SUPPORT VECTOR MACHINE SCHEDULER for UNCERTAIN COMMUNICATION DEMAND in COMPUTATIONAL GRID

LAKSHMIPRIYA.C[1] ,SATHYALAKSHMI.S[2]

SCHOOL OF COMPUTING SCIENCES, HINDUSTAN UNIVERSITY, CHENNAI

Abstract— **Grid has evolved as heterogeneous resource sharing environment. The resources are from different management domain. Finding a suitable resource for a given job is called scheduling. The highly dynamic and heterogeneous nature of grid makes scheduling a big challenge. To achieve maximum resource utilization and minimum makespan grid needs effective and efficient scheduling algorithm. Grid scheduling normally considers the availability of the grid resources and application demands. These demands are not known in advance for the highly dependent application and require more data transfer during execution. Without proper consideration of these demands at the time of scheduling leads to poor system performance and increase in makespan of the application. This unpredictable data transfer is called Uncertainty in communication. IPDT-FUZZY uses some parameters to find the uncertainty in given applications. When the degree of uncertainty is less, the IPDT-FUZZY scheduler behaviour is similar to the classical scheduler. In this paper Support Vector Machine (SVM) based scheduler is proposed to consider the uncertainty in communication demand, and also to compare its performance with other classical schedulers.**

Keywords—— **Uncertainty, Communication demand, IPDT-FUZZY, SVM, QOI, IP-FULL-FUZZY,ILP-FUZZY, Makespan, DAG, GIS, RMS**

I. INTRODUCTION

Grid is a shared environment which is implemented by standard based service infrastructure that supports resource sharing [1]. Here the resources are from different management domain so finding a suitable resource for given job is challenging one. Grid scheduling is important to efficient management of grid resources.

Information about the status of available resources is very important for a grid scheduler to make a proper schedule. Grid scheduler must select the suitable resources from the GIS (Grid Information Service).

Basically grid scheduler receives application from grid users select the suitable resources based on the application requirement. Basically the resource information is present in the GIS [1]. Resource utilization is more important in grid and utilization is measured only when the tasks are properly scheduled. To utilize the resources in grid the given   application is divided in to smaller pieces called "Tasks" [2].

Then these tasks are allocated to resources based on the application demands and resource availability. Grid has many types of scheduler for resource allocation to application. Once tasks are allocated to resources according to the schedule they are executed until all have been finished their execution. Since the resources are from multiple organization scheduler will take more time to execute the task so the makespan of the task will be increased. In grid computers are connected by shared communication links and the time required to transfer data across them significantly affect the applications execution time.

Basically uncertainty is related to both communication and computational demand. Uncertainty is the lack of assurance of definite value. Uncertainties of schedule can result from many reasons such as resource availability, communication demand between the dependent tasks and computational demands. Such uncertainty can also affect the decision making process on the dynamic environment [4]. Uncertainty in communication related to data transfer between dependent tasks.

This paper proposes a SVM (Support Vector Machine) scheduler to address uncertainty during the communication in grid [6]. SVM is a binary

classifier. Here SVM performance will be compared with other scheduler like FIFO, Priority and IPDT-Fuzzy.

II. RELATED WORK

ILP-FUZZY proposed in [2] which consider the uncertainty of computation and communication demands. These demands of application can cause unpredicted performance and increase the makespan. The author introduced a scheduler based on fuzzy optimization for dealing with uncertainties of the demands of applications. Here the scheduler accepts as input a set of dependent tasks described by the DAG (Direct Acyclic Graph). When the task dependency is increasing the unpredicted amount of data to be exchanged among tasks. This scheduler used to find the application and speedup and execution time. Fuzzy approach is attractive if schedulers are designed to handle a high level of uncertainty.

IP-FULL-FUZZY explained in [3] which consider the uncertainty of both application demand and resource availability during the scheduling. The IP-FULL-FUZZY scheduler is based on a fuzzy optimization formulation. The effectiveness of the proposed scheduler is compared to that of three non-fuzzy schedulers, as well as to that of a fuzzy scheduler which considers only uncertainties of application demands. The performance of IP-FULL-FUZZY was compared to that of a scheduler which considers only application demand uncertainties and that of a non-fuzzy scheduler (RANDOM). IPFULL-FUZZY scheduler is robust for both types of uncertainty, producing speedup values 33% and 21% higher than those given by IP-APP-FUZZY and RANDOM.

IPDT-FUZZY described in [4] which consider the communication demand between the dependent tasks. It follows a proactive method so it does not consider the overhead of task migration and re-scheduling. IPDT-FUZZY uses some parameters to find the uncertainty in given application such as QOI (Quality of Index). When QOI value is equal to 100% the application does not contain any uncertainty.

SVM (Support Vector Machine) explained in [5] for task scheduling. It is machine learning approach for task scheduling in heterogeneous computational grids. SVM scheduler is a multiclass scheduler it maps any tasks to a resource present in a machine and the queue of tasks waiting to be dispatched. It uses the dynamic learner to adapt the dynamic and heterogeneous nature of grid application.

A. Existing System

In 2008 ILP-Fuzzy describes

➤ Application demand can cause unpredicted performance which leads to ineffective scheduler.

➤ In proposed system fuzzy optimization is used to deal with uncertainty.

➤ Here Schedulers can be executed parallel to compute the scheduling.

➤ Linear programming method reduces the number of variables needed in fuzzy formulation.

In 2010 IP-Full-Fuzzy observes

➤ Avoid the task migration which is necessaries to improve the makespan and also avoid the misleading information about the resources.

➤ Fuzzy Triangular method was used to express the IP-Fuzzy.

➤ It is robust for both application demand and resource availability.

➤ It will reduce the network utilization when the level of uncertainty in data transfer demand is high.

In 2011 IPDT-FUZZY describes

➤ IPDT-Fuzzy scheduler considers the demand of grid application with   uncertainties.

➤ It was used fuzzy optimization. Fuzzy numbers are used to express the application demand.

III. SYSTEM ARCHITECTURE

A. Support Vector Machine

This paper uses the technique called SVM (Support Vector machine) it is a binary classifier [5].SVM is used to schedule the task in heterogeneous computational grid.SVM consists of two components: Training set and dynamic learner.

B. Training Set

The incoming tasks are stored in a training set.

C. Dynamic Learner

Dynamic learner is the machine learning technique. It can easily adapt the dynamic and heterogeneous nature of grid application, because in grid the applications are coming from different management domain. In this paper which is used to check uncertainties of grid application.

D. Basic Architecture
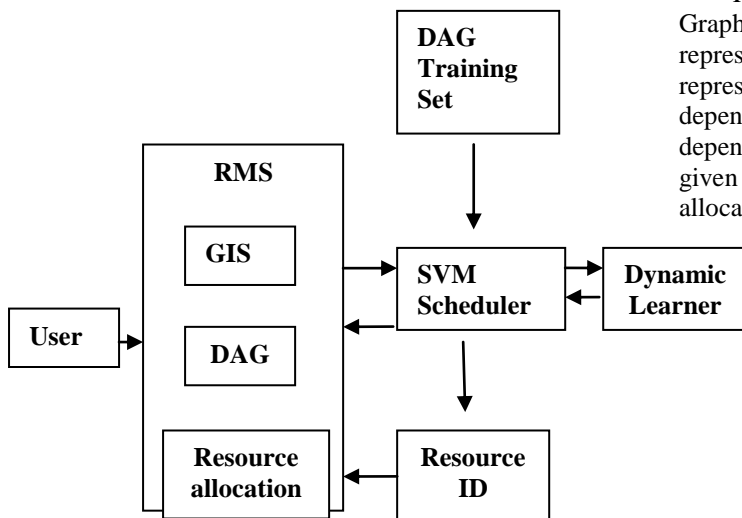
Figure1shows the basic architecture of SVM Scheduler.

Fig.1. SVM Scheduler Architecture

SVM is a machine learning technique which is used to find uncertainty in communication demand. User sends their application to the RMS. Management service will create the DAG of tasks and also get resource information from the GIS. Now RMS sends the tasks to SVM scheduler. Scheduler will compare the incoming DAG with already present DAG in a training set. If there is a match, scheduler directly allocates the task to resources otherwise the incoming DAG is sent to the dynamic learner. This learner will check the task uncertainty then it will send to the scheduler and includes in to training set. Finally scheduler allocates the resources to the task and this information is sent to RMS then RMS send task resource allocation list to user.

1) Resource Management System:

RMS provides resource availability information and also it creates the DAG which is given by user. RMS consist of

GIS (Grid Information Service)
DAG (Direct Acyclic Graph)

RMS provides the resource availability and also it divides the given application in to number of tasks.

2) Grid Information Service:

Grid Scheduler get the resource information from a Grid Information Service [1], it will collect the information from heterogeneous domain. Each domain consists of one resource manager, local applications and resources. Resource manager gives the first preference to local applications to utilize the resources.

3) DAG (Direct Acyclic Graph):

Here the scheduler accepts the DAG (Direct Acyclic Graph) as a input. DAG contains the Node and Edges. Node represents the number of instruction to be executed [2]. Edge represents the number of bits to be exchanged by two dependent tasks. Host capacity= $(\text{inst/unit of time})^{-1}$.Task dependency directly proportional to uncertainty. In the Fig.2. given job is decomposed in to 11 tasks. Now each task is allocated to separate resource to complete, the execution.
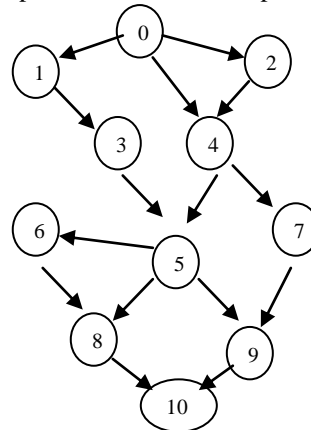

Fig.2. DAG Representation

E. SVM Scheduler

SVM (Support Vector machine) is a binary classifier [5].SVM is used to schedule the task in heterogeneous computational grid. SVM consist of two components:

DAG Training Set
Dynamic learner

SVM Scheduler receives the input (DAG) from the RMS (Resource Management Service) and received DAG is compared with the already trained DAG in Training Set. Normally Training set consist of number of trained DAGs these DAGs may have high dependency. If the received DAG is not similar to the DAGs trained set, then it is sent to the dynamic learner and it will find the task dependency then the task is sent to the SVM scheduler.

Resource id defines the number of available resources in grid environment. Now the SVM scheduler allocates the resources to the task. SVM Scheduler is a list which provides information about the host on which each task should be executed, the starting time of the task, and the time when data transfer should take place. Finally the task and resource mapping is send to user.

1)   Estimation of the job completion time:

To compute the completion time the task is assigned to particular resource in grid.

2)   Estimation of the Data transfer between tasks:

Before going to estimate the completion time, have to find data transfer rate to be found between the dependent task. Edge will connect the two nodes and its weight represents number of bits to be exchanged by two dependent tasks, and the labels of the nodes represents the quality of instruction to be processed. For example to compute the time taken to transfer data from t0 to task t1, located respectively in the host h0 and h1.

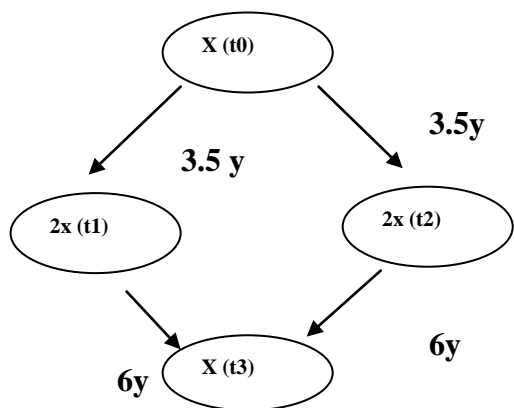Fig.3. DAG vertices and edges

$$3.5*8/y=28 \text{ units of time}$$

F. Grid Resources

Fig.4. describes the available resource in grid [5]. Main aim is to allocate the given task to resource. Here node represents hosts, with labels expressing the inverse of host capacity (instruction/unit of time)$^{-1}$ and edge represents available bandwidth. The dotted line represents existence of other resource in the grid.
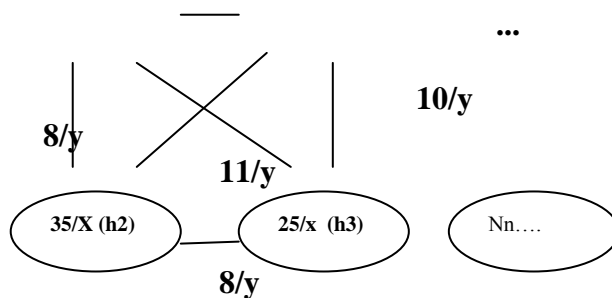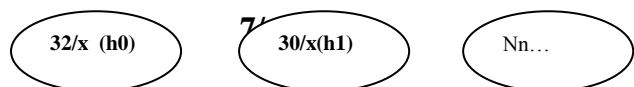

Fig.4. Grid Resources

1)   Estimating processing time of a task:

Now the task is allocated to resource. For example, to compute the processing time of task t0 on host h1

$$X*30/X=30 \text{ units of time}$$

TABLE I
USER GET THE INFORMATION FROM SVM

Resource id	Tasks	Completion time
R1	T0	30 sec
R2	T1	35 sec
R3	T2	37 sec
....	....	...
Rn	Tn	...

The above table describes the resource and task allocation and job completion time for task. Here task T0 is assigned to resource R1. Resource R1 completes the task T0 at 30 secs.

IV. CONCLUSION

This paper proposes a SVM scheduler for uncertain communication demand in heterogeneous grid. SVM is a binary classifier and it includes machine learning technique for dynamic and heterogeneous environment. Finally the performance of SVM is compared with other scheduling techniques like FIFO (First Come First Out), Priority and other schedulers in grid. This performance analysis will check whether the SVM scheduler is suitable for uncertain communication demand.

V. REFERENCES

[1] Schopf J.M," Ten actions when grid scheduling- the user as

    a grid scheduler," Grid Resource Management: State of

    the Art and Future Trends,KAP,2003,ch.2,pp15-23.

[2]Batista.D.M,DrummondA.C,andFonsecaN.L.S.,."Schedulig grid tasks under uncertain demands," in *Proc. ACM SAC*, 2008, pp.2041-2045.

[3] Batista.D.M, Nelson L.S. da Fonseca Robust scheduler for grid networks under uncertainties of both application demands and resourceavailability 2010, CompuNetworks.

[4] Pedram Keyani Neal Sample Gio Wiederhold Scheduling Under Uncertainty: Planning for the Ubiquitous Grid.

[5] Batista.D.M, A.C.Drummond, and N.L.S.da Fonseca SchedulingGridtasksin Face of UncertainCommunication Demand

[6] Yong-Won Park, Kenan Casey and Sanjeev Baskiyar,"A Novel Adaptive Support Vector Machine Based Task Scheduling,".