Energy Constrained Hierarchical Task Scheduling Algorithm for Mobile Grids

Arjun Singh [#1], Prasun Chakrabarti [*2], Surbhi Chauhan [@3]

Dept of Computer Science and Engineering, Research Scholar at Suresh Gyan Vihar University, Jaipur
vitarjun@gmail.com
[*]Dept of Computer Science and Engineering, Sir Padampat Singhania University, Udaipur
prasun.chakrabarti@spsu.ac.in
[@]Dept of Computer Science and Engineering, Master in Technology from Amity University, Noida
surbhichauhan2009@gmail.com

Abstract— In mobile grids, scheduling the computation tasks and the communication transactions onto the target architecture is the important problem when a mobile grid environment and a pre-selected architecture are given. Even though the scheduling problem is a traditional topic, almost all previous work focuses on maximizing the performance through the scheduling process. The algorithms developed this way are not suitable for real-time embedded applications, in which the main objective is to minimize the energy consumption of the system under tight performance constraints. This paper entails an energy constrained hierarchical task scheduling algorithm for Mobile Grids to minimize the power consumption of the mobile nodes. The task is rescheduled when the mobile node moves beyond the transmission range. The performance is estimated based on the average delay and packet delivery ratio based on nodes and flows. The performance metrics are analysed using NS-2 simulator.

Keywords—Energy Dissipation; Job Scheduling; Battery Power; MIN-MAX Algorithm

I. INTRODUCTION

A. *Mobile Grids*

In the past few years, mobile devices revolutionize the world from data-access devices to the one that is capable of processing and storing considerable amounts of data. Mobile grid computing is planned for creating the grid services available and accessible anytime, anywhere from mobile devices. The grid users can utilize the limited resources of mobile devices.

Mostly, the existing grid environments do not consider the mobile devices. Grids are really pervasive and mobile for the next generation. Hence the integration of mobile devices is necessary to influence available resources and widen the range of supply services. Mobile devices are limited to processing power, battery life and storage capacity. These constraints lead to the slow application execution and hinder operability. Mobile devices in grid environment may behave as user of grid resources or as grid resources providers. Their integration into the Grid as resource providers (not just as consumers) is very difficult due to the limitation constraints on energy and processing capacity of mobile devices.

There are three methods of mobile device’s integration into the Grid-

a) The mobile devices are just the interfaces to resources available in the Grid system, and do not provide any services.
b) Raw resources like CPU, memory, and storage in mobile devices are used to finish the tasks in the grid environment. This scheme just considers mobile devices as conventional resources to attain goals.
c) This scheme is to develop services in mobile devices to support the mobile services in a mobile grid and to enable mobile devices to provide services. This model is the one that contributes the most complete integration in which the mobile devices can be both consumers and providers of services. Such integration could open up the possibilities in developing the mobile nature of these devices in a grid computing environment.

A distributed computing infrastructure that is used to solve the advanced scientific and engineering problems is termed as grid. It is differentiated from conventional distributed computing systems in its large-scale resource sharing and innovative applications. Conglomerations of several resources with different owners are computational grids. Recently, this architecture has been enhanced to include various mobile devices in order to offer a seamless source of computational power and storage capacity.

Load balancing or job scheduling problem is the efficient assignment of jobs and utilization of resources of unused devices which is a primary consideration of such Mobile Grid Computing systems. The job allocator assigns the jobs to the available resources and tries to optimize a specified performance metric. For example, time deadline or revenue maximization [1][2][3][4].

B. Energy constrained Mobile Grids

The limited battery capacity becomes a constraint and power or energy management becomes an issue when the resources are mobile. As the devices are heterogeneous, battery capacity may also be heterogeneous. To maximize the performance and the cost-effectiveness of the system, the heterogeneity of the resources and tasks in an HC system is developed. A significant research problem is how to allocate resources to the tasks and to order the tasks for execution on the resources to increase some performance criterion of a mobile grid which is called as mapping or resource allocation. A resource allocation will pay attention in allocating resources on a certain system [8] [9] [10].

These papers detect some parameters and resources to provide a solution for resource management and make it energy efficient. Finally, a suitable calculation is done for further process and then gives min-max algorithm and scheduled resource.

To provide the better understanding, this paper gives a suitable introduction in section-one that is followed by the existing work's description present in section two. Then the paper describes the problem definition and gives the best solution for the problem in section three. Finally, the paper gives an appropriate conclusion in the last section.

II. LITERATURE REVIEW

Li Chunlin et al., [11], have proposed a price based distributed energy constrained resource allocation optimization for mobile grids which is formulated as a utility optimization problem. This problem can be decomposed into two sub-problems. The interaction between the two sub-problems is controlled by using a pricing variable. The application utility based on its allocated resources including computation and communication resources and on the consumed energy which causes a coupled utility model. In this model, the utilities are the functions of allocated resources and consumed energy.

Abdul Aziz et al., [12] have proposed a various heuristics for power-aware scheduling algorithms. This algorithm is used for scheduling jobs with dependent tasks onto the computational grid. A power-aware scheduling scheme is discussed that reduces power consumption by changing the status of the workstation to hibernate or offline. The decision of selecting one is a difficult task as every solution has its own set of pros and cons. The implementation with minimum cost has the merit of the lowest power consumption but the response time of the jobs is high. On the other hand, the minimum response heuristics keep the response time to the lowest with more power.

One can design heuristics that can calculate the trade-offs between the two objective functions and help in choosing an intermediate solution that eliminates both extremes. The main drawback of this method is that it cannot operate on heterogeneous grid environment.

Nikolaos D. Doulamis et al., [13] have proposed a new algorithm for fair scheduling and it is compared with the other scheduling schemes such as the Earliest Deadline First (EDF) and the First Come First Served (FCFS) schemes. This algorithm uses a max-min fair sharing approach for providing fair access to users. When there is no scarcity of resources, this algorithm assigns enough computational power to each task to finish within its deadline. When there is congestion, the main idea is to fairly minimize the CPU rates assigned to the tasks so that the share of resources that each user gets is proportional to the user's weight. The weight may be defined as the user's contribution to the infrastructure or the price he is eager to pay for services or any other socioeconomic consideration.

III.PROBLEM IDENTIFICATION AND PROPOSED SOLUTIONS

Arjun Singh and P. Chakrabarti [14] have proposed an ant based resource discovery and mobility aware trust management for mobile grid systems. Initially the super-grid nodes are selected in the network using ant colony optimization based on the parameters such as distance, CPU speed, available bandwidth and residual battery power. These selected nodes are utilized in the resource discovery mechanism. In order to maintain strong security with mobility management system, a proficient trust reputation collection method has been adopted. By simulation results, author shows that the proposed approach is efficient and offers more security.

One important problem is scheduling the computation tasks and the communication transactions onto the target architecture when a mobile grid environment and a pre-selected architecture is given. This includes-

- Deciding the assignment of tasks and communication transactions onto the different computation and communication resources, respectively, and
- Fixing the order of their execution on these shared resources which is referred as the scheduling problem. The solution to the scheduling problem has a significant impact on the total system energy consumption because Due to the heterogeneity of the architecture, assigning the same task to different processing elements leads to very different computation energy consumption.
- For different task assignments, the inter-task communication volume and the routing path can vary significantly which leads to very different values for the communication energy consumption.

Although the scheduling problem is a conventional topic for research, almost all existing work focuses on maximizing the performance through the scheduling process. The algorithms developed are not suitable for real-time embedded applications, in which a common objective is to reduce the energy consumption of the system under tight performance constraints. Moreover, most of the existing work neglects the inter-processor communication aspects during the scheduling process, or assumes a fixed delay proportional to the communication volume, without considering the subtle effects like the communication congestion which may change dynamically throughout tasks execution. Fig.1 shows the architecture of mobile grid network.

A. Estimation of Power

After time t, the power consumed by the node (P_c (t)) is computed as follows-

$$P_c(t) = DP_{tx} * a1 + DP_{rx} * a2 \quad (1)$$

Where DP_{tx} = Number of data packets transmitted by the node after time t.

DP_{rx} = Number of data packets received by the node after time t.

a1 and a2 are constants in the range of (0,1).

If P_i is the initial battery power of a node, the residual battery power [9] of a node at time t, can be calculated as:

$$P_{res} = P_i - P_c\,(tc) \quad (2)$$

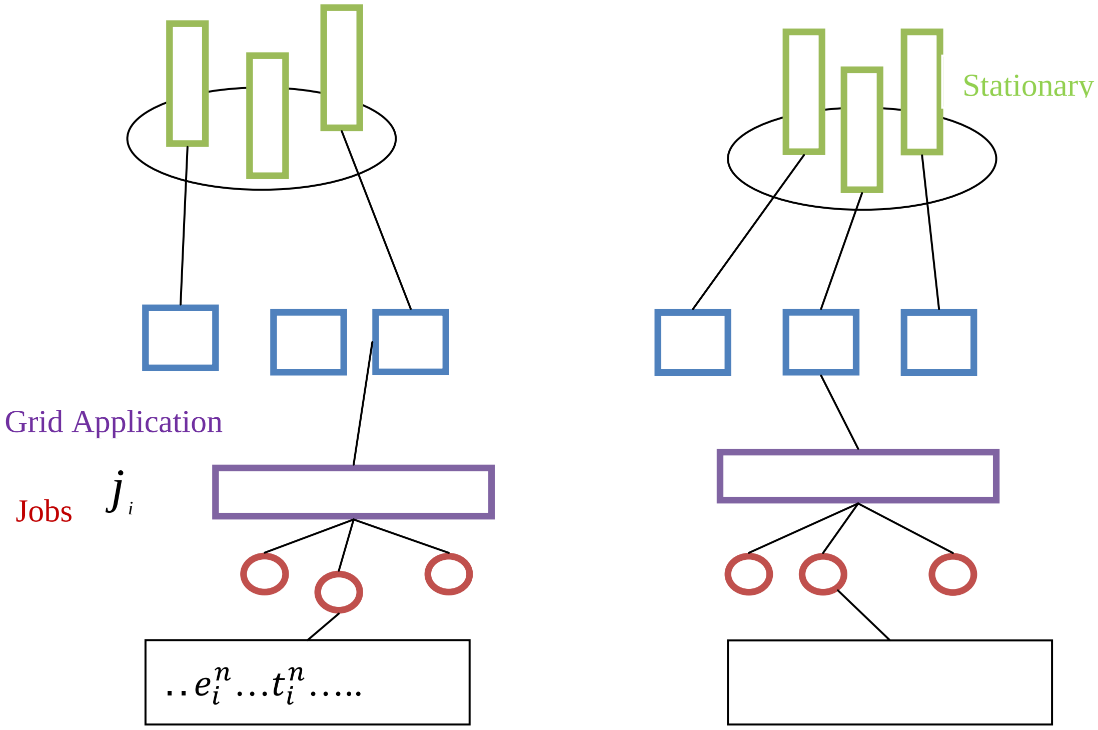

Fig. 1: Mobile Grid Network Architecture

B. Hierarchical Task Scheduling

The profile of MN contains the following fields:

- mnid (Mobile Node Id)
- imgsid (I-MGS Id)
- speed (speed of MN)
- no_tasks (no. of tasks currently executing)
- wload (current work load)
- power (power level of the node)
- ntasks (details of current tasks)
- status (status of the node)
- For each task, the following details are stored.
- tid (Task id)
- tsize (memory size required)

- exectime (execution time)

C. Algorithm

i. Whenever a MN wants to execute a set of tasks {T_j, j=1,2….k}, it submits the task details to its SGN.

ii. The SGN estimates the workload of its MNs {MN_i,i=1,2,,,,n} using the following formula:

$$Wload_i = CWload_i + \left(\sum_{j=1}^{k} tsize_j \, / power_i \right)$$

Where- $CWload_i$ is the current workload of MN_i,

$Wload_i$ is the work load of MN_i,

$power_i$ is the power of the mobile MN_i and $tsize_j$ is the size of the task j.

So total work load is –

$$TWload = \sum_{i=1}^{n} Wload_i$$

i. Let avg is the average workload of the grid and T is the tolerable limit of the work load.

ii. If $Wload_i$ > avg + T, then the status of MN_i is marked as OVERLOADED.

iii. If $Wload_i$ < avg, then the status of MN_i is marked as UNDERLOADED.

iv. If $Wload_i \geq$ avg and $Wload_i \leq$ avg + T, then the status of MN_i is marked as BALANCED.

v. Let n_ul, n_ol and n_bl denote the number of under-loaded MNs, no. of overloaded MNs and no. of balanced MNs, respectively.

vi. Then SGN check the following condition-

If ntasks > (n_ul + n_bl), then, split the tasks into subtasks as

$$stask = ntasks/n\_ul + n\_bl$$

Otherwise, don't split the task.

vii. The SGN assigns the subtasks to the under-loaded and balanced MNs only.

viii. If there is no under-loaded and overloaded MNs, the SGN sends a request to another SGN with the details of the remaining tasks to be executed.

ix. SGN then forwards this request to other SGN.

x. The same process is repeated until all the sub-tasks are successfully assigned.

xi. Once the execution of the sub-tasks is over, the corresponding MNs return the completed task to the requested MN.

xii. If any one of the assigned MN moves out of the range, then its task is again rescheduled to another MN, in the same way described above.

D. Algorithm Description

Whenever a MN wants to execute a set of tasks, it submits the task details to its SGN. The SGN estimates the workload of its MNs using the following formula:

$$Wload_i + \left(\sum_{j=1}^{k} tsize_j / power_i \right)$$

Here Workload is represented in terms of the task size and battery power of the mobile node. So the mobile node with least workload and more battery power can be selected to execute the tasks. The status of mobile node is marked as balanced, overloaded and under-loaded depending on the average work load. The status of mobile node is marked as overloaded, if the work load is greater than the sum of the average workload and the tolerable limit of the work load. The status of mobile node is marked as under-loaded, if the work load is lesser than the average workload.

On the other hand, if the work load is between the average work load and the sum of average work load and the tolerable limit, then the mobile node is considered as balanced. After marking the status of the mobile nodes, the SGN checks whether the number of tasks is greater than the sum of the number of under-loaded and the number of balanced mobile nodes and splits the tasks into subtasks. Otherwise, the tasks are not split into subtasks.

The SGN assigns the subtasks to the under-loaded and balanced MNs only. If there are no under-loaded and overloaded MNs, the SGN send a request to its I-MGS with the details of the remaining tasks to be executed.

The I-MGS then forwards this request to its other SGN. The same process is repeated in other SGN until all the sub-tasks are successfully assigned. Once the execution of the sub-tasks is over, the corresponding MNs return the completed task to the requested MN. If any one of the assigned MN moves out of the range, then its task is again rescheduled to another MN.

IV. ADVANTAGES OF PROPOSED APPROACH

Following are the notable advantages of proposed approach-

i. It minimizes the power consumption of the mobile nodes.

ii. Since it is hierarchical, work can be redistributed or rescheduled.

iii. Since the tasks are split into sub-tasks and executed parallel, it reduces the execution time and memory size.

iv. If a mobile node moves out of range, its task can be rescheduled.

V. SIMULATION AND RESULTS

NS2 [20] is used to simulate the proposed algorithm. Fig. 2 gives the sample network topology used in the simulation. In this simulation, 30 grid nodes are deployed in a 1000 meter x 1000 meter region for 50 seconds simulation time. Among the total 30 grid nodes, 20 nodes act as mobile grid nodes (indicated as “G” and blue colour in figure 4) and 10 nodes act as super grid nodes (indicated as “SG” and red color in figure 4).

It is assumed that each mobile grid node moves independently with the same average speed. The speed of the mobile grid node is 2m/s. All nodes have the same transmission range of 250 meters. For Service discovery Service Level Protocol is used.

A SLP service agent is attached to the nodes for providing the services and SLP user agent is attached to the

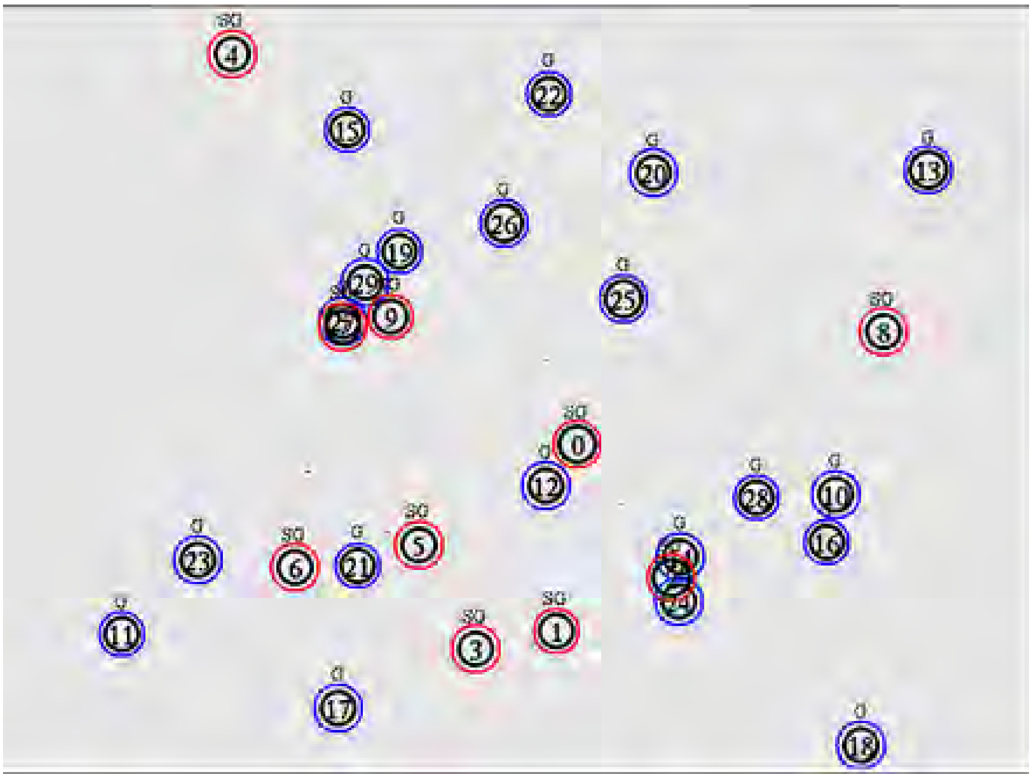

Fig. 2: Simulation Topology

clients for requesting the service. In this simulation, clients send service requests to the super grid nodes. The super grid nodes select the grid nodes matching the service request and assign the tasks as per our algorithm.

The simulation settings and parameters are summarized in table 1.

TABLE I
Simulation Parameters

No. of Grid Nodes	20
No. of Super Grid Nodes	10
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Simulation Time	50 sec
Service Discovery Protocol	SLP
Server Application	SLPsa
Client Application	SLPua
Speed	10m/s
Clients	4
Requested Load	10 to 50kb
No. of Requests	1,2,3,4 and 5.

Energy Constrained Hierarchical Task Scheduling (ECHTS) algorithm compared with EDF technique [13]. According to the following matrices performance of the algorithm is evaluated:

- *Average Delay:* It is measured as the average delay occurred for each client while getting the requested service.
- *Packet Delivery Ratio:* It is the ratio of number of the nodes for the successful data transmission.
- *Energy Consumption:* It is the average energy consumed by all the mobile grid nodes.
- *Throughput:* It is the average throughput obtained at the receiver.

A. Based on Number of Requests

The number of requests is varied from 1 to 5 with load 10Kb.

Figure 3 shows the delay of ECHTS and EDF techniques for different number of flows scenario. We can conclude that the delay of our proposed ECHTS approach has 47% of less than EDF approach.

Figure 4 shows the delivery ratio of ECHTS and EDF techniques for different number of flows scenario. We can conclude that the delivery ratio of our proposed ECHTS approach has 2% of higher than EDF approach.

Figure 5 shows the throughput of ECHTS and EDF techniques for different number of flows scenario. We can conclude that the throughput of our proposed ECHTS approach has 23% of higher than EDF approach.

Figure 6 shows the energy consumption of ECHTS and EDF techniques for different number of flows scenario. We can conclude that the energy consumption of our proposed ECHTS approach has 9.7% of less than EDF approach.

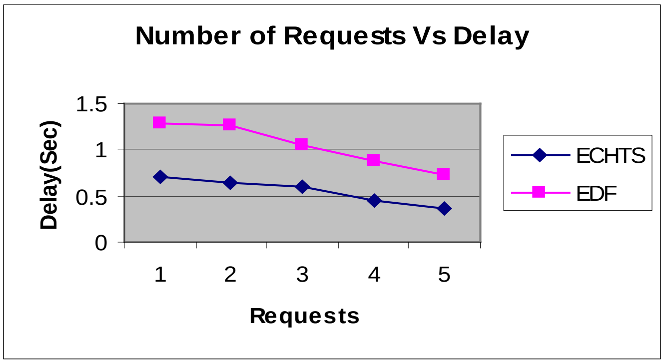

Fig. 3. Request vs Delay

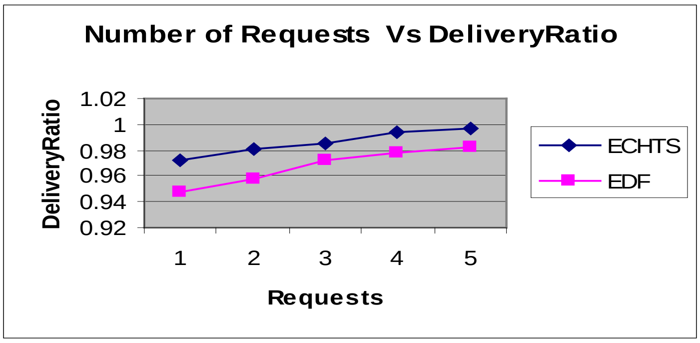

Fig. 4. Request vs Delivery Ratio

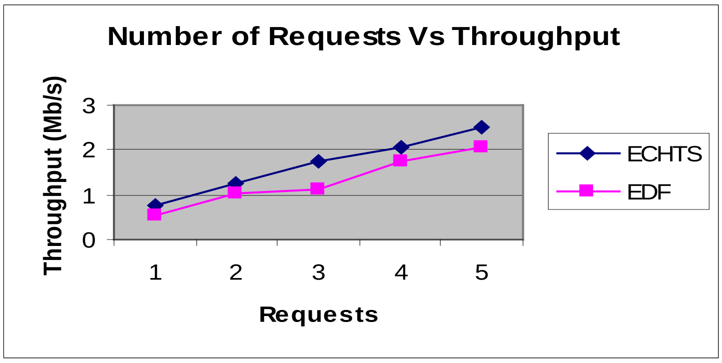

Fig. 5. Request vs Throughput

B. *Based on Load*

C. Vary the load of the request as 10 to 50 kb for 2 requests.

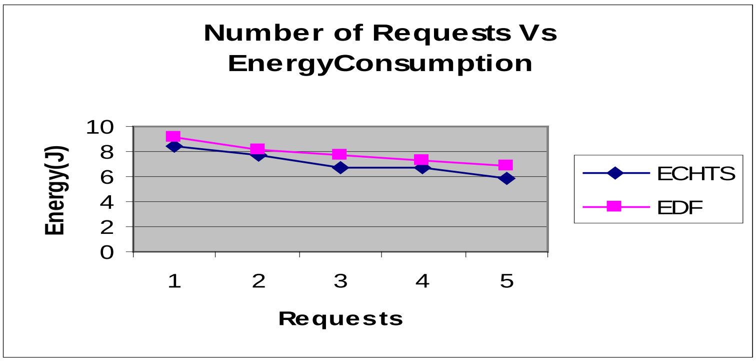

Fig. 6. Request vs Energy Consumption

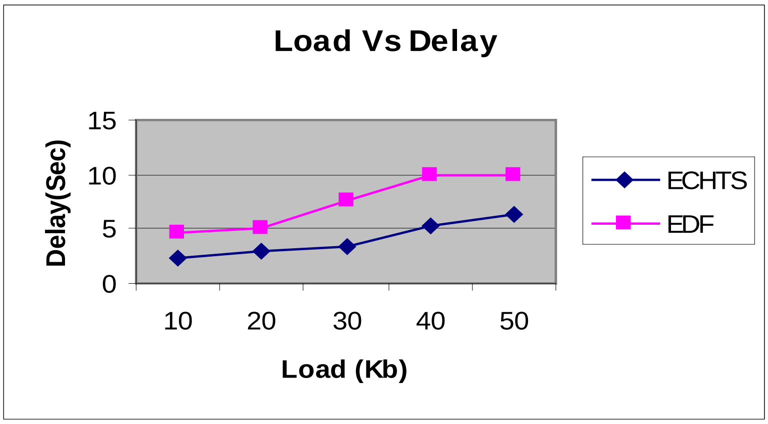

Fig.7. Load vs Delay

Figure 7 shows the delay of ECHTS and EDF techniques for different rate scenario. We can conclude that the delay of our proposed ECHTS approach has 45% of less than EDF approach.
Figure 8 shows the delivery ratio of ECHTS and EDF techniques for different rate scenario. We can conclude that the delivery ratio of our proposed ECHTS approach has 19% of higher than EDF approach.

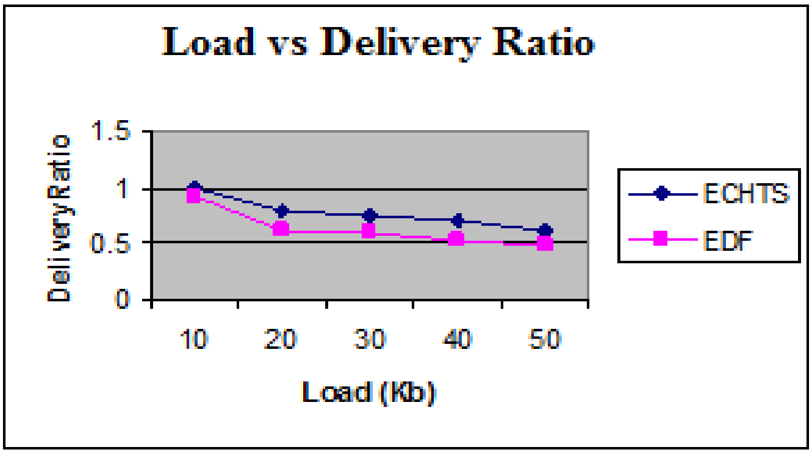

Fig: 8 Load vs Delivery ratio

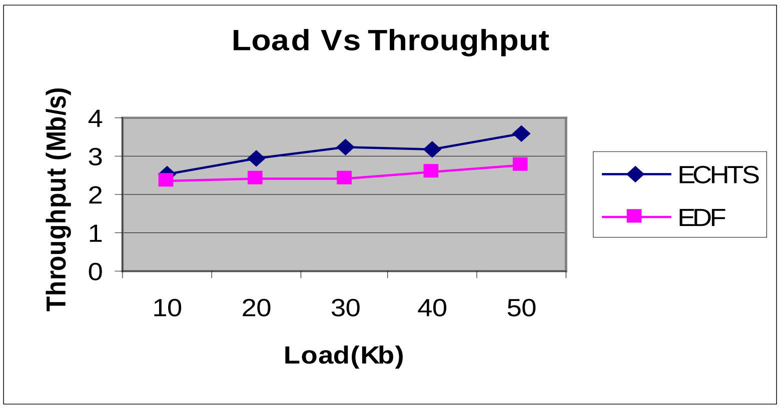

Fig 9: Load Vs Throughput

Figure 9 shows the throughput of ECHTS and EDF techniques for different rate scenario. We can conclude that the throughput of our proposed ECHTS approach has 18% of higher than EDF approach.

Figure 10 shows the energy consumption of ECHTS and EDF techniques for different rate scenario. We can conclude that the energy consumption of our proposed ECHTS approach has 11% of less than EDF approach.

VI. CONCLUSION

In this paper, an energy constrained hierarchical task scheduling algorithm for Mobile Grids is proposed to minimize the power consumption of the mobile nodes. The task is rescheduled when the mobile node moves beyond the transmission range. Scheduling is necessary due to the need to address the deadline constraints and system heterogeneity. The performance is estimated based on the average delay; that is reduced, packet delivery ratio; that is increased, fairness and bandwidth; that is slightly increased, and energy consumption; that is decreased based on nodes and flows. By simulation, the performance metrics are analysed using NS-2 simulator.

REFERENCES

[1] S Rodriguez Juan Manuel, Alejandro Zunino, and Marcelo "Campo 2010 Mobile grid seas: simple energy-aware scheduler," Proc. 3rd High-Performance Computing Symposium-39th JAIIO

[2] Ghosh, Preetam, Nirmalya Roy, Sajal K. Das, and Kalyan Basu 2004 "A game theory based pricing strategy for job allocation in mobile grids," In Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International, p. 82. IEEE.

[3] Li, Chunlin, and Layuan Li 2012 "Simultaneous Optimization of Application Utility and Consumed Energy in Mobile Grid," Computing and Informatics 29, no. 6+ (2012): 1117-1140.

[4] Ghosh, Preetam, et al 2005 "A pricing strategy for job allocation in mobile grids using a non-cooperative bargaining theory framework," Journal of Parallel and Distributed Computing 65.11 (2005): 1366-1383.

[5] Sim, Kwang Mong 2006 "A survey of bargaining models for grid resource allocation," ACM SIGecom Exchanges 5.5 (2006): 22-32.

[6] Khan, Samee Ullah, and Cemal Ardil 2009 "Energy efficient resource allocation in distributed computing systems," International conference on distributed, high-performance and grid computing.

[7] Prosperi, Francesco, Mario Bambagini, Giorgio Buttazzo, Mauro Marinoni, and Gianluca Franchino. "Energy-Aware Algorithms for Tasks and Bandwidth Co-Allocation under Real-Time and Redundancy Constraints"

[8] Kim, Jong-Kook, Howard Jay Siegel, Anthony A. Maciejewski, and Rudolf Eigenmann. "Dynamic resource management in energy constrained heterogeneous computing systems using voltage scaling," Parallel and Distributed Systems, IEEE Transactions on 19, no. 11 (2008): 1445-1457.

[9] Tekbiyik, Neyre, Tolga Girici, Elif Uysal-Biyikoglu, and Kemal Leblebicioglu. "Proportional Fair Resource Allocation on an Energy Harvesting Downlink-Part II: Algorithms," arXiv preprint arXiv:1205.5153 (2012).

[10] Malla, Samip, Birendra Ghimire, Mark C. Reed, and Harald Haas. "Energy efficient resource allocation in OFDMA networks using ant-colony optimization," In Communications and Information Technologies (ISCIT), 2012 International Symposium on, pp. 889-894. IEEE, 2012.

[11] Li, Chunlin, and Layuan Li. "Energy constrained resource allocation optimization for mobile grids," Elsevier, 2009.

[12] Aziz, Abdul, and Hesham El-Rewini, "Power efficient scheduling heuristics for energy conservation in computational grids," Springer, 2011.

[13] Doulamis, Nikolaos D., Anastasios D. Doulamis, Emmanouel A. Varvarigos, and Theodora A. Varvarigou. "Fair scheduling algorithms in grids," Parallel and Distributed Systems, IEEE Transactions on 18, no. 11, 2007.

[14] Singh Arjun, Chakrabarti P. "Ant based Resource Discovery and Mobility Aware Trust Management for Mobile Grid Systems," IACC IEEE 3rd International Conference Feb 2013 page 637-644.

[15] Preetam Ghosh and Sajal K. Das. "Mobility-aware cost-efficient job scheduling for single class grid jobs in a generic mobile grid architecture," Future Generation Computer Systems, In Press, Corrected Proof:–, 2009.

[16] Chunlin Li and Layuan Li. "Energy constrained resource allocation optimization for mobile grids," Journal of Parallel and Distributed Computing, 70(3):245–258, 2010.

[17] W. X. Shen, C. C. Chan, E.W. C. Lo, and K. T. Chau. "Estimation of battery available capacity under variable discharge currents," Journal of Power Sources, 103(2):180 – 187, 2002.

[18] Ming-Chiao Chen, Jiann-Liang Chen, and Teng-Wen Chang. "Android/Osgi-Based Vehicular Network Management System," Computer Communications, In Press, Corrected Proof 2010.

[19] Ashish Agarwal, Douglas O. Norman, and Amar Gupta. "Wireless Grids: Approaches, architectures, and technical challenges," Working papers 4459-04, Massachusetts Institute of Technology (MIT), Sloan School of Management, 2004.

[20] Network Simulator: http:///www.isi.edu/nsnam/ns