International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

SCRUTINIZING GRID RESOURCES AND FORECASTING BASED ON MACHINE LEARNING IN A HOOKED-UP SYSTEM

AMUDHA BHARATHI. B[1], AISHWARYA. R[2]

1 PG Student, ME Computer and Communication, Sri Sai Ram Engineering College,Chennai 44.
1 PG Student, ME Computer and Communication, Sri Sai Ram Engineering College,Chennai 44.

amudha.bhaskaran@gmail.com ,
ramanathan.aishwarya@gmail.com

Abstract

Grid computing is a term referring to the combination of computer resources from multiple administrative domains to reach a common goal. The two types of functions available in the grid Computing are job scheduling and Resource allocation. The main purpose of using Grid computing in a hooked-up system to utilize the system resources in an efficient manner, improving it's performance and also it eliminate bottleneck, diagnosing fault and maintaining dynamic load balancing. Job scheduling and Resource allocation are mainly depend on sufficient information of available resources. The process of monitoring means it response for the state of execution, distribution of load and Prediction in the sense it focuses on the running track of resources in grid system. The system design principle describe about the Responsiveness and Robustness, Extensibility and Efficiency. In Existing system it scrutinize and forecast the grid resources alone. In Proposed system it will scrutinize and forecast the Grid task, classification using Genetic Algorithm(GA) and Particle Swarm Optimization(PSO) based on machine learning.

Keywords-- **GA, PSO**

I. Introduction

Grid aggregates heterogeneous resources distributed across Internet, regardless of differences between resources such as platform, hardware, software, architecture, language, and geographical location. Such resources, which include computing, storage, data, communication bandwidth resources and other resources, are combined dynamically to form high performance computing capability of solving problems in large-scale applications. We present the design and implementation of an infrastructure that enables monitoring of resources, services, and applications in a computational grid and provides a toolkit to help manage these entities when faults occur. This infrastructure builds on three basic monitoring components: sensors to perform measurements, actuators to perform actions, and an event service to communicate events between remote processes. Thus timely acquiring resource status information is of great importance in ensuring overall performance of grid computing. There are mainly two mechanisms for acquiring information of grid resources: grid resource monitoring and grid resource prediction.

Grid resource state monitoring cares about the running state, distribution, load and malfunction of resources in grid system by means of monitoring strategies. Grid resource state prediction focuses on the variation trend and running track of resources in grid system by means of modeling and analyzing historical monitoring data. Historical information generated by monitoring and future variation generated by prediction are combined to gather to feed grid system for analyzing performance, eliminating bottleneck, diagnosing fault, and maintaining dynamic load balancing, thus to help grid users obtain desired computing results by efficiently utilizing system resources in terms of minimized cost, maximized performance or tradeoffs between cost and performance. Performance monitoring in Computational Grid setting is widely recognized to be an essential capability. There are a variety of systems available for taking measurements to this end, particularly for network performance. For the Grid, this information is used primarily to optimize application execution.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Applications a redirected (either by a human user or an automatic scheduler) to use the resource (network, CPU, memory system, etc.) that exhibits the best measured performance. Frequently, monitor data gathered from these tools is used as a prediction of future performance. That is, an observation of past performance implies that future performance levels will be similar.

II. PREDICTION PATTERN

The prediction pattern is schematically shown in Fig.1.A historical data set is divided into three parts: training, validation and test sets. The training set is used to build prediction model, which is optimized using validation set and evaluated using test set. The model takes historical data as input and generates prediction for future variation. Our research goal is to design a distributed system that seamlessly integrates various cooperating components to achieve high performance in grid resource monitoring and prediction.

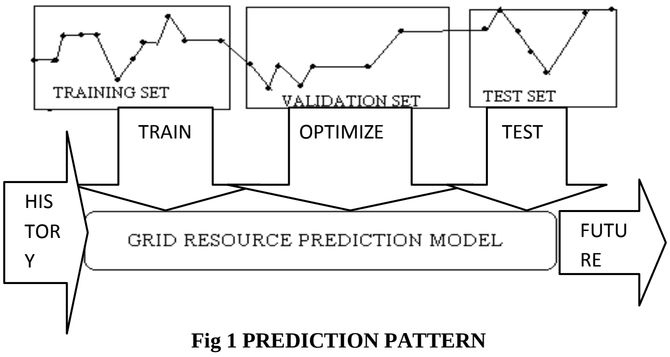

Fig 1 PREDICTION PATTERN

III. SYSTEM ARCHITECTURE DESIGN

In this section, we present the architecture of our grid resource monitoring and prediction system. We first introduce the principles used in our design. Then, we illustrate the service distribution and work flow of our system in details.

3.1 System design principles

The system design principle describe about the Responsiveness and Robustness, Extensibility and Efficiency.

Responsiveness and robustness

In our system, resource sensors and prediction models are periodically executed to generate up-to-date information for users. Function independent and starlike distribution are introduced in our system. Since grid resource states vary dynamically, the information monitored or predicted has to be timely updated, to guarantee online reflection of resource conditions.

Modularity and extensibility

To achieve tight cohesion as well as loose coupling, the components embedded in our system have independent functions and can be integrated into most grid computing environments as an independent subsystem.

Efficiency

The fundamental function of Grid System is Executing jobs, so embedded monitoring or prediction components should minimize overhead to guarantee grid's normal service. The resource sensors on computing nodes is inevitable.

Transperency

The Users who are using the Grid does not need traversal of all the nodes. In Our System We design a uniform and friendly interface component for accessing the information monitored or predicted.

3.2 Service distribution

We propose to build the whole system that consists of two subsystems: resource monitoring subsystem inside the computing environment, and resource state prediction subsystem outside the computing environment.

Monitoring service

It manages resource sensors and generates resource monitoring data.

Prediction service and evaluation service

Prediction service and evaluation service are deployed on each prediction node. Corresponding to a prediction request customized by a grid user, one prediction service takes charge of the whole prediction procedure and manages resource prediction models.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Information service

Information service is deployed on information node. Two types of mechanism are defined for information acquisition: localregister and group register. Local register timely collects information, from resource sensors to monitoring service, and from prediction model to prediction service. group register timely collects information from both services and aggregates them for storage or publication.

3.3 System work flow

Monitoring is the precondition of prediction, thus we enclose the monitoring work flow into the prediction work flow for a more compact description. The sequence diagram of system work flow is illustrated in Fig.2. In what follows we provide description in details.

(a) Grid user logs on information node, customizes three terms before sending a monitoring/ prediction request: which node, which resource type, and how long the prediction will last. A prediction request is then created accordingly and sent to information service.

(b) Information service launches a monitoring work flow by sending a monitoring request to monitoring service; a resource sensor is activated as requested, timely updated monitoring data is then sent back to information service through local and group register; historical records are stored in database for achieving prediction.

(c) Information service chooses a prediction service and sends a customized prediction request. The chosen prediction service then takes charge of the whole prediction procedure.

(d) Prediction service acquires historical monitoring data from information service, builds a set of candidate prediction models of different types and parameters. It combines each candidate model with historical data as an evaluation subtask.

(e) Prediction service sends subtasks to evaluation services for feedback, then fixs the model with best performance out of comparison on evaluation results. If an evaluation service is time out, prediction service will redirect the subtask to another one.

(f) Prediction service feeds the fixed prediction model with timely updated monitoring data, then timely updated prediction data is sent to information service through local and group register; historical prediction records are stored in database for checking prediction error.

(g) Grid user gets the resource information monitored or predicted from browser. If upon request or prediction error exceeds certain threshold, prediction service will reload the latest historical data, and go over step **(d)** for model optimization.

(h) Information service terminates the monitoring or prediction procedure when the customized time is used out.

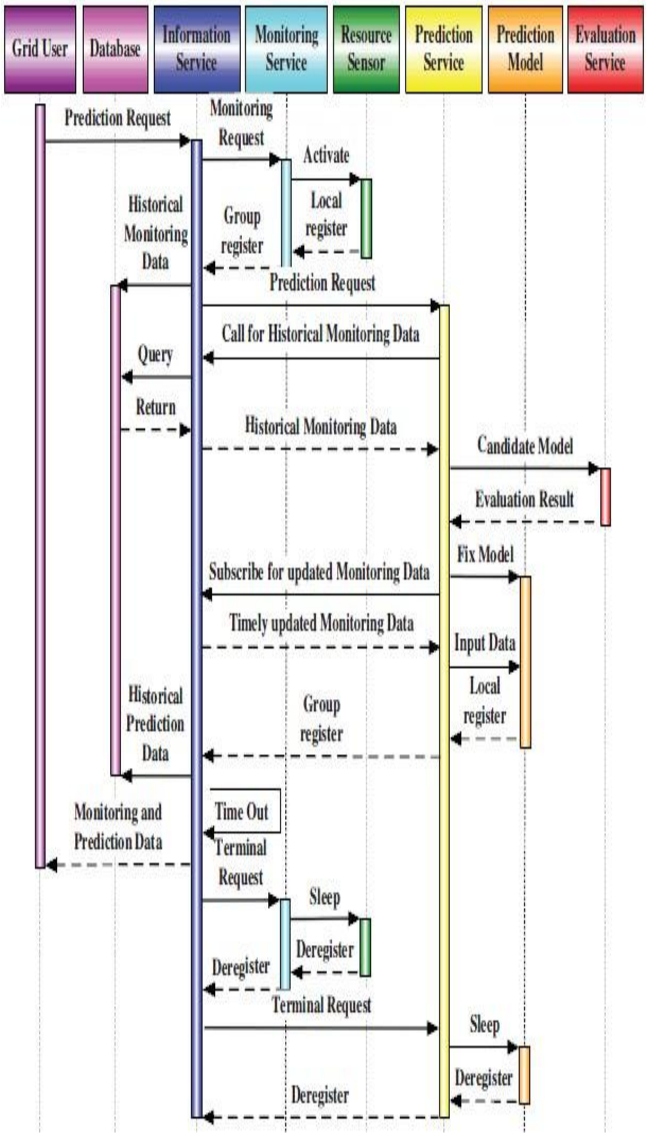

Fig 2 Sequence diagram of monitoring and prediction work flow

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

IV. Algorithm Used

In Our Project we are using Genetic Algorithm(GA) and Particle swarm Optimization(PSO) for Prediction and Monitoring the Grid task and it's classifications

4.1 Genetic algorithm (GA)

Genetic Algorithm(GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solution to optimization and search problems.

4.2 Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a solution with regard to a given measure of quality.

V. Conclusion

This system consists of a set of distributed services to accomplish all required resource monitoring, data gathering, and resource state prediction functions. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for prediction model's hyper-parameter selection. Comparative simulations indicate that the PSO achieves lower error and costs less optimizing time than GA. We plan to move on our research further in the following aspects: monitoring and prediction of grid tasks classifications and evaluation of grid resources, classification and evaluation of grid tasks.

REFERENCES

[1] "Online System for Grid Resource Monitoring and Machine Learning Based prediction"- Liang Hu, Xi-Long Che, Member, IEEE, and Si-Qing Zheng, Senior Member, IEEE

[2] "Grid Harvest Service:A performance System of Grid Computing" MingWu, Xian-He Sun

[3] L. F. Bittencourt, and E. R. M. Madeira, "A Performance-oriented Adaptive Scheduler for Dependent Tasks on Grids," Concurrency and Computation: Practice and Experience, vol. 20, no. 9, pp. 1029-1049, June 2008.

[4] "Alchemi: A .NET-based Enterprise Grid Computing System"- Akshay Luther, Rajkumar Buyya, Rajiv Ranjan, and Srikumar Venugopal

[5] "An Infrastructure for Monitoring and Management in Computational Grids" Abdul Waheed, Warren Smith, Jude George, and Jerry Yan

[6] "Real-time Performance Monitoring, Adaptive Control, and Interactive Steering of Computational Grids" Jeffrey S. Vetter Daniel A. Reed

[7] L. T. Lee, D. F. Tao, and C. Tsao, "An Adaptive Scheme for Predicting the Usage of Grid Resources," Computing Computers and Electrical Engineering, vol. 33, no. 1, pp. 1-11, 2007

[8] P. A. Dinda, and D. R. O'Hallaron, "Host load prediction using linear models," Cluster Computing, vol. 3, no. 4, pp. 265-280, 2000.

[9] P. A. Dinda, "Design, Implementation, and Performance of an Extensible Toolkit for Resource Prediction in Distributed Systems," IEEE Transaction on Parallel and Distributed Systems, vol. 17, no. 2, pp. 160-173, 2006