Computer Science & Engineering Research Journal, CUET ■ Vol. 07 (2010-2011) ■ ISSN: 1990-4010

Vol. 07 (2010-2011) ISSN: 1990 - 4010

Computer Science & Engineering Research Journal

Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong-4349, Bangladesh

ISSN: 1990-4010

Computer Science and Engineering Research Journal

Vol. 07 (2010-2011)

Department of Computer Science and Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong-4349, Bangladesh

ISSN: 1990-4010

Computer Science and Engineering Research Journal

Vol. 07 (2010-2011)

Department of Computer Science and Engineering

Chittagong University of Engineering & Technology (CUET)

Chittagong-4349, Bangladesh

ISSN: 1990-4010

Computer Science and Engineering Research Journal

Published By:
Department of Computer Science and Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong-4349, Bangladesh

Published one or two volume annually. All communications to the Associate editor, Computer Science & Engineering Research Journal, Department of Computer Science & Engineering, Chittagong University of Engineering & Technology (CUET), Chittagong-4349, Bangladesh.

ISSN: 1990-4010

Computer Science and Engineering Research Journal

Vol. 07 (2010-2011)

Contents **Page**

ON-DEMAND SCHEDULING STRATEGIES IN ROAD SIDE UNITS (RSUs)-BASED VEHICULAR AD HOC NETWORKS (VANETs)

G. G. Md. Nawaz Ali[1,2]

[1]Dept. of Computer Science, City University of Hong Kong, HK SAR

[2]Dept. of Computer Science and Engineering, Khulna University of Engineering & Technology, Bangladesh

Abstract: Recently, the use of Road Side Units (RSUs) has been proposed as a mechanism to handle the connectivity issues in VANETs for data dissemination. In this paper, we provide a model where RSU deals with both download and upload queues. In VANETs, since vehicles are highly mobile, if an RSU fails to receive the updated information from a vehicle; all the subsequent vehicles receive the stale data from that RSU which substantially decreases the main objective of data dissemination. To find an efficient data dissemination procedure in this circumstances, we propose three different scheduling algorithms for scheduling the requests from the both queues and study the performance of a number of different on-demand scheduling algorithms using simulation experiments with various parameter settings and high workload. Finally, we recommend with which scheduling algorithm is sustainable in this environment and which parameters should consider to achieve the best result from our proposed different scheduling algorithms.

Keywords: *VANETs, Road Side Unit (RSU), on-demand scheduling algorithm, on-demand broadcast etc.*

1. INTRODUCTION

Data dissemination in Vehicular Ad Hoc Networks (VANETs) received considerable attention by the researchers from the past decade. In VANETs, as many vehicles may request the same data item, so broadcasting is a popular approach for data dissemination. Recently, researchers have proposed the use of Road Side Units (RSUs) for supporting on-demand data broadcasts, particularly where strict time constraints are involved.

However, in such a scenario using RSU, when many vehicles need to upload and download data in the same RSU, an efficient scheduling strategy is required. Time constraint is an important issue here, because a RSU's transmission range is not large, and failure in reaching the vehicles while they are in the range will result in wasted transmission; similarly poor scheduling might prevent updates regarding time-sensitive data that are useful to other vehicles from being uploaded in time. For example, If a vehicle has observed a road accident while it approaches an RSU, it then can provide this information to that RSU. The RSU updates its database and provides this updated information to other vehicles, and upon getting this information these vehicles may change their routes or take appropriate actions.

A number of researchers have studied scheduling issues recently. Nadeem et al. [1] use periodic broadcast approach for data dissemination. Zhang et al. [2] consider both the upload and download service and try to balance the adaptivity of these two services according to the fluctuation of workload but they do not maintain the time constraint and update the database with most updated information. Yi et al. [5] consider reliability and fairness of information distribution among Mesh Road Side Units (MRUs) but their approach does not deal with the strict time constraints of VANETs data dissemination from RSUs to vehicles.

We study the performance of different scheduling algorithms for both uploading and downloading requests considering different constraints. In this paper our main contributions are:

- We propose three different scheduling algorithms:

1. **Fairness of Service Scheduling Algorithm** for maximizing the fair service to both upload and download queue.

2. **Service Utilization Scheduling Algorithm** for maximizing the channel bandwidth utilization of an RSU.

3. **Data Freshness Scheduling Algorithm** for maximizing the freshness of the downloaded data.

- We study the performance of number of different on-demand scheduling algorithms in each of our proposed scheduling algorithms and analyze the joint scheduling performance.

- We find out by analysis which on-demand algorithm has the best performance and recommend the appropriate parameters to achieve the best result in the RSU-based VANETs environment.

The rest of this paper is organized as follows. Section 2 surveys the related work. Section 3 describes about our system model and preliminaries, section 4 shows our proposed scheduling algorithms and section 5 describes the simulation model and experimental results. Finally, we conclude with the a discussion of our results and future work.

2. RELATED WORK

Broadcasting maximize the channel bandwidth utilization because by a single broadcast many outstanding requests can be served, unlike unicasting where for every requested data needs to be disseminated individually. To get the maximum benefit from the broadcast, scheduling algorithm plays an important role.

A number of push based model have been proposed by researchers. Wong and Ammar [6] investigate the First Come First Serve (FCFS) algorithm in video-tex systems. Acharya et al. [7] introduce asymmetric communication environments where downstream link has greater capacity than upstream link. Vaidya and Hameed [8] formulate the square root rule for minimizing response time from the broadcast server. Other researchers propose pull based (also known as on-demand). Wong [9] uses the Longest Wait First (LWF) algorithm to find the next item for scheduling. Xuan et al. [10] propose a Broadcast on Demand (BoD) service model and study a number of algorithms based on EDF. Aksoy and Franklin [11] propose the R×W algorithm for large scale on-demand data broadcast, which incorporate popularity and request urgency for making scheduling decision. Fernandez and Ramamritham [12] propose a hybrid algorithm called Time Critical Adaptive Hybrid Broadcast (TC-AHB) for time-critical asymmetric communication. For heterogeneous workload Acharya and Muthukrishnan [13] introduce a new metric called stretch which is the ratio of response time to the service time and a corresponding algorithm called Longest Total Stretch First (LTSF). Wu and Cao [14] reduce the calculation overhead of LTSF algorithm. [3] Proposes Slack Time Inverse number of Pending requests (SIN) for time critical on-demand broadcast. Chen et al. [4] introduce Preemptive Temperature Inverse Slack Time (PTIS) for handling multi-item data requests.

However none of the work considers the scheduling issue for both upload and download requests along with clients mobility and strict time constraints.

3. BACKGROUND AND PRELIMINARIES

3.1 System Model

In our model, we assume that VANET services are provided to the vehicles at the hot spot zones such as gas station or intersection of the roads where number of vehicles gather or pass naturally higher than the other areas. When a vehicle is in the transmission range of an RSU it can generate requests. Request type can be either download or upload. A download request means a vehicle wants the latest updated data from the RSU server and upload means vehicle wants to upload the updated information of a data item to the RSU server.

A RSU has two queues as shown in Figure 1, one for upload requests referred as *Upload*

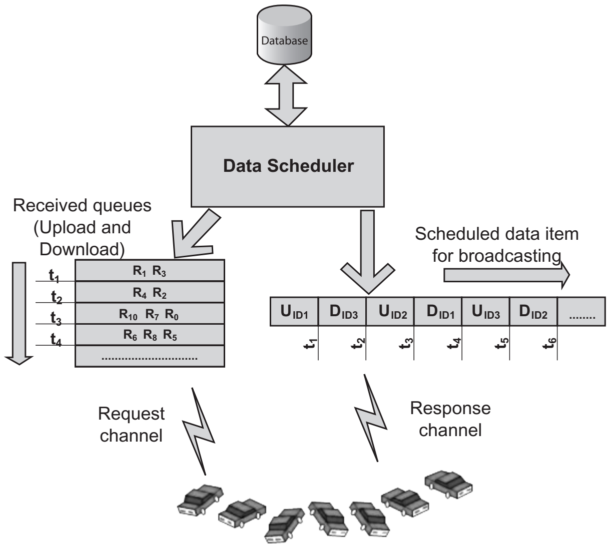


Fig. 1 Scheduling in an RSU.

Queue and the other for handling download requests referred as *Download Queue* from the vehicles. During each scheduling decision time, our model uses an existing on-demand scheduling algorithm for finding the best candidate from the uploading and downloading queue and then our proposed scheduling algorithm to make final scheduling queue for next service cycle. Each decision time at most scheduling window size requests are taken in the scheduling queue. A vehicle can generate request or receive response only while it is within the transmission range of an RSU.

3.2 Notation and Assumptions

When a vehicle submits a request, it submits some information; we denote this information as a request tuples. Each submitted request R_i carries the following tuples:

$$R_i = (NO_i, ID_i, SIZE_i, TYPE_i, T_i^{in}, T_i^{out}, T_i^{r}, T_i^{stamp}, T_i^{deadline}, T_i^{serv})$$

NO_i : the number of the request; the ID_i : of the requested data item; $SIZE_i$: the size of the requested data item; $TYPE_i$: taking values in {upload, download}, indicating the type of uploading/downloading operations; T_i^{in}: the time the vehicle enters the communication range of the RSU; T_i^{out}: the time the vehicle leaves the communication range of the RSU; T_i^{r}: the time the request is generated; T_i^{stamp}: the time the updated information is generated by a vehicle; $T_i^{deadline}$: the deadline assign by a request, beyond this time the request will be dropped; T_i^{serv}: the time for uploading/downloading the data item, it can be evaluated by $SIZE_i$ divided by the available channel bandwidth. Assume there are n requests at time t . The set of requests is denoted by $R^t=\{R_1, R_2, \dots , R_n\}$.

Schedule: When a vehicle submits a request, the request needs to be scheduled. Assume at time t, a set of requests R^t reside in the RSU received queue to be scheduled. The schedule to the requests should follow the following principles. **First**, since each request needs to occupy the communication channel for data transmission, it should make sure that the uploading/downloading operation finishes before the vehicle moves out of the communication range. **Second**, since the uploading operation will update data content, if there exist both uploading and downloading requests for the same data, the uploading request should be served first to avoid the downloading of a stale data item.

If a request $R_i \in R^t$ is scheduled to be served at t, we call R_i is *satisfiable* if it meets the following 2 conditions: (1) $t \geq T_i^{in}$ & $t + T_i^{serv} \leq T_i^{out}$; and (2) there is no uploading request to data ID_i in R^t. If either of them is violated, we call it *unstaisfiable*.

Due to the broadcasting nature of wireless communication, for the downloading operation, when a data is broadcast, a set of requests waiting for the same data can be satisfied at the same time. We call such set of requests *shareable* requests.

Request's Life Time:

Assume the radius of the transmission range of an RSU is R meters and average speed of vehicle within the transmission range of RSU is S meter/sec. So, if a vehicle reaches at the transmission range of an RSU at time t = 0 and generate the first request at the same time, then the average deadline of the first request is, AVERAGEDEADLINE=$\frac{2R}{S}$. If we consider the speed variation factor of a vehicle from the average speed, the compensated *DEADLINE* is, $DEADLINE = \frac{2R}{S} * uniform(\psi_{max}, \psi_{min})$, here ψ is random number use for compensate the speed variation. So, anytime T, the deadline of a request is Deadline

$= \frac{2R}{S} * uniform(\psi_{max}, \psi_{min}) - T_i^r$,

where T_i^r is request generation time. Suppose, the request R_i wants to update the server data item ID_i and the server's and vehicle's data item time stamp are $ID^{T\,stamp}_{server}$, $ID_i^{T\,stamp}$, respectively. Then, the update request will be receive by the upload queue if and only if, $ID_i^{T\,stamp} < ID^{T\,stamp}_{server}$. However, Any type request R_i will be discarded from the scheduler when $Current_{Clock} > T_i^{deadline}$.

Scheduling Window Size: Scheduling window size defines, during each scheduling decision time, the maximum number of requests taken in the scheduling queue to choose the appropriate requests for the next service cycle for serving from the both upload and download queue.

3.3 Performance Metrics

To compare and analyze the performance of our experimental result we use the following metrics:

1. **Deadline Miss Rate**: It defines percentage number of requests misses the deadline to the total number of requests received by an RSU. So,

$$Deadline\ Miss\ Rate = \frac{DMN}{DMN + SN} * 100\%$$

Where DMN and SN are Deadline Missed Number and Satisfied Number respectively.

2. **Throughput:** It is the number of requests successfully served by the scheduler per unit time

3. **Average Response Time:** It measures the average time needed for a request to get the response after submitting that request to an RSU.

4. **Computational overhead:** The number of times a scheduler needs to make the scheduling decisions to serve all the vehicles' requests until all the vehicles pass the transmission range of an RSU.

4. SCHEDULING SCHEMES

4.1 Necessity of two-step scheduling

The conventional one-step scheduling algorithm only deals one type request, mostly download requests [4, 7, 8, 9, 11, 13]. However, as in VANETs we also need to get the updated information from the vehicles, here we have to deal with two types requests. Both the requests compete for the bandwidth of the RSU response channel. Hence we need to formulize a model which can set the rules which type request will get higher priority over other type, because different type requests purpose is different, we can't compare straightforwardly between them. Zhang et al. [2] proposes a two-step scheduling model where they do the comparing between request type queue levels not request level. However, some requests in the lower priority queue may have higher urgency than some requests in the higher priority queue and these requests may miss their deadline because of residing in the lower priority queue. Considering this issue, in our proposed model, we schedule in the request level, not in the queue level, so each request will get their reward based on their metrics value (deadline, size, and/or popularity). As at a time we can only serve one request, hence first we have to find out the suitable candidates from the both queues (first-step scheduling) and then we need to finally schedule from these both queues suitable candidates (second-step scheduling).

4.2 Two-step Scheduling Procedures

First-step: we sort the requests of upload and download queue to assign them priority according to the principle of used existing on-demand scheduling algorithm. **Second-step:** Using our proposed scheduling algorithms, we form the scheduling queue from the first-step sorted queues. For performing first step, we use some popular on-demand algorithms and in second-step, we apply all of them in each of our proposed second-step scheduling algorithms to measure the overall system performance

4.3 Existing On-demand Scheduling Algorithms

For the first step scheduling, we use the following on-demand algorithms:

1. **First Come First Served (FCFS):** It selects the request from the waiting queue according to their arrival time. Request having highest arrival time will get the highest priority in this approach.

2. **Most Request First (MRF):** It selects request according to the popularity of the requested data item. Data item requested by the maximum number requests will be selected first.

3. Earliest Deadline First (EDF): EDF selects requests according to the deadline of the requests, which requests has the minimum deadline will be selected for being served first.

4. Deadline Size Inverse number of pending Requests (DSIN): DSIN algorithm incorporates the deadline of the request, size and popularity of the requested data item. It selects the data item which has the minimum DSIN_Value. DSIN_Value of a request is:

$$DSIN_{value} = \frac{Deadline\ x\ Size\ of\ the\ requested\ data\ item}{Number\ of\ pending\ requests}$$

4.4 Our Proposed Scheduling Algorithms

After first-step scheduling, we get the sorted upload and download queue where requests in the front side of the queues are the best candidate for being serviced. For second-step scheduling, we propose the following three different scheduling algorithms:

1. Fair Service (FS) Scheduling Algorithm:

Fair Service Scheduling Algorithm gives the fair service to the both queues. It takes one update request from the front end of uploading queue, pop this request from the uploading queue and insert it in the scheduling queue, then takes one download request from the front side of the downloading queue, pop it and insert it in the scheduling queue. Continue this procedure until either scheduling queue filled up or both the upload and download queue become empty. As we need to provide the updated information to the vehicles so before taking downloading request to scheduling queue it checks whether there is any upload request for that data item in the upload queue, if there is then it schedules the upload request first then the download request.

2. Service Utilization (SU) Scheduling Algorithm:

Usually, download requests need more bandwidth for transmission data item than upload requests. So, for utilizing the channel bandwidth, the Service Utilization algorithm takes request first from the downloading queue and see whether there is any request for that data item in the upload queue, if there is, then first schedule that upload request then the download request. This procedure continues until either the scheduling queue filled up or downloading queue becomes empty.

3. Data Freshness (DF) Scheduling Algorithm:

This is a greedy approach. This procedure greedily decides what percentage of the window size will be used by the update requests and the rest of them by download requests. This is defined by the freshness factor and this factor is adjustable.

Due to space limitation, we could not add the pseudo code of our above proposed three algorithms.

5. PERFORMANCE EVALUATIONS

5.1 Experimental Setup

Our simulation environment is similar to Figure 1. We assume the inter-arrival time of request is exponentially distributed, a data item can be requested by more than one vehicle and a vehicle can request for services until it exceeds the transmission range of an RSU. The request data access pattern is the commonly used Zipf [16] distribution with θ ranging from 0 to 1. We perform our simulation experiment using CSIM19 [15], other than the default parameters, we use parameters shown in the Table 1. For experimental data traces, we let all the vehicles pass the RSU transmission range repeatedly in similar fashion until we get the stable data from the same parameter settings.

Before collecting a traced data, we use 100 times iteration for the same settings while every time with different seed value.

5.2 Effect of Deadline Miss Rate

Fig. 2 shows the deadline miss rate, throughput and average response time of different on-demand algorithms; FCFS, MRF, EDF and DSIN are applied in the Fair Service scheduling algorithm by varying number of vehicles. When the number of vehicles increases, more requests are in the received queue, and predictably more requests miss their deadlines for every algorithm. However, in the case of throughput when scheduling algorithms has better option to choose the more suitable requests among many requests their throughput rises. Average response time has no much fluctuation except in the initial stage. Considering all metrics DSIN algorithm outperforms the other algorithms. Service Utilization and Data Freshness algorithms also exhibit the same result (not shown in this paper).

5.3 Effect of θ

Figure 3 shows the effect of varying the Zipf parameter θ value from 0 to 1 for Fair Service

(FS) scheduling algorithms. When θ value is 0, vehicles requests are uniformly distributed; hence popularity does not dominate for requests selection. However, with the increasing of θ values clients request pattern becomes more skewed, popular data items been requested by many vehicles' requests, then by a single broadcast many requests are served. So, for increased θ value deadline miss rate decrease, throughput increase and average response time decreases. As MRF and DSIN algorithm use popularity as requests selection criteria they get the most benefit for increasing θ. However, in respect of the entire performance criteria DSIN algorithm performs better than all others. Due to space limitation, we could not add the result of SU and DF scheduling algorithms, though they give the identical result of FS algorithm.

Table 1 Simulation parameters.

Parameter	**Default**	**Range**	**Description**
Num Vehicle	40	10 - 100	Number of vehicles
IGTM	0.5	0.1 - 1	Request generation interval of each vehicle
DBSize	100	100 - 500	Number of data items in the database
DownItemsize	-	10 - 512 K bytes	Size of each download data item
UploadItemsize	-	5 - 256 K bytes	Size of each upload data item
BrodcastBandwidth	100	-	Channel broadcast bandwidth K bytes / s
RSU Range	350m [5]	350 - 400 m	RSU communication range
MAXSCHEDULE	20	1 - 40	Scheduling window size
FRESHNESSFACTOR	0.7	0.0 - 1.0	Scheduling window size for upload requests
THETA	0.8	0 - 1	Zipf distribution parameter
GRate	-	0.5 - 1.3	Updated upload requests generation parameter

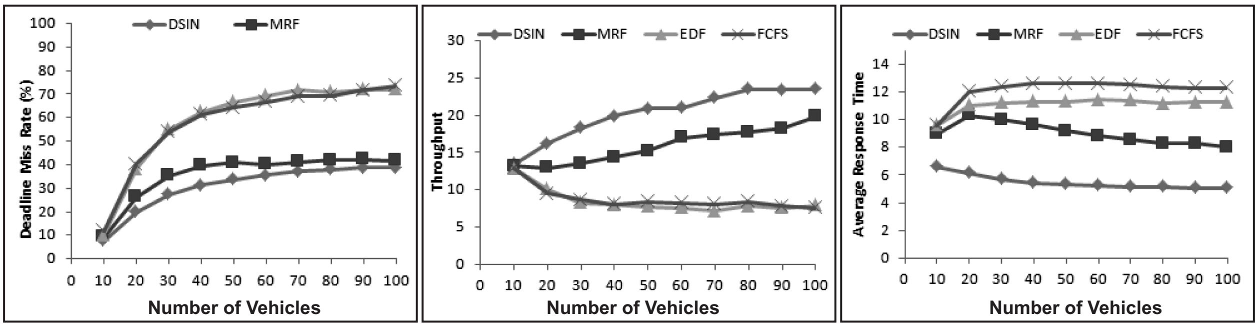


Fig. 2 Performance of different on-demand first-step scheduling algorithms for varying number of vehicles while using Fair Service (FS) scheduling algorithm in the second-step: (a) Deadline miss rate, (b) Throughput and (c) Average response time.

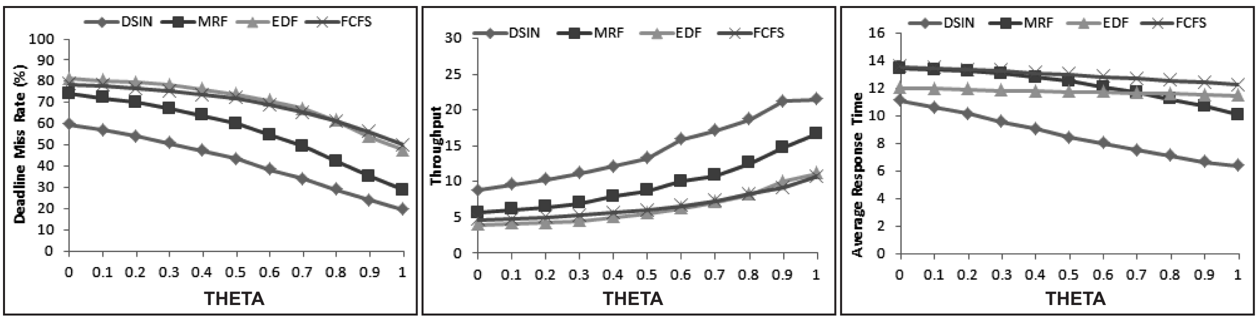


Fig. 3 Performance of different on-demand first-step scheduling algorithm for varying θ while using Fair Service (FS) scheduling algorithm in the second-step: (a) Deadline miss rate, (b) Throughput and (c) Average response time.

5.4 Effect of Parameters (Freshness Factor and Window Size)

Form the above experimental result, we find that in terms of deadline miss rate, throughput and average response time DSIN algorithm performs better than all other on-demand scheduling algorithms. So, we apply the DSIN algorithm as a representative of all the on-demand algorithms to analyze the effect of the freshness factor in Data Freshness scheduling algorithm and window size in Fair Service, Service Utilization and Data Freshness. This analysis part is not included in this paper.

5.5 Complexity Analysis of Our Proposed Scheduling Algorithms

In the first-step we sort the both upload and download queue according to the used first-step scheduling algorithm (EDF, DSIN etc.). Assume we have total n number requests in each queue and scheduling queue window size is w. As we know based on the used sorting algorithm, one queue sorting complexity is $0(n^2)$ (e.g. bubble, selection sort etc.) or $0(n \log n)$ (e.g. Merge, Heap sort etc.). For our proposed first second-step algorithm FS, say we schedule k number download requests among w (rest $w - k$ number from upload queue), hence we need to search k times in the upload queue (size n) to find those requests which ask for the same data item as those k number download request do. Hence, total complexity of FS algorithm is $0(2n^2 + k \times n) \equiv 0(n^2)$, the second algorithm SU also yields the same complexity. However, for the third algorithm DF, we don't need to search in the upload queue for some specific upload requests, it just takes greedily some upload and download requests, hence the total complexity of DF algorithm is $0(2n^2 + 2) \equiv 0(n^2)$. Hence, we can say that for considering upload requests along with download requests for providing updated data item to the vehicles, our two-step scheduling procedure doesn't create more complexity than that for sorting a queue. That's why; the scheduling decision taking computational overhead is almost same for all the algorithms in fig 4 (d).

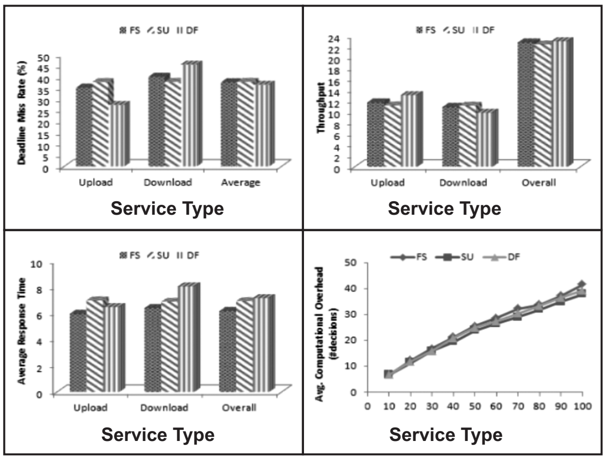


Fig. 4 Performance comparison of our proposed different second-step scheduling algorithms while using DSIN as the first-step scheduling algorithm: (a) Deadline miss rate, (b) Throughput, (c) Average response time, and (d) Average computational overhead.

5.6 Performance Comparison of Our Proposed Scheduling Algorithms

Figure 4 shows the performance comparison of Fair Service, Service Utilization and Data Freshness scheduling algorithm in terms of service type. For this experiment, we use DSIN for first-step scheduling and set the number of vehicles to 100. Figure 4(d) shows the average computational overhead is almost same for all our proposed scheduling algorithms with increasing workload, however, 4(a), 4(b) and 4(c) show the difference in terms of deadline miss rate, throughput and average response time respectively. Data freshness algorithm has the lowest deadline miss rate for upload requests. For upload requests Fair Service algorithm performs better than Service Utilization algorithm in respect of all the metrics. For handling download requests Fair Service and Service Utilization algorithm achieve almost the same throughput. In terms of average response time Fair Service algorithm outperforms the other two.

6. CONCLUSION AND FUTURE WORKS

To update the RSU's database with the most updated information is an important issue for effective data dissemination in VANETs. In this paper, we propose two-step scheduling operation. In the first-step, we use several existing on-demand algorithms to form the sorted uploading and downloading queues. In the second-step, we propose three different scheduling algorithms: Fair Service, Service Utilization and Data Freshness to ensure fairness for both upload and download requests, improve utilization of the channel bandwidth and provide updated data to the clients respectively.

After performing a series of simulation experiments, we observe that the DSIN algorithm, which incorporates deadline, data item size and popularity of the data, performs the best among the different on-demand scheduling algorithms for first-step scheduling. For second-step scheduling, the Data Freshness algorithm performs better when freshness factor lies in the range of 0.5 to 0.7, it means when work load is high, upload requests should use 50% to 70% of total scheduling window size and rest of it for download requests. All scheduling algorithms have better performance when the scheduling window size is in the range of 15 to 25. Among these three scheduling algorithms, Fair Service has the lowest average response time value, Data Freshness has the lowest deadline miss rate regarding uploading service and Service Utilization maximize the channel bandwidth utilization by servicing more download requests along with decent throughput value. From the experimental result, we also conclude that, Data Freshness algorithm should be effective when the number of update requests is large, Service Utilization is suitable when needs to broadcast a lot of static information and Fair Service algorithm should be used during normal rush hours respectively.

In the future study, we plan to apply our scheduling algorithms to multiple RSUs with co-operative data access along with load transferring and balancing capabilities among them.

REFERENCES

1. Nadeem, T., Shankar, P., Iftode, L. *"A Comparative Study of Data Dissemination Models for VANETs"*, In proceeding of the Vehicular Ad Hoc Network, 2006.
2. Zhang, Y., Zhao, J., Cao, G., *"On Scheduling Vehicle-Roadside Data Access"*, In proceeding of the Vehicular Ad Hoc Network, 2007.
3. Xu, J., Tang, X., Lee, WC., *"Time-Critical On-Demand Data Broadcast Algorithms, Analysis and Performance Evaluation"*, Journal of IEEE Transactions on Parallel and Distributed Systems, vol. 17(1), pp. 3-14, 2006.
4. Chen, J., Lee, V.C.S., Chan, E, *"Scheduling Real-time Multi-item Requests in Wireless On-demand Broadcast Networks"*, In proceeding of the 4th International Conference of Mobile Technology, Applications and Systems, 2007.
5. Yi, L. Z., Bin, L., Tong, Z., Wei, Y, *"On Scheduling of Data Dissemination in Vehicular Networks with Mesh Backhaul"*, In proceeding of the IEEE ICC, 2008.
6. Wong, J. W., Ammar, M. H., *"Analysis of Broadcast delivery in Videotex System"*, Journal of IEEE Transactions on Computers, vol. C-34(9), 1985.
7. Acharya, S., Alonso, R., Franklin, M., Zdonik, S., *"Broadcast Disks: Data Management for Asymmetric Communication Environments"*, In proceeding of the ACM SIGMOD Conference, CA, 1995.
8. Vaidya, N., Hameed, S., *"Data broadcast in asymmetric wireless environments.* In proceeding of the 1st International Workshop on Satellite-based Information Services (WOSBIS)", 1996.
9. Wong, J. W., *"Broadcast Delivery"*, Journal of IEEE, vol. 76(12), 1988.
10. Xuan, P., Sen, S., Gonzalez, O., Fernandez, J., Ramamritham, K. *"Efficient and timely dissemination of data in mobile environments.* In proceeding of the *3rd IEEE Real Time Technology and Applications Symposium"*, Montreal, Canada, 1997.
11. Aksoy, D. and Franklin, M. R ×W, *"A Scheduling Approach for Large-Scale On-demand Data Broadcast"*, Journal of IEEE/ACM Transactions on Networking, vol. 7, 1999.
12. Fernandez, J., Ramamritham, K. *"Adaptive Dissemination of Data in Time-Critical Asymmetric Communication Environment"*, Journal of Mobile Networks and Application, 2004.
13. Acharya, S., Muthukrishnan, S., *"Scheduling On-demand Broadcasts: New Metrics and Algorithms"*, In proceeding of the MOBICOM, 1998.
14. Wu, Y., Cao, G., *"Stretch-Optimal Scheduling for On-Demand Data Broadcasts"*, In proceeding of the ICCCN, 2001.
15. Schwetman, H., *"CSIM19: A powerful tool for building system models"*, In proceeding of the 33th IEEE Winter Simulation Conference, Arlington, VA, USA, pp. 250-255, 2001.
16. Zipf, G., *"Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology"*, Addison-Wesley, Cambridge, MA, 1949.

GENERATION OF DOUGHNUT SHAPE OPTICAL VORTEX BY MEANS OF COMPUTER GENERATED PHASE DIFFRACTIVE OPTICAL ELEMENTS

M. O. Faruk[1] and D. Cojoc[2]
[1]Dept. of Physics, Shahjalal University of Science & Technology (SUST), Bangladesh
[2]LILIT Group, TASC-INFM Nanolithography Beamline at Elletra Synchrotron Light Source, Area Science Park, Trieste, Italy

Abstract: The aim of this work is to generate the Laguerre-Gaussian beam, often referred to as optical vortex that cannot be performed using common laser source. This beam can be obtained by converting the Gaussian beam generated by common laser source, by means of properly designed Computer Generated Phase Diffractive Optical Elements (CGPDOE). Laguerre-Gaussian beam is exhibiting doughnut like transversal intensity distributions and carrying orbital angular momentum about their axis. When viewed this beam in a perpendicular plane along the propagation axis, the vortex appears as a dark in the center region surrounded by a bright concentric ring of light. This paper gives detailed instructions for generating doughnut shaped optical vortex and optical vortex structure using Spatial Light Modulator (SLM) by means of CGPDOE.

Keywords: *Optical Vortex, Spatial light modulator, Computer Generated Phase Diffractive Optical Elements, Topological charge.*

1. INTRODUCTION

There has recently been increasing interest in the study of optical fields with wave-front dislocations. The wave front dislocation refers to a continuous line on which the wave phase is undefined (singular) and its field amplitude vanishes or, in a mathematical form, they are zero of complex functions with a nonzero phase circulation around them. If the light wave is represented by complex numbers of the form $Ae^{il\theta}$, where A is the amplitude of the field, θ is the angle measured in the plane transversal to the direction of propagation and l is the beam's topological charge which indicates the number of phase twisting around the dark region. The wave front dislocation occurs due to this twisting, the light wave along the z-axis cancels out because the value of the angular coordinate θ is not well defined at $r = 0$, giving rise to a phase singularity [1] where the amplitude vanishes. It is well known that such a wave front dislocation beam carries orbital angular momentum L_z. For light beams the value of L_z per photon is given by

$$L_z=-ihl \qquad (1)$$

Where Planck's constant is divided by 2π. From Eq. (1) it is evident that the continuity condition at $\theta=2\pi$ implies the quantization of the angular momentum. The value of L_z indicates that there is a rotation of the momentum vector P around the dark hole as the beam propagates. This analogy with the velocity fields of fluids suggest that the wave front dislocation observed in light beams should be called an optical vortex.

In recent years several important applications of optical vortices have been developed. For example, the absence of the gradient force in the central hole can be used to make optical tweezers that can trap neutral particulates [2], which receive the angular momentum associated with the rotation of the phase [3]. In nonlinear optics, wave guiding can be achieved inside the central hole [4,5]. In astronomy the singularity can be used to block the light from a bright star to increase the contrast of astronomical observations using optical vortex coronagraphs [6], which are useful for the search of extra solar planets. Applications in quantum information employing quantized properties of the angular momentum of vortex light beams have also been proposed [7]. Vortices are topological objects, which are important in many branches of physics such as fluid mechanics, Bio-Physics, superconductivity, Bose–Einstein condensates [8], and superfluidity [9].

Nowadays, optical vortices can be generated by means of Diffractive Optical Elements (DOE's). DOE's are optical devices which influence the wavefield by diffraction. A specified illumination wave can be converted to a diffracted (reflected or transmitted) wave with a desired distribution of its amplitude, phase or polarization. The various methods of producing DOE's are mode conversions [10,11], spiral phase plates [12], multilevel spiral phase plate [13,14], optical wedge [15,16] and computer generated DOE's [17]. Nanolithography may be used to create spiral Phase Plate having sub-micrometer features; however, such masks are formidably expensive and are not readily available. Contact lithography may also be used, but this approach is labor intensive because many well-aligned master masks are needed to achieve adequate phase resolution.

From the working principle point of view, there are two principal types of DOE's: Amplitude Diffractive Optical Elements (ADOE's) and Phase Diffractive Optical Elements (PDOE's). ADOE's, which change only the amplitude of the incident wave, have very low efficiency and hence are rarely used. Another important type of DOE's is represented by the phase diffractive elements (PDOE's), which are designed to implement the required transformations of the light field without the loss of light energy. Since they act only on the phase of the incident wave, PDOE's demonstrate a high efficient value, as the transforms of the wave field being performed without energy loss. In this paper, Computer Generated Phase Diffractive Optical Elements (CGPDOE's) are designed and implement on the Gaussian beams for producing optical vortex. The use of CGPDOE's and the availability of high quality spatial light modulator allow the control of the Gaussian beam which is easier, fast, inexpensive and more versatile of producing optical vortex [18] compared to the conventional arrangement.

2. DESIGN OF THE CGPDOE MASK

The aim of this section is to imprint a phase $e^{il\theta}$ in a Gaussian beam. Thus, The CGPDOE's will be made by the interference of a reference tilted plane wave e^{-ikx} and an object wave, $e^{il\theta}$ which carries the singular phase. When this CGPDOE's is illuminated by a beam containing the reference wave (in this experiment a Gaussian beam is used from a laser), the object is reconstructed and the vortex appears in the output beam.

In general a phase singularity propagating along the z direction will have the form

$$E(r, \theta, z) = E_0 \exp(il\theta) \exp(ikz) \qquad (2)$$

Next, consider a plane wave u, propagating obliquely to the axis

$$u = \exp(ik_x x - ik_z z)$$

Assume the recording device is located at z = 0 for simplicity. The intensity distribution may then be found by squaring the sum of the two amplitude functions:

$$I = 1 + E_0^2 + 2E_0 \cos(k_x x - l\theta)$$

A Fourier transform of this yields the transmittance function actually used to create the diffraction gratings [20].

$$T(r, \theta) = T_0 \exp[i\alpha \cos\{l\theta - (2\pi/\mu) r \cos\theta\}] \qquad (3)$$

where α is the amplitude of the phase modulation, T_O is the constant absorption coefficient of the CGPDOE's, and μ is period of the grating (fringe spacing).

The implement of these mathematics in computer is needed to use codes in Matlab. Cojoc and his group [18] developed codes and algorithms to calculate the suitable CGPDOE's. The code of designing CGPDOE's is based on the iterative algorithm to calculate the phase function. Efficient iterative techniques, based on the thin-element approximation and the scalar diffraction theory, have been proposed to design phase CGPDOE's with surface relief modulation in laser light. There are two approaches are used for CGPDOE's calculation, 1) Phase Retrieval using Iteration Algorithms (PRIA) followed by genetic algorithms for optimization [21], and 2) spherical wave propagation [22]. In this paper, PRIA approach is used for CGPDOE's calculation.

3. CGPDOE DESIGN ALGORITHMS

The Phase Retrieval using Iteration Algorithms (PRIA) [18] approach starts from the desired intensity pattern $I_f(\text{x})$ that the CGPDOE's should generate in the output plane, being $\mathbf{x}$ = (x, y) the vector position in a two-dimensional Cartesian system. This pattern is determined by the electric field of a collimated laser beam that is shaped by diffraction with a CGPDOE's and then transferred to the output plane by a suitable set of lenses.

When illuminating the CGPDOE's with a collimated laser beam, the diffracted field at a distance z after the CGPDOE's will be:

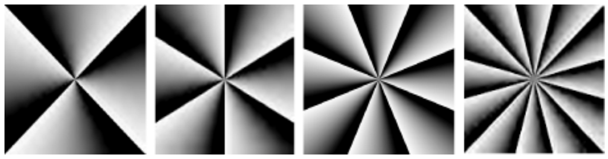
$$E_g(\mathbf{x},z)=P[A_g(\mathbf{x})\exp\{i\ddot{o}(\mathbf{x})\}] \quad (4)$$

Where $A_g(\mathbf{x})$ is the amplitude of the illuminating beam and P[*] is the propagation operator and ö($\mathbf{x}$) is the phase function.

The approach used in the implementation of PRIA is based on iterative algorithms that explore the space of phase distributions, ö(x), to retrieve the phase function that generates the desired intensity distribution. PRIA starts with the initialization of the diffracted field, E_g(x; z). The phase function ö(x) can be initialized with different phase distributions (e.g. random, constant, Gaussian). A randomly generated distribution was found to give the best result. One cycle of the iterative algorithm follows this sequence:

1. back-propagate the field E_g(x; z) to the CGPDOE's plane;
2. replace the amplitude of the resulting field with the amplitude of the illuminating beam, while the phase remains unchanged;
3. propagate the field at a distance z;
4. replace the amplitude of the resulting field with the desired amplitude, $|E_g(x;\ z)|$, and calculate the Mean Square Error (MSE) between the desired and obtained intensities.

Laser beam shaping for different purposes (i.e. Optical Tweezers) by repeating this cycle, the MSE decreases monotonically until the change of the MSE becomes insignificant and the algorithm is stopped. The phase calculated in the second step of the last iteration cycle is the phase function of the CGPDOE's. The CGPDOE's mask is generated by using PRIA algorithm as shown in Fig. 1. 1(a)-1(d). When one of these CGPDOE's is illuminated with Gaussian beam, the resulting far-field Fraunhofer diffraction pattern will be the doughnut shape optical vortices.



(a) (b) (c) (d)

Fig. 1 CGPDOE's for generating optical vortices with different topological charge *l*. The CGPDOE's shown in (a) imposes one 2ð phase change for one revolution around the beam axis, i.e. *l* = 4, (b) *l*=6, (c) *l* = 8 and (d) *l* = 12.

4. OPTICAL SETUP

The optical setup as shown in Fig. 2 was used for experiments relying on CGPDOE's. For the purpose of investigating the intensity profile, a vibration-isolated table was used as a platform.

A laser source was used to produce a Gaussian beam (a 1024 nm single mode diode laser with a spherical collimator to have the diameter of the beam of 5 mm exit). The power of the laser beam was 5 mW. A beam attenuation, ND (neutral density filter), was employed to decrease the laser beam intensity so that only an appropriate beam brightness was used for CCD. Two lenses L_1 (focal length 75 mm) and L_2 (focal length 400 mm) were used as a beam expander, so that the beam filled the LCD of the SLM which has an area of 20×20 mm. The distance between these two lenses is 475 mm and the magnification factor of the expander was 5.33 (m = 400/75 = 5.33). The beam splitter BS separated the incident Gaussian beam into two parts, one was reflected and another transmitted through the BS. The transmitted beam is received by the LCD of the SLM. The SLM modulates the phase of the incoming plane wave (the modulation depends on the pattern loaded into its RGB port). The SLM which we used in our laboratory was a reflection type Hamamatsu PPM X8267 series.

If one can make a phase mask by controlling LCD molecules in such a way to produce, for example, a gradual phase change from 0 to $2\pi l$ in a circular fashion across the incoming beam wave front (Fig. 5), then a helical wavefront of topological charge l will result.

An LCD based element works by applying an electric field between two walls of a cell containing appropriately oriented liquid crystals. The applied electric field causes LCD molecules to tilt, which results in a change in refractive index. By controlling refractive index change one can provide an appropriate phase shift to the incoming wavefront in order to produce a doughnut laser beam.

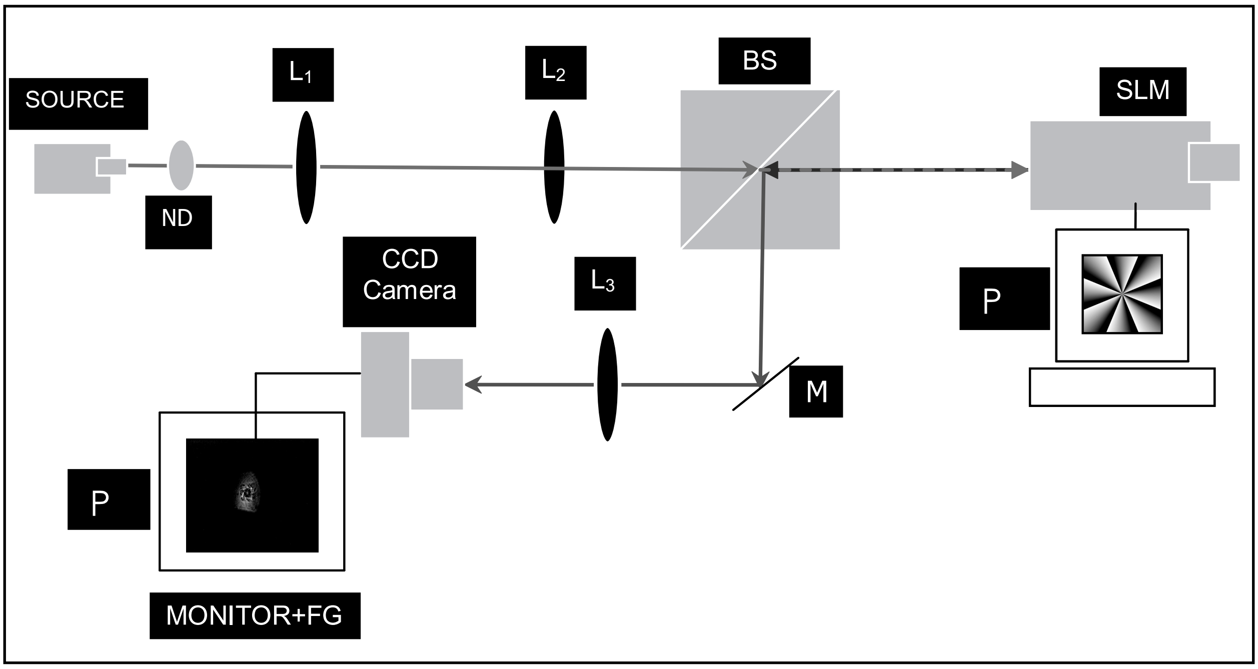


Fig. 2 Experimental setup for testing CGPDOEs consisting of Source - single mode diode laser (1024 nm, 5 mW); ND- neutral density filter; beam expander (lenses L1 and L2); BS- beam splitter, SLM- Spatial light modulator; M- Mirror; L3- Fourier lens , CCD- Charge coupled device, monitor with frame grabber (FG).

When an appropriate voltage is applied to the LCD, molecules inside the cell tilt according to the electric field strength. The amount of the tilt depends on the voltage giving rise to the corresponding change in refractive index for the light polarized along the long axis of the liquid crystal. For a given voltage such as 10V, one can select a proper thickness of LCD cell to produce a phase change from 0 at the first slice to 2π at the second slice in Fig. 5, so that the outgoing wave front has a helical shape with topological charge l=8. The advantage of using an SLM to modulate the phase according to the phase ramp pattern is in its case of use and ability to use a greater number of phase levels. For this reason such a device is increasingly used in the novel laser trapping and manipulation experiments[18].

The reflected modulated light is again passed through the beam splitter BS and the resultant modulated beam was then passed through the Fourier lens L_3 of focal length 200 mm.

It is known that the Fourier spectrum of an object is its diffraction pattern after having passed through a convergent lens in the focal plane of the lens. This modulated output laser beam is the LG beam. To investigate the different properties of this LG beam, a charge coupled device (CCD) camera was used. A CCD camera with a frame grabber was used to monitor the image of the input CGPDOE's. The images taken by the CCD camera were analyzed by a PC and software for processing the CCD images of the vortices.

5. EXPERIMENTAL RESULTS

Once the CGPDOE's mask has been made, a laser beam is passed through it, and vortices may be observed for different topological charge of the resultant diffraction pattern. A sketch of the experiment is depicted in Fig. 3

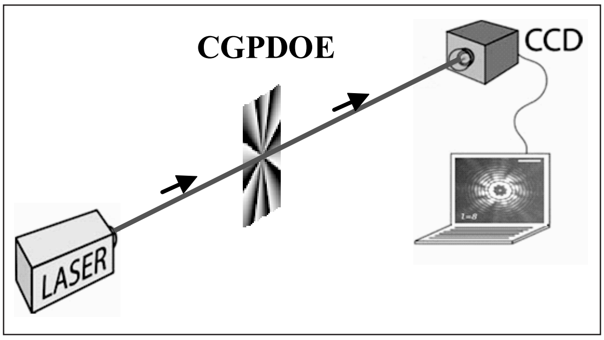


Fig. 3 Sketch of the generation of optical vortex by a CGPDOE's. A laser beam illuminates the CGPDOE's creating a diffraction pattern that can be observed with a CCD camera.

The phase pattern displayed on its LCD can be changed via computer allowing fast modification of the diffracted pattern. This gives dynamic control on the position, number and shape of the traps by simply projecting on the SLM the suitable sequence of CGPDOE's. The result of the experiment is illustrated in Fig. 4 where the intensity patterns in the focal plane of a microscope objective are shown for different topological charges *l*.

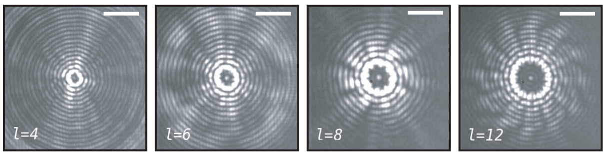
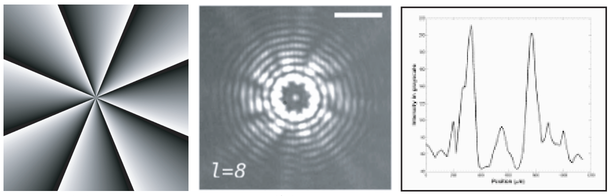


Fig. 4 The intensity patterns in the focal plane of a microscope objective are shown for different topological charges *l*. The diameter of the "doughnut" can be precisely controlled by tuning *l*. Magnification of the objective 60× and scalebar: 10μm.

Fig. 4 shows, the conversion of a 1064 nm Gaussian beam into a Laguerre- Gaussian mode is performed using CGPDOE's such as those shown in Fig. 1, implemented on a computer driven SLM.

The CGPDOE's can be implemented by a SLM driven by a computer. CGPDOE's modulate the phase of the input beam to produce the desired wave front, by imposing a pattern on the phase delays



(a) (b) (c)

Fig. 5 (a) CGPDOE's for generating optical vortices with topological charge *l*=8, (b) The intensity patterns in the focal plane of the microscope objective are shown for same topological charges and (c) Intensity vs. position curve for the doughnut beam of same topological charge.

The intensity distribution analysis of the doughnut shape optical vortex of topological charge *l*=8 is illustrated in Fig. 5.

A unwanted center spot of the doughnut is shown in the above Fig. 5(b). Practical CGPDOE's only diffract a portion of the incident light into the intended modes and directions. The maximum theoretical efficiencies of CGPDOE's is 40.4% [18]. The undiffracted portion of the beam typically forms an unwanted central spot.

Furthermore, CGPDOE's have phase profiles that vary continuously between 0 and 2π, which need to be discredited prior to implementation. This step necessarily introduces errors. Moreover, the mismatch of the phase levels implemented on the LCD from the calculated values may exist a small contribution of zero order. The sum over the CGPDOE's of the phase modulation errors introduced by the LCD gives a constant term in the CGPDOE's plane, which is converted in the zero order peak in the observation plane.

6. CONCLUSION

In conclusion, it is shown that, an optical vortex generated by means of a CGPDOE's which is exhibiting doughnut like transversal intensity distribution. This kind of optical vortex generation demonstrates easy fabrication, compactness and very low cost. These results show promising application of phase CGPDOE's mask in generating optical vortex.

7. ACKNOWLEDGEMENT

The authors express their regardful thanks to the

Abdus Salam International Centre for Theoretical Physics, Trieste, Italy and UNESCO for the financial support and TASC Laboratory for their hospitality at the center.

REFERENCES

1. Nye, J. F. and Berry, M. V., *"Dislocations in wave trains"*, In proceeding of the R. Soc. London, Ser. A, vol. 336, pp. 165–190, 1974.
2. Gahagan, K. T. and Swartzlander,G.A., *"Optical vortex trapping of particles"*, Opt. Lett., vol.21, pp. 827–829, 1996.
3. Simpson, N. B., Dholakia, K., Allen, L., and Padgett, M. J.,*"Mechanical equivalence of spin and orbital angular momentum of light: An optical spanner"*, Opt. Lett., vol. 22, pp. 52–54, 1997.
4. Carlsson, A. H., Malmberg, J. N., Anderson, D., Lisak, M., Ostrovskaya, E. A., Alexander, T. J., and Kivshar, Y. S., *"Linear and nonlinear waveguides induced by optical vortex solitons"*, Opt. Lett., vol. 25,pp. 660–662, 2000.
5. Salgueiro, J. R., Carlsson, A. H., Ostrovskaya, E., and Kivshar, Y., *"Second harmonic generation in vortex-induced waveguides"*, Opt. Lett., vol. 29, pp. 593–595, 2004.
6. Foo, G., Palacios, D. M. and Swartzlander, G.A., *"Optical vortex coronagraph"*, Opt. Lett., vol. 30, pp. 3308–3310, 2005.
7. Mair, A., Vaziri, A., Welhs, G., and Zeilinger, A., *"Entanglement of the orbital angular momentum states of photons"*, Nature, vol. 412, pp. 313–316, 2001.
8. Matthews, M. R., Anderson, B. P., Haljan, P. C, Hall, D. S., Wieman, C. E., and Cornell, E. A., *"Vortices in a Bose-Einstein condensate"*, Phys. Rev. Lett., vol. 83, pp. 2498–2501, 1999.
9. Rayfield, G. W. and Reif, F.,*"Quantized vortex rings in superfluid helium,"* Phys. Rev. A, vol. 136, pp. A1194–A1208, 1964.
10. Abramochkin, E., Volostnikov, V., *"Beam transformations and nontransformed beams"*, Opt. Commun. , vol. 83, pp. 123-128, 1991.
11. Beijersbergen, M. W., Allen, L., van der Veen, H.E.L.O., and Woerd-man, J. P.,*"Astigmatic laser mode converters and transfer of orbital angular momentum"*, Opt. Commun., vol. 96, pp. 123-127, 1993.
12. Beijersbergen, M., Coerwinker, R. P. C., Kristensen, M., and Woerdman, J. P.,*"Helical-wavefront laser beams produced with a spiral phaseplate"*, Opt. Commun., vol. 112, pp. 321-329, 1994.
13. Rotschild, C., Zommer, S., Hershcovitz, O., and Lipson, S. G., *"Adjustable Spiral Phase Plate"*, Appl. Opt., vol. 43, pp. 2397-22405, 2004.
14. Kotlyar, V. V. and Kovalev, A. A., *"Fraunhofer diffraction of the plane wave by a multilevel (quantized) spiral phase plate"*, Opt. Lett., vol. 33, pp. 189-194, 2008.
15. Izdebskaya, Y., Shvedov, V. and Volyar, A.,*"Generation of higher-order optical vortices by a dielectric wedge"*, Opt. Lett., vol. 30, pp. 2472-2479, 2005.
16. Yuan, X. C., Ahluwalia, B. P. S., Chen, H. L., Bu, J., and Lin, J., *"Generation of high-quality optical vortex beams in free-space propagation by microfabricated wedge with spatial filtering technique"*, Appl. Phys. Lett., vol. 91, pp. 051103-051109, 2007.
17. Heckenberg, N. R., McDuff, R., Smith, C. P.,*"Generation of optical phase singularities by computer-generated holograms"*, Opt. Lett., vol. 17, pp. 211-233, 1992.
18. Faruk, M. O., *"Multiple Optical Vortices Generated by Means of Difractive Optical Elements"*. P.G.D. Thesis, The Abdus Salam International Centre of Theoretical Physics, Trieste, Italy, 2006.
19. Cojoc, D., Fabrizio, E. D., Businaro, L., Cabrini, S., Romanato, F., Vaccari, L., and Altissimo, M., *"Design and fabrication of diffractive optical elements for optical tweezers arrays by means of e-beam lithography"*, Microelectr. Eng., vol. 61, pp. 963-969, 2002.
20. Guo, C. S., Liu, X., Ren, X. Y., and Wang, H. T.,*"Optimal annular computer-generated holograms for the generation of optical vortices"*, J. Opt. Soc. Am., Vol. 22, pp. 385-390, 2005.
21. Cojoc, D., Emiliani, V., Ferrari, E., Cabrini, S., Proietti, R. Z., and Fabrizio, E. D.,*"Multiple optical trapping by means of diffractive optical elements"*, Japan J. Appl. Phys., vol. 43, pp. 3910-3915, 2004.
22. Gershberg, R. W. and Saxton, W. O., *"A practical algorith for the determination of phase from image and diffraction plane pictures"*, Optik, vol. 35, pp. 237-246, 1972.

GRID COMPUTING TECHNIQUE TO DEVELOP A PORTAL-BASED WEB APPLICATION FOR SEQUENCE ALIGNMENT

Muhammad Ibrahim Khan[1] and Md. Rezaul Karim[1]
[1]Dept. of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh

Abstract: Grid computing combines computer resources from many administrative domains to apply their computational powers to complex tasks, often employing software to divide up the tasks and spread them among many resources. A sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Using the grid computing sequence alignment can be done easily and more efficiently. In this paper we need not be over emphasized on sequence alignment as it is well known to the society of bioinformatics. So the paper mainly concentrates on the significance of lowering the time and space complexity of large volume of sequence alignments by distributing the load to a number of machines using existing Grid and portal based systems.

Keywords: *Grid, Globus, GridSphere, Web portal, Middleware.*

1. INTRODUCTION

The term Grid refers to a heterogeneous network of computers that communicate over the Internet, for processing large data sets and computationally intensive tasks to be used jointly in [1]. The idea is to build a powerful and cost effective meta computing infrastructure without using high performance servers. Importantly, grid computing allows users to control what resources are shared, by whom, at what time and under what security and authentication conditions. GridSphere Portal Framework is an open-source portlet-based Web portal that allows developers to quickly create and package third-party portlet Web applications that can be used on the Grid.

2. RELATED RESEARCH

ABCGrid is the application on bioinformatics computing grid in [2] and GLAD which is also used as tools for large scale bioinformatics grid in [3]. There are several gridsphere portals which is currently active. Such as, University of California Grid Portals (UCGP) in [4], GEON (GEOscience Network) grid portal in [5], CHRONOS Portal in [6], P-GRADE Grid Portal in [7].

3. PROPOSED METHOD

There are several Global Grid technologies such as, Globus, Condor, Nimrod-G, Gridbus. We have worked with Globus technologies and it has some benefits from the other technologies such as, Globus provides the infrastructure that allows applications to handle such dispersed and mixed resources as if they constituted a single, virtual machine. Globus is made up of a layered architecture so that high-level global services are added atop low-level local core services. Its toolkit gives consideration to issues such as security, data and resource management, portability, and information discovery. Globus provides an array of services which application and tool developers can customize to their specific needs without using the Globus communication libraries in [8].

Now there are several Globus technology. Such as, Globus toolkits, Open Grid Services Architecture (OGSA), Globus Resource Allocation Manager (GRAM), Monitoring and Discovery Service (MDS), GridFTP. We have choosen the Globus toolkits because Globus Toolkit mechanisms are being used in dozens of major international Grid projects employed at hundreds of sites.

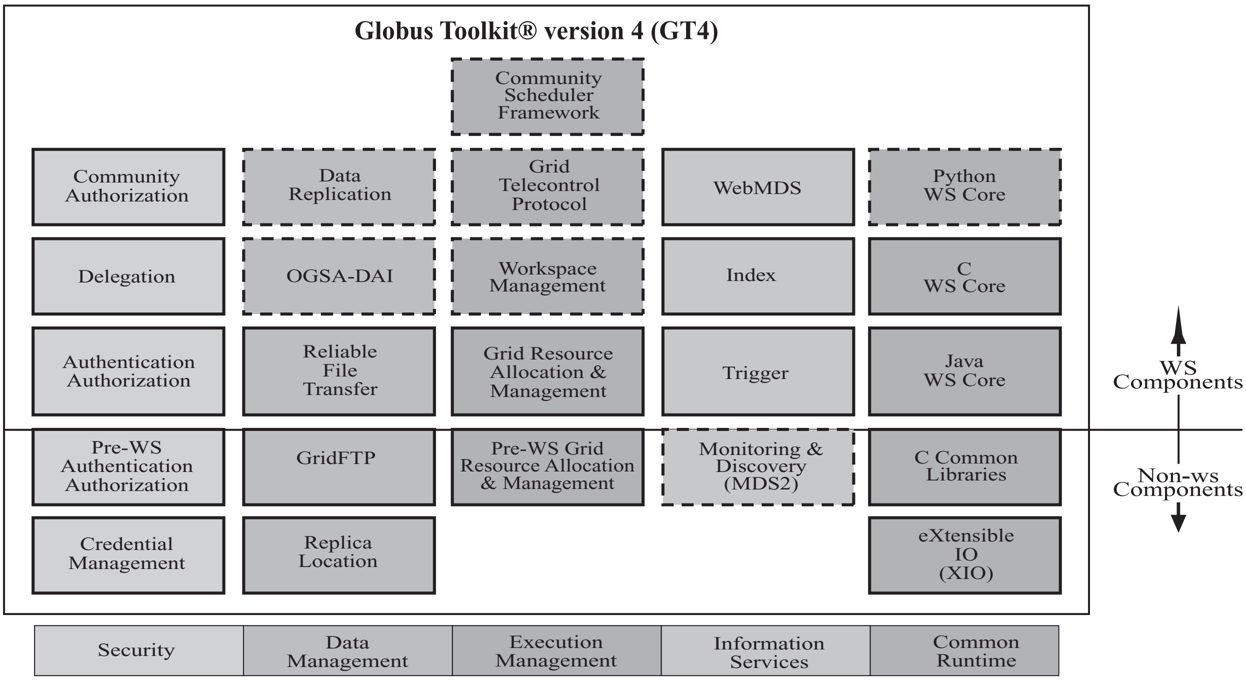


Fig. 1 Layered architecture of Globus middleware.

Core GT Component: public interfaces frozen between incromental releases; best effort support
Contribution/Tech Preview: public interfaces may change between incremental releases
Deprecated Component: not supported: will be dropped in a future release

The Globus Toolkit v 4.0.3 provides A set of basic facilities needed for grid computing, better Security, Resource Management, Data Management, Information Services, Application Programming Interfaces (APIs), C bindings (header files) needed to build and compile programs. It also provides a rapid development kit known as Commodity Grid CoG), which supports technologies such as Java, Python, Web services, CORBA, and so on[9].

A portal is a doorway to another reality or realities. The word portal carries similar connotations in the context of information technology. By portal, we mean any experience with a browser that has a point of origin or gateway that the user must pass through in order to get to his or her destination, which would be any kind of information the user is seeking. The portal is the single entry point to a world of information geared toward a specific audience.

There are several Existing Major Portal Tools and Technologies such as, GridSphere, Commodity Grid (CoG) Kits, GridPort, Liferay, Stringbeans, uPortal, Open Grid Computing Environment (OGCE), Pluto, IBM's WebSphere. We have worked with the GridSphere portal framework which provides an open-source portlet based Web portal.

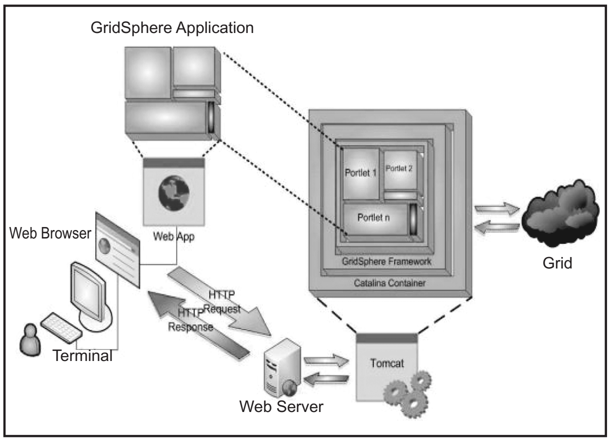


Fig. 2 GridSphere portal framework's architecture.

GridSphere enables developers to quickly develop and package third-party portlet web applications that can be run and administered within the GridSphere portlet container.

GridSphere's Portal Framework has a series of useful features, including fully JSR 168-compliant portlet API implementation, a higher-level model that enables construction of sophisticated portlets using the GridSphere User Interface tag library and visual beans, support for simple development and integration of new portlet applications, support for RBAC (Role Based Access Control), a comprehensive portlet service model to encapsulate reusable portlet logic to be shared between several portlets, flexible portable presentation based on XML that can be quickly changed to design custom portal layouts; database support and persistence of data using Hibernate JDO/OQL[9].

A. Setting up the cluster

- **Cluster architecture :** Initially, we have set up our grid so that it will appear like a grid client (NodeA) connecting to a grid that appears as a "cluster" of Torque (PBS) job managed machines represented by NodeB and a "cluster" of Sun Grid Engine (SGE) job managed machines represented by NodeC. The Globus toolkit will be configured to use either of these "clusters" to offload process intensive jobs to, to expedite completion of a task and show a direct benefit of a compute grid.
- **Setting up Red Hat Linux on 3 machines :**

Packages to select:
Xwindows
Gnome
Applications:
Editors
Graphical Internet
Text based internet
Server:
Server configuration tools
Web server
Windows File Server
Postgres SQL Database
Development:
Development tools
Java Development
System:
Admin tools
System tools
Printing Support
nodeA
IP: 172.16.31.39
Netmask: 255.255.240.0
Hostname: nodeA.iut
Gateway: 192.168.31.1
Primary DNS: 192.168.31.10
nodeB
IP: 172.16.31.40
Netmask: 255.255.240.0
Hostname: nodeB.iut
Gateway: 192.168.31.1
Primary DNS: 192.168.31.10
nodeC
IP: 172.16.31.41
Netmask: 255.255.240.0
Hostname: nodeC.iut
Gateway: 192.168.31.1
Primary DNS: 192.168.31.10

- **Configure Linux for Globus :**
  - Edit /etc/hosts

Sample Host file:

```
Do not remove the following line, or various programs
# that require network functionality will fail.
172.16.16.60 nodeA.iut nodeA
localhost.localdomain localhost
::1 localhost6.localdomain6 localhost6
```

- Edit /etc/profile

Sample profile file:

```
# /etc/profile
# System wide environment and startup programs, for login setup
# Functions and aliases go in /etc/bashrc
pathmunge () {
if !echo $PATH | /bin/egrep -q "(^|:)$1($|:)" ; then
if [ "$2" = "after" ] ; then
PATH=$PATH:$1
else
PATH=$1:$PATH
fi
fi }
# ksh workaround
if [ -z "$EUID" -a -x /usr/bin/id ]; then
EUID=`id -u`
UID=`id -ru`
```

```
fi
# Path manipulation
if [ "$EUID" = "0" ]; then
pathmunge /sbin
pathmunge /usr/sbin
pathmunge /usr/local/sbin
fi
# No core files by default
ulimit -S -c 0 > /dev/null 2>&1
if [ -x /usr/bin/id ]; then
USER="`id -un`"
LOGNAME=$USER
MAIL="/var/spool/mail/$USER"
fi
HOSTNAME=`/bin/hostname`
HISTSIZE=1000
if [ -z "$INPUTRC" -a ! -f "$HOME/.inputrc" ];
then
INPUTRC=/etc/inputrc
fi
JAVA_HOME=/usr/java/jdk1.5.0
ANT_HOME=/usr/local/ant-1.6.4
CATALINA_HOME=/usr/local/tomcat-5.0.27
JUNIT_HOME=/usr/local/junit3.8.1
PBS_HOME=/var/spool/PBS
GLOBUS_LOCATION=/usr/local/globus-4.0.3
GRID_SECURITY_DIR=/etc/grid-security
GRIDMAP=/etc/grid -security/grid -mapfile
GLOBUS_HOSTNAME=`/bin/hostname`
SGE_ROOT=/usr/local/sge
PATH=$JAVA_HOME/bin:/bin/tar:$ANT_HOME/
bin:$PATH:$CATALINA_HOME/bin:$GLOBUS_
L
OCATION/bin:$GLOBUS_LOCATION/sbin:/usr/
local/bin
CLASSPATH=$JAVA_HOME/lib/tools.jar:$CLAS
SPATH:$JUNIT_HOME/junit.jar
LD_LIBRARY_PATH=$GLOBUS_LOCATION/li
b:$LDLIBRARY_PATH
export PATH USER LOGNAME MAIL
HOSTNAME HISTSIZE INPUTRC
JAVA_HOME CLASSPATH
ODBCINI CATALINA_HOME JUNIT_HOME
PBS_HOME GLOBUS_LOCATION
GRID_SECURITY_DIR
GRIDMAP GLOBUS_HOSTNAME
ANT_HOME LD_LIBRARY_PATH SGE_ROOT
#export PATH USER LOGNAME MAIL
HOSTNAME HISTSIZE INPUTRC
for i in /etc/profile.d/*.sh ; do
if [ -r "$i" ]; then
. $i
fi
done
unset i
unsetpathmunge
```

B. Setting up Environment

➢ Installing Globus toolkits :

1) Download the software bundle GTAI1.0.1.tar.gz and copy the bundle to thelocation usr/local
2) Untar the bundle using the following steps
   a. As 'root' user, change to the directory /usr/local

   cd /usr/local

   b. Untar the GTAI1.0.1.tar.gz

   tar-zxvfGTAI1.0.1.tar.gz
3) Edit auto.cfg and set the hosts IP address.
4) Run the command source /etc/profile to set up the required environment variables.
5) Download & Install two rpm packages
   a. perl-XML-Parser-2.34-6.1.2.2.1.i386.rpm
   b. xinetd-2.3.14-8.i386.rpm
6) Run the script GTAI_install.sh

   /bin/sh GTAI_install.sh

➢ Installing/Configuring Java CoG-kit

We have not faced any kind of problems since we've followed the instruction from the official Java Cog-kit website carefully. Installing and configuring java CoG-kit requires some simple and basic steps which can be done very easily.

➢ Installing/Configuring Grid sphere

For this development work we have used the latest Gridsphere having version number 3.1. The installation of GridSphere exhibits some major problems. The guidelines documented on the GridSphere website were carefully followed but one problem was configuring the database. We have solved it by changing some configuration files of post GRESQL database. This documentation can be obtained from the official postgresql database website. Another problem was that GridSphere did not work when a new project was created. As the Gridsphere documentation says, the new project should be visible in the layout tab of the Gridsphere portal framework, but there was none. We have solved it by simply running a commandtouch $HOME/.gridsphere/ portlets/"project_name" 2

➢ Installing/Configuring Eclipes Europa IDE

Eclipse Galileo was chosen as IDE for the development. Sysdeo Tomcat plugin was also used for the Eclipse IDE. Integrating Gridsphere framework with Eclipe enables development and changes to the Gridsphere framework itself in order to be debugged and tested in Eclipse.

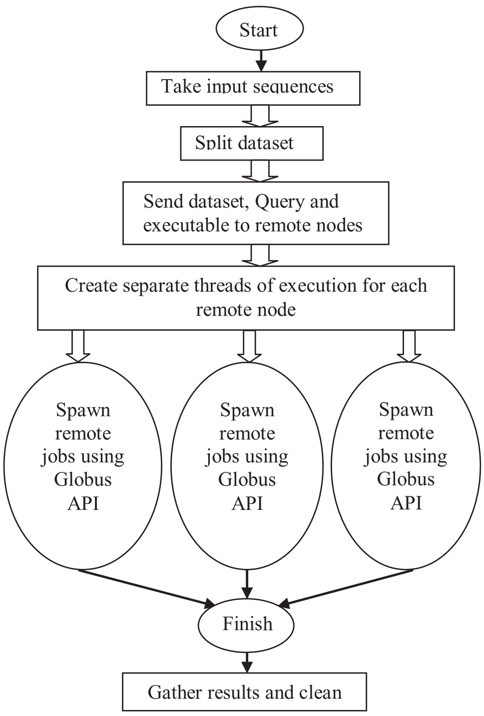


Fig. 3 The flow chart for the head.

C. Developed a Gridsphere portal framework

We have developed a Gridsphere portal framework in where we can take input for bioinformatics sequence such as, local alignment sequences. First, our dataset is divided according to the active node(computer) and then dataset and query have to be sent to that active node. After processing the node the result is gathered and check the approximate result by our node ie. root computer.

The flow chart for the node is given below: It first opens the GASS[11] server and then copy dataset query and executables then runs the executable. After completion the nodes send the result back to head. Then it deletes the remote files.

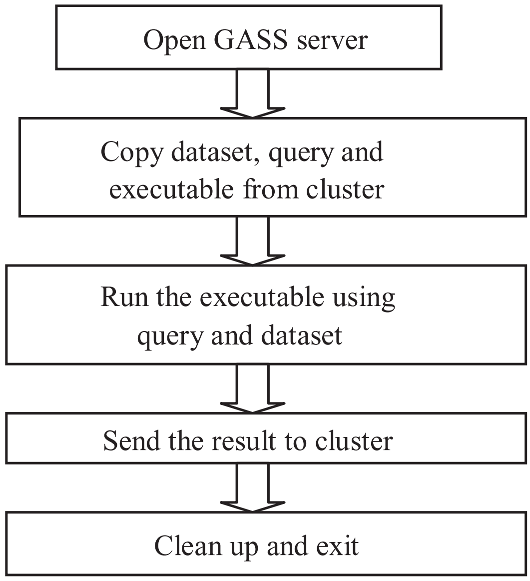


Fig. 4 The flow chart for the node.

4. EXPERIMENTAL RESULTS

We experiment various sequence with different length. We create an input window in which we can put two or more input sequences. We use the Data of type Bacteria DNA sequence that was downloaded from NCBI[12] ftp server. The size of the original dataset is 237Mb. Then applying tokenizing on the dataset we reduce the data size to 91Mb. Several experiments were done on this data From the table 1 we are seeing that for different number of processors there are time variation and the more the number of processor the less the time required.

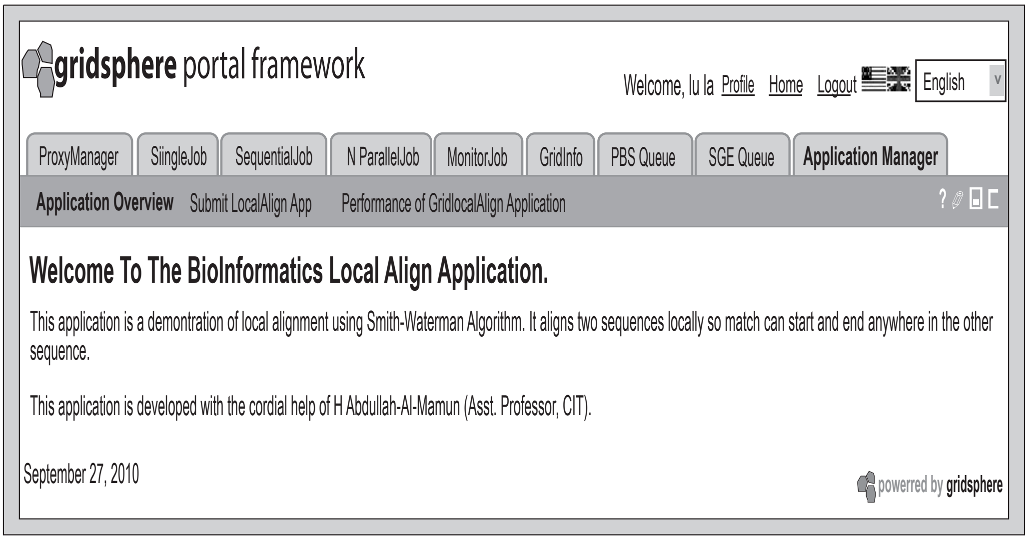


Fig.5 Developed gridsphere portal framework for bioinformatics sequence alignment.

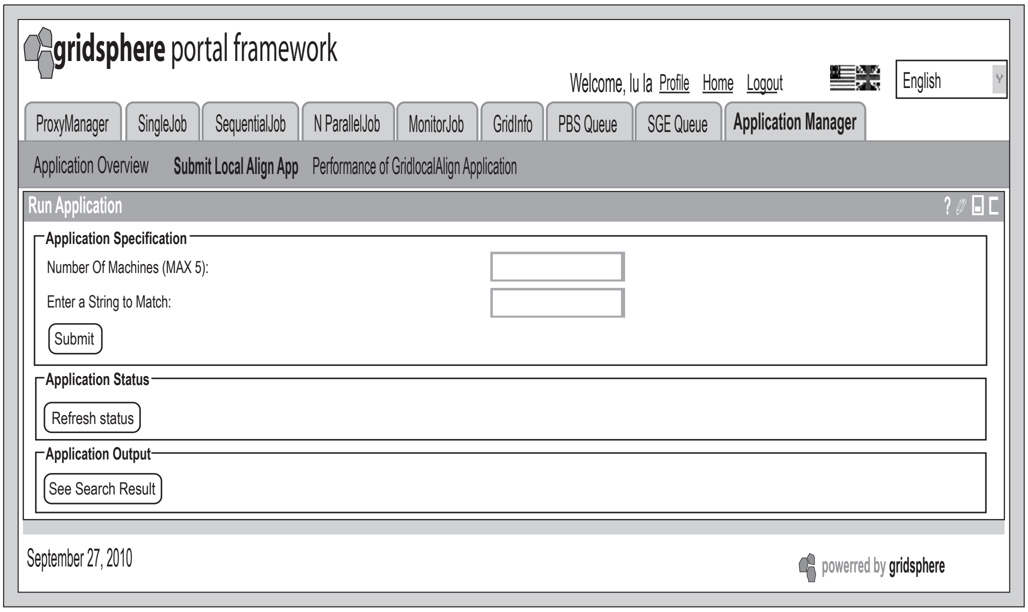


Fig. 6 Developed gridsphere portal framework for bioinformatics sequence alignment input window.

TABLE 1 Analysis according to Time Complexity.

Number of processor	Time (min.)
2	131.00
4	67.99
6	42.05
8	32.41
10	26.55

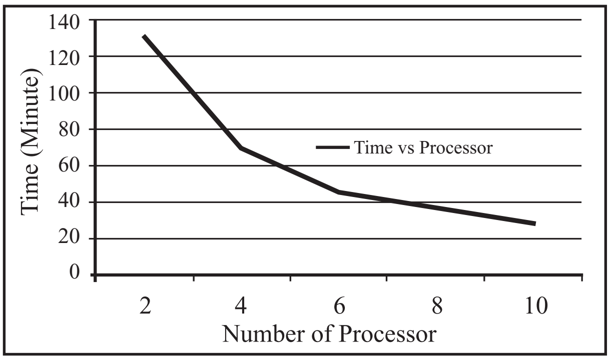


Fig. 7 Time vs. Number of processors graph for Table 1.

From the data the graph representation is given above, in which we are seeing that for more the number of processor the less the time is needed.
From the table and graph we are seeing that it is efficient for large scale bioinformatics sequence alignment as like other grid clustering [13, 14].

5. CONCLUSION

Globus technology is used to share computer resources securely and efficiently on the Grid when managed by various Local Resource Management systems across both a single administrative domain and across many organizations. While other existing portals, UCGrid Portal, D-Grid Portal etc. provide a good spread and convergence of solutions, however most of them do not directly interact with the globus. Globus provides a simple, yet powerful middleware infrastructure for the submission of jobs. So this tool can be leveraged for more flexible use of Grid. The aim of this paper is to develop a Grid portal based on the Globus Toolkit, through which remote job submission, monitoring and scheduling will be possible. The goal of this portal is to provide tight integration between Gridsphere and Globus.

REFERENCES

1. Foster I, Kesselman C (eds.). *The Grid: Blueprint for a Future Computing Infrastructure. Morgan Kaufmann:* San Francisco, CA, 1999.
2. Ying Sun, Shuqi Zhao, Huashan Yu, Ge Gao and Jingchu Luo, "*ABCGrid: Application for Bioinformatics Computing Grid,*" Accepted in Oxford University Journal on February 28, 2007.
3. Yong Meng Teo, Xianbing Wang and Yew-Kwong Ng, "*GLAD: a system for developing and deploying large-scale bioinformatics grid,*" Accepted in Oxford University Journal on September 10, 2004.
4. Last day of access (November 20, 2011) [Online]. Available: http://www.ucgrid.org/
5. Last day of access (November 20, 2011) [Online].Available: http://www.portal.geongrid.org/
6. Last day of access (November 20, 2011) [Online]. Available: http://www.portal.chronos.org/
7. Last day of access (November 20, 2011) [Online]. Available: http://www.portal.p-grade.hu/
8. Foster I, Kesselman C. Globus, *"A metacomputing infrastructure toolkit,"* International Journal of Supercomputer Applications 1997; 11(2):115-128.
9. Last day of access (November 20, 2011) [Online]. Available: http://www. globus.org
10. Last day of access (November 20, 2011) [Online]. Available: http://www.gridsphere.org
11. Last day of access (November 20, 2011) [Online]. Available: http://www.globus.org/gass
12. Last day of access (November 20, 2011) [Online]. Available: http://www.ncbi.nlm.nih.gov/
13. Clifford R, Mackey AJ. Disperset, "*A simple and efficient approach to parallel database searching,*" Bioinformatics 2000; 16:564-565.
14. Last day of access (November 20, 2011) [Online]. Available: http:// www.interscience.wiley.com

SOCIAL NETWORK ASSISTED PERSONALIZATION WITH USER CONTEXT FOR RECOMMENDER SYSTEMS

Kazi Masudul Alam [1], Aysha Akhtar [1] and Md. Rashadul Hasan Rakib[2]
[1]Computer Science & Engineering Discipline, Khulna University, Bangladesh
[2]Dept. of Computer Science & Engineering, Mawlana Bhashani Science & Technology University, Bangladesh

Abstract: In recommender systems, social networks are considered as a trusted source for user interests. In addition, user context can enhance users' decision making. In this paper, we design a new architecture for user personalization which combines both social network data and context data. Our system aggregates a user's preference data from various social networking services and then builds a centralized user profile which is accessible through public Web services. We also collect user's contextual information and store it in a central space which is also accessible through public Web services. Based on Service Oriented Architecture, recommender systems can flexibly utilize users' preference information and context to provide more desirable recommendations. We present how our system can integrate both types of data together and how they can be mapped in a meaningful way.

Keywords: *Centralized personalization, Management, Recommender system, User context, User profile.*

1. INTRODUCTION

Recommender systems are used in various E-commerce sites to provide their customers with more customization options [8]. Recommender systems suggest products that a customer would like the most based on the overall sales of a product, demographics of the customer, analysis of past buying behavior, and/or collaborative rating history of a customer. Most of the systems rely on user profile to discover proper products relevant to user needs [4]. In recent years, users tend to use various social media or network services. Consequently, a user should maintain several profiles which contain more or less redundant information, update them to reflect his/her interest changes, which is cumbersome work. This tedious labor can be reduced by managing a central profile for every user.

Although a number of researchers have studied various personalization methods for recommender systems, the above-mentioned issue is not addressed well [3]. Walter et al. [5] suggested a combination of distributed recommendation systems with trust and reputation mechanisms. Seth [9] shows that the use of sociological theory helps to simplify the system design, to improve scalability, and to understand the behavior of the user because people generate many data of interest in their network. Hence, a collection of users' interest and activity data from various social networking services can be a good source for updating their profile information. In addition to a user profile, current and historical context of a user can also influence his/her preference [1, 2, 10]. For example, a person would prefer a particular place for the summer vacation, but he/she may not prefer the same place for the winter vacation. In this case, season context can affect the user's decision making. Context of a user can be of various types such as current time, current location, current weather, history context, user event/schedule list, physical, cognitive status, etc. Since social networking is getting more popular, it is worthwhile combining social network data with context in a way that can be highly beneficial to personalization. In this paper, we design a new recommender system architecture where personalization information is stored separately from the recommendation engine. In this architecture, one's social network activity data are collected using Service Oriented Architecture (SOA) and aggregated in a separate central server. On the other hand, our system also considers users' current and historical context that can be utilized by recommender systems. Context data is collected through user's Smartphone and aggregated in another separate central server. For every user, the system maintains only one centralized activity, preference profile as well as one centralized context data profile. Our compilation procedure for generating user profiles is intuitive and does help users to maintain their profiles with less effort. We present how such system architecture can integrate both types of activity and context data together and how they can be mapped in a useful way. Additionally, since the profiles and context data can be easily accessible based on SOA technique, any existing and new E-commerce website can flexibly adapt one's profile and context data to provide recommendations. The remainder sections are organized as follows: Section 2 presents previous work closely related to our system. Section 3 describes our system architecture while Section 4 presents a prototype implementation and experimental results. Finally Section 5 concludes the paper with possible future work.

2. RELATED WORK

In this section, we briefly present previous work related to our study. Walter et al. [5] presents a trust-based recommendation system that adopts online social networks to use trust relationships between users. They investigated how the dynamic trust affected the performance of the system. However, it only dealt with one social network data. As users are getting to maintain various online social networking, it is worthwhile considering data from various social networks. Chung et al. [4] described an integrated personal recommender system which assists individuals make evaluations about entities in meaningful ways. But, the model and scenario management part in their architecture put extra overhead on the user. Adomavicius et al. [1] demonstrated that the contextual information increased the quality of recommendations in certain settings. They considered contextual information, such as time, companion, and weather, during the recommendation process. As a result, they showed that in terms of recommendation accuracy and user's satisfaction, context-aware approach significantly outperformed non-contextual approach. In [2], Adomavicius et al. discussed the general notion of context and how it can be modeled in recommender systems. Setten et al. [10] proposed an application for Context-aware Mobile Personal Assistant that serves a tourist with information and services depending on a specific context. They integrated a recommender system with a context-aware application platform to provide context-aware recommendations.

Though there are several works related to social network-based recommender system as well as context-aware recommender system individually, there are few works which combine both of them and suggest a generic architecture to map between them. The work in [6] is closely related to our proposed system in which both social network and user context were considered. Nevertheless, their social affinity calculation only considered the information available in a low end mobile device such as phone call records, SMS records, and phone address book data. Hence, such system is not applicable to a highly scalable scenario where a user consumes various online social networking services. Additionally, users' context is only limited to location and time. On the contrary, in our approach, social network data can be gathered from various social networking/media services by using Web service technology. Moreover, we take into consideration diverse types of context for users, such as location, time, weather, physical/cognitive status, and so on.

3. PROPOSED SYSTEM

In this section, we present an overall architecture of the proposed system in brief. Our system considers both context awareness and social networks for centralized personalization. Our goal is to facilitate the practice of profile management as well as to integrate a user's current context with his / her history context for recommender systems to make more acceptable recommendations.

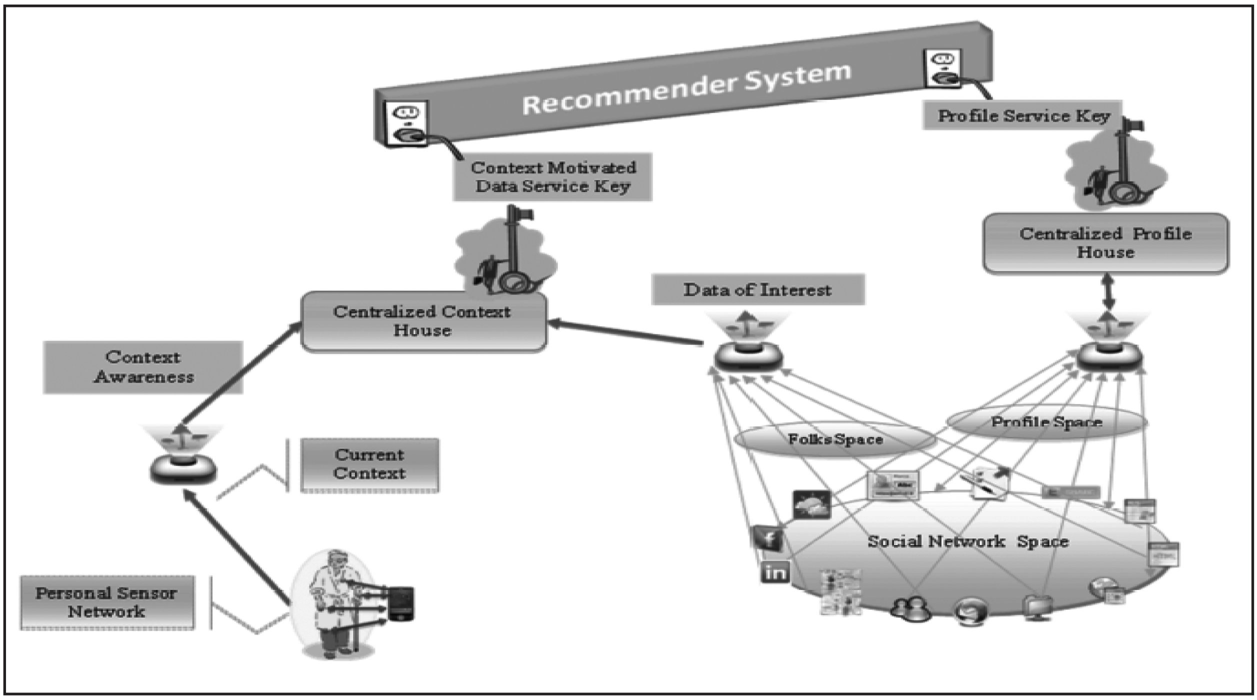


Fig. 1 High level architecture of the proposed approach.

Fig.1 shows the high level architecture of the system. The system is divided into two separate web service keys: a Central Profile Service Key and a Context Motivated Data Service Key. In our system we design two central storages, called a centralized profile house (CPH) and a centralized context house (CCH). We aggregate information which a user produces on heterogeneous services into CPH whereas collecting available context data in CCH. The user only has the authority to share his/her public web services. Recommender systems can collect information related to a user through these houses and thus generate recommendations for the user dynamically.

3.1 Centralized Profile House

Every user is maintaining his/her online social network life. We call it a social network space of users. In this social network space, users consume heterogeneous services from various providers, such as Facebook[1], Twitter[2], Linkedin[3], and Delicious[4]. While using these social networking tools, users generate a huge amount of data such as their status, tweets, comments, tags, and writings, through daily additions. These data represent their personal preferences. Since these activities are generally performed in different services, it is worthwhile collecting those data and thus compiling a central information center, CPH. All this information will be extracted by using users' consent i.e. authentication through the available Application Programming Interface (API) of the respective social networking site. After gathering recent information from various social network services, we employ standard clustering algorithms so that related information is grouped together.

To support interoperability for heterogeneous services, we represent this information using a standard format. In our system, we adopt Attention Profile Markup Language (APML[5]), which is an open standard that encapsulates a summary of user interests in a simple, portable way, to represent users' profiles. A current profile contains, up-to-dated key texts represented by the concept block (**Fig. 2**). Under the concept field, there is type property which denotes the type of the text such as twitter status and Facebook status. e.g., {enjoying cricket match Australia vs England}, {planning a trip to

[1] http://www.facebook.com/
[2] http://twitter.com/
[3] www.linkedin.com
[4] http://www.delicious.com/

[5] http://apml.areyoupayingattention.com

```
<?xml version="1.0" ?>
- <APML xmlns="http://www.apml.org/apml-0.6" version="0.6">
  - <Head>
      <Title>Example APML file for Social network based central profile data</Title>
      <Generator>Written by Hand</Generator>
      <UserEmail>test@uottawa.ca</UserEmail>
      <DateCreated>2011-04-07T20:55:00Z</DateCreated>
    </Head>
  - <Body defaultprofile="current">
    - <Profile name="current">
      - <ImplicitData>
        - <Concepts>
            <Concept id="1" value="love to watch cricket" type="twitter status" tag="games,friend" public="true" datetime="2011-03-11T01:55:00Z" />
            <Concept id="2" value="need to take medicine every day" type="facebook status" tag="medicine, health, heart" public="true" datetime="2011-03-12T01:55:00Z" />
            <Concept id="3" value="want to go in a trip to France" type="facebook status" tag="trip,family,valentine" public="true" datetime="2011-03-13T01:55:00Z" />
            <Concept id="4" value="planning to buy a car for my office" type="linkedin status" tag="car,office" public="true" datetime="2011-03-14T01:55:00Z" />
            <Concept id="5" value="bookmarked romantic movie" type="delicious tag" tag="movie,romantic,valentine" public="true" datetime="2011-03-15T01:55:00Z" />
            <Concept id="6" value="like action movie" type="facebook like" tag="move,action,friend" public="true" datetime="2011-03-12T01:55:00Z" />
            <Concept id="7" value="listening justin bieber" type="facebook status" tag="music,work" public="true" datetime="2011-03-13T01:55:00Z" />
            <Concept id="8" value="looking for baby boy dress" type="twitter status" tag="dress,family," public="true" datetime="2011-03-14T01:55:00Z" />
          </Concepts>
        </ImplicitData>
      </Profile>
    + <Profile name="history">
    </Body>
</APML>
```


Fig. 2 Sample Attention Profile Markup Language file.

France with my family}, {listening justin bieber}. By default, current profile data is public, but user can control privacy of his/her data by setting the profile through Graphical User Interface (GUI). Here public profile data signifies the easy authenticated access through web services. We also keep a history profile which contains old dated data. We update, delete, and reorder the current and history profile in CPH based on the datetime property field. The tag field of the concept block represents the category or tag cloud of the text which is used for mapping with the context data. Tag data set can be of type {medicine, health, heart}, {movie, romantic, valentine}, {trip, family, valentine day}, etc. Finally, the user profile for each user is accessible by utilizing the public Web service URL to the recommender systems.

3.2 Centralized Context House

Context plays an important role for decision making in E-commerce. In our system, we consider both situational and physical context to enable our system to recommend more desirable products to a potential buyer. In our system, a user's Smartphone plays the role of his/her context collection client. Since mobile phones are considered to be the closest personal device and since today's mobile phones are equipped with more processing power, memory, and various sensors, we believe that such assumption is legitimate. In our system, we regard context as location, date, time, weather, events from schedule list, and physical states. Physical states can involve a user's heart beat rate, body temperature, cognitive state, etc. context related to physical states is collected by the devices connected with the user and the value is transferred to the Smartphone in real time. These contexts knowledge are later combined with folks space knowledge and stored in a central information center CCH.

3.2.1 Data of Interest Categorization

In a user's personal social networks, he/she may have one or group of trust people depending on a specific context. For example, for a particular physical state, relevant data of a user would be his/her doctor's recommendations. As total social network friends and folks space of a user is very large and contains a large amount of data, it is important to categorize updated context data with Data of Interest (DOI) tags. This can reduce the candidate information space.

In the social network space, users can participate in heterogeneous services and thus generate various textual contents. In this space, users have already established relationships with their friends who may share common interests, passions, and/or attitudes, e.g., family, personnel, business friends, work colleagues, health support personnel, etc (**Fig. 3**). The friends also have

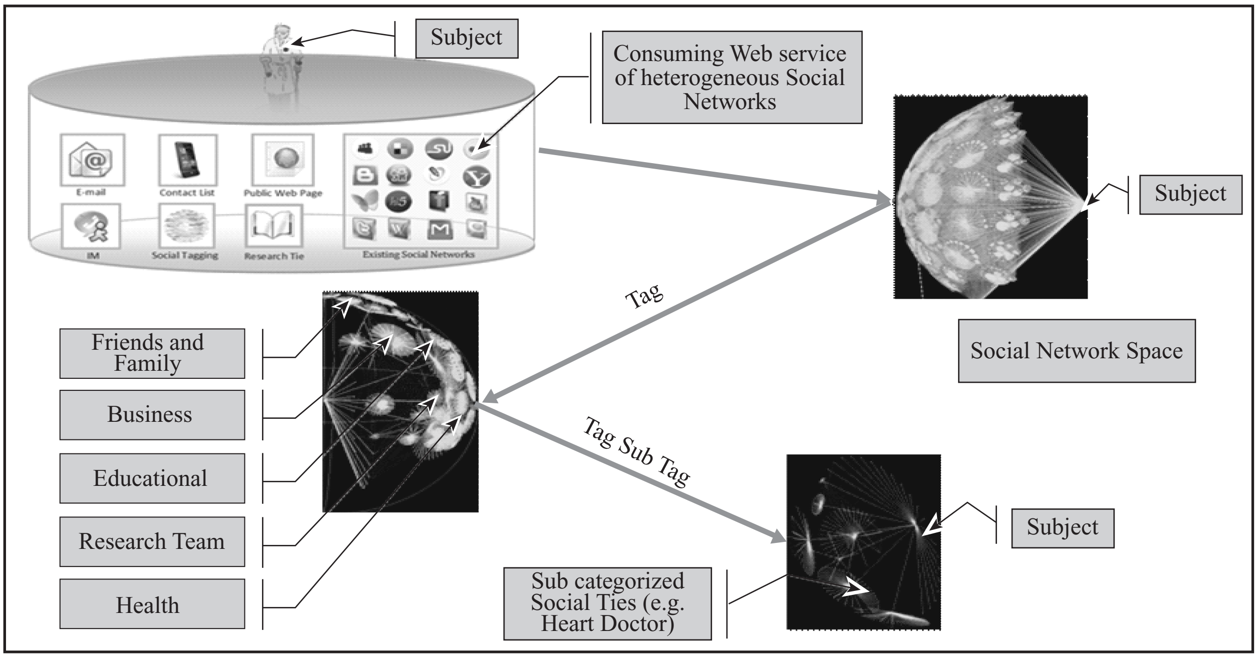


Fig. 3 Data of Interest (DOI) categorization.

their personal profiles to describe their current status. Hence, while categorizing data in friends and folks spaces, we also consider their relationships and current status in terms of a given user to determine specific type of data. For example, conversation with a doctor in Facebook may generate data related to health support; this can be categorized as *Tag: Health, sub tag: Orthopedics, heart, etc*. Tags for DOI generate virtual data taxonomy tree to find proper data relevant to a certain context.

3.2.2 Combining Context and DOI

To capture the meaningful context for users, we map context to appropriate DOI data. For example, the user's health status such as "above average heart rate" can be mapped with appropriate DOI data "health, heart doctor"; the calendar event such as "birthday of my boss" can be mapped with DOI data "friend, office"; the calendar event such as "valentine day trip with X" can be mapped with DOI data "friend, family, valentine"; context like location, date, time, weather can be mapped with any DOI or can be left without mapping based on the selection of the context settings. Finally this context data is represented in a portable XML format described in the following section and published as accessible information through public Web service URL similar to the user profile. Context and DOI data collections are carried out periodically.

3.2.3 Context Data Representation

We also represent context data using a XML based file format (**Fig. 4**), which is motivated from User Context Markup Language (UCML) [7]. Using UCML, we can manage both current and history context. Various contexts related information is stored in its individual format. Every context has a DOI attribute which can map a certain context to possible sets of profile data. We also represent various types of contexts like location, calendar event, physical status, etc. XML-based context representation ensures easy manipulation of data and portability among heterogeneous systems.

3.3 Similarity Matching Between Profile and Context Data

While a user shares his/her profile and context public web service URL with a certain E-commerce service, the corresponding recommender engine extracts the current APML and UCML-based XML file. We measure similarities between concept tags in APML and context DOI in UCML to identify proper context associated to profile data. We can employ a suitable similarity matching algorithm, such as cosine-based similarity measurement. After the similarity matching, if total profile size is

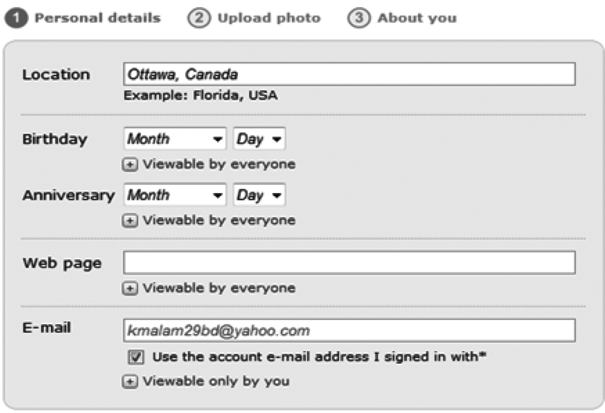
```
- <UCML>
  - <Head>
      <Title>Example UCML file for user context</Title>
      <Generator>Written by Hand</Generator>
      <UserEmail>test@uottawa.ca</UserEmail>
      <DateCreated>2011-02-12T20:55:00Z</DateCreated>
    </Head>
  - <Body defaultcontext="current">
    - <contexts name="current">
      - <context id="1" type="geographic" retrieved="2011-02-11" DIO="">
        + <geographic>
        </context>
      - <context id="2" type="calendar/event" retrieved="2011-02-10" DOI="friend, office">
        - <event>
            <name>need to buy gift for my boss</name>
            <eventdate>12-02-2011</eventdate>
          </event>
        </context>
      - <context id="3" type="physical" retrieved="2011-02-11" DOI="health, heart doctor">
        - <physical>
            <name>above average heart rate</name>
          </physical>
        </context>
      - <context id="4" type="calendar/event" retrieved="2011-02-12" DOI="friend, valentine">
        - <event>
            <name>valentine day trip with X</name>
            <eventdate>14-02-2011</eventdate>
          </event>
        </context>
      </contexts>
    + <contexts name="history">
    </Body>
  </UCML>
```

Fig. 4 Sample User Context Markup Language file.

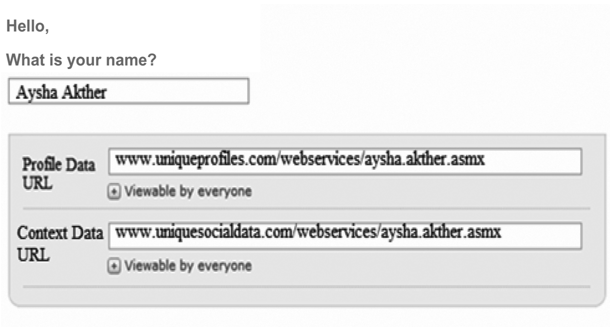
40% of previous size, then we consider it as a fully context motivated dense profile data. Otherwise, we consider both profile and context data as the profile base. This threshold is determined after numerous simulation results. The recommender system could utilize these up-to-date personalization data for the user, in turn making compelling recommendations. These data sets contain multiple domain items. Accordingly, it would be possible for a recommendation engine to provide diverse recommendations.

4. PROTOTYPE IMPLEMENTATION AND SIMULATION

To test our system, we used Facebook and Twitter as social networking sources. And, our context data source was generated by using a J2ME-based event list application running in a Netbeans emulator. The J2ME application periodically created some events and uploaded them to the CCH. On the other hand, Facebook and Twitter client applications extracted status and tweet data, and then stored in the CPH. We invited 30 Facebook and 20 Twitter users, and collected their data, such as status, wall, and profile. The 30 Face book users were X female, Y male aged from xxx – yyy. The Twitter users were X female, Y male aged from xxx – yyy. From the collected data, we constructed a centralized profile base where information is clustered according to their genre. For example, our test subject, 50 persons like different movies, different songs, different athletes, different vacation places etc.



(a)



(b)

Fig. 5 a) Current ecommerce website requires various user information and b) Our proposed system architecture only requires user's public web service URL.

In order to compile a list of movies, songs, athletes, vacation places etc. we used the above described aggregation and clustering method. This list helps the system to easily converge a new status or tweet to find a genre while compiling user wise central profile base. We adopted 3-gram collocations to extract new status/tweet using keyphrase extraction tool and K-Means clustering algorithm is used to cluster an N-gram to most suitable pre-calculated clusters. For simulation purpose, CCH and CPH periodically collect information from context source and social network source, respectively. We predefined a list of DOI tags which were later mapped with appropriate context data. Analogously, in the CPH phase, tag lists were generated based on the status data. We created two local web services, one is to access CPH information and the other one is to access CCH information. We plugged (**Fig. 5**) those two web service URLs to a recommendation engine to provide central profile and central context information of a user to the engine. For a recommendation method, we simply used word-to-word matching between the centralized user profile and attributes of products which were previously stored.

We first measured our system performance by storing time log of every operation stages. **Fig. 6** shows time required to collect information from CPH and CCH through iformation from CPH and CCH through separate Web services. The result showed that data collection time from CPH and CCH is almost same. However, the time required mapping CPH and CCH tags is much lower.

We continued to examine data length collected through Web services. **Fig. 7** shows the result. We reported values that were 10 times simulation results. From this graph, we observed that, after mapping CPH and CCH tags, some simulations provide dense context motivated profile data and some fails to do so. In case of failure to provide dense data we combine both profile and context data which also can be a good profile data base which in turn better reflects user current interests. Since the CPH data collection and the CCH data collection can be conducted by using separated web services in parallel, the total time required to update interest data is *Sum*(*Max(CPH Data Collection time, CCH Data Collection time), time for CPH and CCH data tag mapping)*. For any recommendation engine, this would be the total delay time to get the updated profile data. As in our method, CPH and CCH are separate central update procedures, their update do not cause delay to the recommendation process.

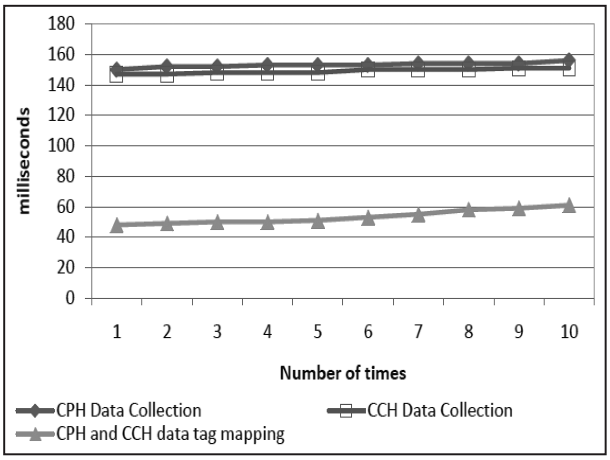


Fig. 6 Time required to collect CPH and CCH data using Web service.

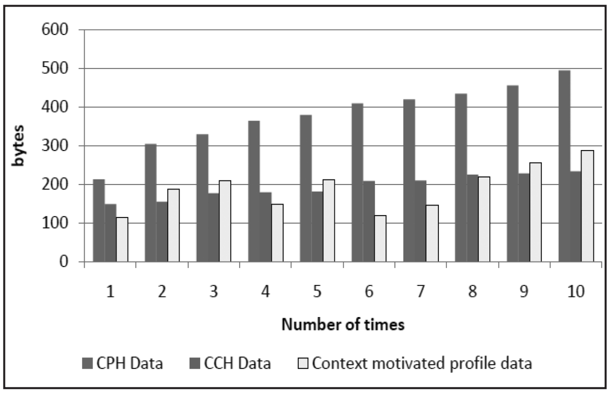


Fig. 7 Various data lengths obtained at recommendation stage.

5. CONCLUSION

In this paper, we design new system architecture for building a personal user model that can be applied to recommender systems. Our system considers not only social network data but also context data. After aggregating possible activity data of a user in social network services, we generate a central profile for the user. In addition, we collect various user contexts and maintain them in a central context house. The major advantage of our system is that it can be easily incorporated with existing or new recommender systems through the Web service architecture. By utilizing aggregated user profile and rich context information, recommender systems could provide more desirable items for their customers.

In future work, we plan to conduct more detailed performance evaluations. In addition, we intend to simulate diverse recommendation techniques to examine how the proposed profiles and context are beneficial to them.

REFERENCES

1. Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A.,*"Incorporating contextual information in recommender systems using a multidimensional approach"*, ACM Transactions on Information Systems (TOIS), 23(1), PP. 103–145, 2005.
2. Adomavicius, G. and Tuzhilin, A., *"Context-Aware Recommender Systems"*, Recommender Systems Handbook, PP. 217-253, 2011.
3. Bedekar, M. V., Deshpande, B., and Joshi, R., *"Web Search Personalization by User Profiling"*, In proceedings of the 1st International Conference on Emerging Trends in Engineering and Technology, PP. 1099 – 1103, 2008.
4. Chung, R., Sundaram, D., and Srinivasan, A., *"Integrated personal recommender systems"*, In Proceedings of the 9th International Conference on Electronic Commerce (ICEC '07). ACM, New York, NY, USA, PP. 65-74, 2007.
5. Frank E. Walter, F.E., Battiston, S., and Schweitzer, F., *"A Model of a Trust-based Recommendation System on a Social Network*", Autonomous Agents and Multi-Agent Systems. Springer. Volume: 16, Issue: 1, PP. 57-74, 2007.
6. Garg, D., Visweswariah, K., and Kalyanaraman, S., *"CAESAR: A Context-Aware, Social Recommender System for Low-End Mobile Devices"*, In proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware (MDM '09). IEEE Computer Society, Washington, DC, USA, PP. 338-34, 2009.
7. Meneses, F., and Moreira, A., *"A flexible location-context representation"*, In proceedings of the 15th IEEE international symposium on personal, indoor and mobile radio communications (PIMRC '04). Volume: 2, PP. 1065–1069, 2004.
8. Schafer, J. B., Konstan, J., and Riedl, J., *"Recommender Systems in E-Commerce"*, In Proceedings of the ACM Conference on Electronic Commerce. PP. 158-166, 1999.
9. Seth, A, *"Design of a Social Network Based Recommender System for Participatory Media Content"*, Ph.D. Thesis manuscript, University of Waterloo, 2008.
10. Setten, M.V., Pokraev, S., and Koolwaaij, J., *"Context-Aware Recommendations in the Mobile Tourist Application COMPASS"*, Springer-Verlag, PP. 235-244, 2004. (www.interscience.wiley.com)

LOSSLESS IMAGE COMPRESSION TECHNIQUE BASED ON SNAKE ORDERING

Kaushik Deb[1], Tauhidul Alam[1], Md. Mahammud Ali[1], and Kang-Hyun Jo[2]
[1]Dept. of Computer Science & Engineering, Chittagong University of Engineering & Technology, Bangladesh
[2]Dept. of Electrical Engineering & Information Systems, University of Ulsan, South Korea

Abstract: Use of image has been increasing in many applications. An effective image compression reduces significant amount of storage space and saves time during transmission. In this paper, a lossless image compression technique is proposed to achieve more compression ratio and less computational cost using snake ordering technique. In the proposed technique, a gray image is divided into the 8*8 block and then snake ordering method is applied on this block which provides the sequence of pixel values. The snake ordering method gives the better compression ratio compared to existing zig-zag method that decreases the transmission time indeed. Entropy coding is employed on this sequence to look for the highest different pixel value (DP) which is concerned with bits per pixel, peak signal to noise ratio and mean square error. The experimental results illustrate that the proposed compression framework provides 20% more compression ratio and less computational cost than other methods [10, 11].

Keywords: *Image compression, Snake ordering technique, and Entropy coding.*

1. INTRODUCTION

Even though the memory capacities of computers have increased as new technologies are emerging, the requirement for more storage space is also increasing as more data are needed to be stored. In the case of image data, the spatial and color resolutions are increased for the betterment of image quality, thus it requires more space to store images and more time while sending over network. Image compression minimizes the size of image without degrading the quality of image. There are different techniques available for compressing images. These techniques are classified into two categories called lossless and lossy compression techniques. In lossless compression techniques, no information regarding images is lost which means that the reconstructed image from the compressed image is identical to the original image in every sense. Whereas in lossy compression, some image information is lost i.e. the reconstructed image from the compressed image is similar to the original image but not identical to it. Many works carried out in image compression either lossy or lossless and hybrid one even. In [1], authors proposed an image compression technique using 9/7 wavelet transform based on lifting scheme which was used for both lossy and lossless compression. In another work, researchers used a hybrid (combination of lossy and lossless) compression using daubechies-4 wavelet in combination with the lifting scheme and entropy encoding [2]. In [3], a lossless image compression and decompression technique was applied to reduce the number of bits per pixel and the transmission time. An adaptive data codec for still images that can minimize the energy required for the wireless image compression in a narrow band system was proposed in [4]. In [5], a back propagation based network was used for fractal image compression whereas intrablock variance distribution and vector quantization for that in [8]. A channel splitting and division-free arithmetic encoding technique was proposed in [6] which reduces the precision requirements for the histogram bins. However, near-lossless compression [9] provides pixel difference between original and compressed image above a given value. In [7], authors proposed a method that incorporated progressive transmission and near-lossless compression in a single framework.

The main emphasis of this compression framework is done with the lossless compression technique using snake ordering and entropy encoding. Hence, the proposed framework gives the reduction on compression ratio and peak signal to noise ratio by preserving the vital information and reducing computational cost. The performance of our technique remains better compared to state-of-art technique like context-based adaptive lossless image coding (CALIC)[10] and set partitioning in hierarchical trees (SPIHT) [11].

2. THE PROPOSED FRAMEWORK

The framework of proposed image compression technique is illustrated in Figure 1. In the proposed framework of image compression, the grey images can be compressed. The steps of the proposed framework are as follows. At first step, the source image is taken as input and this image is divided into 8*8 block. As an example, one 8*8 block might be

145	151	159	158	142	129	137	156
142	137	136	139	135	132	142	157
146	132	123	126	129	132	141	153
152	142	137	137	134	129	135	147
145	150	159	161	147	132	135	149
131	145	164	171	155	135	136	149
130	138	153	161	150	130	124	129
140	137	140	147	140	121	107	104

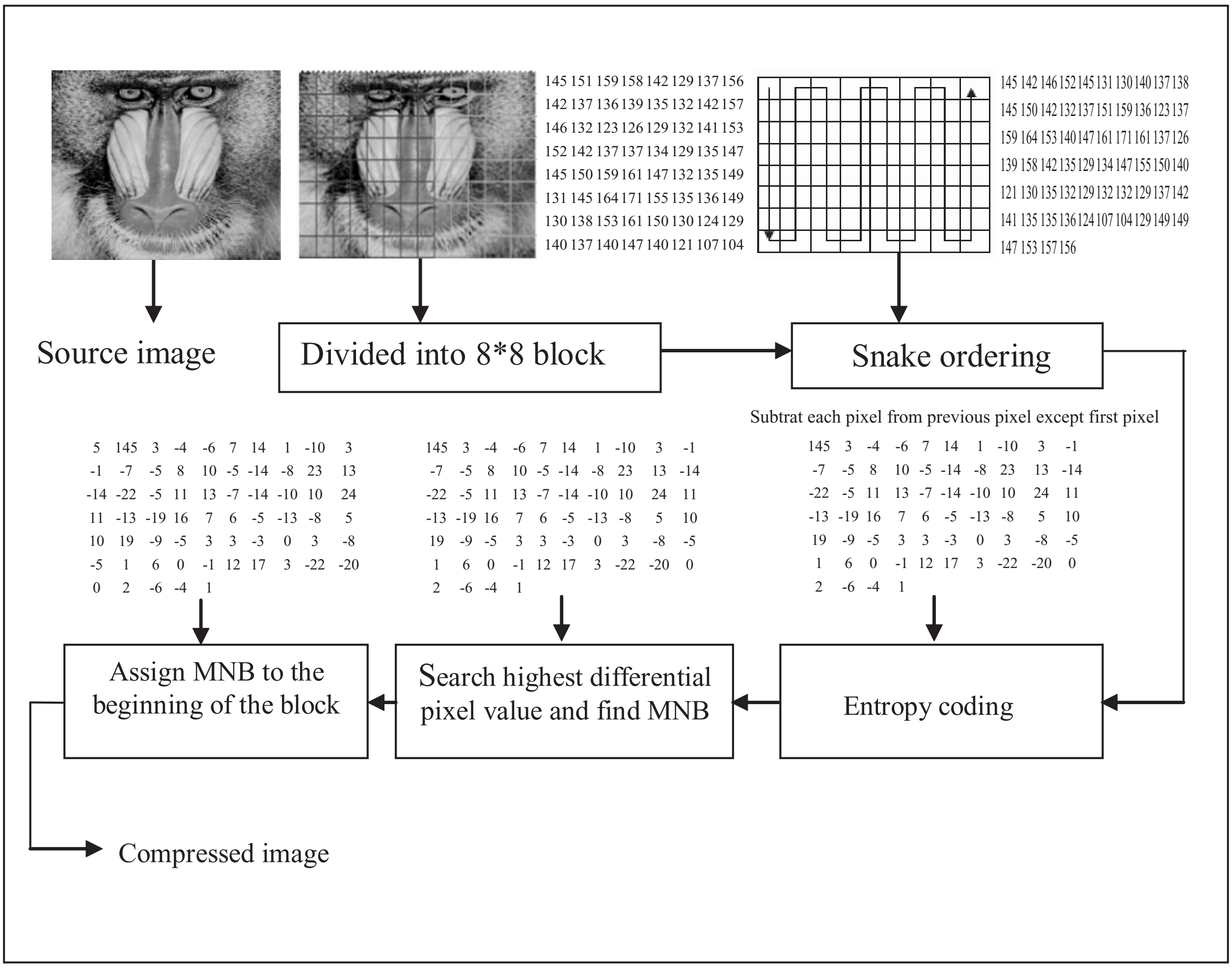


Fig. 1 The proposed framework.

At second step, the snake ordering is applied on this block. Here the block is read starting from leftmost column top to bottom then immediate next column bottom to top. Subsequently, immediate right column is scanned from top bottom and this process continues until block ends. Snake ordering transforms an 8×8 block of input values to a linear combination of these 64 pixel values. This step results the following sequence of values:

145 142 146 152 145 131 130 140 137 138
145 150 142 132 137 151 159 136 123 137
159 164 153 140 147 161 171 161 137 126
139 158 142 135 129 134 147 155 150 140
121 130 135 132 129 132 132 129 137 142
141 135 135 136 124 107 104 129149 149
147 153 157 156

At third step, entropy coding is implemented on this sequence which is actually subtraction of every pixel value from the previous pixel value except the first pixel value. Let us assume the model that subtracts each pixel value by its previous pixel value to generate the residual values

$$r_i = s_{i-1} - s_i \tag{1}$$

Where r_i, s_i and s_{i-1} are residual value, each pixel value and previous pixel value, respectively. This step results the following transformed sequence:

145	3	-4	-6	7	14	1	-10	3	-1	-7
-5	8	10	-5	-14	-8	23	13	-14	-22	-5
11	13	-7	-14	-10	10	$\boxed{24}$	11	-13	-19	16
7	6	-5	-13	-8	5	10	19	-9	-5	3
3	-3	0	3	-8	-5	1	6	0	-1	12
17	3	-22	-20	0	2	-6	-4	1		

At very next step, this transformed sequence where the first pixel value (AC value) remains unchanged is explored to get highest differential pixel value (DP). Here, the highest differential pixel value in this example is 24. Then maximum number of bits in binary representation of DP (MNB) is calculated. For this example, the MNB is 5. At final step, it is assigned to the beginning the block and hence, we get the following compressed image from the source image.

5	145	3	-4	-6	7	14	1	-10	3
-1	-7	-5	8	10	-5	-14	-8	23	13
-14	-22	-5	11	13	-7	-14	-10	10	24
11	-13	-19	16	7	6	-5	-13	-8	5
10	19	-9	-5	3	3	-3	0	3	-8
-5	1	6	0	-1	12	17	3	-22	-20
0	2	-6	-4	1					

Snake ordering works better than the zig-zag ordering. That means it requires fewer bits per pixel after entropy coding. Now, the comparison between two ordering is given in the following:

For the above example,

In case of Snake ordering,

After entropy encoding total bit = 212 bit

Average bit = 212/64=3 bit

In case of Zig-zag ordering,

After entropy encoding total bit = 230 bit

Average bit = 230/64=3.59 bit

3. RESULTS

All experiments were done on Core(TM) 2, 2 GHz with 2 GB RAM under MATLAB environment. In the experiments, different sizes of the images were used as shown in Table 1. The complete testing image compression database consists of 100 digital images and some images are shown in Figure 2. The proposed framework is compared with other methods. The comparison result is shown in Table 2. The results obtained from implementation of the proposed method are compression ratio (CR), bits per pixel (BPP), peak signal to noise ratio (PSNR), mean square error (MSE) and computational cost are calculated for comparing the performance of the proposed framework with the other methods. Output image as shown is Fig. 4. If we see carefully, Fig. 2 and 4 are same that means without a change in the quality of image.

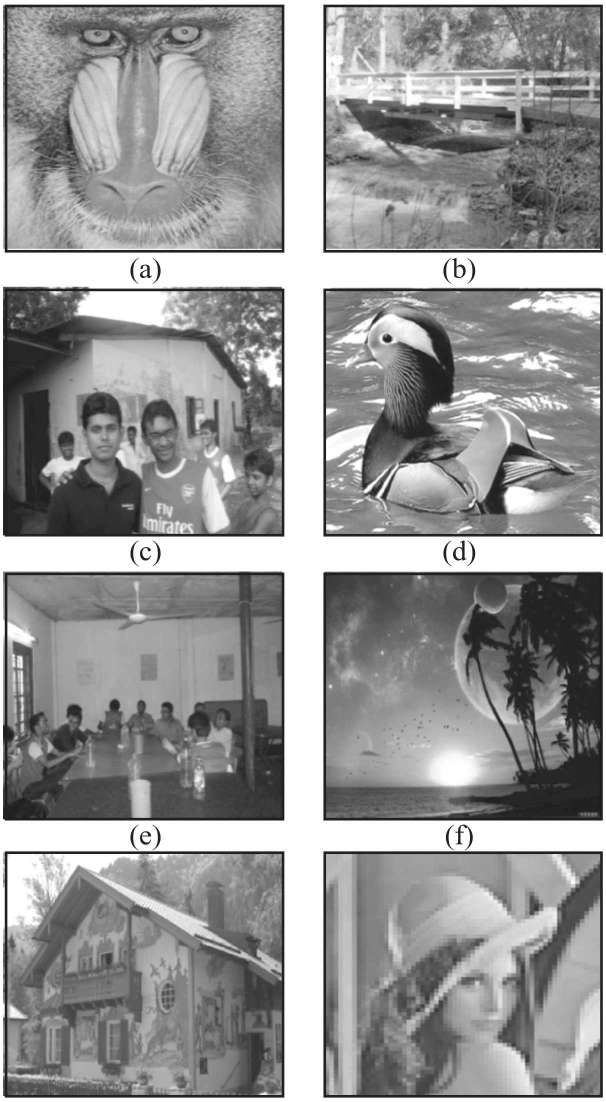


Fig. 2 Some sample input images.

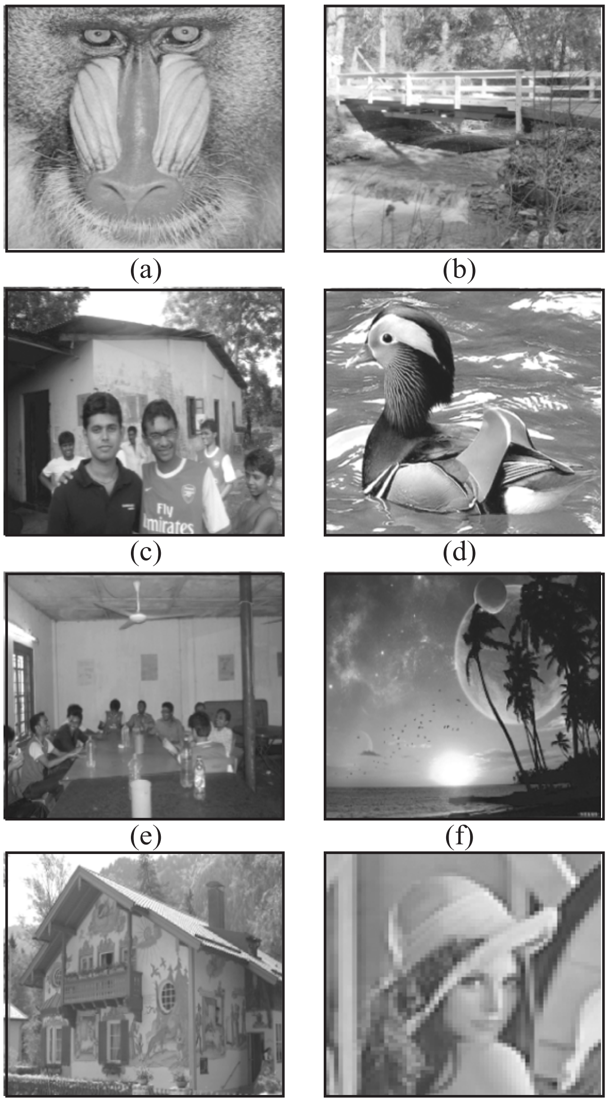


Fig. 4 Some sample output images.

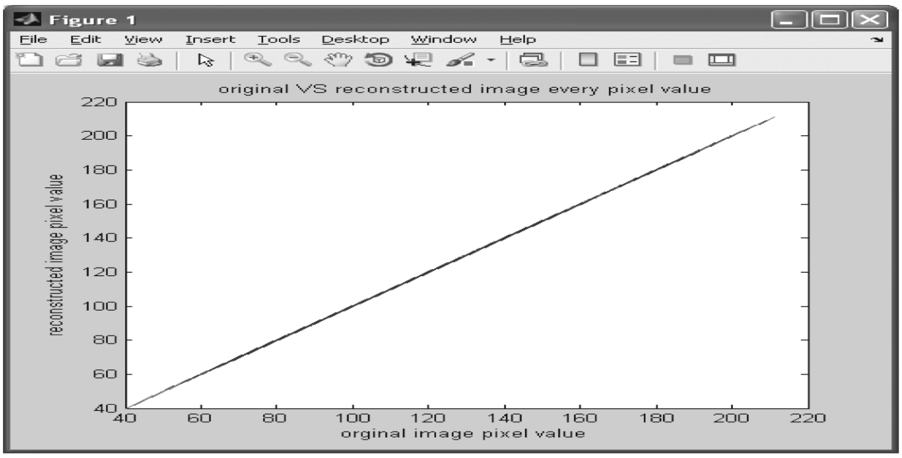


Fig. 3 Comparison of reconstructed and original pixel values.

The proposed framework is lossless because the graph of Fig. 3 shows that it is linear with respect to the x-axis and y-axis which represent the original and the reconstructed image pixel value, respectively.

The proposed framework is compared with SPIHT [11] and CALIC [10] methods as per the following parameters.

$$\text{Avg. comp. ratio} = \frac{\sum_{i=1}^{100} N_i^1}{\sum_{i=1}^{100} N_i^2} \tag{2}$$

Where i, N_i^1 and N_i^2 are no. of images, original image size and compressed image size, respectively.

$$BPP = \frac{8}{Avg.CR} \tag{3}$$

Where BPP is bits per pixel

$$MSE = \frac{\sum_{M,N} [N_1(m,n) - N_2(m,n)]^2}{M * N} \tag{4}$$

Where $N_1(m,n)$ and $N_2(m,n)$ are original image pixel value and reconstructed image pixel value, respectively. In the previous equation, m and n are the number of rows and columns in the input images and M, N are total rows and columns in all images, respectively. Then the block computes the PSNR using the following equation:

$$PSNR = 10 \log_{10}\left(\frac{R^2}{MSE}\right) \tag{5}$$

Where R is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this R value is 255. The computational cost for different types of images is shown in Table 3.

Table 1 Average compression ratio.

Image no.	Original size	Compressed size	Compression ratio
1	257 KB	138.92 KB	1.85
2	257 KB	130.45 KB	1.97
3	2.25 MB	1045 KB	2.15
4	293 KB	114.9 KB	2.55
5	402 KB	118.23 KB	3.40
6	500 KB	139.66 KB	3.58
7	577 KB	139.7 KB	4.13
8	1.83 MB	443 KB	4.23
....	......	......	......
99	385 KB	205.88 KB	1.87
100	385 KB	145.28 KB	2.65
Average compression ratio :			2.65

Table 2 Comparison of the performance analysis.

Method name	BPP	PSNR	MSE
SPIHT method	4.75	α	0
CALIC method	4.63	α	0
The proposed method	3.02	α	0

Table 3 Average computational cost for image compression.

Image type	Size/duration	Band width	Transmission time (s)
Gray image [5]	512 × 512	2.1 Mb	73
Gray image (SPIHT) [8]	512 × 512	2.1 Mb	43.34
Gray image (The proposed method)	512 × 512	2.1 Mb	27.64

4. CONCLUSION

In conclusion, it can be said that image compression is a very important issue for network transmission and storage space. In this work, a compression framework is proposed for gray images to achieve more compression ratio and less computational cost using snake ordering technique. The given gray image is divided into 8*8 block and then snake ordering method is employed. Subsequently, entropy coding is applied on transformed sequence to figure out the highest different pixel value (DP). Eventually, maximum number of bit required to represent differential pixel value in binary is assigned to the beginning of the block. Thus, the proposed framework generates the compressed image for storage or transmission. Experimental result also shows that the proposed framework surpasses than the conventional set partitioning in hierarchical trees (SPIHT) and Context based adaptive lossless image codec (CALIC) methods [10, 11]. The proposed framework yields 20% more compression in comparison to the other methods. The computational cost reduces in our proposed framework. It is sensitive for the color images. So, our future work is to modify the proposed framework so that it can work for color images effectively.

REFERENCES

1. Aree Mhammad and Loay George, *“Intraframe compression using lifting scheme –wavelet based transformation (9/7-Tap Filter)”*, Journal of Zankoy Sulaimani, Vol. 8, Issue 3, No. 1, 2011.
2. Parveen Banu and Dr.Y.Venkataramani, *“An efficient hybrid image compression scheme based on correlation of pixels for storage and transmission of images”*, International Journal of Computer Applications, Vol. 18, No. 3, 2011.
3. Jagadish h. Pujar and Lohit m. Kadlaskar, *“A new lossless method of image compression and decompression using huffman coding techniques”*, Journal of Theoretical and Applied Information Technology,Vol. 8, pp. 18-23, 2010.
4. Nishat Kanvel and DR. Elwin Chandra Monie, *“Adaptive lifting based image compression scheme for narrow band transmission system”*, Journal of Physical Sciences, Vol. 4, pp. 160-164, 2009.

5. Y Chkrapani and K Soundera Rajan, *"Implementation of fractal image compression employing artificial neural networks"*, World Journal of Modelling and Simulation, Vol. 4, No. 4, pp. 287-295, 2008.
6. Lane Brooks and Keith Fife, *"Hardware efficient lossless image compression engine"*, In proceeding of the IEEE International Computer Application Signal System Processing, Vol. 185, pp. 385-388, 2004.
7. Ismail Avcibas and Nasir Memon, *"A progressive lossless/near-lossless image compression algorithm"*, IEEE Signal Processing Letters, Vol. 9, No. 4, pp. 312-314, 2002.
8. Shin-Si Chen, Chang-Biau Yang and Kuo-Si Huang, *"Fractal image compression based on intrablock variance distribution and vector quantization"*, Society of Photo-Optical Instrumentation Engineers, Vol. 41, No. 11, pp. 2824-2830, 2002.
9. R. Ansari, N. Memon and E. Ceran, *"Near-lossless image compression techniques"*, Journal of Electronices Imaging, Vol. 7, No. 3, pp. 486–494, 1998.
10. X. Wu and N. Memon, *"Context-based, adaptive, lossless image coding"*, IEEE Transmission Communication, Vol. 45, pp. 437–444, 1997.
11. A. Said and W. A. Pearlman, *"A new fast and efficient image codec based on set partitioning in hierarchical trees"*, IEEE Trans. Circuits Syst. Video Technol., Vol. 6, pp. 243–249, June 1996.
12. Last day of access (November 20, 2011). [Online]. Available: http://en.wikipedia.org/wiki/MATLAB
13. Last day of access (November 20, 2011). [Online]. Available: http://www.mathworks.com/access/helpdesk/techdoc/ref

UB OPERATOR PRECEDENCE PARSING ALGORITHM

A. H. M. Kamal[1]
[1]Dept. of Computer Science & Engineering, Jatya Kabi Kazi Nazrul Islam University, Bangladesh

Abstract: This paper presents an algorithm for parsing operator grammar that can deal with unary operator. The bulk of this paper is devoted to operator precedence parsing method, especially to unary operator precedence parsing methods that are typically used in compiler. We have algorithm in hand for operator precedence parsing. But they cannot handle unary operators efficiently. Some of the proposals are language dependent link C. In this paper an algorithm has been proposed to deal any right-sentential form. That means, this proposed algorithm can handle (parse) any right-sentential form, where it is containing unary or binary operator or both is not a factor. This proposal is not for any specific compiler. Rather it is a generalize one. Obviously this will help in compiler writing.

Keywords: *Unary operator, Prime phrase, Handles, Right-sentential form, Operator precedence relation.*

1. INTRODUCTION

A grammar gives the precise, yet easy-to-understand syntactic specification of a programming language. The syntax of a programming language constructs can be described by context free grammar [1]. CFG has played a central role in compiler technology. Finally parsing unary/binary operators in an efficient way is presented here.

UB operator stands for unary and binary operator. Parsing is the process of determining if a string of tokens can be generated by a grammar. In compiler, parser takes tokens from scanner as input, then parses these token streams and produces a parse tree or parse table. Most parsing methods fall into one of the two classes, called top-down and bottom-up parsing. These terms refer to the order in which nodes in the parse tree are constructed.

To aid the compiler designer lot of proposal on parsing grammar is already on our hand. In [4] author analyzes the Floyd´s operator precedence grammar algorithm and defined a necessary and sufficient condition for that Floyd’s algorithm. In [5], author introduces Meta-rules for parsing German grammar. However, it does not guide us to parse unary operator. In [6], authors tried to solve that problem. But the proposed solution of that is to use a second dummy operator. However selecting a dummy operator and its type is sometimes difficult for the compiler. In [7] author proposes a technique for parsing unary operator. But the limitation is that it is stated only for C compiler. Another technique in handling unary operator is explained in [8]. It is stated to solve that problem in lexical analysis level. At the time of lexical analysis, lexical analyzer will return two different symbols for unary and binary operator. However indicating binary and unary operator by different symbols at lexical analysis section may create complexity at parser section if proper information of symbols is not passed to parser. In [9] author has just shown a solution for few limited expression. It is not a generalized one for all expressions as well as for compilers. Maintaining a stack and putting a symbol in stack [10] for unary operator can be another solution of handling unary operators. Authors in [11] introduce an algorithm to process Chinese language where they use word as operand and operator as relation. They have also maintained a precedence of operator. There they have avoided the use of unary operator.

Though a few number of research paper is cited in the above which were presented to parse operator grammar, there is a great problem at parsing a unary operator (++, --, -, !, etc). Actually, to the best of my knowledge, there is no standard and machine independent algorithm in our hand for handling unary operator of operator grammar by operator precedence parsing method. To overcome that problem we have presented here an algorithm for parsing unary operator of operator grammar.

The article is structured in few sections. The area of our discussion belongs to operator grammar. Operator precedence parsing is an easy-to-implement parsing technique. Compiler maintains an operator precedence relation table to perform the precedence parsing. Section 2 will describe those three terminologies. Section 3 will explain the present operator precedence parsing algorithm. Section 4 will elucidate the proposed algorithm. Lastly section 5 will demonstrate the result of the algorithm.

2. BASIC TERMINOLOGIES

Before describing our proposed algorithm it would be convenient to define some basic terminologies like operator grammar, operator precedence parsing, operator precedence relation.

2.1 Operator Grammar

Operator grammars have the property (among other essential requirements of CFG) that no production right side is $\in$ or has two adjacent non-terminals.

Definition: a grammar G is an operator grammar if it contains no rules of the form $V \rightarrow aXYb$, where $X, Y \in V_N$, that is, the occurrence of consecutive nonterminals in the right part of any rule is forbidden. Also empty rules are not permitted. The language generated by an operator grammar is called an operator language [2].

Observe the following example.

E → EAE | (E) | -E | id … … … (i)

A → + | - | * | / | ↑

The above grammar (i) is not an operator grammar, because the right side EAE has more than one consecutive nonterminals. However if we substitute for A each of its alternatives, we obtain the following operator grammar:

E → E + E | E – E | E * E | E / E | E ↑ E | (E) | - E | id … … … (ii)

2.2. Operator Precedence Parsing

Operator precedence parsing is an easy-to-implement parsing technique, which involves establishing precedence relations between the operator (terminal) symbols of a grammar. This parsing method constructs a parse essentially in a left-to-right manner. The parse is not a converse rightmost derivation. The method, however, is a bottom-up technique, which involves no backup [2].

As a general parsing technique, operator precedence parsing has a number of disadvantages. For example, it is hard to handle tokens like the minus sign, which has two different precedences (depending on whether it is unary or binary). Few compilers support some unary operator like ++, -- (in c, c++). They may be postfix or prefix of an operand. So it is also hard to handle such token like ++, --. The proposed algorithm, presented below, is able to handle such kind of operator.

2.3. Operator Precedence Relation

In operator precedence parsing we have three disjoint precedence relations, <: , ≅ and :>, between certain pairs of terminals. These precedence relations guide the selection of handles and have the following meanings:

Relation	Meaning
a <: b	**b** will be reduced before **a**
a ≅ b	**a** has the same precedence as **b**
a :> b	**a** will be reduced before **b**

The intention of the precedence relation is to delimit the handle of a right–sentential form, with <: marking the left end, ≅ appearing in the interior of the handle, and :> marking the right end.

According to the definition of operator grammar the right-sentential form may be written as $B_0 a_1 B_1 \dots \dots \dots a_n B_n$, where each B_i is either $\in$ (the empty string) or a single nonterminal and each a_i is a single terminal. Suppose that between each a_i and a_{i+1} exactly one of the relations <: , ≅ and :> holds.

From the above discussion a formal definition of operator precedence relations can be build. The operator precedence relations associated with an operator grammar are defined over its terminal alphabet as follows:

1. $S_1 \cong S_2$ if and only if there exists a production $U \rightarrow aS_1S_2b$ or $U \rightarrow aS_1XS_2b$, where $S_1, S_2 \in V_T$, and $X \in V_N$
2. S_1 <: S_2 if and only if there exists a production $U \rightarrow aS_1Xb$ such that $X \rightarrow^+ S_2 \ldots \ldots$ or $X \rightarrow^+ YS_2 \ldots \ldots$, where $S_1, S_2 \in V_T$, and $X, Y \in V_N$.
3. S_1 :> S_2 if and only if there exists a production $U \rightarrow aXS_2b$ such that $X \rightarrow^+ \ldots \ldots S_1$ or $X \rightarrow^+ \ldots \ldots S_1Y$, where $S_1, S_2 \in V_T$, and $X, Y \in V_N$.

According to the roles of operator precedence relation an operator precedence relation table can be defined. For this, consider the following grammar.

E→E+E|E–E|E*E|E/E|(E)|-E|E-E|++E|E++|E--|--E|id ... ... (iii)

Let us use $ to mark each end of the string, and define $ <: b and b :> $ for all terminal b. Now following is operator precedence relation table for the above grammar.

	+	-	*	/	↑	id	(	)	$
+	:>	:>	<:	<:	<:	<:	<:	:>	:>
-	:>	:>	<:	<:	<:	<:	<:	:>	:>
*	:>	:>	:>	:>	<:	<:	<:	:>	:>
/	:>	:>	:>	:>	<:	<:	<:	:>	:>
↑	:>	:>	:>	:>	:>	<:	<:	:>	:>
Id	:>	:>	:>	:>	:>			:>	:>
(	<:	<:	<:	<:	<:	<:	<:		
)	:>	:>	:>	:>	:>			:>	:>
$	<:	<:	<:	<:	<:	<:	<:		

Fig. 1 Operator precedence relation Table.

Now consider one initially have the right-sentential form id + id * id and the above precedence relation table. Then the string w with the precedence relations is:

w = $ <: id :> + <: id :> * <: id :>$ (iv)

for example <: is inserted between the leftmost $ and id since <: is the entry in row $ and column id.

3. OPERATOR PRECEDENCE PARSING ALGORITHM

We have few operator precedence parsing algorithms in our hand which are mentioned in section 1. They were proposed to solve some specific objectives. The most common and famous one is explained here. This algorithm has two parts.

1. Processing input string: This step reduces all the terminal by nonterminal without operator

2. Reducing to Start: This step reduces all terminals (operator) and parses to the root (start).

3.1 Processing Input String

The handle can be found by following process [1].

i) Scan the string w from the left and until the first :> is encountered. In above i) expression (iv), this occurs between the first id and +.

ii) Then scan backwards (to the left) over any ≅'s until a <: is encountered. In (iv), we scan backwards to $.

iii) The handle contains every thing to the left of the first :> and to the right of the <: encountered in step (ii), include any intervening or surrounding nonterminals. In expression (iv), the handle is the first id.

By using this process prime phrase (id) can be reduced by E and then the right-sentential form E + id * id can be reached. After scanning subsequence input and reducing prime phrase by handle the right-sentential form E + E * E can be obtained.

Algorithm: Operator-Precedence Parsing Algorithm

Input: an input string **w** and a table of **precedence relations**.

Output: reduction step to start symbol or root of a parse tree; otherwise an error indication.

Method: initially stack contains $ and input buffer contains w$.

a. set *ip* to point the first symbol of w$;
b. repeat forever
c. if $ is on top of the stack and *ip* points to $ then
  - return
d. else
  - **a** be the topmost terminal symbol on the stack and **b** be the symbol pointed to by *ip*;
e. if a <: b or a ≅ b then
  - push b onto the stack ;
  - advanced *ip* to the next input symbol;
f. else if a:> b then
  - repeat
  - pop stack
  - until the top stack terminal is related by <: to the terminal most recently popped.
g. else error ();
h. End

3.2 Reducing to Start

At this point a string of only non-terminals and operators is produced. This obtained right-sentential form is E + E * E. Consider now the string is $ + * $ obtained by deleting the nonterminals. Inserting the precedence relations, the string $ <: + <: * :> $ can be found. This indicates that prime phrase is E * E and its reduction leads to a string E + E. Similarly, E is found reducing prime phrase by handle. This is the start symbol.

4. PROPOSED ALGORITHM

In our proposed algorithm we have modified the precedence relation table. Then we have processed the input string to have a formatted input string. This formatted input string and modified precedence relation table is used in our algorithm to parse a grammar.

4.1 Policy

Before entering to the explanation of our proposed algorithm consider the following right-sentential form.

a + (–b) + c
a – (b++) +c
a – (++b) +c
a – (– –b) + c
a – (b– –) + c
a && (¬ b) && c

It is not possible to parse the above right-sentential form using the algorithm described above. Here is a proposal that will handle all kind of operators if an exact table of precedence relation is provided. Actually this proposed algorithm uses two precedence relation tables (that is define bellow).

	+	–	*	/	↑	Id	(	)	$	@
+	≅	:>	<:	<:	<:	<:	<:	:>	:>	
–	:>	≅	<:	<:	<:	<:	<:	:>	:>	
*	:>	:>	≅	:>	<:	<:	<:	:>	:>	
/	:>	:>	:>	≅	<:	<:	<:	:>	:>	
↑	:>	:>	:>	:>	≅	<:	<:	:>	:>	
id	:>	:>	:>	:>	:>			:>	:>	:>
(	<:	<:	<:	<:	<:	<:	<:			<:
)	:>	:>	:>	:>	:>			:>	:>	
$	<:	<:	<:	<:	<:	<:	<:			
@						<:		:>		

Fig. 2 Operator precedence relation Table.

Firstly the unary operator along with its operand (enclosing unary operator with its operand within '( )') will be grouped. Consider the input (right-sentential form) a+b++–c. We will group it as a+(b++)– c. How the group is formed is not a matter of our discussion. Any conventional algorithm can be used.

Second task is to insert $ at the beginning and end of input string. Next insert @ before the operand that contains postfix unary operator or insert @ after the operand that contains prefix unary operator. So for a right sentential form a+(b++)–c our input string will be w=$a+(@ b++)-c$ (Similarly for a+(-b)-c our input string will be w=$a+(-b@)-c$). Now insert preceding relation to string w from table of precedence relations (Fig 2).

Then the string will be
w=$<:a:>+<:(<: ≅ <:b:>+ ≅ + :>):> – <:c:>$
This string w will be termed as a formatted string. This formatted string and the modified precedence relation table is used in our proposal to parse a string successfully.

4.2 Our Proposal

The proposed algorithm also has two parts:

(i) Processing input string
(ii) Reducing to root (start)

4.2.1 Processing Input String

Here formatted input string is processed. Processed means reducing (prime phrase) all terminals including grouped non-terminal (expression enclosed by first parenthesis), without operator by handle.

Algorithm: UB Operator Precedence Parsing

Input: a formatted string w

Output: a right-sentential form after reduction of all terminal including group, without operators.

Method: Initially the stack contains $ and the input buffer the formatted string. To parse, we execute the following program:

(i) set *ip* to point to the second symbol of w.
(ii) repeat forever
(iii) if $ is on top of stack and *ip* points to $ then return;
(iv) Else
(v) let **a** be the topmost symbol on the stack and **b** the symbol pointed to by *ip*;
(vi) if **b= :>** then
(vii) repeat
(viii) Pop the stack
(ix) until the stack top is <:
(x) This popped string in reverse is prime phrase. Now pop <: from stack and push the handle of prime phrase to stack.
Advance *ip* to the next input symbol
(xi) else if **b**='**@**' or **b**= '≅' then
(xii) Advance input pointer *ip* to the next input symbol.
(xiii) else
(xiv) Push **b** onto the stack
Advance *ip* to the next input symbol;
(xv) end

4.2.2 Reduction to Root:

Now follow the following precedence relation table to complete bottom-up parsing

	+	-	*	/	↑	$
+	:>	:>	<:	<:	<:	:>
-	:>	:>	<:	<:	<:	:>
*	:>	:>	:>	:>	<:	:>
/	:>	:>	:>	:>	<:	:>
↑	:>	:>	:>	:>	:>	:>
$	<:	<:	<:	<:	<:	

Fig. 4 Precedence relation.

At this point we will get a sentential form that contains no unary operators, no group. So nothing is new in this section. This method is exactly as the discussion in 5.2.

5. RESULT ANALYSIS

This algorithm works very well and can successfully parse any operator, unary or binary. It is tested with various sample strings and in our experiment it is found that it successfully parses all the strings without error or illegal termination. The result of parsing a sample string is outlined here in the following.

Consider the right-sentential form is a+(b++)-c. So according to proposed algorithm, w will be as:
w= $a+(@b++)-c$

After inserting precedence relation:
w=$<:a:>+<:(<:@<:b:>+ ≅ +:>):>-<:c:>$

Now this w is the formatted string.

Step	Unexpended input string	Stack Contents	Prime Phrase	ip pointed to
0	$<:a:>+<:(<:@<:b:>+≅+:>):> –<:c:>$	$		
1	a:>+<:(<:@<:b:>+≅+:>):>-<:c:>$	$<:		<:
2	:>+<:(<:@<:b:>+ ≅+:>):>-<:c:>$	$<:a		A
3	+<:(<:@<:b:>+ ≅+:>):>-<:c:>$	$E	id	:>
4	<:(<:@<:b:>+ ≅+:>):>-<:c:>$	$E+		+
5	(<:@<:b:>+ ≅+:>):>-<:c:>$	$E+<:		<:
6	<:@<:b:>+ ≅+:>):>-<:c:>$	$E+<:(		(
7	@<:b:>+ ≅+:>):>-<:c:>$	$E+<:(<:		<:
8	<:b:>+ ≅+:>):>-<:c:>$	$E+<:(<:		@
9	b:>+ ≅+:>):>-<:c:>$	$E+<:(<:<:		<:
10	:>+ ≅+:>):>-<:c:>$	$E+<:(<:<:b		B
11	+ ≅+:>):>-<:c:>$	$E+<:(<:E	id	:>
12	≅+:>):>-<:c:>$	$E+<:(<:E+		+
13	+:>):>-<:c:>$	$E+<:(<:E+		≅
14	:>):>-<:c:>$	$E+<:(<:E++		+
15	):>-<:c:>$	$E+<:(E	E++	:>
16	:>-<:c:>$	$E+<:(E)		)
17	-<:c:>$	$E+E	(E)	:>
18	<:c:>$	$E+E-		-
19	c:>$	$E+E-<:		<:
20	:>$	$E+E-<:c		C
21	$	$ E+E-E	id	:>

Fig 3: Step-by-step operation of proposal algorithm.

So after executing the algorithm we will get the right-sentential form as $E+E-E$ in stack.

7. PERFORMANCE ANALYSIS

The algorithm performs very well in parsing right-sentential forms. It uses linear data structures and simple calculation methods. Performance analysis will be perfect if we explain here the space and time complexity used by our algorithm.

Space Complexity: It first expands the string by inserting the precedence relations which makes the string double in length. A stack of same size is used to handle the functionality. This way the handled data structure becomes four times of original expression. However the same data structure can be used to verify multiple expressions. As a result space complexity is negligible.

Time Complexity: The original expression is expanded first and it becomes the double after conceiving the precedence relation. Every times it receives a symbol from processed input, w, and compares with the stack top. After the stack operation it advances the input pointer. Thus it takes 2n comparison where n is the length of original expression. So the complexity of that algorithm is n in O(n) whereas its functionality is machine independent. So it is a well performing one.

8. CONCLUSION

Obviously the proposed algorithm will help in compiler writing. It is already mentioned that there is no existing efficient generalized algorithm, to our best of knowledge, for parsing operator grammar with unary operator(s). This proposed algorithm can be a solution in parsing unary operator(s). It is a completely new idea in parsing unary grammar. It is very easy to understand and to implement. It parses string efficiently and correctly. The complexity is similar to present algorithm and it is n in O(n). However the functionality of that algorithm is machine independent. So it can be a nice choice to the compiler writer and code generator to parse unary operator(s).

REFERENCES

1. Alfred V. Aho, Ravi Sethi, Jeffrey D. Ullman; *"Compilers Principles, Technique, and Tools"*; Addison- Wesley, pp. 40-47, 181-190, 195-200, 203-214, 1999.
2. Jean-Paul Tremblay, Paul G. Sorenson; *"The Theory and Practice of Compiler Writing"*; McGraw-Hill, pp.278-280, 286-300, 1985.
3. John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman; *"Introduction to Automata Theory, Language and Computation"*; Person Education Asia; 2nd Edition, pp. 131-145, 169-190, 2002.
4. Peter Ruika, *"Validity test for Floyd's operator-precedence parsing algorithms"*, Journal of Lecture Notes in Computer Science, Volume 74/1979, ISSN: 0302-9743, pp 415-424, 2006.
5. Wilhelm WEISWEBER, Susanne PREUB, *"Direct Parsing with Metarules"*, In proceeding of the COLING-92, NANTES AUG, 23-28, 1992.
6. Leo Liberti, *"Performance Comparison of Function Evaluation Methods"*, oai:arXiv.org:cs/0206010 (2007-08-16); 2002.
7. Steve Hill, *"Continuation Passing Combinators for Parsing Precedence Grammars"*; {http://www.cs.kent.ac.uk/pubs/1994/70}, Technical Report 24-94*, University of Kent, UK; 1994.
8. Last day of access (October 19, 2011) [Online]. Available: http:// "203.208.166.84/ dtanvirahmed / cse309N / CSE309-4.5-4.6.ppt"
9. Last day of access (October 19, 2011) [Online]. Available: http:// "homepage. smc.edu/ kennedy_john/ parsetrees.pdf
10. Last day of access (October 19, 2011) [Online]. Available:"http://www.csc.lsu.edu /~gb/csc1351/Ass ignments/proj4.pdf"; 2006.
11. Xiao Yang, Jiancheng Wan, *"A parsing algorithm of natural language based on operator precedence"*, In proceedings of the 12th IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE '05) (IEEE Cat. No. 05EX1156)", ISBN: 0-7803-9361-9, pp 73- 78, 2006.

AUTOMATIC SPEECH RECOGNITION TECHNIQUE FOR BANGLA CONSONANTS

A.H.M. Ashfak Habib[1] and Sanjib Biswas[1]

[1]Dept. of Computer Science & Engineering, Chittagong University of Engineering & Technology, Bangladesh

Abstract: Automatic recognition of spoken letter is one of the most challenging tasks in the field of computer speech recognition. The difficulty of this task is due to the acoustic similarity of many of the letters. Accurate recognition requires the system to perform fine phonetic distinctions. This paper presents a technique for recognizing spoken letter in Bangla. In this study we first derive feature from spoken letter. Mel-frequency cepstral coefficient (MFCC) has been used to characterize a feature. Dynamic time warping (DTW) employed to calculate the distance of an unknown letter with the stored ones. K-nearest neighbors (KNN) algorithm is used to improve accuracy in noisy environment.

Keywords: *Delta and double delta, Mel-frequency cepstral coefficient (MFCC), Dynamic time warping (DTW), K-nearest neighbors (KNN), Speech recognition.*

1. INTRODUCTION

Bangla (can also be termed as Bengali), which is largely spoken by the people all over the world, has been performed a very little research where many literatures in automatic speech recognition (ASR) systems are available for almost all the major spoken languages in the world. Although Bangla speakers' number is about 230 million today, which makes Bangla the seventh language [1], a systematic and scientific effort for the computerization of this language has not been started yet. Some efforts are made to develop BangIa speech corpus to build a Bangla text to speech system [2]. However, this effort is a part of developing speech databases for Indian Languages, where Bangla is one of the parts and is spoken in the eastern area of India (West Bengal). But most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Although the written characters of standard Bangla in both the countries are same, there are some sounds which are produced differently in different pronunciations of standard Bangla.Therefore, there is a need to do research on main stream of Bangla, and we confined this study on recognition of spoken letter only.

Some developments on Bangla speech processing or Bangla ASR can be found in [3]-[10]. For example, Bangla vowel characterization is done in [3]; isolated and continuous Bangla speech recognition on a small dataset using hidden Markov models (HMMs) is described in [4]; recognition of Bangla phonemes by Artificial Neural Network (ANN) is reported in [7]-[8]. Continuous Bangla speech recognition system is developed in [9], while [10] presents a brief overview of Bangla speech synthesis and recognition. Most of the research effort on recognizing Bangla speech is performed using the ANN and LPC based classifier. No research work has been reported yet that uses the DTW and K-NN technique and MFCC based feature extraction.

In this paper, we build an ASR system for Bangla consonants. We first develop a small size of database for bangle consonants to achieve the goal. Then mel-frequency cepstral coefficient (MFCC) is used to extract feature from the input speech. After that extracted feature are saving as reference templates. Then real time input coefficient are compared to the reference templates using dynamic time warping (DTW) algorithm. And finally the output of DTW are inserted into the k-nearest neighbors (K-NN) based classifier for obtaining the word recognition performance.

2. SPEECH RECOGNITION SYSTEM

Speech is input via microphone and its analog waveform is digitized. The job of the recognition system is to derive necessary information from the waveform needed to make the correct decision. The process of recognition system operation typically consists of two phases first one is training and finally recognition. In the training phase, data for known classes are fed to the system. In the recognition phase, the system computes the features of pattern for unknown input and identifies the input with the class whose reference pattern matches these features most closely.

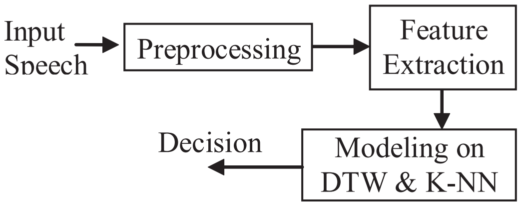


Fig. 1 Speech recognition system.

There are several alternative candidates those can be used as feature for speech. In the current research, mel-frequency cepstral coefficient (MFCC) is used for extracting features. Patterns are vectors of features. In pattern recognition DTW and K-NN are used for take decision.

2.1 Processing

In this stage, the first step is to record the speech data by a microphone in a specified format (wav file, 16000Hz and 16 bits). This wav data will be converting into a form that is suitable for further computer processing and analysis through a series of process that involves noise elimination and the speech end point detection process.

2.1.1 End Point Detection

We used the generalized end point detection algorithm which accepts an audio sample as input and returns a trimmed down version with non-speech sections trimmed off. Also known as voice activity detection, it utilizes the algorithm due to Rabiner & Sambur (1975).

2.2 Feature Extraction

In this paper the most important thing is to extract the feature from the speech signal. The speech feature extraction in a categorization problem is about reducing the dimensionality of the input-vector while maintaining the discriminating power of the signal. In this project we are using the Mel Frequency Cepstral Coefficients (MFCC) technique to extract features from the speech signal. Fig. 2 below shows the complete pipeline of Mel Frequency Cepstral Coefficients (MFCC).

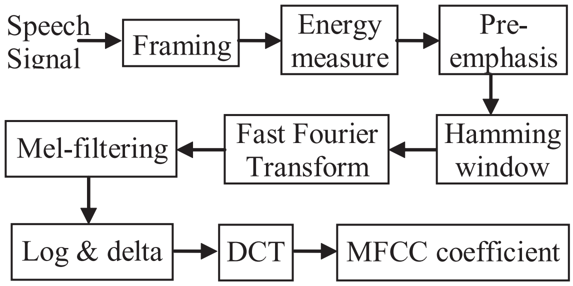


Fig. 2 The Procedure of MFCC.

2.2.1 Framing

The input speech signal is segmented into frames of 20~30 ms with optional overlap of 1/3~1/2 of the frame size. In this thesis we used 25 ms fo frame and overlap of 10 ms. i.e. 400 samples in each frame and 160 samples of overlap. For each of the frame we calculate 39 MFCC coefficient ((Energy+12) MFCC + Delta MFCC + Delta-Delta MFCC)

2.2.2 Energy Measure

The logarithmic frame energy measure (logE) is computed after the offset compensation filtering and framing for each frame:

$$\log E = \ln \{\Sigma^{N}_{1} S(i)^2\}$$

Here N is the frame length and S_{of} is the offset-free input signal. A floor is used in the energy calculation which makes sure that the result is not less than -50. The floor value (lower limit for the argument of ln) is approximately 2e-22.

2.2.3 Pre-Emphasis

The speech signal s(n) is sent to a high-pass filter:

$$s_2(n) = s(n) - 0.97.s(n-1)$$

Where s2(n) is the output signal and the value of a is usually between 0.9 and 1.0. We used 0.97 as the value of a. The goal of pre-emphasis is to compensate the high-frequency part that was suppressed during the sound production mechanism of humans. Moreover, it can also amplify the importance of high-frequency formants.

2.2.4 Windowing

When signal or any other function is multiplied by a window function, the product is zero valued outside the interval. The windowing is done to avoid problems due to truncation of the signal. Frame signal is tapered by hamming window to avoid discontinuities at the ends. If the signal in a frame is denoted by s(n), n = 0,…N-1, then the signal after Hamming windowing is s(n)*w(n), where w(n) is the Hamming window defined by:

$$w(n,\alpha) = (1-\alpha) - \alpha\cos(2\pi n/(N-1)),\ 0 \leqq n \leqq N-1$$

In practice, the value of α is 0.46.

2.2.5 FFT & Mel-Filtering

A Fast Fourier Transform (FFT) is an efficient algorithm to compute the Discrete Fourier Transform (DFT) and its inverse. A DFT decomposes a sequence of values into components of different frequencies but the frequency bands are not positioned logarithmically. As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system. We multiple the magnitude frequency response by a set of 24 triangular bandpass filters to get the log energy of each triangular bandpass filter. The positions of these filters are equally spaced along the Mel frequency, which is related to the common linear frequency f by the following equation:

$$mel(f)=1125*\ln(1+f/700)$$

Mel-frequency is proportional to the logarithm of the linear frequency, reflecting similar effects in the human's subjective aural perception. The reasons for using triangular bandpass filters are Smooth the magnitude spectrum such that the harmonics are flattened in order to obtain the envelop of the spectrum with harmonics. This indicates that the pitch of a speech signal is generally not presented in MFCC. As a result, a speech recognition system will behave more or less the same when the input utterances are of the same timbre but with different tones/pitch. And it Reduce the size of the features involved

2.2.6 Log & Delta Cepstrum

The energy within a frame is also an important feature that can be easily obtained. Hence we usually add the log energy as the 13rd feature to MFCC. Delta Cepstrum is used to catch the changes between the different frames. Delta Cepstrum is used to catch the changes between the different frames. It is also advantageous to have the time derivatives of (energy+MFCC) as new features, which shows the velocity and acceleration of (energy+MFCC). The equations to compute these features are:

$$\Delta C_m(t)= [\Sigma\pi=-M^M C_m(t+\pi)\pi] / [\Sigma\pi=-M^M \pi^2]$$

The value of M is usually set to 2. If we add the velocity, the feature dimension is 26. If we add both the velocity and the acceleration, the feature dimension is 39.

2.2.7 DCT

In this step, we apply DCT on the 24 log energy Ek obtained from the triangular bandpass filters to have L mel-scale cepstral coefficients. The formula for DCT is shown next.

$$C_m=\Sigma_{k=1}^{N}\cos[m*(k-0.5)*\pi/N]*E_k,\ m=1,2,\ ...,\ L$$

Where N is the number of triangular bandpass filters, L is the number of mel-scale cepstral coefficients. Usually we set N=24 and L=12. Since we have performed FFT, DCT transforms the frequency domain into a time-like domain called quefrency domain. The obtained features are similar to cepstrum, thus it is referred to as the mel-scale cepstral coefficients, or MFCC.

2.3 DTW & K-NN

Speech is a time-dependent process. Hence the utterances of the same word will have different durations, and utterances of the same word with the same duration will differ in the middle, due to different parts of the words being spoken at different rates. To obtain a global distance between two speech patterns (represented as a sequence of vectors) a time alignment must be performed. In this type of speech recognition technique the test data is converted to templates. The recognition process then consists of matching the incoming speech with stored templates. The template with the lowest distance measure from the input pattern is the recognized word. If D(i,j) is the global distance up to (i,j) and the local distance at (i,j) is given by d(i,j)

$$D(I,j)=min[D(i-1,j-1),D(i-1,j),D(i,j-1)]+d(i,j)$$

The final global distance D(n,N) gives us the overall matching score of the template with the input. The input word is then recognized as the word corresponding to the template with the lowest matching score. In order to improve the recognition accuracy in noisy environment we use K-NN algorithm. K-nearest neighbor algorithm (K-NN) is a method for classifying objects based on closest training examples in the feature space. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). We use k=3 in our thesis.

3. EXPERIMENTS AND RESULTS

In this research work, we give emphasis to the inclusion of DTW and K-NN technique for recognizing Bangla speech as no such works have been seen and also to evaluate the performance from several aspects.

The experiment was done using digital computer with 2.8GHz speed and 512MB memory machine. The sound samples were taken in a room environment and different speech processing techniques were applied to make the samples suitable for feature extraction and recognition. We have taken a vocabulary of10 consonants and test samples from 2 different speakers to observe the performance. The recognizer is capable of recognizing each spoken consonants existing in the database only when the words are spoken by the same speaker and the mood of the speaker is same. However for different speaker the performance decreases to almost 10 15%.Recognizing continuous speech with ANN classifier has average accuracy rate of 73.36% (K. J. Rahman, 2003), for three layer Back- Propagation Neural Network the maximum accuracy rate is 86.67% (M. R. Islam, 2005), and spoken letter recognition by measuring Euclidian distance, which can recognize only the vowels, has an 80% accuracy rate (A H M. Rezaul Karim, 2002). In comparison, the recognizer presented in this paper has an average accuracy rate of 90%. Spoken letter recognition by using only DTW is 80% but use of K-NN it increases almost 10%. Table-1 shows the performance of using K-NN.

Table. 1 Accuracy rate.

Classifier	Speaker dependent	Speaker independent
DTW	78%	60%
DTW & KNN	86%	75%

The performance analysis reveals the importance about the improvement of the recognition with different speaker. Several studies on SR system emphasizes on the training data with varieties of speakers to increase the performance. So, next we should put our effort on collecting the training data from different speaker and observe the performance.

4. APPLICATION

The entire domain where speech recognition technology can be applied are automatic translation, automotive speech recognition, dictation, hands-free computing: voice command recognition computer user interface, home automation, interactive voice response, medical transcription, mobile telephony, pronunciation evaluation in computer-aided language learning applications and robotics. In our research work we are considering the isolated speech recognition for commands & control, data entry, mobile telephony and home automation task.

5. CONCLUSION

In this paper, we concentrated on the research and development of a Bangla Speech Recognizer using the appropriate technique and tools. We have studied the past works and to the best of our knowledge this work is the first reported attempt to recognized Bangla speech using DTW and K-NN Technique with the assist of template language model. Scientists achieved remarkable success in speech recognition for many languages. In English, continuous speech can be recognized with accuracy rate more than 95% [11]. Unfortunately in Bengali accuracy is about 85%[4]. In this paper, we discussed how a spoken letter can be recognized. And the level of accuracy is almost 90%. By ensuring the perfection (i.e. noise free) of the recorded signal it is possible to increase the accuracy of recognition. Hope our effort will help to take the research on Speech Recognition one step toward continuous speech recognition in Bangla with higher accuracy.

REFERENCES

1. Last day of access (October 20, 2011 [online]. Available: http://www.bangla pedia.org/ httpdocs/ HT/B_0137.HTM
2. Hossain, S. A., Rahman, M. L. and Ahmed, F., *"Bangla vowel characterization based on analysis by synthesis"*, In proceedings of the WASET, vol. 20, pp. 327-330, April 2007.
3. Hasnat, M. A., Mowla, J. and Khan, M., *"Isolated and Continuous Bangla Speech Recognition: Implementation Performance and application perspective"*, In proceedings of the International Symposium on Natural Language Processing (SNLP), Hanoi, Vietnam, December 2007.
4. Karim, R., Rahman, M. S. and Iqbal, M. Z., *"Recognition of spoken letters in Bangla"*, In proceedings of the 5th International Conference on Computer and Information Technology (ICCIT02), Dhaka, Bangladesh, 2002.
5. Houque, A. K. M. M., *"Bengali segmented speech recognition system"*, Undergraduate thesis, BRAC University, 2006.
6. Roy, K., Das, D. and Ali, M. G., *"Development of the speech recognition system using artificial neural network"*, In proceedings of the 5th International Conference on Computer and Information Technology (ICCIT02), Dhaka, Bangladesh, 2002.
7. Hassan, M. R., Nath, B. and Bhuiyan, M. A., *"Bengali phoneme recognition: a new approach"*, In proceedings of the 6th International Conference on Computer and Information Technology (ICCIT03), Dhaka, Bangladesh, 2003.
8. Rahman, K. J., Hossain, M. A., Das, D., Islam, T. and Ali, M. G., *"Continuous Bangla speech recognition system"*, In proceedings of the 6th International Conference on Computer and Information Technology (ICCIT03), Dhaka, Bangladesh, 2003.
9. Hossain, S. A., Rahman, M. L., Ahmed, F. and Dewan, M., *"Bangla speech synthesis, analysis, and recognition: an overview"*, In proceedings of the NCCPB, Dhaka, 2004.
10. Stuckless, R. *"Special Applications of Automatic Speech Recognition (ASR) with Deaf and Hard-of- Hearing People"*. American Annals of the Deaf.

DESIGN OF AMPLIFIERS WITH HIGH LINEARITY

R. M. M. Hasan[1], K. T. Ahmmed[2] and Anik Saha[3]
[1]Dept. of Electrical & Electronics Engineering, Chittagong University of Engineering & Technology, Bangladesh
[2]Dept. of Applied Physics & Electronics & Communication Engineering, Chittagong University, Bangladesh
[3]Dept. of Computer Science & Engineering, Chittagong University of Engineering & Technology, Bangladesh

Abstract: CMOS implementation of a novel amplifier design technique based on the negative impedance compensation is presented. The simulation results have shown that this technique is suitable for linearising amplifiers with low open-loop gain, which is likely to be the case in very high frequency amplifier design and therefore appropriate for RF applications. It has also shown that the circuit configuration using the novel technique is relatively simple and high linearity and high gain accuracy is achievable.

Keywords: *CMOS, IMD, Negative impedance, OP-AMP.*

1. INTRODUCTION

The amplification of signals in electronic devices is nonlinear, which cause the processing of signals in electronic circuit or equipment to be also nonlinear. Usually, even very small deviations from linearity may result in combination effects, greatly reducing the interference immunity and sensitivity of electronic systems, especially when they are used in high frequency applications such as multi-carrier telephony, wireless and mobile communication. Therefore, linearization techniques have become essential in the amplifier designs due to the rigorous performance requirements of modern communication systems.

A number of linearization techniques have been proposed such as feedforward [1] and predistortion [2] that offer different degree of performance at expense of circuit complexity. Unfortunately, most of these methods require costly and bulky RF circuitry that is not suitable for mobile terminals. Recently a low frequency method [3, 4] has also been described, which offers the potential of greatly simplifying the design of linearization systems and feedforward system has been used in many applications because of its unconditionally stable characteristics and ability to produce a broad-band and highly linear amplifier [5]. But the feedforward approach is very sensitive to component tolerance and drift, and requires adaptive control [5, 6]. In predistortion signal approach phase and amplitude must be accurately set to achieve the desired cancellation. This setting, usually, is difficult to be established because of the high sensitivity of the predistorters to variations of temperature, ageing and other external stimuli and also the predistorter and amplifier nonlinearities are required to be known [7, 8].

This paper describes a new approach of using negative impedance compensation in view of minimizing non linear distortion introduced by the amplifier in high frequency [9]. The approach is based on adding an impedance circuit on the input of the amplifier to reduce the non-linearity. This approach has two advantages 1) It increases the gain accuracy 2) Reduces the effect of nonlinear terms generated by the active device. As will be seen, by using the new technique with fully CMOS technology, both the gain accuracy and linearity of the amplifier can be improved significantly without loss in gain. The research aims at the realisation of highly linear amplifiers with fully CMOS technology in high-integration RF transceivers used for wireless communications.

2. OPERATIONAL AMPLIFIER

2.1 Effect of Non-zero Op-Amp Input Admittance and Output

The fundamental configuration in which the operational amplifier is used as an inverting amplifier is shown in Fig. 1. For a practical op-amp let the op-amp in Fig. 1 have finite voltage gain A, finite input impedance Z_i and non zero output impedance Z_0.

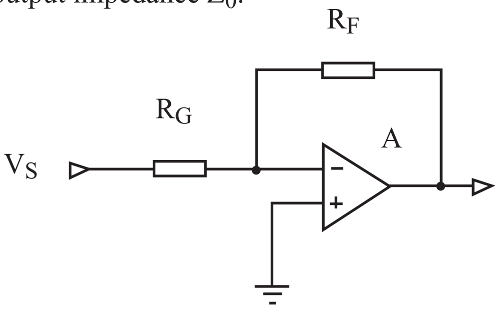


Fig. 1 Amplifier with feedback.

In this case the voltage gain expression [9]:

$$\frac{V_0}{V_s} = -\frac{R_F}{R_G}\frac{1}{1+\frac{R_F}{A}\left(\frac{1}{R_G}+Y_i+\frac{1}{R_F}\right)\frac{1+(G_F+Y_L)Z_0}{1-\frac{1}{A}G_F Z_0}} \tag{1}$$

Where, $Y_i = \frac{1}{Z_i}$, $Y_L = \frac{1}{Z_L}$, $G_F = \frac{1}{R_F}$

From the equation (1), it can be seen that there are few possible solutions of obtaining precision gain and high linearity: 1) by making op-amp gain A very large still applies in principle even with finite Y_i and Z_0 2) by making the terms $\frac{1}{R_G}+Y_i+\frac{1}{R_F}$ or $1+(G_F+Y_L)Z_0$ small. We may supplement Y_i and Z_0 by the addition of circuit admittances Y_i and Z_0 as follows:

$$Y_i \xrightarrow{\text{yields}} Y_i + Y_N \tag{2}$$

$$Z_0 \xrightarrow{\text{yields}} Z_0 + Z_N \tag{3}$$

The two terms in (1) become zero under the conditions:

$$Y_N = -\frac{1}{R_G}-\frac{1}{R_F}-Y_i = -G_G - G_F - Y_i \tag{4}$$

$$Z_N = -\frac{1}{G_F+Y_L} - Z_0 \tag{5}$$

This leads to the idea of adding compensating negative elements to the circuit in Fig. 1 such that the burden of obtaining a low error term falls not only on having a high A but is assisted by making one of the other terms in (1) small [9]. Since Z_N is a series parasitic element and Y_N is a shunt element, and a shunt negative element is easier to implement than a series one, it is opted for adding a compensating admittance in parallel with the op-amp input port. The realization of the negative element will require another op-amp with finite gain and parasitic admittances and it also have to be taken into account the input admittance of the amplifier A according to (2). In order to do this, it has to be considered the precise form for the compensating circuit that may be used [9].

2.2 Case of Finite Op-Amp Input Capacitance

Compensating negative impedance required for the circuit in Fig. 2 can be implemented using the topology of the form shown in Fig. 3 [9] that contains an op-amp. Assuming that the op-amp has finite gain A', and input admittance Y'i, the input admittance of the circuit in Fig. 3 can be obtained as [9]:

$$Y_{in} = -\frac{G_A G_C}{G_B}\frac{1-\frac{1}{A'}\left(1+\frac{G_B}{G_C}\right)+Y'_i\frac{G_B}{G_A G_C}-\frac{Y'_i}{A}\left(\frac{1}{G_A}+\frac{1}{G_C}\right)}{1+\frac{1}{A}\left(1+\frac{G_C}{G_B}+\frac{Y_i}{G_B}\right)} \tag{6}$$

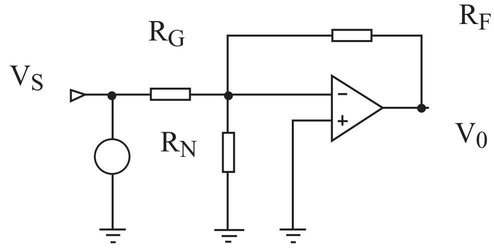


Fig. 2 Amplifier with negative admittance compensation.

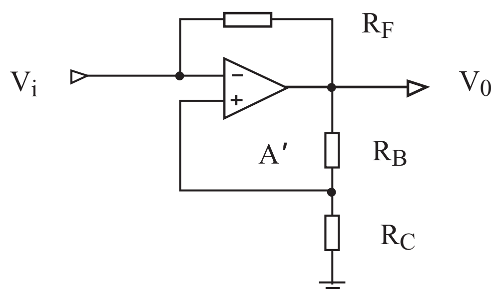


Fig. 3 Implementation of compensation impedance.

Where, $G_A=1/R_A$, $G_B=1/R_B$ and $G_C=1/R_C$.

Assuming a high frequency at which the effect of the op-amp input capacitances Y_i and Y_i' in Figures 2 and 3 may be neglected and assuming op-amp output impedance Z_0 is negligible, (1) become [9]:

$$\frac{V_0}{V_i}=-\frac{R_F}{R_G}\frac{1}{1+\frac{R_F}{A}\left(\frac{1}{R_G}+\frac{1}{R_F}+Y_N\right)} \qquad (7)$$

$$Y_{in}=-\frac{G_AG_C}{G_B}\frac{1-\frac{1}{A'}\left(1+\frac{G_B}{G_C}\right)}{1+\frac{1}{A'}\left(1+\frac{G_C}{G_B}\right)} \qquad (8)$$

Clearly, G_A may be chosen to achieve an arbitrarily low value for the precision gain and high linearity.

3. CMOS IMPLEMENTATION AND TEST

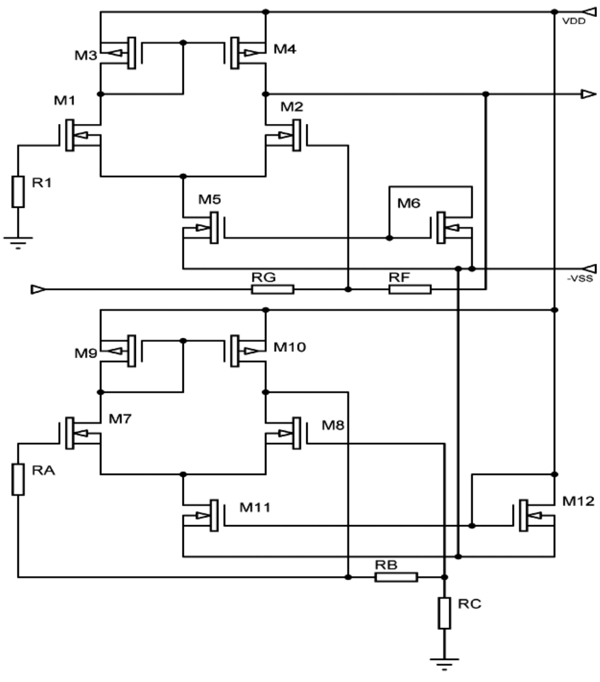


Fig. 4 Evaluation Amplifier with distortion correction.

The amplifier with compensation circuit, shown in Fig. 4 was designed to realise the circuit shown in Fig. 2. In this design, both of the main and complementary amplifiers are built around a conventional long-tail pair differential amplifier with almost the same topology. The amplifier to be linearised, in the upper part of this circuit, consists of transistor M_1, M_2 with biasing supplied by M_3 and M_4. A resistor R_1 is inserted between the non-inverting input and ground reducing the input offset voltage due to different voltage drops due to bias current, and may reduce distortion in some amplifiers. The complementary amplifier configured to give the negative resistance correction circuit comprised transistors M_7 - M_{12} and two equal resistances, R_B & R_C and R_A can be adjusted to obtain the desired negative resistance. To test various performance characteristics including the DC transfer characteristic, the frequency response and the intermodulation distortion (IMD) software simulations were carried out for this designed amplifier. Microwave Office AWR 2006 has been used for the simulation.

For DC transfer characteristic test, the parameters of R_F and R_G in the main amplifier were chosen to be 2kΩ and 500Ω respectively so the closed-loop gain is -4. The resistors values of R_B and R_C have been chosen as 1.5kΩ each. The simulations have been performed with the value of R_A=370Ω. The actual value of R_A was realised lower than calculated value ($R_N=R_F//R_G$=400Ω) due to the finite gain and finite input resistance of the complementary amplifier. The simulation results shown in Fig. 5 clearly indicate the linearity improvement of the amplifier with compensation compared with the original amplifier without compensation. Fig. 5 shows that the linear range of the DC transfer characteristic with compensation is approximately between -1.077V and +0.4479V. On the other hand the linear range of the original amplifier is approximately between -0.4927V and +0.3418V. So the linear range of the compensated amplifier is much better than the original amplifier. For a sinusoidal input signal the largest linear amplitude possible is 1.077V for the compensated amplifier, and only about 0.4927V for the original amplifier. The improvement figure for the linear range is almost 82.73%.

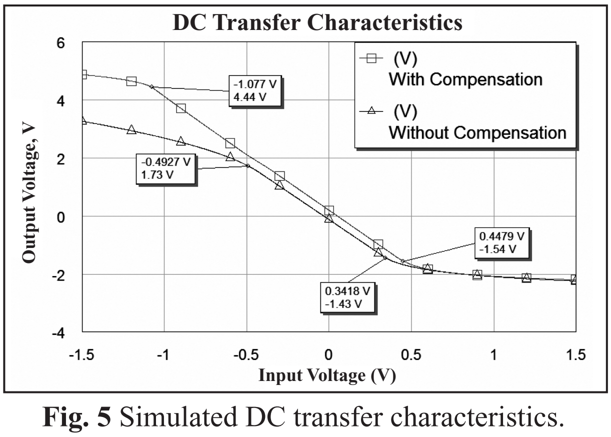


Fig. 5 Simulated DC transfer characteristics.

In practice the exact value of R_A may not be obtained due to tolerances on either the gain of the complementary amplifier or R_A or both. This can be observed by looking at the change of DC characteristic with variation of R_A alone. The simulation results show that the affect on the DC transfer characteristics can effectively be neglected when R_A has ±5% tolerance on the required value of 370Ω, which means that the allowed range for the resistance of R_A is from 360Ω to 380Ω .

Intermodulation distortion (IMD) can be evaluated by performing a two-tone test on the amplifier. In this two-tone test the two tone input signals were set to 0 dBm power at frequencies f_1=1MHz and f_2=1.001MHz. The value of R_A was 400Ω. Fig. 6 and Fig.7 show the frequency spectrum for the both amplifiers. It can be clearly observed that there was an improvement of IMD in the amplifier after compensation. The improvements in IMD and IP3 with compensation are calculated as follows:

$$\mathrm{IMD}\big|_{\mathrm{improvement}} = \mathrm{IMD}\big|_{\mathrm{with}} - \mathrm{IMD}\big|_{\mathrm{without}} = 4.63\mathrm{dB}$$

$$\mathrm{IP3}\big|_{\mathrm{improvement}} = \mathrm{IP3}\big|_{\mathrm{with}} - \mathrm{IP3}\big|_{\mathrm{without}} = 4.75\mathrm{dBm}$$

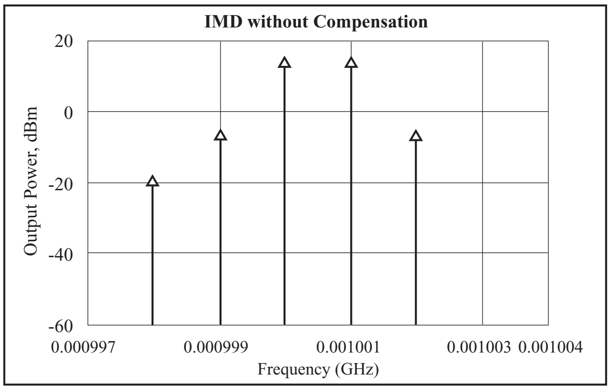


Fig. 6 Frequency spectrum of the amplifier without compensation.

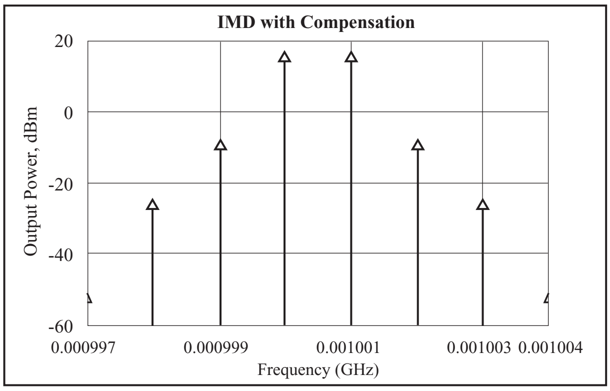


Fig. 7 Frequency spectrum of the amplifier with compensation.

In order to examine the effect of different values of R_A on the linearity, simulations have been carried out. Table 1 shows the relationship between the actual R_A and the improvement of IMD. As can be seen, significant improvements for IMD can be obtained under the value of R_A from 320Ω to 450 Ω.

It is noticed that a maximum improvement, 6.09dB, occurs when R_A=370Ω. This is due to the effect of finite input resistance which means that a lower value for R_A is required than the nominal value of R_A = R_G//R_F = 400Ω that was used to obtain the IMD and IP3 figures in Fig. 6 and Fig.7. From Table 1 it can be observed that the IMD was worse with the value of R_A less than 320Ω and slightly improved after the value of 450Ω. However, if the practical value of R_A falls below the optimum value for best IMD and IP3, it is noticed that both IMD and IP3 degrade significantly. It is suggested therefore that R_A should be chosen to be slightly higher than the optimum value due to resistor tolerance, say 373Ω ±1% in this example.

Table 1 Effect of negative resistance on IMD.

R_A (Ω)	IMD before compensation (dB)	IMD after compensation (dB)	Improvement of IMD (dB)
280	20.80	16.951	-3.849
320	20.80	23.78	2.98
370	**20.80**	**26.89**	**6.09**
400	20.80	25.43	4.63
450	20.80	23.10	2.3
550	20.80	21.642	0.842
650	20.80	21.197	0.397
750	20.80	21.019	0.219

Table 2 Effect of input and feedback resistance on bandwidth.

R_G (Ω)	R_F (Ω)	Gain	Bandwidth (MHz)
100	400	4	219.06
200	**800**	**4**	**225.67**
500	2000	4	218.40
1000	4000	4	188.90

The bandwidth of the amplifiers was varied with the changing of the input resistance (R_G) and feedback resistance (R_F) values. The bandwidth was measured at the -3dB point (half power point). From Table 2 it can be seen that the highest bandwidth (225.67MHz) was occurred with R_G=200Ω and R_F=800Ω.

The simulation showed that using the negative resistance compensation method can degrade the bandwidth of amplifier. The simulation results in Fig. 8 show that the degradation was nearly 60MHz and the gain magnitude has been improved by 6%. There were several methods for bandwidth-enhancement whose can be used to overcome the bandwidth degradation due to the introduction of negative impedance compensation. The frequency response can be considered if R_A is replaced by a parallel element $Y_A=1/R_A+j\omega C_A$. From the simulation, the optimum value for C_A was found to be 5pF and the optimum value for R_A was found 265Ω. Simulation results in Fig. 9 show that the bandwidth has been improved by 28.68MHz.

It can be observed in Fig. 9 the gain of the compensated amplifier has been reduced from -3.9 to -3.74 after using bandwidth enhancement parallel circuit. But still there was a gain improvement of 2% in the compensated amplifier.

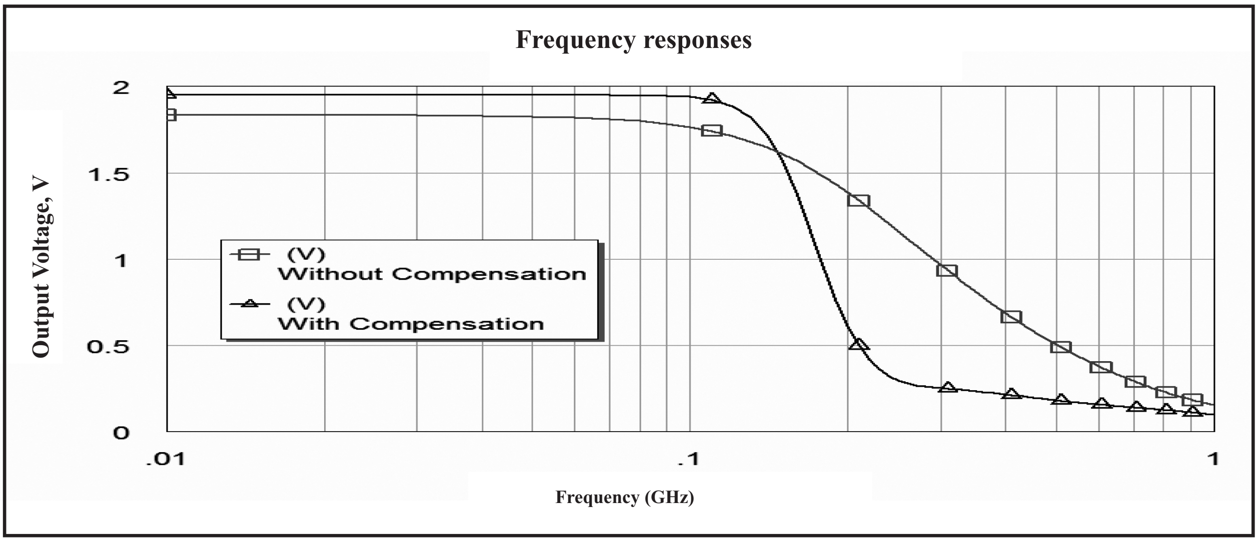


Fig. 8 Frequency responses without bandwidth-enhancement.

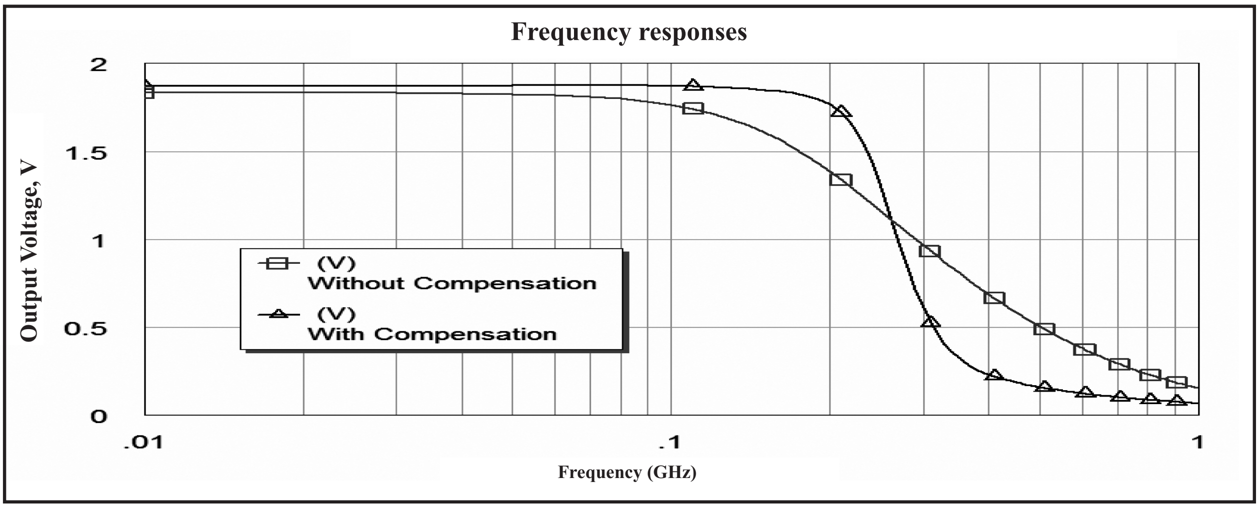


Fig. 9 Frequency responses with bandwidth-enhancement.

4. CONCLUSION

A novel linearization technique based on negative impedance compensation has been presented. In this technique, a negative impedance is used at the inverting input of the main amplifier and the negative impedance circuit is implemented using a complementary amplifier. Both amplifiers are implemented in CMOS technologies. It is seen that compensated amplifier can have some advantages over the traditional linearising methods. Firstly, the main and complementary amplifiers can be of the same design, a desirable feature in manufacture. The complementary amplifier need not be highly linear as it is only handling small signals. Secondly, this method differs from the traditional predistortion and feedforward techniques in that a high precision auxiliary amplifier is not necessary as it is only part of the negative resistance circuit.

5. ACKNOWLEDGEMENT

Authors thank Dr. Ruiheng Wu for his help and suggestions.

REFERENCES

1. D. Myer, *"Ultra linear feedforward amplifier design,"* In proceedings of the IEE MTT-S Int. Microwave Sysmp. Dig., pp.1125-1127, 1998.
2. A. Katz, A. Guida, R. Dorval, and J. Dragone, *"Input adaptive lineariser system"*, In proceedings of the IEEE MTT-S Int. Microwave Symp. Dig., pp.1499-1502, 2000.
3. Y. Hu, J.C. Mollier and J. Obregon, *"A new method of third-order intermodulation reduction in nonlinear microwaves systems"*, IEEE Trans. Microwave theory Tech., Vol. MTT-34, pp. 245-250, 1986.
4. Y. Yang and B. Kim, *"A new linear amplifier using low-frequency second-order intermodulation component feedforwarding"*, IEEE Microwave Guidew Wave Lett., Vol. 9, pp. 419-421, 1999.
5. J. K. Cavers, *"Adaptation behavior of a feedforward amplifier linearizer"*, IEEE Trans. Veh. Technol., vol. 44, pp. 31-40, 1995.
6. Y. Kim, Y. Yang, S. Kang, and B. Kim, *"Lineraization of 1.85 GHz amplifier using feedforward predistortion loop"*, In proceedings of the IEEE MTT-S Dig, Baltimore, MD, 1998.
7. J. Namiki, *"An automatically controlled predistorter for multilevel quadrature amplitude modulation"*, IEEE Trans. Commun., pp. 707-712, 1983.
8. P. B. Kenington, *"Methods to linearize RF transmitters and power amps-linearization techniques"*, Microwaves & RF, pp. 79-89, 1999.
9. R. Wu, *"Design of amplifiers with high gain accuracy and high linearity"*, In proceedings of the Circuits and Systems, MWSCAS 2007, 50^{th} Midwest Symposium on Volume, Issue, pp 269 - 272, 2007.

DEVELOPMENT OF LOCATION BASED MOBILE APPLICATION ON ANDROID PLATFORM

Mir Md. Saki Kowsar[1] and Ratna Halder[1]
[1]Dept. of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh

Abstract: Location Based Systems (LBS) are becoming popular rapidly with the wide adoption of GPS-enabled smart phones. Most of the LBS applications are used for capturing the user's location and presenting some information and driving direction to the desired location. Our Location Based Mobile Application enables the users to visualize their relative geographic location and to arrive on their desired destination point without need of human help. It uses the searching and navigation facilities of Google map. Additionally, the proposed system introduces a reminder alarm in a bid to alert the commuter when he/she passes over the particular location. However, in Bangladesh, location based systems are currently underutilized for guiding the users in commute. One of the prime reasons behind the limitation is that Google map is not properly updated for our country. That's why, to implement our system, we have to maintain a separate web-based database in order to store location related data. In this paper, we discuss the problems related to implementation of LBS System and its possible relations. Moreover, we present the various features of the proposed system that performs on Android based smartphone.

Keywords: *Location based service, GPS, Google map, Navigation, Android.*

1. INTRODUCTION

LBS are the mobile services in which the user location information is used to guide the commuters. Currently many service providers develop and deploy value-added services: providing proximity information, navigation directions, or tracking and many more utilities.

Our Location Based Mobile Application is based on android platform and Google map. Firstly, we are going to explain why we picked Android platform and Google map to be our field of interest in this project.

Android is a mobile operating system using a modified version of the Linux kernel. It is developed by the Open Handset Alliance led by Google [1]. It allows developers to write managed code in the Java language, controlling the device via Google-developed Java libraries. Google Map is a web mapping service application that offers street maps, a route planner for traveling by foot, car, bike or public transport and an urban business locator for numerous countries around the world [2].

In Bangladesh, Google map is not properly updated for all locations. That's why, we have to maintain separate web based database for storing required location data.

When a tourist or a person goes to a new country or a new location and if he is not familiar with this location, simply he will face some problem. Our Location Based Mobile Application makes the people familiar with the unknown location of an area. This application helps people to find his/her location easily through toast message and also helps to reach his/her destination using the map and routing direction. This application not only just supports static direction but also act as a mini navigation system that updates its direction dynamically which is similar to global navigation system. A reminder alert can also be set up for a specific location. Now, when the mobile barrier passes this location or nearby location then an alert message will be displayed.

Since, the application developed for Android based smartphones it has to be designed efficiently both in terms of user interface (UI) and in terms of source code.

This paper presents an overview of the system architecture as well as the development environment used to create a LBS application. Section two deals with theoretical background and related work performed in location based systems. Section three presents the application architecture and the tools used. Implementation process of our application discuss in section four. Finally, we conclude our discussion in section five with future work.

2. THEORETICAL BACKGROUND AND RELATED WORKS

2.1 GPS

Global Positioning System, a worldwide MEO satellite navigational system formed by 24 satellites orbiting the earth and their corresponding receivers on the earth. The GPS satellites continuously transmit digital radio signals that contain data on the satellites location and the exact time to the earth-bound receivers. The system enables a GPS receiver to determine its location, speed and direction [8].

2.1.1 Calculate Current Position

GPS satellites circle the earth twice a day in a very precise orbit and transmit signal information to earth. GPS receivers take this information and use triangulation to calculate the user's exact location [9]. Essentially, the GPS receiver compares the time a signal was transmitted by a satellite with the time it was received. The time difference tells the GPS receiver how far away the satellite is. Now, with distance measurements from a few more satellites, the receiver can determine the user's position and display it on the unit's electronic map. A GPS receiver must be locked on to the signal of at least three satellites to calculate a 2D position and track movement [10]. When three satellites are measured simultaneously, the intersection of the three spheres reveals the location of the receiver as shown in Fig. 1. With four or more satellites in view, the receiver can determine the user's 3D position (latitude, longitude and altitude). Once the user's position has been determined, the GPS unit can calculate other information, such as speed, bearing, track, trip distance, distance to destination, sunrise and sunset time and more.

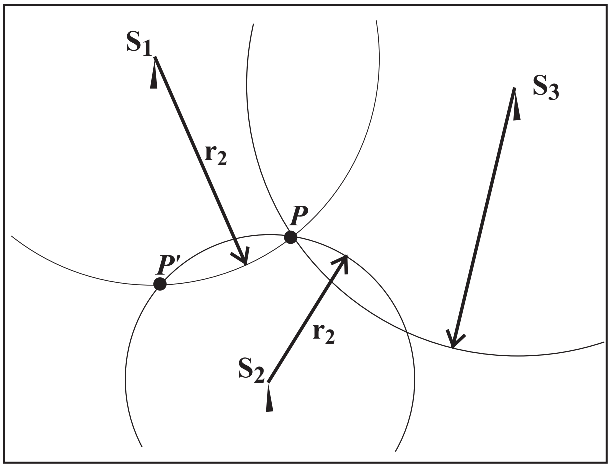


Fig. 1 GPS positioning system.

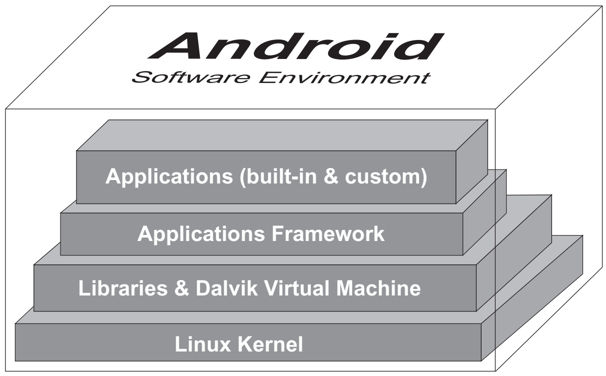


Fig. 2 Android software environment.

2.2 Android Platform

Android is a software environment and not a hardware platform, which includes an OS, built on Linux kernel-based OS hosting the Dalvik virtual machine (Fig. 2). The Dalvik virtual machine runs Android applications as instances of the virtual machine. Android contains a rich user interface, application framework, Java class libraries and multimedia support [7]. Android also comes with built-in applications containing features such as short message service functionality, phone capabilities and an address book.

2.3 Coordinate Format

The coordinates that are retrieved from the GPS can be represented as one of the following formats: DD°MM'SS.SS, or DD.DD°, DD°MM.MM where D is for Degrees, M is for Minutes and S for Seconds [11].

2.4 Calculation Distance between Two Coordinate

Haversine formula is used to calculate the great-circle distance between two points - that is, the shortest distance over the earth's surface - giving an 'as-the-crow-flies' distance between the points. If the distance between point A ($Long_1$, Lat_2) and the point B ($Long_1$, Lat_2) wanted to be calculated, then the distance in meters is [12]:

R = earth's radius (mean radius = 6,371km)

$\Delta lat = lat_2 - lat_1$

$\Delta long = long_2 - long_1$

$a = \sin^2(\Delta lat/2) + \cos(lat_1) * \cos(lat_2) * \sin^2(\Delta long/2)$

$c = 2 * \operatorname{atan2}(\sqrt{a}, \sqrt{(1-a)})$

$d = R * c$

Where d: the distance in meters, lat1: Latitude of point A, long1: Longitude of point A, lat2: Latitude of point B, long2: Longitude of point B, R: the radius of the earth in meters

2.5 Android Location API

Android support Location based service APIs. Location service allows finding out the device current location. The application can ask for periodic update of the device location [13]. The application can also register a intent receiver for proximity alerts like when the device is entering and existing from an area of given longitude, latitude and radius.

2.6 Google Map API

Android also provides an API to access the Google MAPs. So with the help of the Google MAPs and the location APIs the application can show required places to the user on the MAP [13]. Android defines a package called com.google.android.maps. The package contains classes related to rendering, controlling and overlaying information on the Google maps on the android devices.

3. RELATED WORKS

There are many research works already exist for tourist and student with map technology. But most of these researches have some limitation, for this they might not be suitable for the developing countries. In our project "Development of Location Based Mobile Application on Android Platform", we have tried to overcome these problems with some modification so that user can use this application easily in our country. Some existing researches and their limitation are discussed in below.

a. One of the existing researches is Rapid Prototyping of a Mobile Location-Based Tour Guide [3]. This project implements a location-based tour guide for the University of Ontario Institute of Technology campus. The guide would potentially be used to supplement or replace existing human tour guides by showing the user important features of the campus along with relevant associated data. This would free visitors of the campus to arrive on their own schedule without the need to occupy human resources.

   This application shows some important feature as a tourist guide but no navigation facility is available in this application. Also the user needs to know the name of the specific location to find his destination point. In our application, user does not face any problem when the location is fully unknown to him and navigation facility is available and then provides the warning during passing the destination location. Moreover, our application is not bounded within any limited location such as university campus.

b. Another existing research is Location-Based Service-A mobile implementation approach [4]. This project has developed for helping the student to find departments, library or any location on the campus. They just identify the current location and destination location and then add markers on the both the coordinate points.

c. Another mentionable existing project is Sdsumap in Sdsubuddy Android Based Application [5]. SDSUBuddy application is a complete package of the university directory which also includes a little navigation facility. This navigation allows the user to show routing direction using SDSU map.

This is only suitable for the student of San Diego State University (SDSU) to identify departments or any location in his/her campus and display some information about the location. But they have used their own campus map. This application use Google search facility to search nearby location which is not possible for Bangladesh.

4. DEVELOPMENT OF APPLICATION PLATFORM

Our application platform depends on location server in the mobile communication networks. It processes location information by using global positioning system and database. After receiving latitude and longitude coordinates, those are used to identify location on Google map which can be easily understood by mobile users. This application also needs android based smartphone. Our LBS system use toast message and add a marker in current location to notify mobile users. Some location related coordinate points and information is stored in the website. This database server helps to improve the location positioning effect and to draw routing path. The information is stored in the website as a XML file. XML is suitable to link database by internet. For this reason, it is easy to develop different location service application based on actual needs. In addition, it provides location information service and positioning function and so on. There are many techniques to retrieve information from webstored XML file. In our application SAXParser technique is used to obtain information from XML file. Our application can find out maximum three nearest location using haversine formula. Different markers are added to different founded location. These markers and location information are used to help mobile users to identify desire location. Also an alert is used to notify the user. The architecture of our Location Based Mobile application platform is shown in Fig. 3.

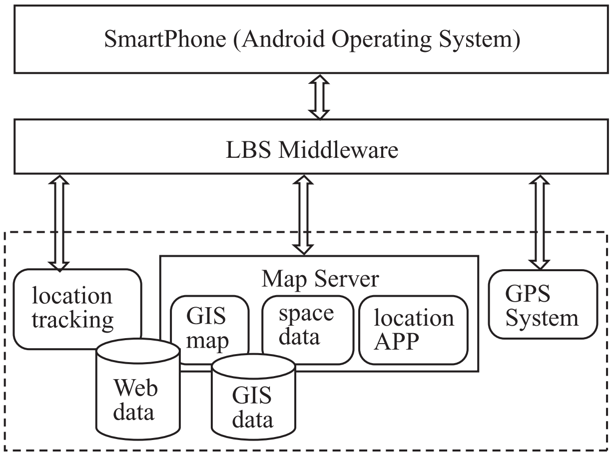


Fig. 3 Architecture of our application platform.

The architecture of our Location Based Mobile application platform is shown in Fig. 3. This architecture contains a number of components including maps and Geographic Information System (GIS) information, GPS services and LBS application-specific subcomponents.

The map server system is the core of LBS application platform. In our application platform, we use Google map. GPS performs location collection to get a latitude and longitude for a specific user. Depending on the technology, this component may be accessed via the LBS Middleware or directly to the mobile phone. Web data means web stored database which is already mentioned. Location tracking system allows a user's route to be determined from current location point to destination point. LBS Middleware helps to provide a consistent interface to LBS applications.

Our routing process is geographic based routing principle that relies on geographic position. In our application, we have used Google map routing system information and it uses an http protocol. Google map routing planning URL structure is "http://maps.googleapis.com/maps/api/directions/output?parameters". Certain parameters are required while others are optional. Required parameters are origin and destination point. Optional parameters are mode, waypoints, alternatives, unites etc. Fig. 4 displays the main steps to draw a route on the map displayed on the mobile terminal.

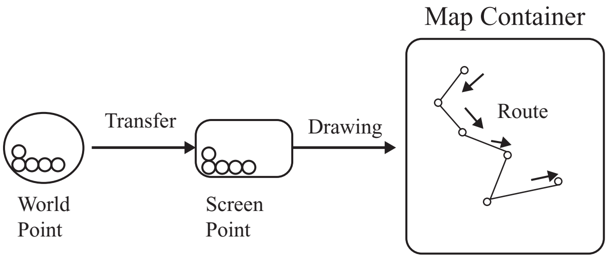


Fig. 4 Drawing route on map of mobile phone.

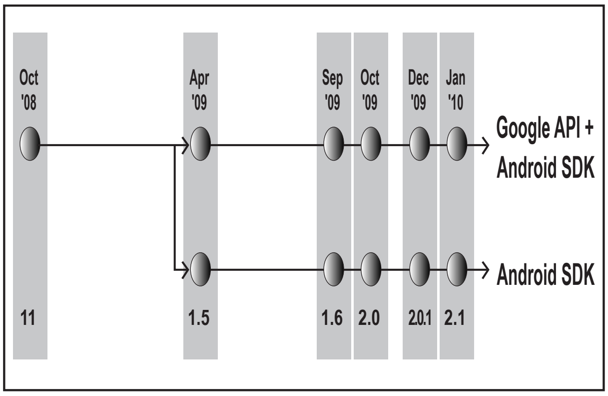


Fig. 5 A timeline of Google API and Android SDK.

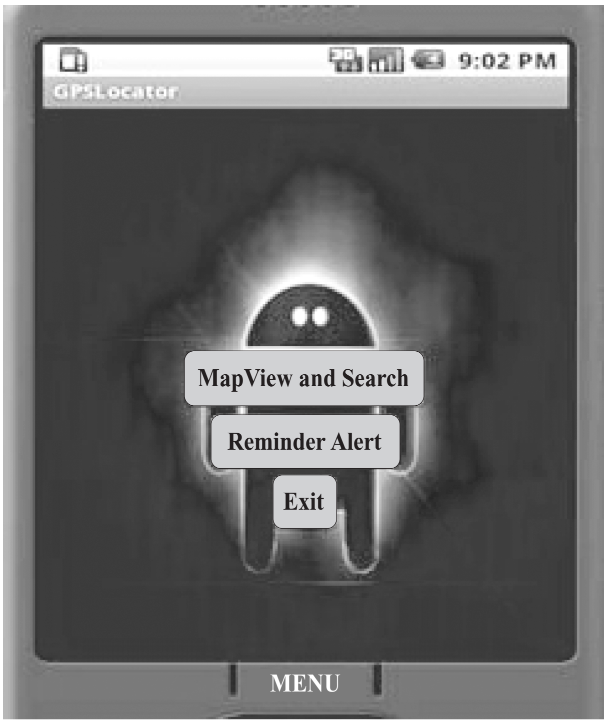


Fig. 6 Main menu of our application.

5. IMPLEMENTATION

5.1 Application Environment

Our location based application is based on Google API and android SDK (Fig. 5). So, this application runs on android supported mobile phone. Also mobile phone must be GPS enable.

5.2 Application Overview

Our application has three option i) MapView and Search ii) Reminder iii) Exit which is shown in Fig. 6

1) Our lbs application used Google maps high-resolution satellite images which a user can understand easily. The MapView service shows a map of only the area immediately surrounding the place where the user is a currently located. As shown in the Fig. 7, the user's location is indicated by a yellow marker in the middle of the screen, and the map auto-scrolls and refocuses to reflect the user's current location in real-time. This provides an up-to-date picture of the user's current location context at all times. It has built in zoom control as well as two finger zoom to make easier view of the map see (Fig. 7). Users can move the map to anywhere using touch functionality.

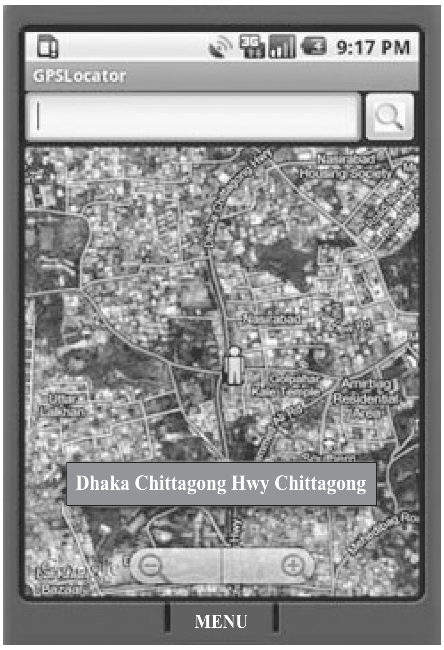


Fig.7 Current location detection of the mobile user.

2) Search nearby location features a powerful integrated search option that allows users to find specific places such restaurants, bookstores, mosque, library, hotel, market, coffee-shops, gas-stations, etc, near their current geographic location using keywords or specific location name (Fig. 8).

Fig. 8 Search nearby market.

9a. Routing path for nearest founded location.

9b. Another figure for routing path.

Fig. 9 Routing Path.

It shows maximum three most nearest places around current location. The results are displayed as a textual message, showing the name, street address, and direction from the current location of each place found. Also different color markets are added on the found location. Each different color defines the places respectively to their location starting from nearest. The user would usually want to know something further about what the each marker signify. Then, location related information will be displayed on the screen depending on which marker is tapped.

3) Another important feature of the application is routing. A routing path is drawn from user own location to the most nearest found location which helps the user to reach his destination location easily see (Fig. 9a and 9b). This routing path act as a mini navigation system that updates it direction dynamically based on location changed.

4) Generally reminders are based on time. However, time is not sufficient to capture the context in which the user wants to be reminded. A location aware reminder system can allow users to set reminders based on where they are located, and whether they are entering or exiting a particular location.

Fig. 10 Reminder system.

In our location based application, a reminder is automatically added to the most nearest location which is obtained by search.

User can also add a reminder on his/her own location or any other location see (Fig. 10).

6. CONCLUSION

The main objective of our application is to explore different aspects of location-based services. Our "location based mobile application on android platform" is implemented in Bangladesh which is act as location guider. But there had been number of challenges while developing this application on android framework.

Within this paper we have given an overview about some of the most important processes and services with integrated searching and driving direction facility which is necessary to build location based services for nomadic users.

But our application works in open space areas only since it depends on GPS. It could also be adapted to work indoors with an appropriate method of geolocation. In future, we will try to overcome this limitation. Also we will try to add some important feature like highly advanced voice based processing, whether to search information or to find out other details we can add this voice processor.

REFERENCE

1. Last day of access (November 20, 2011). [Online]. Available: http://en.wikipedia.org /wiki/Android_operating_system
2. Last day of access (November 25, 2011). [Online]. Available: http://en.wikipedia.org/ wiki/Google_Maps
3. M.Mior, *" Rapid Prototyping of a Mobile Location-Based Tour Guide"* Undergraduate Thesis Report, Faculty of Science, University of Ontario Institute of Technology, December 12, 2008.
4. L.Li, *"Location Based Service"* Undergraduate Thesis Report, Faculty of Applied Science, University of Otage, November, 2006.
5. H.Jain, *"Sdsumap in Sdsubuddy Android Based Application"* Thesis Report for Master of Science in Computer Science, March, 2011.
6. Last day of access (October 20, 2011) [Online]. Available:http://www.ibm.com/ developerworks/xml/library/x-xmlphp2/ index.html
7. Last day of access (October 25, 2011). [Online]. Available: http://developer.android. com/guide/basics/what-is android. html
8. Last day of access (December 12, 2011). [Online]. Available: http://www.webopedia. com/TER M/G/GPS.html
9. Last day of access (November 26, 2011). [Online]. Available: http://www8.garmin. com/aboutGPS/
10. A. Devaraju, *"Mobile Location Based Service"*, Undergraduate Thesis, September 2004.
11. Rick Broida, "*How to Do Everything with Your GPS*", McGraw Hill/ Osborne, 2nd edition, 2004.
12. Last day of access (December 12, 2011). [Online]. Available: http://www.jaimerios.com /?p=39
13. Last day of access (October 20, 2011). [Online]. Available: http://www.androidcom petencycenter.com/2009/01/android-location - api/

Reviewers Acknowledgement

We would like to thank the following reviewers who have taken part in the peer-review process of the Computer Science and Engineering Research Journal (CSERJ).

Dr. Sajal Chandra Banik
Faculty Member
Department of Mechanical Engineering, CUET

Dr. Quazi Delwar Hossain
Faculty Member
Department of Electrical & Electronic Engineering, CUET

Dr. Muhammad Ahsan Ullah
Faculty Member
Department of Electrical & Electronic Engineering, CUET

Dr. Asaduzzaman
Faculty Member
Department of Computer Science & Engineering, CUET

Abu Hasnat Mohammad Ashfak Habib
Faculty Member
Department of Computer Science & Engineering, CUET

Md. Monjur-Ul-Hasan
Faculty Member
Department of Computer Science & Engineering, CUET

Thomas Chowdhury
Faculty Member
Department of Computer Science & Engineering, CUET

Nipa Chowdhury
Faculty Member
Department of Computer Science & Engineering, CUET

Kazi Zakia Sultana
Faculty Member
Department of Computer Science & Engineering, CUET

Dr. Kaushik Deb
Faculty Member
Department of Computer Science & Engineering, CUET

And also thanks for active co-operation to **Priyam Biswas,** Faculty member,
Department of Computer Science & Engineering, CUET

Important Dates:

Full-paper submission : August 31, 2012
Notification of acceptance : October 16, 2012
Camera-ready submission : November 01, 2012

Submission Guideline:

* The contribution must be written in English and describe original research, which is not published nor currently under review by other journals or conferences.
* Paper Format: Please use our template (available in our journal's website).
* Authors will receive final proofs and carefully correct and return those as instructed. No new matters should be inserted in the text of final proof without the intimation to the editorial board.
* Manuscripts should be submitted electronically directly to the associate editor **cserj@cuet.ac.bd**, **debkaushik99@gmail.com** in pdf and doc. format.

Call for Papers

Vol. 8 (2012)
Computer Science & Engineering Research Journal (CSERJ)
ISSN: 1990-4010

http://www.cuet.ac.bd/cserj

Computer Engineering has been the key force behind the recent technological development in the world. The Computer Science and Engineering Research Journal (ISSN: 1990-4010) aims to publish the latest research results, current practices, future trends and interchange ideas with a view to solving and locating different problems in all aspects of Computer Science and Engineering and its related fields. The journal aims to publish original research papers after being properly reviewed by the respective experts. This journal is published every year by the Department of Computer Science and Engineering, ***Chittagong University of Engineering and Technology (CUET).***

Scope:

CSERJ seeks papers on all aspects of Computer Science and Engineering, Major areas of interest include, but are not limited to:

Artificial Intelligence and Machine Learning.
Biomedical Engineering.
Biometrics and Bioinformatics.
Computer Vision and Pattern Recognition.
Computer Graphics and Multimedia.
Computer Network and Data Communication.
Cryptography, System and Information Security.
Database Management System.
Digital Signal Processing and Image Processing.
Digital System and Logic Design.
Distributed and Parallel Processing.
Data Mining
E-commerce and E-Governance.
Fuzzy Logic and Neural Network.
Graph Theory and Computing.
Human-Computer Interaction.
Internet, Web Application and GIS.
Information System.
Microprocessor-Based Design.
Machine Translation and Computational Linguistics.
Natural Language Processing.
Optical Computing and Software Engineering.
Robotics and Mechatronics.
VLSI Design and Fabrication.
Wireless, Optical and Mobile Communication.

Correspondence:
Associate editor, Computer Science and Engineering Research Journal (CSERJ),
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong-4349, Bangladesh.
E-mail: cserj@cuet.ac.bd, debkaushik99@gmail.com
Tel: 880-31-723336, Fax: 880-31-714910

ISSN: 1990-4010

Computer Science and Engineering Research Journal

Vol. 07 (2010-2011)

Contents **Page**

Published By

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET), Bangladesh

E-mail: cserj@cuet.ac.bd, Tel: 880-31-723336, Fax: 880-31-714910, Web: http://www.cuet.ac.bd/cserj