorithm towards process planning in software development. According to the common conceptual base facilitated by ontology and the characteristics of knowledge, the concepts and the entities are represented into meta-model and endeavor prospects. The relations between these prospects and its linkage knowledge are used to construct an ontology based reuse algorithm. In addition, our experiment illustrates realization of process planning in software development by practicing this algorithm. Subsequently, its benefits are delineated.**

Keywords- Ontological Knowledge Modeling; OntoReuseAlgo; Ontolayering Principle.

I. Introduction

Reuse practice has become a significant issue in the field of software engineering. The increasing demand for progressively larger and complex systems is one of the causes that reuse has been preferred when dealing with such tribulations in software development. Reuse also responds to an increasing insist for highly reliable, high excellence and less expensive systems.

Accordingly, knowledge reuse will benefit and improve the process planning in software development greatly. Process planning is an intermediate phase between design and implementation. Lucidity and prescribed specification of concepts play a key role in the inclusion of reuse in the process planning. Therefore, a most important issue is to build a common conceptual base characterized by knowledge. Our exploration focuses on this task through the use of defined system elements and layering principal, illustrating various prospects. We have proposed a knowledge based process planning method and it seems indispensable and potential in real applications.

As to knowledge-based process planning, a significant prose already exists for general software implementation processes. Ontology is the basis for the sharing of resources, which provides a common understanding of field knowledge. The scope of ontology application is widespread, and its pursuit is large and whole, and applicable to entire software engineering environment [1] [2]. Therefore, we construct and use a fixed ontology oriented procedure through locating the linkage knowledge. When executing it, irrelevant knowledge items would consequently be avoided in knowledge retrieval so as the effectiveness and efficiency of knowledge reuse may be improved. This may help train novice users to quickly become familiar with the process and achieve efficient and high quality builds.

Ontology has been defined as the vocabulary of unambiguous domain-related concepts and meanings of the concepts anchored in consensus domain knowledge [3]. Also, ontology may provide a reusable and sharable information model. It is popular in engineering knowledge modeling and retrieval for its devastating searching ability [4]. Thus, we have proposed an ontological knowledge modeling in Section II that illustrates the meta-level descriptions to divulge the building blocks of it. Based on ontological knowledge modeling, Section III confers ontology based reuse algorithm towards process planning in software development. Section IV exemplifies a case study to substantiate the efficacy of ontology based reuse algorithm. Finally, we conclude with the advantage of practicing it towards process planning in software

II. Ontological Knowledge Modeling

We have proposed an ontological knowledge modeling for knowledge integration and reuse towards process planning in software development. It constitutes *System Element Classification*, *Ontolayering Principle* and *Knowledge Reuse Scheme* to provide reusable and shareable engineering applications. S*ystem Element Classification* is developed to capture important characteristics to reduce the growing complexity of information and increasing need to exchange it among various software applications. Subsequently, the *Ontolayering principle* uses the ontology in a resource usage manner, specifically by understanding and dissimilating the information comprised by entities. Conversely, natural language is too ambiguous to provide enough and clear definition in engineering applications. Thus, *Knowledge Reuse Scheme* is defined consistent interpretation between different users and different applications. These building blocks are defined as follows:

A. System Element Classification

The classification includes abstract concepts such as work units, stages, work products, model units and people as shown in Fig. 1. Work units are the tasks or activities that software developers perform, and have a start and end time as well as duration. Stages are major time frames, helps the wok units in giving some temporal structure. Work products, such as documents or software, are intangible results of performing work units and have creations and last change times with status. The status of work product is always one of the enumerated list such as Initial, Complete, Accepted or Approved. Finally, the producers are the people and teams that actually perform the work units in order to create work products.

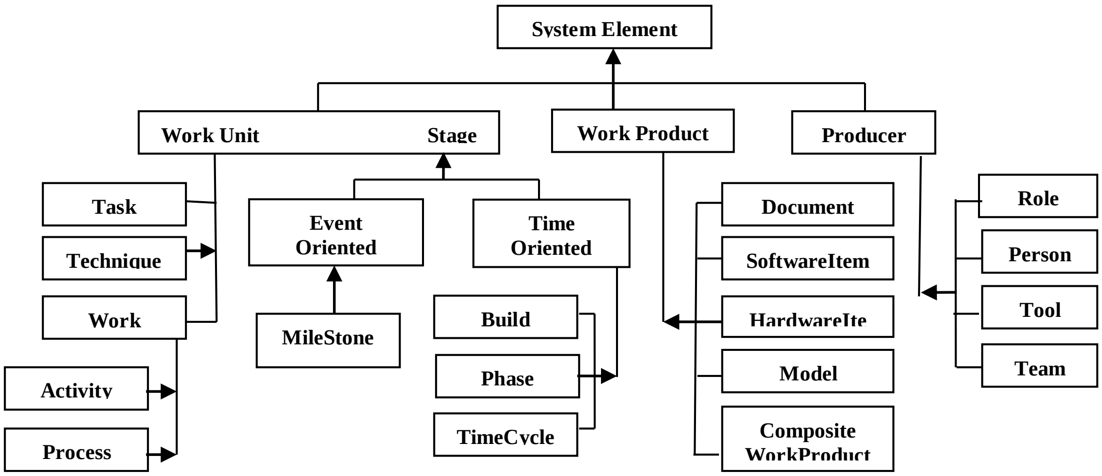


Figure 1: System Element Classification

B. *Ontolayering Principle*

The principle uses the ontology in a resource usage manner specifically by understanding and dissimilating the information comprised by entities [5]. The three prospects namely: Metamodel, Process and Product prospects have been defined around the communities that network with the ontology as shown in Fig. 2. The meta-model prospect acts as a common standard determining the other prospects. The meta-model is intended to be used as an origin by method engineers so that the methodologies can be developed. Method engineers typically uses the concepts in the meta-model prospect by sub typing and instantiation, thereby creating new concepts (subtype of existing ones) and entities (instances of concepts). All these new concepts and entities created by method engineers are seized to form a process prospect. Software developers use it by creating the instances of concepts in it and also, by following the guidance explained by entities. Thus, the instances created by software developers are apprehended to form the product prospect.

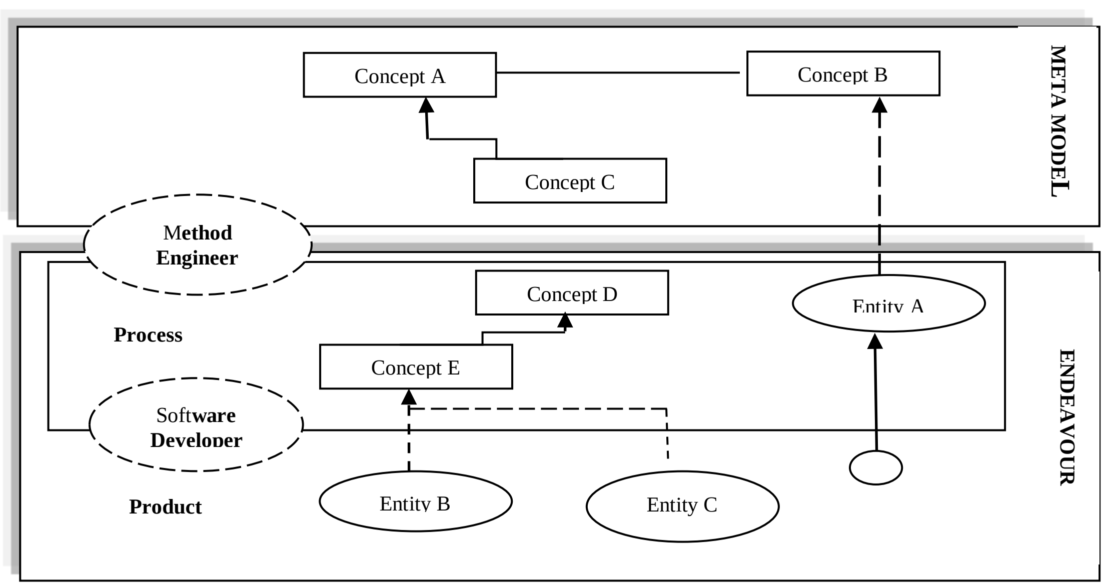


Figure 2: *Ontolayering* Principle

C. *Knowledge Reuse Scheme*

The scheme starts with formalizing the system element requirements according to the representation approach of System Element Classification. Subsequently, identification of the related process concepts and entities that need to be revised according to the *Ontolayering* principle stored in the knowledge base. Consequently, modification of the producer entities, associated concepts based on Ontolayering principle and revision of the influencing product attributes conceded. Finally, simulation of the results is done if the reuse requirements are satisfied; else consider changing a different process concept and entities.

III. ONTOLOGY BASED REUSE ALGORITHM (*ONTOREUSEALGO*)

We have proposed an Ontology Based Reuse Algorithm (*OntoReuseAlgo*) using ontological knowledge modeling approach to aid the product redesign of process plans. The key idea is that a new process plan under new implementation requirements may be obtained by modification of certain concepts and entities of the current process. Ontological knowledge modeling is used to give a uniform representation of the involved information. It starts with understanding the system elements. It includes identifying process concepts and entities that need to change followed by altering with *Ontolayering* principle and modify the producer entity and the associated concepts which help in revising the influencing product attributes for simulating the final process plan. Fig. 3 shows the overall procedure of the proposed approach and the knowledge reuse strategy in process planning.

Different systems may use different concepts and terminology to express the same thing while the same words may be used to represent different things by different systems. Both situations hinder information communication. Therefore, step 1 uses the *System Element Classification*. Conversely, the process and product concepts that need to change could be properly identified and revised through mappings of related *Ontolayering Principle* to certain process and product prospects in step 2, 3 and 4. Subsequently, natural language is too ambiguous to provide enough and clear definition in engineering applications. In addition, text-based definitions are too rigid and insufficient in information retrieval because it necessitates text-match searching. Thus, *Knowledge Reuse Scheme* is applied to step 5 for providing vocabulary and the specification of the meaning or semiotics of the terminology within this vocabulary.

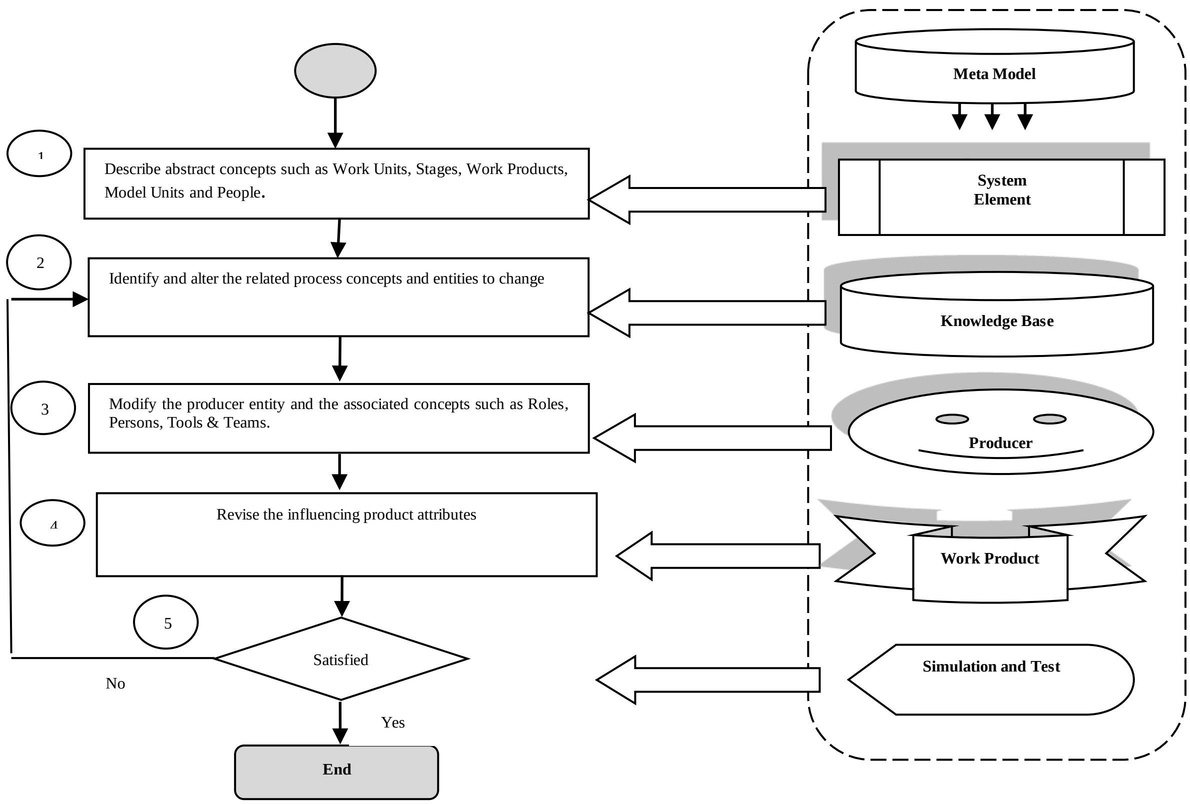


Figure 3: *OntoReuseAlgo*: Reuse Approach for Process Planning

There are 5 steps from the beginning to the end of process planning task, as shown in Figure 1

Step 1: Formalize the system element requirements according to the representation approach of system element classification

Step 2: Identify the related process concepts and entities that need to be revised according to the layering principle stored in the knowledge base

Step 3: Modify the producer entities and associated concepts based on layering principle

Step 4: Revise the influencing product attributes

Step 5: Simulation of the results in step 4 if the reuse requirements are satisfied, then shift to step 5; or else, shift to step 2 and consider changing a different process concept and entities

IV. CASE STUDY

In this section, the design of process planning for redistribution between warehouses is described, based on the proposed reuse approach an *OntoAlgoReuse*. Various people are responsible for carrying out different activities such as Foreman is responsible for one warehouse though Warehouse worker works in a warehouse for loading and unloading subsequently Truck driver is accountable for transportation. In addition, Forklift operator drives a forklift in one warehouse and Official personal receive orders and requests from customers.

Step1- We have analyzed following set of system element requirements, when a foreman performs redistribution of items between the warehouses.

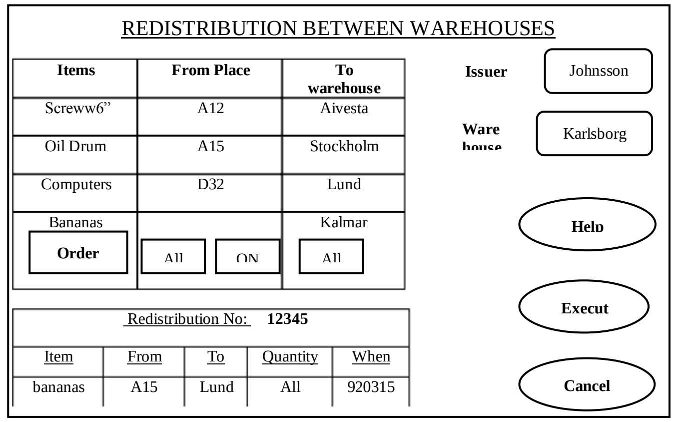


Figure 4: Redistribution between Warehouses Window

TABLE I : SYSTEM ELEMENT REQUIREMENTS OF RETRIEVED CASE

S.No	System element Requirement	Description
1	**Initialization** (when foreman request for the redistribution)	1. The foreman gives a command for redistribution between warehouses
		2. The window in Fig. 4 is presented to the foreman
		3. The items can be ordered in a number of ways with ORDER menu such as alphabetical, index, turnover of the items and storing order.
		4. In the 'From place' table we may choose to view either all places in the current warehouse or, if we have selected an item, the place where the item exists
		5. In the 'To warehouse' table we may select all warehouses or the warehouses that we have to transport to this week
		6. The 'Issuer' and 'warehouse' fields are automatically filled when the window pops up.
		7. The foreman selects an item by pointing to it and dragging it to the Redistribution form then selects from which place to take the items and to which warehouse to transport them.
		8. The foreman then gives the quantity to be moved and the date.
		9. It is possible to change the information when the form has been edited. When the foreman EXECUTES the redistribution, the transport is planned. It is also possible to CANCEL the redistribution. Selecting HELP shows window of information about the current window.
2	**Loading** (when truck fetches the item from the warehouse)	1. A Truck driver asks for a transportation request. The request is marked as ongoing.
		2. Give an appropriate request to the Forklift operators to have the items ready when and where the truck is expected.
		3. When the Warehouse Worker gets a request to fetch items at appropriate time, orders Forklift operators to move the items to the loading platform
		4. When the Truck driver arrives the items are loaded. The Truck driver tells the system when the truck is loaded and when it is expected to be at the new warehouse.
		5. Decrease the number of items in this ware house and mark the transport request as on transport
3	**Unloading** (when a truck delivers the items to the new warehouse)	1. When the truck has arrived at the new warehouse, the items are unloaded
		2. The Truck driver tells the system that the transport to this warehouse has been done.
		3. The Warehouse workers receive the items and determine a place for them in the warehouse
		4. Forklift operators are told to move the items to the new place in the new warehouse
		5. When the Truck driver confirms the insertion, the system updates the new place for the items
		6. The transportation time is recorded and stored in the system
		7. The Redistribution and the transport request are marked as performed.

TABLE II : PROBLEMS WITH RETRIEVED CASE

S.No	Erroneous courses	Description
1	A request is not executable	The execution is interrupted and the Foreman issuing the request is informed
2	Redistribution is wrong	1. The warehouse place does not have enough items to move
		2. The destination warehouse is not appropriate to the item
3	No truck available	When performing loading, and unloading, there may not be any truck available at an appropriate time. Then notify the Foreman who should either delete the request or change it.

Step2, 3 and 4- According to the system element requirements mentioned in Table I, the layering principle has identified following erroneous courses in the service of redistribution between warehouses as shown in Table II.

Step 5: Table III shows the results after executing the knowledge reuse scheme for the task of process planning. Thus, planning endow with new transport requests and therefore change existing transport requests previously in the system

TABLE III : RESULTS

	Planning	Description
1	To coordinate transports and issue transport requests	1. When the redistribution is executed the items to be moved are marked as move-pending
		2. Minimize the use of trucks on condition that all delivery dates should be held and the trucks should be compatible with any delivery requirements for the items.
		3. The transport requests are connected to a specific truck's transportation plan.

V. CONCLUSION

An *OntoReuseAlgo* aims to improve the knowledge reuse in process planning for software development. It supports the application from three aspects: system element classification, representation on the basis of layering principle and knowledge reuse scheme for process planning. The system element classification and the layering principle are adjoined for the workflow control of the knowledge reuse. We have observed the following significant benefits and are as follows:

- Through organizing and modeling the knowledge towards the characteristics of design processes, unnecessary search time can be avoided on irrelevant knowledge items
- It allows explicit credentials for analysis and comparison of different domain theories
- It describes knowledge acquisition approach to structure the entities and relations that need to be acquired in the domain;
- It provides a metalevel view (vocabulary and structure) on their domain which facilitates adequate system documentation and constructs reusable knowledge-system design
- It can be used to define assumptions that enable knowledge exchange between different agents.

REFERENCES

[1] Cossentino, M., Burrafato, P., Lombardo, S., Sabatucci, S., Introducing Pattern Reuse in the Design of Multi-Agent Systems, AITA02, Workshop at NODE 02, Erfurt, Germany, pp. 8-9, October 2002.

[2] De Almeida Falbo, R.; Guizzardi, G. and Duarte, K. C., "An Ontological Approach to Domain Engineering", In Proceedings of the International Conference on Software Engineering and Knowledge Engineering (SEKE´92), Ischia, Italy, July 15-19, 2002.

[3] Dileo, J., Jacobs, T., Deloach, S. Integrating Ontologies into Multi-Agent Systems Engineering, Proceedings of 4th International Bi-Conference Workshop on Agent Oriented Information Systems (AOIS 2002), pp. 15-16, Bologna (Italy), July 2002.

[4] DeLoach, S.: Multi-Agent Systems Engineering: A Methodology and Language for Designing Agent Systems In Proceeding of the 1st Bi-Conf. Workshop on Agent-Oriented Information Systems (AOIS'99), Heidelberg, Germany,1999.

[5] Faria, C. "A Technique for the Acquisition and Construction of Ontology-based Domain and User Models in Multi-Agent Domain Engineering", Master degree dissertation, Federal University of Maranhão – CPGEE, 2004.

[6] Ferreira, S. "A Technique and Tool for Domain Design in Multi-Agent Domain Engineering", Master degree dissertation, Federal University of Maranhão – CPGEE, 2004

[7] Ferreira, S. and Girardi, R. "Specification of a Generic Ontology for Domain Design", Proceedings of III Workshop of Software Engineering (WIS 2003) at Chilean Computing Week 2003 . Chilán, Chile. November 3rd to 8th, 2003. (In Portuguese)

[8] Fleming, M. and Cohen, R., User Modeling in the Design of Interactive Interface Agents, Proceedings of UM99 (User Modeling), 1999

[9] International Workshop on Component Engineering Methodology (WCEM'03) at Second International Conference on Generative Programming and Component Engineering, Ed. Sven Overhage and Klaus Turowski, pp. 41-50. Erfurt, Germany. September 22-25, 2003.

[10] Gruber, T. R. "Towards Principles for the Design of Ontologies used for Knowledge Sharing". International Journal Human-Computer Studies. Nº 43, pp. 907-928, 1995

[11] Guarino, N. "Formal Ontology and Information Systems", Proceedings of FOIS'98, Trento,Italy. Amsterdan, IOS Press, pp. 3-15, 6-8 June, 1998

[12] Harsu, M. "A Survey on Domain Engineering", Report 31, Institute of Software Systems, Tampere University of Technology, December 2002, 26 pp

AUTHORS PROFILE

Shilpa Sharma, Senior Lecturer at SIMS, Indore, India. She has 5 years of teaching experience. She is pursuing Ph.D. in Computer Science and is MCA from Vaishnav Institute of Management, Rajiv Gandhi Vishwavidhyalaya, Bhopal, India. She has 6 papers published to her name in various National and International journals and conferences.

Dr. Maya Ingle is Professor & Senior System Analyst at School of Computer Science and Information Technology, Devi Ahilya University, Indore since 25 years. Presently, she is Principal and Professor at Indore Institute of Computer Application, Indore on extra ordinary leave. She is also Chair Assessment & Research committee. She is Ph. D. in Computer Science and M. Tech in Computer Science from IIT, Kharagpur. Her research interest includes Software Engineering, NLP, Speech Recognition, Usability Engineering and Agile Computing. She has published more than 80 papers in national and international journals and conferences.

Room Impulse Response Modeling in the Sub-2kHz Band using 3-D Rectangular Digital Waveguide Mesh

Zhixin Chen
ILX Lightwave Corporation
Bozeman, Montana, USA

***Abstract* — Digital waveguide mesh has emerged as an efficient and straightforward way to model small room impulse response because it solves the wave equation directly in the time domain. In this paper, we investigate the performance of the 3-D rectangular digital waveguide mesh in modeling the low frequency portion of room impulse response. We find that it has similar performance compared to the popular image source method when modeling the low frequency portion of the early specular reflections of the room impulse responses.**

Keywords - Embedded Systems; Digital Signal Processing; Room Acoustics.

I. INTRODUCTION

The image source method (ISM) [1-3] and the 3-D rectangular digital waveguide mesh (DWM) [4-7] are two relatively simple but very efficient and straightforward methods in modeling room impulse responses for small rectangular rooms. In the image source method, the sound propagation from the source to the receiver is modeled as the superposition of the direct sound and a number of reflected sounds. The image source method results in an impulse response with fine time resolution and only has modest computational complexity for small rooms. The digital waveguide mesh provides a numerical solution to the wave equation in multiple dimensions, and thus is straightforward to model room impulse responses because it solves the wave equation directly in the time domain.

Both the image source method and the 3-D rectangular digital waveguide mesh methods have some limitations. A major drawback of the image source method is that it cannot handle the diffuse reflections and it is not sufficiently flexible to handle the irregular room surfaces. Consequently, the image source method is only practical for modeling the specular reflections in rooms with relatively simple geometries. The 3-D rectangular digital waveguide mesh suffers from severe frequency and direction dependent dispersion at high frequencies. So, the 3-D rectangular digital waveguide mesh is suitable only for modeling the low frequency sound if the oversampling technique with very high computational complexity is not used.

Despite their limitations, both the image source method and the 3-D rectangular digital waveguide mesh are suitable for modeling the low frequency portion of the specular reflections in the small rectangular room. A lot of research has been done to use either image source method [1-3] or digital waveguide mesh [4-7] to model room impulse responses. However, none of these existing works compare these two methods numerically. In this paper, we investigate the consideration of the 3-D rectangular digital waveguide mesh in modeling the low frequency of the specular reflections of the room impulse responses and show that the image source method and the digital waveguide mesh have a very similar numerical performance in modeling low frequency room impulse responses.

The remaining part of the paper is organized as follows. First, the principle of room impulse response models using the image source method and the 3-D rectangular digital waveguide mesh are reviewed. Next, the considerations of the low frequency of room acoustic response modeling using 3-D rectangular digital waveguide mesh are provided. Then, the performances of these two methods are compared for an example source-receiver position. Finally, the paper concludes with a summary of the results and suggestions for future work.

II. BACKGROUND

A. Principle of Image Source Method

The image source method was first used by Gibbs and Jones in 1972 to calculate the sound pressure level distribution within an enclosure [1]. It was then used by Allen and Berkley in 1979 to simulate the impulse response between two points in a small rectangular room [2]. Currently, the image source method is widely used in modeling room impulse responses for small rooms with simple geometry because it has high time resolution and it is relatively easy to implement.

The simulation of sound propagation from a source to a receiver within a rectangular room using the image source method is shown in Fig. 1. The sound propagation is modeled as the superposition of the direct sound and a number of reflected sounds from the source to the receiver. The response to an impulse at the source is then the sum of the delayed impulses. Here the delay time is equal to the sound propagation path length divided by the sound velocity. The impulse response calculation can also be interpreted from a system point view. If the input of the system is an impulse signal, the output of the system will be the impulse response of the system, that is, the modeled room impulse response between the source and the receiver. The system consists of several subsystems. Each subsystem corresponds to a sound propagation path. In reality, the property of the subsystem is affected by the source orientation, receiver orientation, room surface conditions, and the length of the sound propagation path.

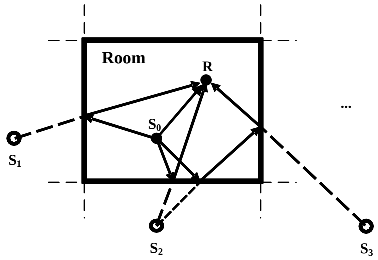


Figure 1: Simulation of sound propagation between two points in a rectangular room using image source method.

B. Principle of 3-D Rectangular Digital Waveguide Mesh

The digital waveguide mesh was first proposed by Van Duyne and Smith in 1993 [4]. It provides a numerical solution to the wave equation in multiple dimensions, and thus has the benefit of incorporating the effects of diffraction and wave interference [4]. The 2-D and 3-D digital waveguide mesh schemes have been used to simulate wave propagation in acoustic spaces [5-7].

The rectangular mesh is the original structure of the digital waveguide mesh. It is a regular array of 1-D digital waveguides arranged along each perpendicular dimension, interconnected at their crossings by scattering nodes with unit delay elements. A 2-D rectangular digital waveguide mesh is illustrated in Fig. 2 [8].

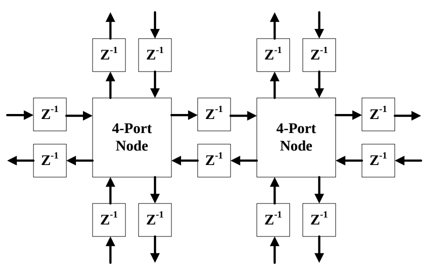


Figure 2: In the 2-D rectangular digital waveguide mesh each node is connected to four neighbors with unit delays.

Similar to the 2-D case, in the 3-D rectangular digital waveguide mesh each node is connected to six neighbors with unit delays. The difference wave equation for the unbounded 3-D rectangular mesh can be expressed using the node values as

$$P(n,i,j,k)=\frac{1}{3}[P(n-1,i+1,j,k)+P(n-1,i-1,j,k)+P(n-1,i,j+1,k)+P(n-1,i,j-1,k) + P(n-1,i,j,k+1)+P(n-1,i,j,k-1)]-P(n-2,i,j,k), \quad (1)$$

where P represents the sound pressure of a node at rectangular coordinates (i,j,k) at time index n. Here i, j, k, and n are arbitrary integers.

The updating frequency or sampling frequency, f_s, of a mesh and the wave propagation velocity, c, are related to each other as [9, 10]

$$f_s=\frac{c\sqrt{N}}{\Delta x}, \quad (2)$$

where N is the dimensions of the mesh and Δx is the spatial sampling interval.

One limitation of the rectangular mesh is that it has numerical dispersion, due to the difference in the physical distances and signal path lengths. For example, in a 2-D rectangular mesh, the spatial distance from a node to its first diagonal neighbor is $\sqrt{2}$ times the distance to an axial neighbor. However, the signal travels along the orthogonal steps from a node to its axial neighbors in one time-step and to a diagonal neighbor, as well as to a second axial neighbor, in two time-steps. Thus, the signal reaches locations at physical distances of $\sqrt{2}$ spatial steps and 2 spatial steps away from the source at the same time instant. The difference in the physical distances and signal path lengths causes a numerical error dependent on both the direction of wave propagation and the frequency of the signal. The error is largest at high frequencies and in the axial directions of the mesh.

Another way to explain why the 3-D rectangular digital waveguide mesh is not appropriate for simulating high frequency sound is that the frequency response is very uneven and the sound propagation dispersion error is large over the whole frequency range, which is shown in Section 3. If we want to model the room impulse response over the whole audio frequency range, we need to oversample the impulse response (for example, by a factor of 10 or more) so that the frequency response will be flat for the desired audio frequency range (20 Hz – 20 kHz). Thus, the node size in the mesh would be reduced to one-tenth of the original size, and there will be as many as one thousand times the number of nodes compared to the original case. So, a digital waveguide mesh with large sample rate requires denser meshes, more computer memory and hence takes on the order of 1000 times longer to run. This is not suitable for current computer workstations. Thus, it is currently not appropriate to use the 3-D rectangular digital waveguide mesh in the simulation of high frequency sounds.

Although the 3-D rectangular digital waveguide mesh suffers from frequency and direction dependent dispersion for high frequency sound **[7]**, the frequency response is nearly flat and the sound propagation dispersion error is small at low frequencies when the updating frequency is high enough. Thus, it is feasible to use the 3-D rectangular digital waveguide mesh for the simulation of low frequency portion of the room impulse response for small rooms. A testing low frequency modeling with a fourth-order Linkwitz-Riley filter with cutoff frequencies of 37 Hz and 1800 Hz was used in this paper.

III. CONSIDERATIONS ON RIR MODELING USING 3-D RECTANGULAR DWM

To apply the 3-D rectangular digital waveguide mesh in the simulation of low frequency part of the room impulse response, the frequency response, the sound propagation direction dependent dispersion, and the performance of the boundary condition of the 3-D rectangular mesh in the desired low frequency range needs to be tested.

A. Frequency Response in the 3-D Rectangular DWM

In the first simulation, the frequency responses for several arbitrary source-receiver positions using the difference wave Eq. (1) for an unbounded 3-D rectangular mesh were tested. The simulations using Eq. (1) were made by constructing a $200 \times 200 \times 200$ node 3-D rectangular mesh and applying an impulse signal into the source node $(80, 100, 120)$. The impulse responses were then simulated at the receiver nodes. In the simulation, the sound velocity was set to 343.5 m/s and the spatial sampling interval was set to 0.0124 m, giving an updating frequency of 48 kHz according to Eq. (2). The impulse response simulations were run for 48 time steps. This ensured that the sound wave would not propagate out of the mesh boundary and therefore the simulated mesh is effectively unbounded. The impulse responses for the simulated receiver nodes were obtained, and also numerically transformed into the corresponding frequency responses using an FFT length of 64 samples.

Fig. 3 shows the impulse responses and the corresponding magnitude frequency responses for three arbitrary receiver nodes $(83, 103, 123)$, $(84, 103, 120)$, and $(85, 100, 120)$. In every magnitude frequency response the spectrum is symmetric about one quarter of the updating frequency. This symmetry characteristic in the frequency domain corresponds to the phenomenon that every other sample is zero-valued in the time domain. The frequency responses are very uneven over the whole frequency range due to the dispersion error of the 3-D rectangular mesh. Theoretically, a real acoustic system would show a flat frequency response since there are no frequency dependent losses in the simple mode of acoustic propagation in the air. The fact that the rectangular DWM simulation gives a non-flat response is an indication of the limitation of the use of 3-D rectangular DWM in the frequency range.

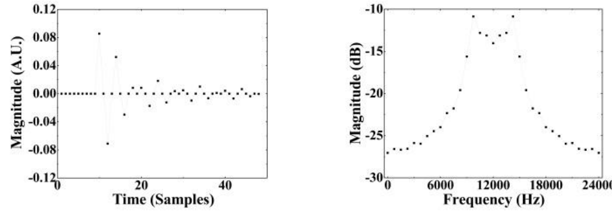
However, Fig. 3 also shows that the frequency responses are quite uniform over the desired low frequency range from 37 Hz to 1800 Hz. For this frequency range, the magnitude variation is very small, with only 0.4 dB for Figure 3(a), 0.1 dB for Figure 3(b), and 0.7 dB for Figure 3(c).

B. Propagation Dispersion in the 3-D Rectangular DWM

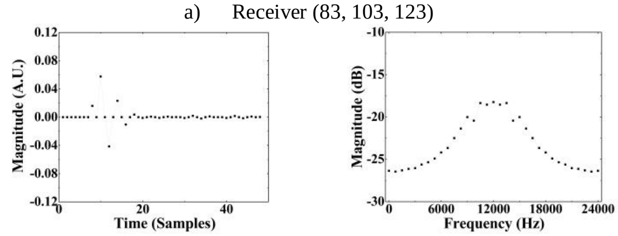
The 3-D rectangular digital waveguide mesh suffers from serious direction dependent dispersion for high frequency sound. According to our simulation, however, the direction dependent dispersion is very small at low frequencies. When a synthesized bandpass signal (37 Hz – 1800 Hz) was fed into a source node in the unbounded 3-D rectangular mesh, the sound propagation within the mesh is nearly omni-directional.

To obtain the sound propagation direction dependent dispersion, the source was located in a fixed position and the receivers were placed in different positions on a sphere with the source as the center.

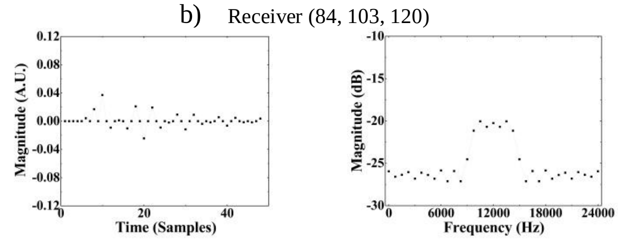
In a second simulation, a $150 \times 150 \times 150$ node 3-D rectangular mesh was constructed and a synthesized bandpass signal (37 Hz – 1800 Hz) was fed into the source node $(75, 75, 75)$.



a) Receiver (83, 103, 123)



b) Receiver (84, 103, 120)



c) Receiver (85, 100, 120)

Figure 3: Impulse response (left figure) and magnitude frequency response (right figure) for three example source-receiver positions in an unbounded 3-D rectangular digital waveguide mesh. Here A.U. represents arbitrary unit.

The updating frequency was set to 48 kHz. The responses for 231 different receiver nodes on a sphere were simulated. The distances between the source and the receivers were 60 nodes.

In the simulation, every response simulation was run for 130 time steps. This ensured that the peaks located in time index 117 of the responses can be obtained. Since the receiver nodes can only have integer spatial indices in the simulation, the distance between the source node and the receiver node may not exactly equal 60 nodes. So, the magnitude of the peaks in the responses was adjusted to that with the distance of 60 nodes according to the inverse square law. The sound propagation in every direction was represented using the magnitude of the peaks of the corresponding simulated response. The simulation results show that the magnitudes of the peaks in the 231 simulated responses are almost equal.

The low frequency sound propagation in the 3-D rectangular mesh is nearly omni-directional in different directions on a circle, as shown in Fig. 4. The largest magnitude is -70.5 dB while the smallest magnitude is -70.8 dB. The small dispersion is due to the fluctuation of the digital waveguide mesh frequency response in different directions. Since the frequency response is nearly flat and the sound propagation dispersion is small, the 3-D rectangular mesh appears to be suitable for the room impulse response modeling in the desired low frequency range.

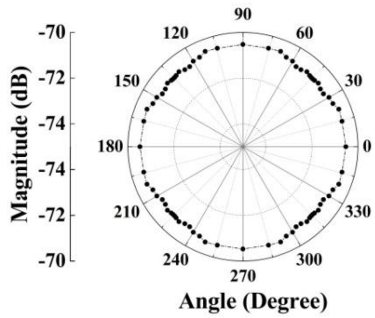


Figure 4: Sound propagation in different directions on a circle in the 3-D rectangular digital waveguide mesh.

C. *Boundary Condition Performance Tests in the 3-D Rectangular DWM*

In real rooms, sound reflections occur when sound waves reach room surfaces. The complex changes of the magnitude and the phase of the sound waves at the room surfaces depend on the surface material and structure as well as the level, frequency, and propagation direction of the sound. Correspondingly, there should be a representation of the boundary condition in the digital waveguide mesh.

Currently, a lot of research is being conducted on modeling boundary conditions in digital waveguide mesh with adjustable shape, diffusion, and reflection characteristics [11-14]. However, there has not yet been a perfect boundary condition implementation for the digital waveguide mesh.

For the 3-D rectangular digital waveguide mesh, there exist only two suitable frequency and angle independent boundary conditions. One is the basic boundary condition based on the simple impedance matching technique while the other is a modified boundary condition proposed by Kelloniemi et al. in 2004 [11], which is based on an improved absorbing boundary condition proposed in [14]. The performances of these two boundary conditions were tested in another simulation and are shown as below.

The first step is to test whether or not these two boundary conditions converge. To save simulation time, without loss of generality, a $40 \times 40 \times 40$ node small 3-D rectangular mesh was constructed and a bandpass signal (37 Hz – 1800 Hz) was fed into the source node $(20, 20, 20)$. The responses were then simulated at three arbitrary receiver nodes $(25, 35, 29)$, $(17, 25, 16)$, and $(3, 31, 27)$ using the reflection coefficients $r = 0.0$, $r = 0.5$, and $r = 1.0$ for both boundary conditions. Every response simulation was run for 24000 time steps, which is long enough to ensure that the convergence test is believable. According to the test, both boundary conditions converge for these three arbitrary receiver nodes.

The second step is to test how well the simulated reflection coefficients match the theoretical values. A test procedure proposed in [11] is used here. For reflective boundaries, the simulated reflection coefficient is equal to the ratio of the magnitude of the simulated reflected sound signal to that of the perfectly specular reflected sound signal. Here the magnitude of the simulated reflected sound signal is equal to the difference between the magnitude of the received sound signal and that of the direct transmission sound signal between the source and the receiver. The magnitude of the perfectly specular reflected sound signal is equal to that of the direct transmission sound signal between the source and the mirror receiver.

A $200 \times 200 \times 200$ node 3-D rectangular mesh was constructed. The simulation was made for 11 source-receiver positions, where the incident angle from the source to the tested boundary ranges from 0º (normal incidence) to 45º. To facilitate the representation, the incident angle of 45º is used as an example.

A bandpass signal (37 Hz – 1800 Hz) was applied as an input over 96 time steps at the source node S1: $(90, 100, 10)$ in the mesh. The response simulation was made at the receiver node R1: $(110, 100, 10)$. The set-up of the source node and the receiver node made sure that the input was fed into a node located close to the tested mesh boundary and far from any other boundary to avoid additional reflections influencing the results.

The output signal was received at the node located at equal distance from the tested boundary as shown in Fig. 5 (a). The tested boundary was simulated using both the basic boundary condition and the modified boundary condition.

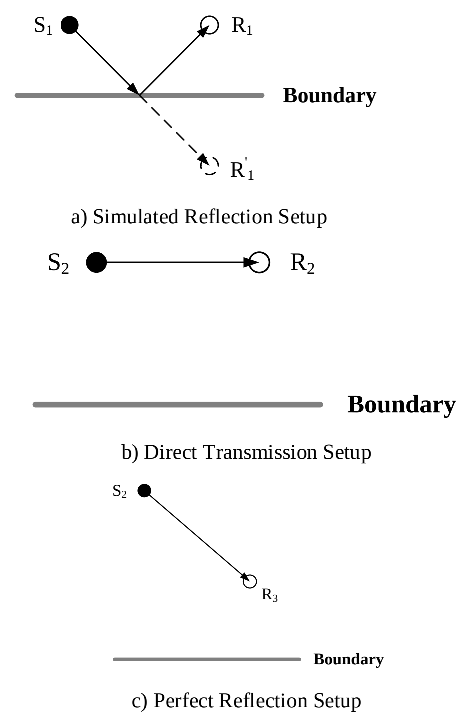


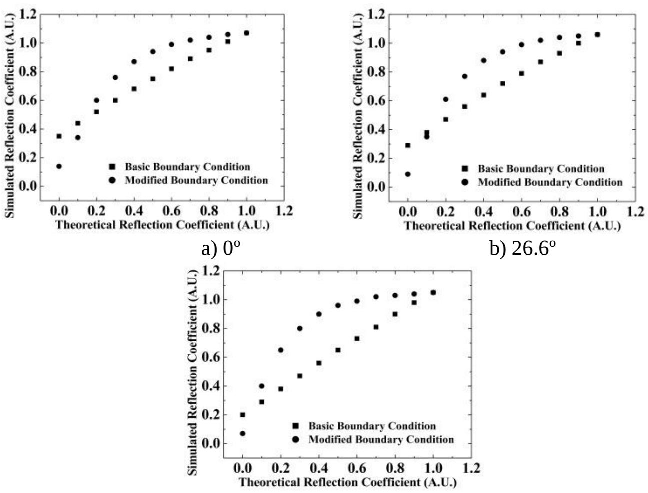
Figure 5: Performance tests of boundary conditions in the 3-D rectangular digital waveguide mesh.

The direct transmission sound signal from the source node to the receiver node was simulated in the same 3-D rectangular mesh. The response between the source S1 and the receiver R1 is equivalent to that between the node S2: $(90, 100, 100)$ and the node R2: $(110, 100, 100)$, as shown in Fig. 5 (b). The direct transmission sound signal was subtracted from the output signal received at the node R1 to get the simulated reflection signal.

In addition, the perfectly specular reflected sound signal was calculated using the simulated response between the node S2: $(90, 100, 100)$ and the node R3: $(110, 100, 80)$, as shown in Fig. 5 (c). The simulated reflection coefficient is then calculated using the ratio of the magnitude of the simulated reflection signal to that of the perfect specular reflection signal.

Fig. 6 displays the comparison of the simulated reflection coefficients using the basic boundary condition and the modified boundary condition to the theoretical values in three incident angles 0º, 26.6º, and 45º. The simulated reflection coefficients calculated using the modified boundary condition are closer to the theoretical values for the lower reflection coefficients while those using the basic boundary condition are closer to the theoretical values for the higher reflection coefficients. The basic boundary condition was chosen in the following room impulse response model using digital waveguide mesh because the room surface reflection coefficients are usually at least as large as 0.3 in real rooms.

It can also be seen from Fig. 6 that there is a nearly linear relation between the theoretical reflection coefficients and the simulated values using the basic boundary condition. But the relation varies with incident angles. A correction function for the simulated reflection coefficients was not used in this paper.



a) 0º b) 26.6º

c) 45º

Figure 6: Comparison between the performances of the basic boundary condition and the modified boundary condition in three incident angles.

IV. COMPARE ISM AND DWM

After the frequency response, the sound propagation dispersion error, and the performance of the boundary conditions of the 3-D rectangular mesh in the low frequency range were tested, the 3-D rectangular digital waveguide mesh based simulation results of the low frequency impulse responses between a source and a receiver in a small simulation room were then compared to that using the image source method.

Before the low frequency room impulse response was modeled, the room dimensions, the source position, and the receiver position were set. The rectangular room was 1.24 meters long, 1.49 meters wide, and 1.74 meters high. The source position was (0.62, 0.62, 0.62) meters and the receiver position was (0.372, 0.496, 0.62) meters. The sound velocity was 343.5 m/s. The source and the receiver were assumed to be omni-directional. Without loss of generality, the reflection coefficients were set to 0.7 for the ceiling and the floor and 0.6 for the walls. A synthesized bandpass signal (37 Hz – 1800 Hz) was fed into the source and the response in the receiver was simulated using both the image source method and the 3-D rectangular digital waveguide mesh.

In the simulation using the image source method, the above-mentioned room dimensions, source position, receiver position, sound velocity, and reflection coefficients were used directly in a MATLAB program developed in [15-16].

In the simulation using the 3-D rectangular digital waveguide mesh, the room was divided into several cubic nodes with the size of 0.0124 meters. Thus, the room dimension was (100, 120, 140) nodes in the mesh. The source was at the node (50, 50, 50) and the receiver was at the node (30, 40, 50). In the simulation, the sound velocity was set to 343.5 m/s, giving an updating frequency of 48 kHz. A bandpass signal (37 Hz - 1800 Hz) was fed to the source and the response at the receiver node was simulated.

Several arbitrary peaks were extracted from the two modeled responses. Since the nodes in the digital waveguide mesh can only have integer spatial indices, there may be small time shift of the delay time in the mesh. So the delay time of the peaks in the two modeled responses may not exactly be the same. Thus, a satisfactory match is judged to occur if the delay time difference is within 0.10 ms. According to this rule, if the delay time of a magnitude peak in the DWM-based modeled impulse response is within 0.10 ms of the ISM based modeled value, the magnitude peak is verified. Otherwise, the sample in the ISM based impulse response having the same delay time as the modeled value will be chosen. 13 peaks from the two modeled responses are extracted and shown in Fig. 7. It can be seen that they have a very small discrepancy. Note that the peak magnitude in the modeled response using the digital waveguide mesh is greater than that using the image source method. This shows that the reflection coefficient calculated using the basic boundary condition in the digital waveguide mesh is actually higher than the theoretical value, as shown in Fig. 6 in the previous section.

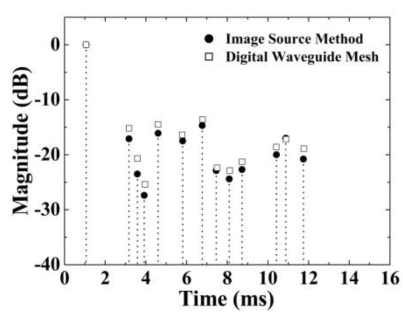


Figure 7: Delay time and magnitude of 13 arbitrarily chosen peaks in the image source method and the 3-D rectangular digital waveguide mesh based models of the low frequency room impulse responses for a small rectangular simulation room.

To test the validity and the consistency of the comparison between these two methods, the comparison was also made in different room sizes, different room surface reflection coefficients, different source positions, and different receiver positions. The tests showed that the image source method and the 3-D rectangular digital waveguide mesh have similar results in these setups.

V. CONCLUSION

In this paper, we verified that the 3-D rectangular digital waveguide mesh is suitable for modeling the low frequency room impulse response because the frequency response is nearly flat and the sound propagation dispersion error is small in the low frequency range in this mesh structure. We also showed that the low frequency impulse response simulation using the image source method and the 3-D rectangular digital waveguide mesh is approximately equivalent to that using the image source method for an example source-receiver in a rectangular simulation room.

There is some improvement that can be made in the future. In the model using the digital waveguide mesh, the simulated reflection coefficient using the basic boundary condition has some discrepancies with the theoretical value. A correction function for a variety of reflection coefficients (for example, range from 0.0 to 1.0 with the step of 0.1) in different directions (for example, range from 0º to 180º with the step of 1º) can be used to improve the performance of the boundary condition. The performance comparison between these the image source method and the 3-D rectangular digital waveguide mesh will be conducted in a room with more complicated geometry.

ACKNOWLEDGEMENT

The author would like to thank for the guidance and assistance from Dr. Robert Maher in Electrical and Computer Engineering Department at Montana State University. The author would also like to thank for the support from Advanced Acoustic Concepts, Inc., and from Benjamin Ph.D. Fellowship at Montana State University - Bozeman.

REFERENCES

[1] B. M. Gibbs and D. K. Jones, "A simple image method for calculating the distribution of sound pressure levels within an enclosure," Acustica, vol. 26, pp. 24-32, 1972.

[2] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am., vol. 65, no. 4, pp. 943-950, 1979.

[3] U. R. Kristiansen, A. Krokstad, and T. Follestad. "Extending the image method to higher-order reflections," J. Applied Acoustics, vol. 38, pp. 195-206, 1993.

[4] S. Van Duyne and J. O. Smith, "Physical modeling with the 2-D digital waveguide mesh," Proc. Int. Comput. Music Conf., Tokyo, Japan, pp. 40-47, Sept. 1993.

[5] L. Savioja, T. Rinne, and T. Takala, "Simulation of room acoustics with a 3-D finite difference mesh," Proc. Int. Computer Music Conf., Aarhus, Denmark, pp. 463-466, 1994.

[6] L. Savioja, J. Backman, A. Järvinen, and T. Takala, "Waveguide mesh method for low-frequency simulation of room acoustics," Proc. Int. Congr. Acoustics, Trondheim, Norway, vol. 2, pp. 637-640, June 1995.

[7] D. T. Murphy and D. M. Howard, "Physically modeling the acoustics of a room using a digital waveguide Mesh," Proceedings of the 25th Euromicro Conference, Milan, Italy, pp. 82-89, Sept. 1999.

[8] Zhixin Chen and Robert C. Maher, "Analytical expression for impulse response between two nodes in 2-D rectangular digital waveguide mesh," IEEE Signal Processing Letters, vol. 15, pp. 221-224, 2008.

[9] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vaananen, "Creating interactive virtual acoustic environments, " J. Audio Eng. Soc., vol. 47, no. 9, pp. 675–705, Sept. 1999.

[10] A. Kelloniemi, Room Acoustics Modeling with the Digital Waveguide Mesh – Boundary Structures and Approximation Method. Ph.D. dissertation, Helsinki University of Technology, 2006.

[11] A. Kelloniemi, D.T. Murphy, L. Savioja, and V. Välimäki, "Boundary conditions in a multi-dimensional digital waveguide mesh," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada, vol. 4, pp. 25-28, May 2004.

[12] A. Kelloniemi, "Improved adjustable boundary condition for the 3-D digital waveguide mesh," Proc. 2005 IEEE Workshop Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 191-194, Oct. 2005.

[13] A. Kelloniemi, "Frequency dependent boundary condition for the 3-D digital waveguide mesh," Proc. 9th Int. Conf. Digital Audio Effects (DAFX06), Montréal, Canada, pp. 161-164, Sept. 2006.

[14] D. T. Murphy and J. Mullen, "Digital waveguide mesh modelling of room acoustics: Improved anechoic boundaries," Proc. Int. Conf. Digital Audio Effects (DAFx), Hamburg, Germany, pp. 163-168, Sept. 2002.

[15] Z. Chen, An Investigation of Acoustic Impulse Response Measurement and Modeling for Small Rooms, Ph.D. dissertation, Montana State University, 2007.

[16] Z. Chen, R.C. Maher, "Addressing the Discrepancy between Measured and Modeled Impulse Response for Small Rooms," Proc. 123rd AES Convention, New York, 2007.

AUTHORS PROFILE

Zhixin Chen is a Firmware Engineer in ILX Lightwave Corporation. He holds a BS degree from Xiamen University in China, a MS degree from Xiamen University in China, and a Ph.D. degree from Montana State University in USA, all in Electrical and Computer Engineering. His research interest and working experience are in the area of acoustics, audio, and speech processing, multimedia communication, and embedded system design for high power current source, temperature controller, and optical power meter.

E-learning as a Research Area: An Analytical Approach

Sangeeta Kakoty
NECRD, IGNOU, Guwahati
Guwahati, India.

Monohar Lal
School of Computer & Inf. Sc.
IGNOU, New Delhi, India

Shikhar Kr. Sarma
Dept. of Computer & IT
Gauhati University, Guwahati, India

***Abstract*— The concept of E-learning is very broad. It was coined in late 90s as the technological enhanced learning mechanism through Internet. Now it captures a broad range of electronic media like Internet, Intranets, Extranets, satellite broadcast, audio/video tape, interactive TV and CD-ROM to make the learning procedure more flexible and user friendly. Because of the flexible nature of E-learning, it has got more demand among the people of our country and the demand is increasing day by day. As the demand is increasing, this is the time to standardize the whole e-learning system in a proper way and the time to increase the quality of existing standards. Though many standards are already there and has accepted by many academia, institutes and organisations, still there are some gaps and works are going on to make them more practicable and more systematic.This paper analyses the current e-learning procedure and showing the new dimension of research work on this area that follows the important and most neglected research areas till today in this domain. It also analyses the importance of e-education system and recent market of e-learning procedure.**

Keywords- E-learning; learning standards; technology enhanced learning procedure; e-education system; dimension of research areas.

I. INTRODUCTION

This is the age of WWW and we are living in a globalized era, where the world is massively being connected. The E-learning initiatives have connected the whole world and have removed the barrier of age, place, time and socio-economic nature. The technological revolution has created a new dimension in whole education scenario. With the amazing development of Internet, the field of education has tried to exploit web as a communication channel to connect distant learners with their learning resources. Tom Kelly quoted that "E-learning is about information, communication, education and learning" [20]. It is a platform with flexible learning using Information Technology and Communication (ITC) resources, tools and applications, and focusing on interactions among teachers, learners and online environment [1]. E-learning usually refers to structured and managed learning experiences, and may involve the use of Internet, CD-ROMs, software, other media and telecommunications. Because of the flexible nature of E-learning and since it provides the right information in right time and in right place, students are now more familiar and feel more comfort in this new education system.

As the technology is advancing, the demand of online learning is also increasing. The technologies, tools, techniques, methodologies and standards are advancing in such a way that it has to overwhelm the ability of educationists to isolate, study, and report on the best methods to be used for any given audience [2]. With all these advances, the prospects for E-Learning are clearly bright and many.

Though E-learning concepts have become more popular and many standards have been developed and used by many institutions and organizations, the real challenges facing many problems [6]. It is sure that many of standards developed in appropriate theoretical frameworks and formal approaches which guarantee that people do not stop merely at creating technical solutions, but can force on to document the impact of technology on learning, and can pass along with lessons learned. Further, in a developing country like India, the digital divide is still with us, and appropriate use of technology requires consideration of a wide range of low-tech and high-tech solutions [4]. Many researches in this domain are still trying to make the content representation standard. They are trying to make the learning style more powerful from the learner's point of view which is more important area of research where learner's preferred pace is getting more interest than other areas.

Through this paper, author is trying to establish the line and type of research work in E-learning environment that people are working on. It focuses mainly the prospective areas of research in this domain. Though E-learning is a very broad area, we are concentrating and focusing the education system only which is the more promising area of whole E-learning scenario. We start this paper with an introductory note on the relationship between e-learning and education system. It followed by the recent market and its growth in this developing country. Section IV and section V is all about the discussion on prospects and promising research areas of e-learning domain.

II. THE EDUCATIONAL SYSTEM AND E-LEARNING

The popularity of Internet along with the extensive development of standard protocol and services creates a new dimension in the whole education scenario. It makes the online education more attractive. Everyday new approaches are coming and bringing new prospects in education and trying to refining the system towards personalized self-learning [3]. The benefits of E-learning are mainly the cost efficiency, accessibility and flexibility. However, whilst much has been made of the benefits to the organization of e-learning, there has been little, if any, qualitative investigation into the attitudes and views of the users themselves [14]. According to a research study in Europe, it has been seen that E-learning course effectiveness is 93.5%, and 92.5% says that course completion in this system is not critical. The United States Distance Learning Association says, "Distance learning encompasses all technologies and supports the pursuit of lifelong learning for all" [13]. Canadian educator George Siemens launched a weblog and newsletter called **eLearnSpace** covering developments in e-learning technology and deployment. The website also includes comprehensive guides for educators that want to integrate e-learning tools and techniques into their institutions. **Training Watch** is an E-learning database where numerous volunteers contribute for the corporate e-learning community. The front page of this site offers visitors a rolling feed of e-learning news and analysis from a variety of sources. Readers provide instant feedback and ratings, highlighting some of the critical debates within the e-learning industry. It also contains numerous product reviews and field reports, ideal for managers that want to evaluate potential e-learning solutions for their organizations. Now-a-days there are different online portals; various websites are available for online news letters, online education, latest research findings to support both corporate and education sectors.

"*E-Learning Centre Resource Guide*" is the UK's most prominent online source for e-learning information offers a comprehensive guide to commonly available tools, software, and other resources. In addition to listing highly specialized resources, this e-learning guide offers tutorials for creating educational and professional development content using common, off-the-shelf computer software packages. There are different e-learning consultants like *Jay Cross* that maintains daily weblogs to share recent findings, *William Horton Consulting inc*. to help businesses, schools and governments in implementing effective e-learning programs, *Maxine Morse*, a leading learning consultant of Europe to help large organizations to develop effective e-learning strategies both for students and employees.

Now this is the time to standardize the whole e-learning system and to do this there are some standards bodies like Advanced Distributed Learning (ADL), IEEE Learning Technology Standards Committee and Instructional Management Systems (IMS) Project. They are the key players and mainly taking initiatives for this. Apart from them, another two organizations are also working on specifications related to e-learning are Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) and Aviation Industry CBT (AICC). The most popular SCORM standard is developed to focus on the opportunity to create learning content, which may be reused, accessible on multiple levels, interoperable. It was developed to incorporate several, disparate efforts to develop standards for learning technology initiatives.

III. MARKET OF E-LEARNING

The whole online learning education industry we can divide into three major market segments:

- Content organisations: firms that furnish course structure, multimedia, simulations, testing and assessment; both off-the-shelf solutions as customized applications.
- Learning services firms: firms that provide needs assessments, program-building components, content-design, development and programming, technical and systems integration, site management and hosting, maintenance and online mentoring.
- Delivery solutions companies: firms selling technologies associated with e-learning, including training, authoring tools, course management systems, collaborative software and virtual classrooms and add-on tools.

Growing demands for e-learning require a combination of methodologies, tools, and technologies to effectively scale by e-learning development throughout the organization. IDC forecasts that the e-learning market, which was about $6.5 billion in 2003, is increased to more than $21 billion by 2008, and will hit $52.6B by 2010. The 2007 U.S. e-learning market is presently $17.5 billion [15]. The U.S. enterprise e-learning adoption accounts for 60 percent of the market, while Europe's accounts for 15 percent, overall usage of e-learning in Asia are expected to reach a compound annual growth rate of 25 percent to 30 percent through 2010. According to an education sector report by CLSA Asia Pacific Markets, the e-learning market size is estimated at $27 million or Rs. 105crore, which is projected to grow to $280 million or Rs. 1,092crore by 2012 [17]. Comparing this to the size of the US e-learning market valued at $4 billion or Rs 15,600 crore, and it may appear rather small, but the potential for growth in India given the huge population, lack of educators, etc, is much more. Meanwhile, the current e-learning global market size is over $20 billion (Rs 78,000 crore) grown ten-fold since 2000, and is expected to surpass $52.6 billion (Rs 2,05,000 crore) by 2010. Online tutoring, on the other hand, is a $4 billion (Rs 15,600 crore) industry and is growing at a rate of 10%-15% per annum according to *Financial Express* [15]. Computer-based learning on the other hand, still has a market size of $150 million (Rs 590 crore), in India which is expected to grow to $750 million (Rs 2,900 crore) by 2012, according to the CLSA Asia Pacific market report [26]

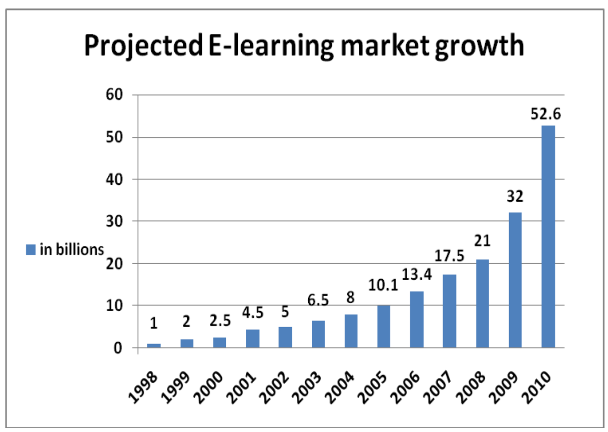


Fig 1: Year-wise E-learning market growth

If we project the growth of e-learning from the year 1998 to 2010, according to the value cited by different sites in different times, it has been seen that it always follows an increasing graph. This increasing growth is clearly viewed in the above graph.

According to Training Magazine report, training organizations are shifting their staffing models away from a dominant focus on trainers and are now more focused on design, e-learning, and service and support activities. In addition, they are now outsourcing much of the delivery. A few years ago, more than 70 cents of every training dollar went to payroll. Today the figure is about 65 cents. Training magazine also reported that e-learning now accounts for 15 percent of all training delivered, which is a two-fold increase from just one year ago, and signifies that e-learning is here to stay as a mainstream training delivery vehicle. Meanwhile, classroom training has dropped from 70 percent to 62 percent; however, it still remains the dominant form of training [26].

ASTD's 2010 *State of the Industry Report* shows that e-learning is still increasing as it now accounts for 27.7 percent of corporate training, its highest level since ASTD began collecting data on the use of technology for this report 14 years ago and in 2008, it was 23.1 percent [8].

In the 2010 market, corporations are the top buyers of Self-paced eLearning. By 2015, corporations will still be the top buyer, followed by higher education and the PreK-12 buyers [5].

IV. Prospects of E-Learning

The current promising application areas of e-learning are content management and web security. Guild research report 2006 also focus that "Improving the quality of learning contents" gets the highest priority than other areas [22]. According to Blezu and Popa (2008), E-learning has lots of prospects in various sectors like: in dynamism, in real time, in collaboration, global reach and delivery of speech [10]. This is because in this domain:

- Learners can access information that is correct and up to date through the web, information databases or university or company intranets.
- Learners are able to meet in a virtual space with other members and practitioner experts to discuss issues, answer questions and even participate in simulations and management games without having to leave their office or home.
- Learners benefit from learning when required, learners are able to access the right sort of training at the right time with the right people.
- Learners have access when they want it.
- Learners have access to the same materials.
- Learners regardless of where they are receive the same message and are able to engage other learners and practitioners globally.

Guild Annual Research Report, April 2006 was on the subject of future directions in e-learning. From their survey sample it was noticed that "designing and developing e-learning content" activity will get more focus and attention in future. The second highest priority, according to this report is the "addressing learner requirements and preferences" [22]. For an organization, "Extend the global reach of the E-learning content" is the focusing priority area to get the content out beyond the geographical limit. In that report it was clearly written that "Blended learning" will grow significantly in the year ahead.

V. Research in E-Learning

E-learning is the education through distance mode where many technologies can employ. Though e-learning and distance education are both separate terms, they are quite same. Only difference is that distance education separates the students from traditional classroom environment whereas in e-learning environment it is not. So, research in distance education can help indirectly the whole e-learning procedure. It is seen that, research in distance education has been subject to consistent critique [9] [19] [24]. It has even been characterized as "theoretical and predominantly descriptive" [21]. Richter argues in his paper that in the field of distance education, a validated meta-structure of research topic is lacking, i.e., a map of research areas that would help to organize the body of knowledge in the field. The structure of a research discipline forms the foundation for identifying gaps and priority areas for researchers [23].

Ardil introduced an e-learning collaborative circle for teachers and students to announce their research subjects/projects through a cyberspace where members can interact creatively and freely to consider relevant issues, sharing individual experiences and providing and gaining support from each other [7]. Designing and developing distance education programs requires specialized training and skills. This is more so in case of online learning [11]. Because of this, researchers are still trying to improve and manage the learning contents and trying to give one standards of it. SCORM of ADL is the popular standard of content packaging and communication than other standards like IMS content packaging application and IMS simple sequencing specification. Though SCORM is accepted by many organizations and academia, work is still going on to improve the standards. Though the learning content is sequenced in SCORM standard now, the course creator must understand the use and manipulation of its complicated sequencing rules which is the main barrier for most of the course creator [18].

The learning standards and specifications of whole e-learning procedure can be organized into five sub-branches: metadata management, content packaging and communication, learner profile, learner registration and security [16]. For each and every branch there should be some standards. Since user can purchase based on their quality and appropriateness and with confidence that they will work on it effectively, we should have to think according to their interest and accordingly have to design their standards. They should be user friendly and more interactive. Several initiatives have been taken into account in making the e-learning standards more attractive and assistive.

The metadata standard named "Learning Object Metadata (LOM)" of IEEE is adopted by many organizations. The initiatives "dealing with content packaging" are the IMS content packaging specification, IMS simple sequencing specification and ADL SCORM model which is more popular than other two. Learner profile is the most important section of e-learning procedure and effort is going on to make it standardize. One another important effort in standardization of whole e-learning scenario is the IMS learner information package (LIP) specification. For learner registration, there are two initiatives currently dealing with: the IMS Enterprise Specification and the Schools Interoperability Framework which supports on the exchange of the data into K-9 environment (*http://www.imsproject.org*). When content is launched, it needs to communicate learner's data and previous browsing activity and information. There should be some system to monitor all these activities. Work is going on in this context and is applying in the popular standard model SCORM of ADL. The latest release of ADL is the SCORM 2004. It includes the intelligent tutoring system (ITS) that works as natural human tutors [12]. ADL is actively engaging in research and implementation of digital knowledge environment of future in the area of standards and authoring tools. AICC's new release Package Exchange Notification Services (PENS) can define an interface between authoring tools and LMS systems to automate publishing and testing of learning materials accessed through LMS system. Kakoty and Sarma, 2011 argues in their previous paper that by integrating Expert System technology the whole e-learning system will be more learners centric and assistive for learner [14].

There should be a functional model to understand how different systems will work in e-learning environment and for that SCORM is the highly generalized model that manages the delivery and tracking of learning content to a learner, but does not specify functionality within the learning management system (LMS). So, Liu *et el.* in their paper proposed one advanced functional model of SCORM to define which information would interchanged among each component [16]. Likewise it is very much important to improve the functions of existing standards by applying different ICT tools.

Richter (2009) has done one Delphi study on possible research areas in distance education for which he classified whole distance education research into three broad meta-levels:

A. *Macro level : the system and theories*

B. *Meso level : management, organisation and technology*

C. *Micro level : teaching and learning*

These three levels he again classified into 15 sub areas according to the importance of the research issues and has given rating of importance from 1 to 10 as 1 labeled for "very low importance" and 10 labeled for "very high importance" [23].

VI. Analysis and Recomendation

From the study it has seen that content packaging and content managing is got the highest priority in e-learning research where yet lots of development has to be made. Learner's prospect and interest is increasing very rapidly as the technology is growing. Now researcher has to design the learning methodology according to learner's interests and preferences. For that learner characteristics, their behavior, their learning style has to study which is possible by recording their browsing history at their learn time. That is why now the research on learner characteristics has got widest and highest ranges of rating. Again, the two areas like "access, equity and ethics" and "quality assurance" are of growing interest.

Based on the Delphi report of Richter, the most neglected areas of e-learning are the cultural differences in global distance learning programs and cooperation which should receive much more attention. So from this study it has been seen that globalization of education, cross-culture aspects and culturally complex student support system in distance education as well as in e-learning environment is a prospective research area. We can improve these areas by integrating new technologies and ICT tools. There is also a need for international comparative research in this environment.

A promising technology for realizing e-learning requirements is the Semantic web that can provide flexible and personalized access to the learning materials. In fact, semantic web can be exploited as a very suitable platform for implementing an e-learning system, because it provides all means for e-learning: ontology development, ontology-based annotation of learning materials, their composition in learning courses and proactive delivery of the learning materials through e-learning portals [25]. Since e-learning involves a number of forms, number of complex steps, varying level of interest of learner, there should be a system to take decision at every level and at every stage of learning which can make the system more interactive for learner. A Decision Support System (DSS) is an interactive information system that provides information, models and data manipulation tools to make help decisions in semi-structured and unstructured situation [28].

VII. Conclusion and Future Prospects

As e-learning is definitely a growing field in the educational and training market and e-learning standard is a new emerging area, there are many challenges in implementation of undergoing technological changes and developments. The security of services, the encryption of messages and the common taxonomies to describe services and service access points in e-learning systems environments are all in need of consideration. However, Supporters of e-learning are always looking forward some new developments. Technology advancements will continue to reshape learning over the Internet with increasing use of advanced tools and techniques.

So by employing the new technology in e-learning environment, one can make the system more attractive and interactive for learner that may help to build a learner centric platform in this environment.

VIII. Acknowledgement

The first author acknowledges the academic and financial support given by North East Centre for Research and Development (NECRD), IGNOU, Guwahati for the research work.

References

[1] S. Codone, "An E-learning Primer". Pensacola, Florida, 2001.

[2] "The E-learning e-revolution in colleges and universities". A Pan-Canadian Challenge, 2001.

[3] A.E. Blackhurst, and D.L. Edyburn, "A brief history of Special Education Technology" Knowledge by Design Inc., 2000.

[4] "E-learning quality: Aspects and criteria for evaluation of e-learning in higher education", Report 2008:11R, Swedish National Agency for Higher Education, 2008.

[5] S. Adkins, "The US Market for Self-paced eLearning Products and Services: 2010-2015 Forecast and Analysis" Ambient Insight, LLC, January 2011

[6] I. E. Allen, J. Seaman, "Online Nation: Five Years of Growth in Online Learning". Needham, MA: Sloan Consortium, 2007.

[7] C. Ardil, "E-learning Collaborative Circles" , International Journal of Humanities and social Services, Vol. 1, No. 4, 2007.

[8] ASTD State of the Industry Report. Retrieved December 1, 2010

[9] R. M. Bernard, P. C. Abrami, Y. Lou, and E. Borokhovski, "A methodological morass? How we can improve quantitative research in distance education", Distance Education, 25(2), 175-198, 2004.

[10] C. Blezu, and E. M. Popa, "E-Learning and its Prospects in Education" 12th WSEAS International Conference on COMPUTERS, Heraklion, Greece, July 23-25, 2008.

[11] J. Dikshit, A. Gaba, S. Bhushan, S. Garg, and S. Panda, "learning Attitude, Motivation and Preferences of Online Learners" Indian journal of Open Learning 12(3), 149-167, 2003.

[12] K. R Ellis, "E-learning Standards Update" learning circuits, 2005.

[13] P. Gledhill, "Latest e-Learning research in Europe Middle East and Africa" (page 1 of 3), 2004.

[14] S. Kakoty, and S.K Sarma. "Expert System Applications in E-learning Environment: Analysis on Current Trends and Future Prospects" International Journal of Internet Computing (IJIC), Vol. 1, 90-93, 2011.

[15] D. Kopf, e-Learning Market to hit $52.6B by 2010. The Journal: July 2007.

[16] X. Liu, A. E. Saddik, and N. D. Georganas "An implementable architecture of an e-learning system" Proceedings of Canadian Conference on Electrical and Computer Engineering, May 4-7, 2003, 2, pp.943-946., 2003.

[17] M. McGee, "E-Learning Predicted To Recover--Again". Information Week: Nov, 9, 2004.

[18] M. Melia, , Barrett, R. and Pahl, C "A Model-based Approach to SCORM Sequencing" In Proceeding of the Sixth Annual Irish Educational Technology User's Conference (EdTech06), 2006.

[19] M. G. Moore, "Some observations on current research in distance education" Epistolodidaktika, 1, 35-62, 1985.

[20] Namahn. "E-learning: A research Note" http://www.namahn.com/resources/documents/note-e-learning.pdf

[21] H. Perraton, "Rethinking the research agenda", International review of Research in Open and Distance Learning, 1(1), 2000.

[22] J. Pulichino, "Future directions in e-Learning research Report 2006", The Learning Guild Research 2006.

[23] O. Z Richter, "Research Areas in Distance Education: A Delphi Study", International Review of Research in Open and Distance Learning, Vol. 10, No. 3, 2009.

[24] F. Saba, "Research in Distance Education: A status report", International review of Research in Open and Distance Learning, 1(1), 2000.

[25] L. Stojanovic, S. Staab, and R. Studer, "E-learning based on semantic web" WEBNET 2001, 2001.

[26] Training Magazine "Training Industry Report". Retrieved August 7, 2007.

[27] "E-learning for Innovation: Executive Summary", Helios yearly report 2007.

[28] M. S. Velmurugan and K. Narayanasamy "Application of Decision Support System in E-commerce", Communications of the IBIMA, Volume 5, 2008

WEBSITES REFERRED:

[1] http://www.imsproject.org.

[2] http://www.adlnet.org.

[3] **http://www.worldwidelearn.com/elearning-industry/articles.htm**

[4] http://www.LlearningGuild.com

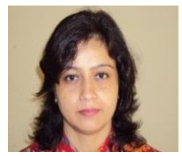
[5] http://www.namahn.com/resources/documents/note-e-learning.pdf

[6] http://www.nwlink.com/~donclark/hrd/elearning/growth.html

[7] http://www.astd.org/TD/Archives/2010/Nov/Free/1110_2010+State+of+the+Industry.htm

[8] http://www.trainingmag.com/managesmarter/images/pdfs/IndRep06.pdf

AUTHORS PROFILE



Sangeeta Kakoty received the B. E. Degree in Computer Science and Engineering from Jorhat Engineering College, Assam in 1995. She received M.Sc. and M. Phil degree in Computer Science from Gauhati University, Assam and from Vinayaka Mission University, Tamil Nadu in 2004 and 2007 respectively. She was working as a lecturer in The Department of Computer Sc., Jagiroad College, Assam since 2005. She got the Research and Teaching Assistant (RTA) fellowship from IGNOU, New Delhi. Now she is a research fellow of IGNOU and working as a RTA in NECRD, IGNOU since 2008. Her main interest includes Expert Systems and Data Mining.

Prof. Manohar Lal is a product of IIT Kanpur, IIT Delhi and University of Delhi. He completed his MTech in Computer Science and Engineering (CSE) from IIT Kanpur; pursued his second PhD in CSE from IIT, Delhi. Earlier, he completed his master's and PhD programmes in Mathematics from University of Delhi. He has vast teaching experience of more than 42 years at various universities including University of Delhi and JNU. He has been Professor and Head of Computer Science Department and Director, Computer Centre/Division at HP University, Shimla as well as MD University, Rohtak. Prof. Lal has long research experience. Earlier he worked in the area of 'Error - Correcting Codes' (a branch of networking/Data Communications). Currently he is working in the areas of 'Automation of Reasoning' (a branch of Artificial Intelligence) and Computer Networks. He has large number of publications in these areas in national and international journals of repute and has been guiding research students in these areas. Also he is associated with large number of academic bodies in various capacities. Dr. Lal is with IGNOU as Professor since July 1998 and as Director, SOCIS from January 03, 2000.

Prof. Shikhar Kr. Sarma got his Ph.D. in the field of computer networks from Gauhati University, India. He is currently holding the position of Professor and Head of Department of Computer and Information Technology at Gauhati University, India. His earlier assignment includes Indian Institute of Technology Guwahati, India, Rajiv Gandhi Central University, India, and Debub University, Ethiopia. His current research areas are NLP and Expert Systems. He holds many major research funding including the Language Technology Development Project, Local Language Search Engine Development Project of Ministry of Communication and IT, Govt. of India etc. He has written numerous research papers, and currently guiding 12 Ph.D. Students.