IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555
Vol. 2, No. 1, 2012

Agent based: QoS Resource Management for Grids

B.S. Bindhumadhava
Real Time Systems and Smart Grid
C-DAC Knowledge Park,
Bangalore, India

Ankit Kumar
Real Time Systems and Smart Grid
C-DAC Knowledge Park,
Bangalore, India

***Abstract*—Grids provide access to distributed resources. When resources are used for critical services, the resources must adhere to stringent Quality of Service (QoS) requirements. This paper presents an approach to resource allocation and management in the grid by proposing a grading system for the resources based on QoS parameters. In this approach, the resource performance is analyzed, in an agent based automated negotiation framework with respect to general effectiveness in completing the assigned task. The paper also presents a schema for Service Level Agreement for grid resources.**

Keywords- Grid computing, Agents, Automated Negotiation, Auction, Quality of Service

I. INTRODUCTION

Grid computing is the “in-thing” in the computing world today. As defined in [1]

“Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources dynamically at runtime depending on their availability, capability, performance, cost, and users' quality-of-service requirements”.

In the world of open distributed systems, individual resources and components work together to achieve a common goal. Such components are meant to dynamically cope with changes in the nature of the environment both in terms of quantity and quality. These systems are very useful in various scenarios ranging from scientific computing to large enterprise systems because of the enormity of the processing and storage powers available. Such resources come together to form communities called virtual organizations (VOs) [2]. The grid, thus, provides us with huge amount of resources. Grid computing tackles compute intensive tasks by aggregating geographically and architecturally dispersed hardware and software resources into large virtual super-resources [23]. To harness the power of the resources, it is essential to have an effective resource management system. The problem is compounded by the fact that the resources operate in a heterogeneous and complex environment [3]. In other words the key issue here is to anticipate, detect and diagnose problems arising out of change in the state of a resource. The quality of service (QoS) delivered by the resource becomes an important aspect in differentiating between similar resources [4]. QoS based grid resource scheduling maximizes the global utility of scheduling system [15].

Another technology, which is working in the area of open distributed systems, is agent technology. Agent technology has been involved in the development of methodologies for autonomous problem solving in dynamic environments. Mobile agents provide a suitable framework for supporting resource and service discovery in grid platforms [18]. The reason is mobile agents offer technical advantages such as reduced communication cost, reduced bandwidth usage, the possibility of using remote interfaces and the support for off-line computation.

II. QUALITY OF SERVICE

Quality of service has been studied extensively in various contexts [5, 6, 26]. There are two types of QoS characteristics namely qualitative characteristics and quantitative characteristics [7]. Quantitative characteristics refer to issues like network latency, CPU performance or storage capacity. On the other hand qualitative characteristics refer to aspects like reliability, task completion ratio, resource discovery time [19] and trust. Qualitative characteristics are difficult to measure and are addressed in this paper. Quantitative characteristics have been dealt with by Rashid Al-Ali et al. [7] and Roy [8].

In grid computing, QoS aims to provide assured availability of resources. QoS in a grid scenario relies on a QoS repository, which contains information about the qualitative characteristics of the resources. In a grid scenario since the resources are heterogeneous and distributed, a Service Level Agreement (SLA) becomes essential between the end-user requesting for grid services and the grid resources. Sahai et al. [9] have proposed a SLA management entity to support QoS in a commercial grid. This work is interesting but at a very preliminary stage. Rashid Al-Ali et al. [7] talk of allocation strategies for running user applications-based on whether they require a Time-domain or Resource-domain allocation strategy.

Deora et al. [10] talk about systems, which are centered, on the use of qualitative measures utilizing user ratings and expectations to calculate the QoS. We also focus on the qualitative characteristics of QoS and present a scheme for measuring and grading resources based on the qualitative characteristics using grading parameters like confidence factor and response time factor. The idea here is to implement an agent based resource negotiation and management approach, where the end user specifies qualitative QoS requirements in the SLA.

III. AGENTS

Agents are autonomous software programs, which act in an environment using their in-built knowledge and the environmental conditions to achieve certain goals [11]. Some of the properties associated with agents are:

Autonomy: Agents operate without direct human intervention and have control over their actions and internal state.

Reactivity: Agents respond to changes in the environment in which they operate.

Proactivity: Agents perceive some environmental changes and act accordingly.

Temporal continuity: Agents run during the entire lifetime of the system either in an active running state or sleep and wait for events.

Goal orientedness: Agents work towards a goal either alone or in conjunction with other agents.

Agents are thus intelligent pieces of code, which move around in the environment performing some action on behalf of the user. Systems involving these agents are particularly useful in scenarios like the grid where the environment and the conditions vary over time. Agents when combined with ontology can provide an important paradigm for supporting dynamic services and collaborative activities such as resource management in Computational, Data and Service Grids.

IV. AUTOMATED NEGOTIATION

Agent systems have been employed in eMarketplace scenarios for automated negotiations between the buyer and the seller [22]. A negotiation is a process of arriving at a mutually acceptable solution for some matter; in this case between the buyer and the seller. It is a key form of interaction in multi-agent systems in which disputing agents decide how to divide the gains from cooperation [16]. An automated negotiation model has three components, namely, Negotiation Protocols, Negotiation Objects and Agents' Decision Making Models.

A. Negotiation Protocols

These protocols lay down the rules which govern the interaction. This covers the permissible types of participants (e.g. the negotiators and any relevant third parties), the negotiation states (e.g. accepting bids, negotiation closed), the events which cause negotiation states to change (e.g. no more bidders, bid accepted) and the valid actions of the participants in particular states (e.g. which messages can be sent by whom, to whom, at what stage) [25]. They prescribe the ordering and type of messages exchanged during the agreement.

A number of negotiation protocols are available namely:

- Contract Net Protocol [12]
- Commodity Market Model
- Auction Model
- Discrete Optimal Control Model
- Game Theory based Model

B. Negotiation Objects

The negotiation objects are the issues over which agreement is taking place. The negotiation objects encode the issues of a contract that agents use to make their decision to accept or reject a proposal for a contract [24]. The issues may be single issues like price or a combination of issues like price, quality, terms and conditions and so on. The negotiation object may also include freebies (like a free music system with a car).

C. Agents' Decision Making Models

The agent may employ a decision-making model based on criteria's such as nature of negotiation objects and the range of operations which can be performed to meet the objectives of the end user. The automated negotiation approach can be used for resource allyocation in the grid. Weiming Shen et al. [13] discuss on agent based adaptive resource negotiation model. We have designed a agent-based resource management framework. However we extend this concept and introduce a QoS based grading system to evaluate and monitor grid resources. In our implementation we use the multi-auction model for negotiation.

The philosophy behind an auction is to let the market determine the true value for the product. A traditional uni-auction model is shown in Fig. 1, where there is one seller and many bidders. The auctioneers act as mediators or negotiating objects between the seller and bidder. In this model, the seller puts up his resource for auction setting a minimum bid price.

The prospective buyers or bidders bid for the product following the bidding rules, which are laid out. In each round, the agent considers bidding if and only if it is not holding an active bid in an auction or it has not already obtained the service. The resource or product is then sold to the bidder based on a predefined set of rules. An auction is an efficient negotiation model. Bidders feel that they correctly evaluate the item they bid on, and auctioneers feel that they sold the product based on the rules laid out by them.

Auctions can be conducted as open or closed depending on whether they allow back-and-forth offers and counter offers.

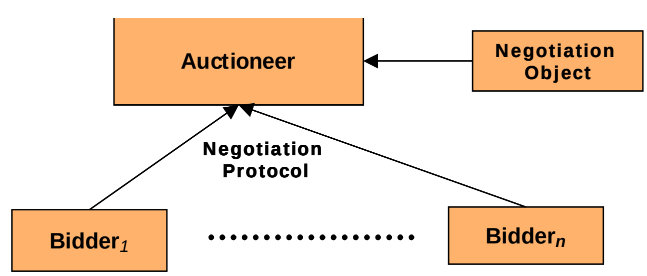

Figure 1. Uni-auction model

The four major types of auction mechanisms are:

- English auction
- Dutch auction
- First-price, sealed-bid auction
- Vickerey auction

In a uni-auction, the auction is auctioneer-centered limiting the choice of the buyer. The buyer does not get an opportunity to compare similar products or resources in this case, and arrive at a reasonable bid price.

V. RESOURCE MANAGEMENT

Resource selection and allocation plays a critical role in improving the overall performance of a grid application. QoS of the resource becomes an important criterion for grid resource selection. For every job on the Grid which has QoS requirements, it is be necessary to negotiate a service level agreement (SLA) to enforce the desired level of service required by the end user.

Some of the QoS based criteria, which need to be assessed and monitored, are:

Confidence – The amount of trust, which can be placed, on a resource in fulfilling the resource requestor's objectives.

Response time – The amount of time taken by the resource to complete the given task.

Based on the outcome of the auction, an agreement is entered into SLA which represents qualitative guarantees placed on service invocations between the user and the resource provider, clearly defining the contractual obligations of the user and the resource. The SLA needs to be regularly monitored during the service lifetime. It provides flexible mechanisms for agreeing on the QoS provided by resources. In case there is a violation, the resource provider should be penalized and an alternate arrangement should be made so that the end user requirements are met.

VI. IMPLEMENTATION AND GRADING SYSTEM

In our approach, the end user looking for grid resources contacts the job agents. The job agents are responsible for negotiating on behalf of the user and getting the end user a grid resource. Weiming Shen et al. [13] proposed a five-layer architecture for the agent based resource management system. Our system is modeled using this architecture (shown in Fig. 2).

The job agents understand the user request and search the grid for resources to carry out multiple concurrent jobs submitted by end users [21]. The job agents (bidders) get information about the auctions from the auction manager and then bid for resources in the auction [19]. The Auction manager agent handles all the job agents which are associated with multiple jobs and forwards the task of resource discovery to Resource Manager Agents. Resource manager agents manage and schedule computations among local resources in the grid. They also provide access to storage devices, network and databases. The resource managers wishing to sell the resources contact the auction manager agent. The auction manager agent spawns off auction agents (auctioneers) to conduct the auctions and negotiates for access to resources at low cost. The auction may follow any one of the auction mechanisms described above.

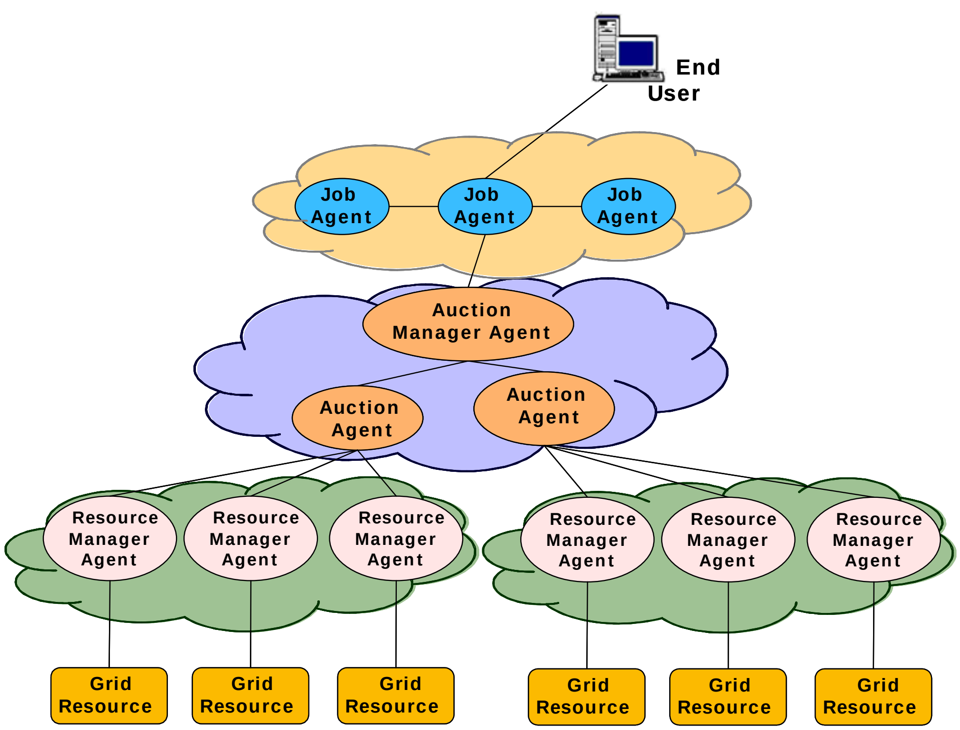

Figure 2. System Architecture

Here we follow a multi-auction model shown in Fig. 3. The bidder (Job agent) incurs the computational cost of a resource by calculating optimal bids given its valuation for items and the expected closing prices. Job agents get information about similar resources by participating in multiple auctions simultaneously. The job agents can then decide to choose the resource most suited to meet the user requirements.

The job agents co-ordinate between the user and the Auction manager agent. For the auction to be meaningful, the job agent must bid the true valuation of the resource. To arrive at the true valuation, the job agent also takes into account the QoS of the resource in order to make predictions of the likely closing prices of various auctions. The job agents maintain a QoS repository.

In this repository, the resources are graded based on their past performance. To illustrate the grading system, consider a resource R.

Here we consider two criteria for grading a resource:

Confidence - The confidence factor is based on the number of times the resource fulfills the task assigned to it. Formally if SCa denotes the service contracts a resource agrees to execute and SCc denotes the service contracts successfully completed by the resource then the confidence factor (CF) of the resource can be shown as:

$$CF = \frac{SC_c}{SC_a}, 0 < CF \leq 1$$

Response time – This is a measure of the time taken to complete a given task. Formally if Te is the time defined in the service contract for task completion and if Ta is the actual time taken to complete the task the response time factor (RF) can be shown as:

$$RF = \frac{T_e}{T_a}$$

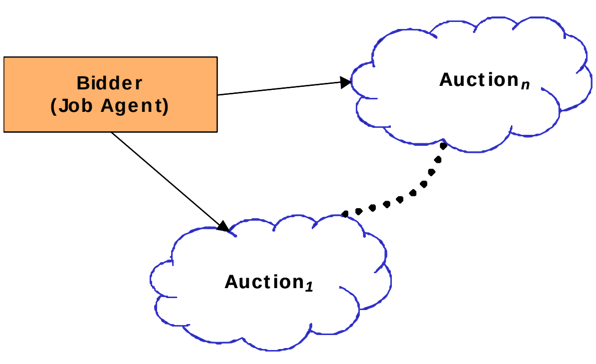

Figure 3. Multi-auction model

The grade of a resource can then be considered as a weighted sum of RF and CF.

$$G_a = w_{rf}RF_a + w_{cf}CF_a$$

where,

Ga is the grade of resource a.

RFa is the response time factor of resource a.

CFa is the confidence factor of resource a.

wrf is the weight vector assigned for the response time factor, which defines the degree of importance, whichneeds to be placed on this aspect of the resource, and is user configurable.

wcf is the weight vector assigned for the confidence factor and is user configurable.

Consider the scenario shown in Table I where a job agent is bidding for three resources A, B and C, which meet the user requirements.

Before bidding for a particular resource, the job agent contacts its peers for information about this resource. The job-agents keep track of the grade for the resources, in the QoS repository and exchange this information with other job agents. Here, we assume that the job agents execute in a trusted unbiased environment. Based on the outcome of the auction, the resource manager and the job agent who bid successfully enter into a SLA. An XML schema is used for SLA formalization.

TABLE I. WEIGHTED PRICE BIDS FOR THE RESOURCES

Resource	Grade	Weighted Price Bid
A	G_a	P_a+G_a
B	G_b	P_b+G_b
C	G_c	P_c+G_c

where,

Pi is the base price of the resource i.

Pi+Gi is the weighted price bid for the resource i.

The XML schema consists of the following components:

Parties – The parties involved in the negotiation are defined here, that is, the details about the buyer and the seller .Other agencies or third parties involved in the negotiation are also defined if one signatory party does not trust the other.

Weight Vector – The weights wcf and wrf assigned for grading the resource are defined here.

Commitment - In the schema the response time commitment can be expressed as a simple 2 element tuple (Value, Criterion), for example, (10 seconds, <) – meaning that the response time should be less than ten seconds.

The job agent also keeps information about the resources, which participated in the negotiation. The resource, associated with the successful bid, is then made available to the user for his jobs. The job agent keeps monitoring the resource performance based on the QoS terms specified in the SLA, while it is handling the user job. In case of failure of the resource, the job agent contacts and negotiates with the next best resource. Since the job agent remembers the resources, which participated in the negotiation, it does not have to go for auctioning all over again unless it is really required .In this model we see that the resources are penalized for poor performance. A resource with a poor grade will get lower and lesser number of bids. The resource will then have to lower its reservation price for bidding which affects its profitability. The resources will therefore aim to improve their QoS.

VII. Experimental Results and Evaluation

A heterogeneous network with 6 grid nodes was considered for experimentation. A common job with Te of 60 seconds was submitted by the Job Agent to all the nodes simultaneously. Table 2 shows the results on each node with its corresponding response time factor. Ta varies due to factors like load on the resource, quality of the resource in terms of speed and internal process scheduling policies.

It can be observed that the resource P3 with Ta of 60 is equal to its Te, which means the task is completed as per the time defined in service contract. The response time factor for P3 is 1 which gives an added weight to its grade as compared with other resources having low response factor.

TABLE II. Effect of Ta on Response Factor

Processor	Ta in seconds	RF
P1	75	0.8
P2	62	0.98
P3	60	1.0
P4	100	0.6
P5	67	0.9
P6	53	1.1

The confidence factor provides a high level of assurance to grade a resource. To illustrate this, the experiment below was conducted. Among a set of grid nodes, 4 different processors with an average load on it were considered as mentioned in Fig. 4. A set of 50 requests were submitted by the job agent to each of the processors with a fixed deadline.

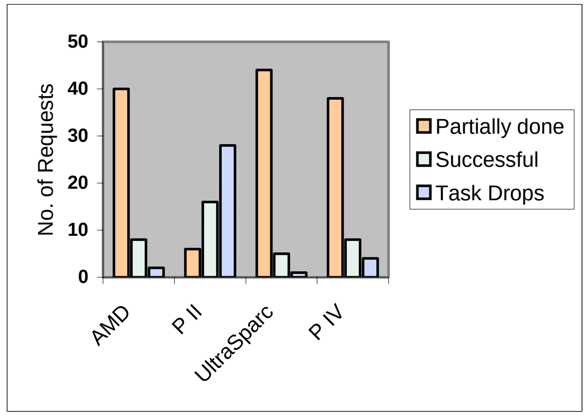

Figure 4. Influence of processor type

The results indicate that a high computing processor successfully completes more number of tasks and the number of task drops increases on a slow processor. This variation is on the parameters which affects the confidence factor and hence the grade of a resource.

Another QoS parameter to measure the confidence factor is the number of processors. For experimentation a set of 3 different jobs were submitted on grid nodes with a varied number of processors as shown in Fig. 5. The experimental results show that the processing time gradually decreases with the increase in number of processors. Hence, the number of processors also accounts in the evaluation of confidence factor and in turn with the grade of a resource. However, there is a slight variation in cases when the processors have more jobs submitted due to their increased computing power. The confidence factor, i.e., the amount of jobs successfully completed by a resource depends on the other factors such as peak hours, resource load, number of processors, time zone and speed of processing.

With these results, it is seen that the grade of a resource is directly proportional to both response time factor and the confidence factor. The job agent can bid for a resource with a higher grade at the cost of varied price in order to meet the deadlines.

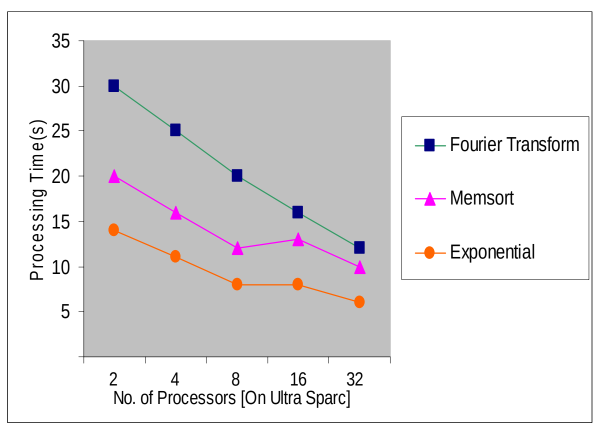

Figure 5. Performance results on Ultra Sparc

VIII. CONCLUSION

Agent based grid computing throws open many issues and challenges. Ian Foster et al. [14] enumerate a list of issues involving grids and agents. In this paper we have presented an agent based resource management scheme for the grid, which adapts itself to the dynamic behavior of the resource. The qualitative characteristics of QoS for a resource are modeled and formalized based on past behavior. We have devised a price-based model for grid resource management where a resource with higher grades can demand a higher price for its services. Experimental results are included to demonstrate various factors which affect the grade of a resource. The evaluation on the grade of a resource is not limited to the confidence factor and response time factor. It may also include other criteria's like resource discovery time required by agents to discover the resources available and agent's overhead due to the number of hops they travel.

The proposed approach can be used for QoS management in other distributed environments as well, say in Production, Planning and Control process for an industrial application. Future work involves considering issues like security, trust etc. which can also be used to formalize the resource behavior.

REFERENCES

[1] http://www.gridcomputing.com/gridfaq.html

[2] Ian Foster et al., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", Int'l J.High Performance Computing Applications, vol. 15. no. 3, Aug 2001,pp.200-222.

[3] Buyya et al., "Economic models for resource management and scheduling in Grid computing", Concurrency And Computation: Practice And Experience Concurrency Computat.: Pract. Exper. 2002; 14:1507–1542 (DOI: 10.1002/cpe.690).

[4] Hanhua Chen et al., "Q-GSM: A QoS Oriented Grid Service Management Framework", Cluster and Grid Computing Lab, Huazhong University of Science and Technology, Wuhan, China.

[5] A. Oguz, A. T. Campbell, M.E. Kounavis and R.F. Liao, "The Mobiware Toolkit: Programmable Support for Adaptive Mobile Networking", IEEE Personal Communications Magazine, Special Issue on Adapting to Network and Client Variability, 5(4), 1998.

[6] G. Bochmann and A. Hafid, "Some Principles for Quality of Service Management", Technical report, Universite de Montreal, 1996.

[7] Rashid Al-Ali, Kaizar Amin, Gregor von Laszewski, Omer Rana, David Walker, Mihael Hategan and Nester Zaluzec, "Analysis and Provision of QoS for Distributed Grid Applications", Journal of Grid Computing, 2(2):163–182, June 2004.

[8] A. Roy, "End-to-End Quality of Service for High-End Applications", PhD thesis, The University of Chicago, August 2001.

[9] A. Sahai, S. Graupner, V. Machiraju and A. Moorsel, "Specifying and Monitoring Guarantees in Commercial Grids through SLA", Proceedings of the 3rd IEEE/ACM CCGrid 2003.

[10] V. Deora, J. Shao, W. A. Gray and N. J. Fiddian, "A Quality of Service Management Framework Based on User Expectations", First International Conference on Service Oriented Computing (ICSOC), Trento, Italy, December 2003.

[11] Luck M. et al., "Agent technology: Enabling Next Generation Computing", AgentLink, 2003.

[12] www.fipa.org, FIPA Contract Net Interaction Protocol Specification, FIPA TC Communication, 2002.

[13] Weiming Shen et al., "Adaptive Negotiation for Agent-Based Grid Computing", Proceedings of AAMAS2002 Workshop on Agentcities: Challenges in Open Agent Environments, Bologna, Italy, July 15-19, 2002, pp. 32-36.

[14] Ian Foster, Nicholas R. Jennings and Carl Kesselman, "Brain meets Brawn: Why Grid and Agents Need Each Other", AAMAS 2004.

[15] Li Chunlin and Li Layuan, "Utility Based Multiple QoS Guaranteed Resource Scheduling Optimization In Grid Computing", IEEE Proceedings of the International Conference on Computing: Theory and Applications, 2007.

[16] Shaheen S. Fatima Michael Wooldridge and Nicholas R. Jennings, "Approximate and Online Multi-Issue Negotiation", Proceedings of the 6th Int. J. Conference on Autonomous Agents and Multi-agent Systems, (AAMAS) May 2007.

[17] Rafael Parra-Hernandez, Daniel C Vanderster and Nikitas J. Dimopoulos, "Resource Management and Knapsack Formulations on the Grid", 5th IEEE/ACM International Workshop on Grid Computing, 2004.

[18] Jian Mia, Qiang Liu, Boyan Cheng and Yuancheng Sun, IEEE Proceedings of International Conference on Geoscience and Remote Sensing Symposium, July 2006.

[19] Yuhui Qiu and Zhixing Huang, "Simulate Grid Resource Trading via Cognitive Agent: A Case Study", Proceedings of the Second International Conference on Semantics, Knowledge, and Grid (SKG 2006).

[20] M. N. Birje, S. S. Manvi and Bhanu Prasad, "Agent based Resource Brokering and Allocation in Wireless Grids", IEEE International Conference on Services Computing 2006.

[21] Sarbani Roy, Sohini Dasgupta and Nandini Mukherjee, "A Multi-agent Framework for Resource Brokering of Multiple Concurrent Jobs in Grid Environment", Proceedings of The Fifth International Symposium on Parallel and Distributed Computing 2006.

[22] Tom Wanyama and Behrouz Homayoun Far ,"A protocol for multi-agent negotiation in a group-choice decision making process", Journal of Network and Computer Applications, Volume 30, Issue 3, Pages: 1173-1195, August 2007.

[23] A. Kertesz and I. Rodero, F. Guim, "BPDL: A Data Model for Grid Resource Broker Capabilities", CoreGRID Technical Report-0074, March 2007.

[24] Rod Moten, "Using Type Theory as a Language for Negotiation Objects in Online Exchanges", Technical Report 02-10, Rensselaer Polytechnic Institute, July 2002.

[25] Matthew Addis, Paul Allen and Mike Surridge, "Negotiating for Software Services", IEEE Proceedings of the 11th International Workshop on Database and Expert Systems Applications, August 2000.

[26] Shaoxiong Hua and Gang Qu, "A New Quality of Service Metric for Hard/Soft Real-Time Applications", IEEE Proceedings of the International Conference on Information Technology: Computers and Communications, 2003.

AUTHORS PROFILE

B.S.Bindhumadhava is Associate Director in Real Time Systems and Smart Grid at Centre for Development of Advanced Computing (C-DAC), Bangalore, India.He obtained his Bachelor degree in Electrical & Electronics Engineering from Bangalore University, Diploma in Systems Management from NIIT and MS (Electronics & Controls) from BITS, Pilani. His interests are in Systems Modelling, Simulation, Fault Tolerant Computing, Agent Computing, Semantic Computing & Cryptanalysis.He is life member of Computer Society of India (CSI), Life member of Advanced Computing & Communications Society (ACCS) and senior member of IEEE.

Ankit Kumar is a Senior Technical Officer in Real Time Systems and Smart Grid at the Centre for Development of Advanced Computing (C-DAC), Bangalore, India. He obtained his Bachelor degree in Computer Science and Engineering from U.P. Technical University, Diploma in Advanced Computing from C-DAC ACTS. His research interests includes Agent Based Computing, Autonomic Computing, Multi-Agent Systems and Energy Efficient Systems. He also actively involved in developing software products and tools for Power Sector.