International Journal of Soft Computing 7 (3): 97-103, 2012
ISSN: 1816-9503

Fuzzy Based Task Scheduling for Hierarchical Resource Allocation in Mobile Grid Environment

[1]S. Thenmozhi, [2]A. Tamilarasi and [3]K. Thangavel
[1]Department of Computer Applications, Chettinad College of Engineering and Technology, Karur, Tamilnadu, India
[2]Department of Computer Applications, Kongu Engineering College, Perundurai, Tamilnadu, India
[3]Department of Computer Applications, Periyar University, Salem, Tamilnadu, India

Abstract: In mobile grid environment, the main challenging issues are scheduling, adaptation, security and mobility. The job scheduling problem becomes more complicated due to the limitations of node mobility. In order to minimize the resource utilization, gaining the maximum profit to be cost effective and satisfying the user constraints, an efficient job scheduling technique is required for mobile grid environment. In this study, researchers propose a fuzzy based task scheduling algorithm for resource allocation depending upon the workload and the resource availability of the grid members. In this scheduling, the computation sensitive task is assigned for grid members with least workload and the communication sensitive task is assigned for grid members with high resource availability. Using the workload and resource availability as input variables, fuzzy decision rule table is created. After defuzzification, the output gives us a perfect matching for scheduling the tasks according to the load and availability. Thus, the algorithm proves to be more effective in task scheduling of mobile grids. From the simulation results, researchers show that the proposed scheduling technique attained maximum throughput and less delay when compared with the existing technique.

Key words: Grid computing, mobile grid, resource allocation, fuzzy task scheduling, India

INTRODUCTION

Resource allocation in grid workflow is required in order to achieve a high performance. Resource selection and resource binding are the two classifications of resource allocation. The common architecture of conventional resource brokering system isolates the resource selection from resource binding (Kumar and Kaur, 2007). Complex resource specification languages and resource selection algorithm are mainly handled by the resource selection. The resource selection algorithm discovers a matching set of resource and negotiates with an individual local resources manager. The selection is based on the provision of the resource specification. The resources are obtained by this application (Thenmozhi and Tamilarasi, 2011).

Managing of jobs include resource allocation for any particular job to schedule tasks equivalently by partitioning the jobs, data management, event correlation and service-level management capabilities. Schedulers are required for effective job management (Litke *et al.*, 2004). Hierarchical structure is formed using meta-scheduler as root and other lower level schedulers are the leaves which provide specific scheduling capabilities. Tasks are allocated appropriately to the processors by minimizing the completion time of a parallel application (Martincova and Zabovsky, 2007).

In order to minimize the resource utilization, gaining the maximum profit and satisfying at the same time the user constraints (security, quality of service, fault tolerance, etc.) an NP-complete job scheduling problem can be applied to allocate jobs efficiently to resources and to be cost effective (He and Ioerger, 2004; Khanli and Kargar, 2007). The scheduling may be done as:

Centralized: Single job scheduler on one instance, all information collected here.

Hierarchical: Two job schedulers, one at global and other at local level.

Decentralized: No central instance, distributed schedulers interact and commit resources. In this study, researchers propose an algorithm for scheduling the nodes in the clusterat the cluster head level enhancing the resource allocation in the mobile grid environment.

Corresponding Author: S. Thenmozhi, Department of Computer Applications, Chettinad College of Engineering and Technology, Karur, Tamilnadu, India

Related research: Bansal *et al.* (2011) have proposed a novel grid-scheduling heuristic that adaptively and dynamically schedules tasks. It doesn't require any prior information on the workload of incoming tasks. The prediction information on processor utilization of individual nodes is used in the grid system in the form of a state-transition diagram, employing a prioritized round-robin algorithm with task replication to optimally schedule tasks.

Katsaros and Polyzos (2008) have formulated and investigated the problem of job scheduling in a mobile grid environment, considering the problems incurred by intermittent connectivity. For inhibiting the mobile and wireless networking environment, instalment scheduling policies were modified. Researchers studied its performance with respect to important performance metrics in a realistic evaluation environment and in comparison with previously proposed policies. Lower resource requirements are provided by partitioning the subtask workload. Moreover, disconnection events affect only the execution of the current workload fragment resulting in improved turn-around times and resource waste.

Lee *et al.* (2009) have presented a novel balanced scheduling algorithm in mobile grid, taking into account the mobility and availability in scheduling. They analyzed users' mobility patterns to quantitatively measure the resource availability that is classified into three types: full availability, partial availability and unavailability. A load balanced technique was proposed by classifying mobile devices into nine groups depending on availability.

Bidgoli and Nezad (2010) have proposed a scheduling algorithm considering the dependencies between tasks and data transfer cost between tasks in grid environments. The best sources are assigned to the scheduler by using this algorithm. It is coupled with optimization of time and costs is necessary for data transfer between tasks. In this algorithm, the future research is to rectify the issue of dynamic grid environment such as errors in the allocated resources.

MATERIALS AND METHODS

Proposed work: In the study (Thenmozhi and Tamilarasi, 2011), researchers had estimated mobility metric and resource availability metric. The Current Workload of the grid nodes (CWL) and the resource Availability (AW) is taken for the scheduling of task.

Estimation of resource availability and workload: The resource availability of the grid nodes can be estimated using this equation:

$$WL_i = \frac{CWL_i + \left(\sum_{j=1}^{k} jobsize_j\right)}{Power_i} \quad (1)$$

Where:

CWL_i = The current workload of n_i
WL_i = The work load of n_i
$Power_i$ = The power of the node n_i
$Jobsize_j$ = The size of the job j

Ptime shows the predicted time for resource availability within the user's range and is calculated by the following equation:

$$Ptime = (UserRange - Distance)/Average\ mobility \quad (2)$$

After estimating the Ptime and WL values, the MA sends these values to its cluster head CH1. The CH1 then schedules the jobs if their grid node satisfies the following condition if:

$$Ptime/WL > Th \quad (3)$$

where, Th is a threshold value (which can be fixed based on the job request). We represent the Availability (AW) using Ptime and work load using WL.

Task scheduling using fuzzy logic: Researchers schedule the tasks according to the Availability (AW) and the current Workload (WL). AW and WL are shown as inputs and depending upon these values, the computation and communication cost is calculated. We build separate tables for the input and output values initially.

Fuzzy set for input values: The input values are the combination of AW and WL. Researchers take three possibilities, high, medium and low for load and availability as described in the Table 1.

Fuzzy set for output values: The output values Computation cost (Cp) and Communication cost (Cm) are assigned with two probabilities low and high. The Table 2 shows the combinations of these values. After determining the input and output values researchers create a fuzzy table to estimate the task according to the AW and WL values.

Table 1: Input values

	AW		
WL	Low	Medium	High
Low	LL,LA	LL,MA	LL,HA
Medium	ML,LA	ML,MA	ML,HA
High	HL,LA	HL,MA	HL,HA

Table 2: Output values

	Cm	
Cp	Low	High
Low	LCp,LCm	LCp,HCm
High	HCp,LCm	HCp,HCm

Fuzzy rules: Fuzzification of the input variables load and availability; from each of the variables, the crisp inputs are taken and the degree is assigned to the inputs to appropriate fuzzy sets:

- Rule evaluation: The antecedents of the fuzzy rules are shown the fuzzified inputs and then it is applied to the consequent membership function
- Aggregation of the rule outputs: The process of fusing the outputs of all rules
- Defuzzification: The aggregated output fuzzy set chance is the input for the defuzzificaion process and a single crisp number is obtained as a output

The QoS related issues can be solved using the fuzzy logic methodology which is pro-active approach. The performance of a highly dynamic nonlinear system can be managed in the absence of a mathematical model using the fuzzy logic (Bidgoli and Nezad, 2010).

Numerous fields including control systems, decision making, pattern recognition and system modelling uses the Fuzzy if-Then rules. The following three steps determine the fuzzy rule based inference:

- Fuzzy matching: the degree of the input is calculated and the fuzzy rules are stated accordingly
- Inference: depending upon the matching degree, the rule conclusion is calculated
- Combination: the final conclusion is the combination of all the conclusions inferred by the fuzzy rules (Thenmozhi and Tamilarasi, 2010)

A fuzzy classification system can be formed by applying a set of fuzzy rules based upon the linguistic values of its features or an object. The rule is applied to the number shown by the antecedent. The weights in this rule are numbered between 0 and 1 in this rule. The input is fuzzified by estimating the antecedent and this is applied for any necessary fuzzy operators. Inference represents the result which is applied to the consequent. A fuzzy classification system is built using a set of fuzzy rules which specifies the classification problem which is to be determined.

In Table 3, load and availability are shown as inputs. Researchers define nine fuzzy sets with the combinations shown in Table 1.

Table 3: Fuzzy Rules for the determining Output

WL	AW	Output
LL	LA	HCp,LCm
LL	MA	HCp,HCm
LL	HA	HCp,HCm
ML	LA	HCp,LCm
ML	MA	HCp,LCm
ML	HA	HCp,HCm
HL	LA	LCp,LCm
HL	MA	LCp,HCm
HL	HA	LCp,HCm

Fuzzy sets: Researchers will now describe the methodology for fuzzy logic approach to schedule the tasks in the resource allocation of mobile grids. For task scheduling, the communication cost and the computation costs are considered. When resource availability of a member is high, then it has a best communication cost and that member can be used for communicating to any distances. Even when the load level is not too low, this member can be used for communication due to its high availability. Similarly, when the load value is low for a member, it can be assigned for the computation purpose. Though the computation level is not too high, the member can participate in the computation due to lower load value.

In assigning the cost, two main input variables are load and availability. With fuzzy logic, researchers assign grade values to the two variables.

$$\text{Fuzzy set} = \{WL, AW\}$$

Fuzzy logic implements human experiences and preferences via membership functions and fuzzy rules. In this research, the fuzzy if-then rules consider the parameters: work load and availability.

The inputs are fuzzified, implicated, aggregated and defuzzified to get the output. The linguistic variables associated with the input variables are low, medium and high. The output variables use two linguistic variables high and low. The first input variable WL can be represented as a fuzzy set as:

$$WL = \text{Fuzzyset}\,[\{Cp, a\}, \{Cm, b\}]$$

Where:

a = The membership grade for computation cost with respect to load

b = The membership grade for communication cost with respect to load

For e.g., if $a = 0.3$ and $b = 0.2$ then possibility is high for computation cost and possibility is less for communication cost. Thus, the member is assigned for computation.

The second parameter resource availability AW can be represented as a fuzzy set as:

AW = Fuzzyset [{Cp, c}, {Cm, d}]

Where:

c = The membership grade for computation cost with respect to availability

d = The membership grade for communication cost with respect to availability

For e.g., if c = 0.1 and d = 0.5 then possibility is high for communication cost and possibility is less for the computation cost. Thus, the member is assigned to communication task. Depending upon the low, medium and high values of load and availability, researchers assign the cost metric. The fuzzy rule set in Table 3 is shown:

- When load is low and availability is high, computation cost is high and communication cost is low
- When load is low and availability is medium, computation cost is high and communication cost is also high
- When the load is low and availability is high then the computation cost is high and communication cost is also high
- When the load is medium and availability is low, the computation cost is high and the communication cost is low
- When the load is medium and the availability is also medium then the computation cost and the communication cost are low
- When the load is medium and the availability is high then the computation cost and communication cost is high
- When load is high and the availability is low then the computation cost and communication cost is low
- When the load is high and the availability is medium then computation cost is low and communication cost is high
- When the load is high and availability is also high then the computation cost is low and the communication cost is high

Defuzzification: Mapping from a space of fuzzy control action defined over an output universe of discourse into a space of non-fuzzy (crisp) control act ions is known as the defuzzification. A crisp control action that best represents the possibility distribution of an inferred fuzzy control action is produced by the defuzzification strategy.

Center of Area (COA): Here, the center of gravity of the output membership function is used for selecting the output crispy value:

$$U_o = \frac{\int w\mu(w)dw}{\int \mu(w)dw}$$

Center of Sums (COS): The contribution of the area of each fuzzy sets is considered while the computation of the union of the fuzzy sets are avoided in the Center of Sums Method:

$$U_o = \frac{\int w\sum_{j=1}^{l}\mu(w)dw}{\int\sum_{j=1}^{l}\mu(w)dw}$$

Height Method (HM): Evaluation of the centroid of each output membership function for each rule is done first and the averages of individual centroids are calculated as the output:

$$U_o = \frac{\sum_{j=1}^{n} w_j \mu(w_j)}{\sum_{j=1}^{n}\mu(w_j)}$$

Middle of Maxima (MOM): The mean value of all local control actions is generated by this MOM strategy. Their membership functions reach the maximum:

$$U_o = \sum_{j=1}^{l} w_j/l$$

Center of Largest Area (COLA): The crisp output value is determined from the convex fuzzy subset with the largest area which is defined as the center of area of the particular subset.

First of Maxima (FM): The smallest value of the domain which has maximum membership degree is taken from the union of fuzzy sets:

$$U_o = \inf\{w \in W \mid \mu(w) = hgt\ (W)$$

Height Weighted Second Maxima (HWSM): Evaluation of the second maximum of each output membership function for each rule is done and the average of individual maxima is calculated as the output (Lee *et al*., 2009):

$$U_o = \frac{\sum_{j=1}^{n} w_j \mu(w_j)}{\sum_{j=1}^{n}\mu(w_j)}$$

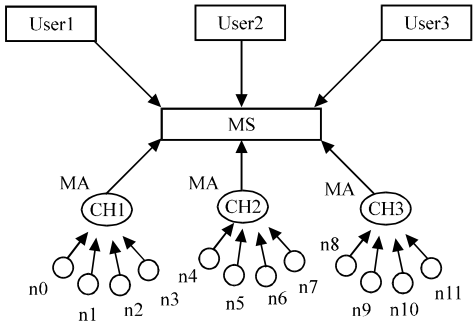

Fig. 1: Function of HRAA

The Center of Area (COA) base on defuzzification method is used in the proposed scheduling algorithm to obtain the desired output.

Overall algorithm for resource allocation: The sequence of operations in Hierarchical Resource Allocation Architecture (HRAA) described in study (Thenmozhi and Tamilarasi, 2011) is shown in Fig. 1. In Fig. 1, the arrows represent the communication messages and the nodes represent the agents/servers. The sequence is as follows:

- The MAs of each node in the local cluster send the resource status information to the CHs
- CH calculates the trust values of its members
- The CHs send this information to the MS
- The MS then create a database which contains information about the status and the price of each resource
- A user submits its job details and the resource requirements to the MS
- The MS sends the job request information to the best CH
- The selected CH allocates the resources to the trusted members depending upon the resource availability and the workload. The task scheduling is performed using the fuzzy based scheduling algorithm
- If any CH is unable to allocate the resources in its cluster then the MS forwards the job request information to other CH
- The process is continued until the job is successfully assigned
- The CHs gather the completed subtasks from the local machines and then send the data back to the MS
- The MS aggregates the completed subtasks and then stores the results in it's own database
- MS also sends back the results to the application of the respective users

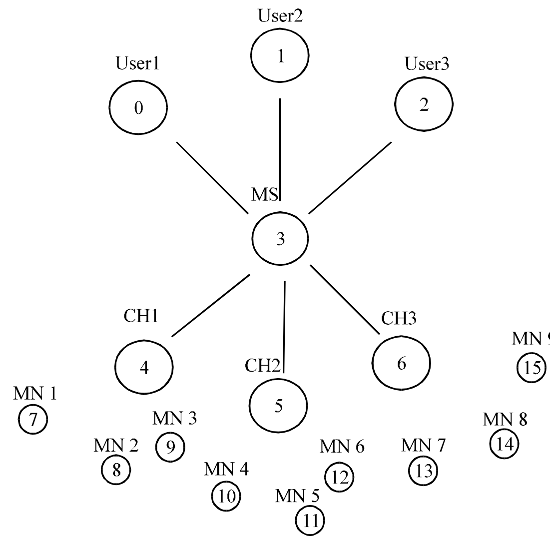

Fig. 2: Simulation setup

Simulation results: In this study, researchers examine the performance of the Fuzzy Based Task Scheduling for Hierarchical Resource Allocation (HRAA-Fuzzy) with an extensive simulation study based upon the Network Simulator Version-2 (Ns-2). The simulation topology is shown in Fig. 2. We compare the results with the previous Hierarchical Resource Allocation Architecture (HRAA) with normal scheduling (Parsa and Entezari-Maleki, 2009).

Performance metrics: In the experiments, researchers measure the following metrics.

Delay: It measures the average end-to-end delay occurred while executing a shown task.

Delivery ratio: It is the ratio of the number of packets successfully received and the total number of packets transmitted.

Drop: It is the total number of packets dropped during the data transmission. Researchers have generated job requests with minimum, medium and maximum workload from the 3 users. User1 submits requests with maximum load and maximum availability and minimum load and minimum availability. User2 submits requests with minimum load and minimum availability and minimum load and medium availability. User3 submits request with medium load and medium availability.

RESULTS AND DISCUSSION

Based on maximum rate: In the initial experiment, researchers vary the maximum load of the job requests,

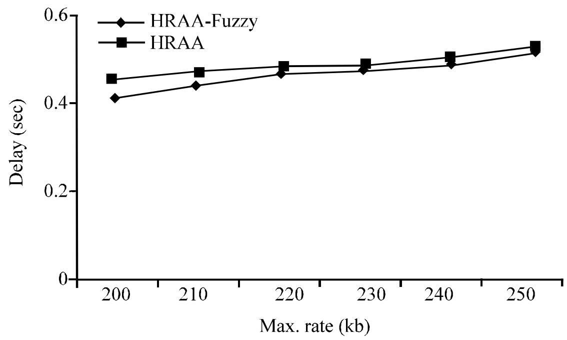

Fig. 3: Maximum rate vs. delay

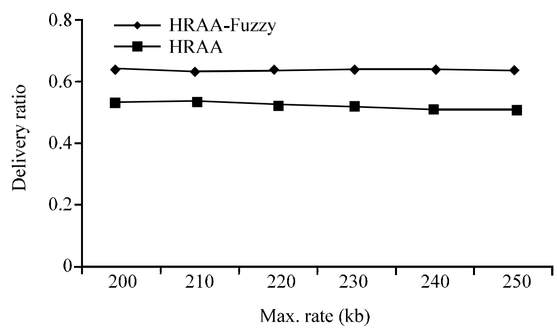

Fig. 4: Maximum rate vs. delivery ratio

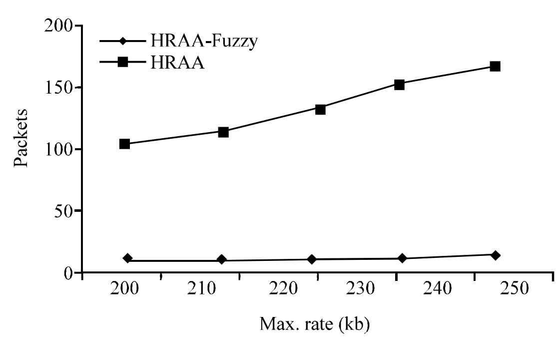

Fig. 5: Maximum rate vs. drop

by varying the rate as 200, 210, 220, 230, 240 and 250 kb. From Fig. 3, researchers can see that the average end to end delay of the proposed HRAA-Fuzzy is less than the existing HRAA protocol.

From Fig. 4, researchers can see that the delivery ratio of the proposed HRAA-Fuzzy is better than the existing HRAA protocol. From Fig. 5, researchers can see that the packet drop of the proposed HRAA-Fuzzy is less than the existing HRAA protocol.

Based on minimum rate: In this experiment, researchers vary the minimum load of the job requests by varying the rate as 10, 20, 30, 40 and 50 kb.

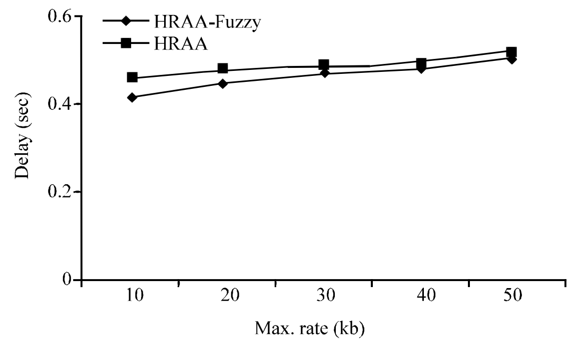

Fig. 6: Minimum rate vs. delay

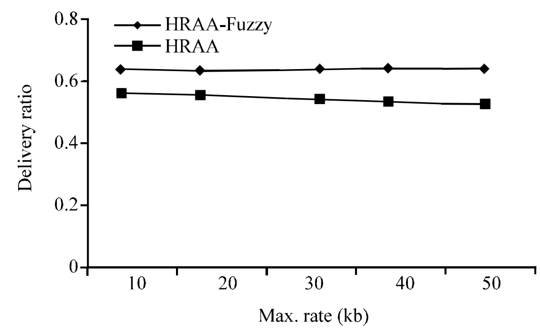

Fig. 7: Minimum rate vs. delivery ratio

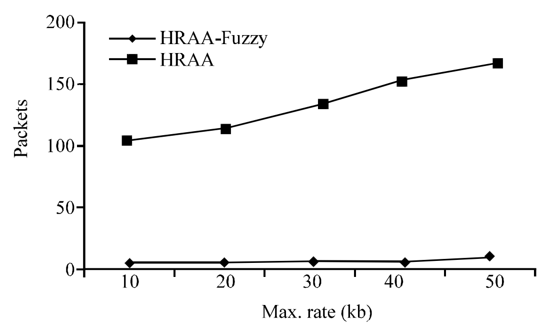

Fig. 8: Minimum rate vs. drop

From Fig. 6, researchers can see that the average end to end delay of the proposed HRAA-Fuzzy is less than the existing HRAA protocol.

From Fig. 7, Researchers can see that the delivery ratio of the proposed HRAA-Fuzzy is better than the existing HRAA protocol.

From Fig. 8, researchers can see that the packet drop of the proposed HRAA-Fuzzy is less than the existing HRAA protocol.

CONCLUSION

In this study, researchers have proposed a fuzzy based scheduling algorithm for resource allocation depending upon the workload and the resource availability of the members. The fuzzy table is created based on the values of two parameters. The fuzzy table takes the value of workload and resource availability as input and researchers consider three linguistic values high, medium and low. The output is shown in the form of computation cost and communication cost and researchers consider two linguistic values, low and high. When the workload of a member is low that member can be assigned for the computation though the availability is less. Similarly, when the resource availability of a member is high, the member is assigned for the communication though it has a higher workload. Depending upon various combinations of the two parameters researchers create the fuzzy table. The output gives us a perfect analysis of scheduling their tasks and thus the algorithm proves to be more effective in task scheduling of mobile grids. From the simulation results, researchers have shown that the proposed scheduling technique attained maximum throughput and less delay when compared with the existing technique.

REFERENCES

Bansal, S., B. Kothari and C. Hota, 2011. Dynamic task-scheduling in grid computing using prioritized round robin algorithm. Int. J. Comput. Sci., 8: 472-478.

Bidgoli, A.M. and Z.M. Nezad, 2010. A new scheduling algorithm design for grid computing tasks. Proceedings of the IEEE World Automation Congress, September 19-23, 2010, Kobe, Japan.

He, L. and T.R. Ioerger, 2004. Task-oriented computational economic-based distributed resource allocation mechanisms for computational grids. Proceedings of the International Conference on Artificial Intelligence, June 21-24, 2004, Las Vegas, Nevada, USA.

Katsaros, K. and G.C. Polyzos, 2008. Evaluation of scheduling policies in mobile grid architecture. Proceedings of the IEEE Performance Evaluation of Computer and Telecommunication Systems, June 16-18, 2008, Edinburgh, pp: 390-397.

Khanli, L.M. and S. Kargar, 2007. Grid-JQA: A QoS guided scheduling algorithm for grid computing. Proceedings of the IEEE 6th International Symposium on Parallel and Distributed Computing, July 5-8, 2007, Hagenberg, Austria.

Kumar, M.R. and N. Kaur, 2007. Job scheduling in grid computers. Proceedings of the National conference on Challenges and opportunities in Information Technology, July 9-12, 2007, Orlando, FL, USA.

Lee, J.H., S.J. Song, J.M. Gil, K.S. Chung, T. Suh and H.C. Yu, 2009. Balanced scheduling algorithm considering availability in mobile grid. Adv. Grid Pervasive Comput., 5529: 211-222.

Litke, A., D. Skoutas and T. Varvarigou, 2004. Mobile grid computing: Changes and challenges of resource management in a mobile grid environment. Proceedings of the 5th International Conference on Practical Aspects of Knowledge Management, December 2-3, 2004, Vienna, Austria.

Martincova, P. and M. Zabovsky, 2007. Comparison of simulated GRID scheduling algorithms. Systemova Integrace, 4: 69-75.

Parsa, S. and R. Entezari-Maleki, 2009. RASA: A new grid task scheduling algorithm. Int. J. Digital Content Technol. Appli., 3: 152-160.

Thenmozhi, S. and A. Tamilarasi, 2010. A hierarchical resource allocation architecture for mobile grid environments. Int. J. Comput. Sci. Inform. Secur., 8: 6-11.

Thenmozhi, S. and A. Tamilarasi, 2011. A Hierarchical Trusted Resource Allocation Architecture (HTRAA) for mobile grid environment. Euro J. Sci. Res., 59: 501-521.