International Journal of Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-3, Issue-2)

Research Article | **February 2014**

Improve Resource Utilization by Task Scheduling in Cluster Computing

Harvinder Singh*
Research Scholar, PTU Jalandhar,
Punjab, India

Dr. Gurdev Singh
Professor, Gurukul Vidyapeeth, Ramnagar,
Banur, Patiala, India

Abstract—

Parallel and distributed computing technique is the solution of many high performance computing requirements in business purpose and research operations. Since parallel computing became useable numbers scheduling algorithms proposed for performance improvement and resource utilization. In this paper we consider scheduling problem for resource allocation to tasks in cluster computing. In addition discussion about most related existing work on task scheduling in cluster computing. Survey concludes existing work require more efficient scheduling algorithm which improve resource utilization and minimize makespan and computation cost. We proposed swarm based scheduling technique which improves resource utilization with best mapping of tasks on resource.

Keywords—cluster computing, task scheduling, high performance, makespan, swarm.

I. INTRODUCTION

Parallel computing is use of concurrently execution of computational problem to reduce work completion time. These computers connected via high speed network devices. Parallel processing used for high performance computing in various scientific and commercial research problems. Parallel computing technique developed from serial computers to overcome single state of work. Parallel computing is a machine of multiple interconnected computers. Performance is improved by application partitioning and scheduling tasks on interconnected nodes to execute concurrently. Parallel execution reduces total execution time by scheduling of tasks on multiple interconnected processing units comparing to serial execution. Parallel computing aims to faster execution and minimize makespan.

Scheduling techniques improve system efficiency by partitioning tasks into sub-tasks and execute separately [1]. It reduces interference, exploitation of specialization work, faster execution work and performed better results. Task partitioning and scheduling get benefits only, if it provided cost effective results. Task dependency and communication cost naturally resides in sub-tasks, minimize communication between them perform best result.

Task scheduling basically in two categories, static and dynamic; static scheduling observe all information related execution cost, communication cost and node power at compile-time; while dynamic do same things at run time. Task scheduling is key factor to achieve high utilization resources, improve performance. Numbers of dedicated nodes execute parallel application concurrently. Task partitioning, grouping, mapping and allocation of tasks, best node selection is challenging problem. Task scheduling performs the parallel execution of tasks onto multiple nodes that minimize overall completion time of execution. Parallel application in which task execution require different computation speed and task scheduling technique fulfill this objective. Optimal task scheduling is required to achieve high performance computing in these applications on cluster environment.

Task scheduling on cluster environment has represents to faster results [2]. Many task scheduling algorithms work have done already in this field and some work requires for more cost effective use of cluster computer to increase performance using new techniques. We proposed swarm based biological method for scheduling problem in cluster system.Task scheduling algorithms decomposes the computational problem and run concurrently on interconnected*n*-nodes. Our present parallel computing model is dedicated and clusterenvironment, where each node has its own processing unit and memory. Our goal of task scheduling is to minimize makespan and high utilization of system resources.

The studies of cluster algorithm reveal a number of challenges, include high resource utilization, dynamic load balancing, and minimize make-span, performance improvement, which direct related with scheduling of task to nodes. In this research study, we use swarm technique on cluster computing environment for parallel application, to achieve high utilization of resources, dynamic load balancing, make-span, performance improvement, with proposer scheduling of task to nodes. Swarm technique is multidimensional search space and keeps track for best solution using intelligent agents. Each agent know its current solution and moving for best solution, it has found during the search.

This paper is organized as described sections: section 2 brief discussion on related work, section 3 problem formulation, section 4cluster computing model, section 4 on task scheduling techniques, section 5 presents existing task scheduling techniques, section 6 proposed methodology for scheduling . In final section 7 conclusion and future work has given.

II. RELATED WORK

In this section, most relevant work presented related to scheduling algorithm in cluster computing performance is based on scheduling scheme, mapping is consider cluster to become profitable platform. Most of work is based on static, dynamic and duplication scheduling techniques.

Kwok proposed classification of scheduling algorithms with different environment and evaluate the results [3]. In case of heterogeneous pc cluster Kwok proposed scheduling algorithm for performance improvement. It reduces task execution start time of critical path task allocate to faster machines using replication to that task [4].Andrei presents modified version of FCP and FLB task scheduling algorithm for heterogeneous system. These are compile-time based and following list scheduling approach. Both perform sort the tasks with fully and partially sorted lists which perform better with low cost in heterogeneous system [5]. Topcuoglu proposed HEFT, which performs task prioritizing and processor selection to minimize finish time. HEFT outperform than list scheduling algorithm to improve task's finish time [6].Boeres proposed clustering for heterogeneous processor (CHP) to reduce communication cost and makespan. It uses cluster mapping phase and processor p choice to choose which processor minimize makespan. It perform better on fine grain or highest number of faster processors [7].Tarek (2004) presents HCPFD for compile-time based task scheduling algorithm perform on bounded number of heterogeneous system. It follows the approach list scheduling and task duplication for scheduling [8].Iliasproposed dynamic load balancing based algorithm based on parallel (n- 1)dimensional hyper grids and scan the work according to relative power of each hyper-grid and entire system load. PSLB scan sender and receiver then migrate tasks to appropriate nodes. This algorithm move overloaded node task ti, from current location vi, to a node vj,it leads to system load balancing [9]. Barbosa proposed list scheduling algorithm on heterogeneous clusters aims to minimize makespan of parallel tasks, whose dependencies are represented by a DAG [10].Prashnth proposed HLTF algorithm, it used meta tasking process to minimize makespan, priority to largest ones but performance affected by communication overheads [11]. Xiao and P-Zhong 2009] The ILTB algorithm is used to balance the load and produce high throughput for the heterogeneous system. The core idea of ITLB algorithm is to move the tasks on the node with latest finish time to the node with earliest finish time [12]. Jiannongpresents the classification of scheduling tools used for parallel environment. Presents cluster computers uses in various fields and application for high performance computing. Task scheduling techniques classify in this paper for cluster computers and emphasis to develop best scheduling strategy for better performance [13]. Ziliang proposed duplication method EETDS on heterogeneous cluster to reduce energy consumption and improve performance [14].Chabbra evaluated two numeric applications on cluster computer and performance change with number of nodes in cluster and high communication affected the performance [15].Ziliang proposed EAD and PEBD scheduling algorithm based on task duplication on homogeneous cluster. These algorithms duplicate tasks to nodes with objectives performance and energy efficiency. Shorten task schedule length and optimize energy consumption on clusters [16].Jing proposed Energy-Aware Scheduling by Minimizing Duplication (EAMD) duplication based algorithm with minimizing extra time and energy consumption on heterogeneous system. It minimizes makespan, improve performance and delete abundant task copies generated by task duplication [17]. Ahmad proposed task scheduling algorithm onto multiprocessor using swarm technology which performed better results than genetic and duplication based techniques [18]. Li proposed SDLS to minimize makespan and improve speedup controlling processing and communication time on clusters [19].

III. PROBLEM FORMULATION

Cluster computing model have different clusters with varies number of nodes and configuration. performance on cluster framework can be achieve only with strategic scheduling plan. In distributed environment nodes runs independently with workflow of applications stand nodes on queue and it is wastage of resource. In this section scheduling problem is formulated as to improve resource utilization and maximize resource profitable. It direct impact on system performance and minimize makespan. Equation1 is used to formulate resource utilization as against to system total execution power.

U : utilization of system resources

Cn : number of machines in cluster

Mn :machines configuration within cluster

Ln : current load of the system

tn : number of tasks currently execute

Cn : communication dependencies

$$U = Cn, Mn + CLn(tn + Cn) \quad (1)$$

For high utilization, minimize makespan requireoptimal solution and we proposed swarm based techniques for scheduling. This optimization technique capable to generate robust solution which improveperformance in cluster environment.

IV. CLUSTER COMPUTING MODEL

Parallel processing is the use of two or more processors working on the principle that large problem divide into smaller ones then solve concurrently. Cluster computing has advances of parallel computing which composed different

configuration of hardware, software and network devices. Hardware: Each node has processor, memory and storage capacity and run as dedicated and parallel. Software: Different OS, development tools, message passing environment. Network: Variety of network cards, switches, protocols.

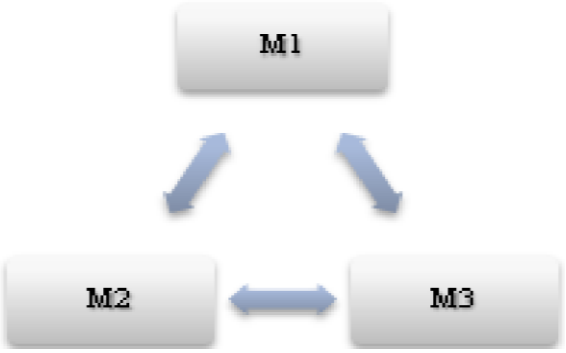

Fig. 1.Cluster Computing Model

Modern scientific and commercial purpose applications required high speed to run and get faster results but single computer consumes too much time to finish work. Proposed application model is run as single program on multiple nodes (SPMD). Partitioning and allocation of tasks perform at compile time and runtime. Optimal task scheduling is required to achieve high performance computing on parallel environment. Parallel application in which tasks execution required different computation speed and task scheduling technique fulfill this objective.

V. SCHEDULING TECHNIQUES

A. Static Scheduling

Static task scheduling performed during compile time. Task partitioning and allocation of tasks to nodes by pre-estimation of task execution time node computation power, task dependency and communication between nodes. It required high memory storage for compile time estimation. It does not perform runtime reallocation and leads node waiting time. Static task scheduling in cluster reduce communication between nodes but leads to node waiting time.

B. Dynamic Scheduling

Dynamic task scheduling in cluster reduce node waiting time. Runtime reallocations tasks assign to nodes minimize node idle time. Large memory space not required for compile time estimation. Dynamic task scheduling performed communication between nodes during runtime to collect tasks and nodes information and allocate tasks to nodes. Cluster nodes always come with different configuration of nodes. So dynamic task scheduling utilizes high system resources and improves performance. Many algorithm proposed for dynamic task scheduling on cluster system.. Dynamic task scheduling is more suitable than static task scheduling in cluster environment.

C. Duplication

Task duplication based technique use in cluster reduces the node idle time. In redundantly run one task on more than one node which control dependent nodes waiting time for input task; duplication technique allocates current execution task from parent node as input to other node. It leads high energy consumption and task length. Many algorithm proposed study for duplication and non-duplication for high performance computing power.

Task duplication based technique uses for high performance computing minimize node waiting time.. It is suitable for dependent task execution where input task required for another node execution. In cluster environment where large and complex applications running task duplication does not perform optimal solution. Duplication techniques minimize communication where task length is minimized.

VI. PROPOSED METHODOLOGY FOR SCHEDULING

The basic approach of this research work is todevelop efficient scheduling algorithm to improve resource utilization in cluster computing. Mapping of tasks on resources in distributed environment is optimization problem which require convergence technique to get best solution. Swarm is collective behavior of agents which are working in decentralized fashion. These agents are capable to generate self-decision as well as group sharing information to reach out on global best solution, known as best mapping solution. Sharing of information between them in dynamic mode and process continuously, which can implement for scheduling to collect information about machine execution capability and load status. Therefore system performance improve drastically, as resource utilization improve with dynamic updation of load and sharing information with other machines.

Swarm technique use the intelligent agentswhich search in *n*- dimensionalas shown in figure 2 for personal best pbi and global best gbi from existing population size with inertia weight. Many cluster computing based methods have been

proposed to solve the scheduling problem. These methods are highly diverse in terms of their assumptions about the structure of the parallel program and the target parallel architecture.

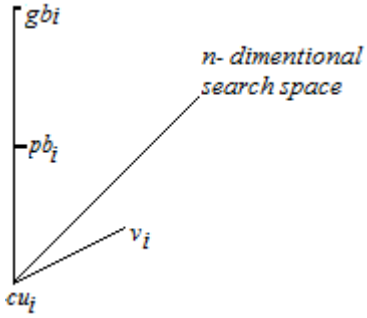

Fig. 2.Swarm agents search for best solution

VII. CONCLUSION

Cluster computing used for high performance, reliability and availability in various fields of scientific and research. Numbers scheduling algorithms proposed for high performance on cluster computing environment. Scheduling methods depends on application nature and cluster environment. Cluster environment is more capable for complex and dynamic application. Scheduling of tasks with various schemes affected the utilization and results.

Future work can be done on task scheduling for resource utilization and performance improvement on cluster computing using swarm technique. Swarm is biological distributed process which search through defined area upto number of maximum generates and upon find mapping tasks on resource with minimum computation cost and time. So, swarm can use to produce results better to become system as profitable resource.

REFERENCES

[1] S. Oliver, "Task Scheduling for Parallel System," John Wiley & Sons,Inc, 2007.
[2] M. Baker, A. Apon, R. Buyya, H. Jin, "Cluster Computing and Applications," Encyclopedia of Computer Science and Technology, vol.45, Marcel Dekker, Aug. 2001.
[3] Y.K. Kwok, Ahmad I., "Static Scheduling Algorithms for Allocating Directed Task Graphs to Multiprocessors," ACM, vol.31, no.4, 1999.
[4] Kwok, "Parallel Program Execution on a Heterogeneous PC Cluster using Task Duplication," IEEE, 2000.
[5] R. Andrei and J.C. Arjan, "Fast and Effective Task Scheduling in Heterogeneous System," IEEE, 2000.
[6] H. Topcuoglu, S. Hariri and M.Y. Wou,"Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing," IEEE trans. Parallel and Distributed System, Vol 13, No.3, p 260-274, 2002.
[7] C. Boeres, J.V. Filho and V.E.F. Rebello, "A cluster-based strategy for scheduling task on heterogeneous processors," pp. 214-221, IEEE, 2004.
[8] H.Tarek and J.Jan, "A High Performance, Low Complexity Algorithm for Compile-Time Task Scheduling in Heterogeneous Systems," IEEE, 2004.
[9] K. Ilias and K. M. Tahar, "Dynamic Task Scheduling in Computing Cluster Environments," Proc. of the ISPDC/heteroPar'04, IEEE, 2004.
[10] J. Barbosa, C. Moris, R. Nobrega and A.P. Monterio, "Static scheduling of dependent parallel tasks on heterogeneous clusters," pp. 1-8, IEEE, 2005.
[11] C S. Prashanth, Baskiyar Sanjeev, "A Low Complexity Algorithm for Dynamic Scheduling of Independent Tasks onto Heterogeneous Computing Systems," ACM, 2005
[12] Q. Xiao and J. Hong, "A dynamic and reliability-driven scheduling algorithm for parallel real-time jobs executing on heterogeneous clusters," Vol.65, no. 8, pp.885-900, Journal of Parallel and Distributed Computing, ELSEVIER, 2005.
[13] C. Jiannong, C. Alvin T.S., S. Yudong, K.D. Sajal , G. Minyi, "A taxonomy of application scheduling tools for high performance cluster computing," Springer, 2006, p. 355-371
[14] Z.Ziliang, Q. Xiao, R. Xiaojun, B. Kiranmai, N. Mais, A. Mohamed, "Energy Efficient Scheduling for Parallel Applications Running on Heterogeneous Clusters," Int'l Conf. on Parallel Processing(ICPP), sept 2007.
[15] C. Amit and S. Gurvinder, "A Cluster Based Parallel Computing Framework(CBPCF) for Performance Evaluation of Parallel Applications," IJCTE, vol.2 no.2 p.226-232, 2010.
[16] Z. Ziliang, M. Adam, R. Xiaojun and Q. Xiao, "EAD and PEDB: Two Energy-Aware Duplication Scheduling Algorithms for Parallel Tasks on Homogeneous Clusters," vol. 60, no. 3, pp. 360-374, IEEE, 2011.

[17] M. Jing and L. Kenli, "Energy-Aware Scheduling Algorithm with Duplication on Heterogeneous Computing Systems," ACM/IEEE 13th Int. Conf. on Grid Computing, p. 122-129, IEEE, 2012.

[18] B. Ahmad Ali and A. Shatnawi, "Static Scheduling of directed acylic data flow graphs onto multiprocessors using particle swarm optimization," Comp. & Operations Research 40, pp. 2322- 2328, ELSEVIER, 2013.

[19] Li. K., Tang. X., Veeravalli. B. and Li. K., "Scheduling Precedence Constrained Stochastic Tasks on Heterogeneous Cluster Systems," vol PP, no.99, IEEE, 2013.

[20] R.C Eberhartt, Y Shi, "Particle Swarm Optimization: Developments Applications and Resources," IEEE ,P. 81-86, 2001.

[21] K. Qinma and H. Hong, "A novel discrete particle swarm optimization algorithm for meta-task assignment in heterogeneous computing systems," pp. 10-17, ELSEVIER, 2011.

[22] Dorigo. M, Birattari M., Stiizle T., "Ant Colony Optimization," IEEE Computational Intelligence Magzine, Nov. 2006.