Duplication Based Hybrid Genetic Algorithm for Multiprocessor Task Scheduling : An Overview

Madhuri Yadav[#1], Sunita Dhingra[*2]
[1]*M.Tech. scholar CSE Department, UIET,Maharshi Dayanad University*
Rohtak (India)
[1]*madhuriyadav.me@gmail.com*
[2]*Assistant Professor CSE Department, UIET,Maharshi Dayanad University*
Rohtak (India)
[2]*sunitadhingramdu@rediff.com*

***Abstract*— Parallel processing of a program refers to the process of dividing the program code and running them to multiple processors. Multiprocessor task scheduling is about distributing the program code to the multiple processors. The problem of task being optimally distributed among processor is Hard in nature. As the problem is NP-Hard, finding an optimal solution is difficult with traditional strategies, so it is decided to use hybrid approach making the system to take advantage of both heuristic and Meta heuristic approach resulting better possible solution.**

***Keywords*- Multiprocessor task scheduling, Genetic Algorithm, Makespan, Lower Bound Algorithm.**

I. INTRODUTION

Scheduling is a process of determining when an activity should start or end, depending on its specification. In parallel processing the objective is to minimize parallel program completion time by properly mapping the task to the processor. This requires balancing the computation load uniformly among the processor while simultaneously keeping the communication costs as low as possible [2]. In multiprocessor task scheduling multiple tasks are to be processed on multiple processors. These tasks are assumed to be non pre-emptive. Multiprocessor task scheduling is an NP-Hard problem with an objective of minimizing the length of schedule, also called makes span. Some heuristic approaches are used to solve the problem of Multiprocessor task scheduling .Multiprocessor task scheduling heuristics are problem specific ,so it is not favourable to use heuristic approach alone.

Directed Acyclic Graph is the well known representation for task assignment problem, where nodes and edges represent the task and communication dependency between them, as there are too many parameters to consider, the problem of scheduling becomes difficult. For NP-Hard problems it is not beneficial traditional approach, but by using hybrid approach (where a combination of heuristic and Meta heuristic will be used to take benefit from both approaches) we can have nearby optimal solution. Some widely used heuristics are Min-Min, Max Min, MCT (Minimum Completion Time), simulated annealing and genetic algorithm.

In problem space there are some problems for which very less information is known, for such type of problems genetic algorithm gives better solution. Genetic algorithm is a Meta heuristic approach tries to solve any problem in large space as it is a general algorithm. Genetic algorithm is useful where there is a large solution space and we are to find optimal solution. Genetic algorithm tries to find out the best possible solution nearest to the optimal solution. Selection and crossover are used to produce solution space. sometimes when there is very large and complex solution space it becomes difficult to find an optimal solution, in this paper it is decided to fed an input (that is the result of heuristic approach) to genetic algorithm, for finding the nearest optimal solution. There is large solution space for the multiprocessor task scheduling and it is expected that randomly generated initial solution provide the weak results.[1] to strengthen the results , duplication based task scheduling is proposed to produce initial population and this solution is fed to genetic algorithm as an input which will give best optimal schedule as an output.

II. RELATED WORK

The basis of research in this area is been parallel processing. Since then many approaches in this field have been developed to solve the Multi-Processor Task Scheduling (MPTS) problem. Some of them are based on evolutionary approaches ([12],[3],[14]); some of them follow on heuristic method ([9],[11]) and some rely on the hybrid approach([10],[15]).

Most of the scheduling algorithm are based on list- scheduling technique([4],[5]).The concept of list scheduling is to first prioritize the nodes then removing the nodes from the list and assigning them to the processor which gives lowest start time to the process. List scheduling is classified as static and dynamic [6]. In static scheduling algorithms the list is not updated with new ordering at run time while the dynamic approaches do. Hu's algorithm for tree structured DAG was proposed by Hu [1961] to construct optimal schedule in polynomial time algorithm for DAG without communication and with unit computational cost [16]**.** some of scheduling algorithm for DAG uses a criteria called ALAP(As Late As Possible) which gives the start time of the task as late as possible for a task to be delayed without effecting the make span [13].A widespread simulation study has been done on level based heuristic [14].some of them are HLFET ((Highest Level First with Estimated Times), HLFNET (Highest Levels First with No Estimated Times), Random priorities to node of the DAG are assigned at random , SCFET (Smallest Co-levels First with Estimated Times) and SCFNET (Smallest Co-levels First with No Estimated Times)[7].a variation of HLFET was also proposed that is CP/MISF (critical path/most immediate

successors first)[8]. A variation of CP/MISF was proposed to make the searching efficient, a depth first approach with heuristic priority assignment called DF/IHS (Depth First with Implicit Heuristic Search) [8].Some other heuristic based on UNC scheduling EZ (Edge-zeroing) algorithm cluster are merged on the base of edge weight [17], LC (Linear Clustering) algorithm, [18], DSC (Dominant Sequence Clustering) algorithm [19]. duplication based scheduling is also done in this area where task are duplicated in multiple processor here the minimum start time of every node is to be minimized so to find optimum schedule .TDB based algorithm are PY(Papadimitriou and Yannakakis) algorithm calculates e-value to approximate the lower bound of a node[20] , DSH(Duplication Scheduling Heuristic) first calculate the start time of nodes without duplication then applies duplication on idle time slot[25], BTDH(Bottom-up Top down Duplication Heuristic) an extension of DSH[21],LCTD(Linear Clustering with Task Duplication)[42],CPFD(Critical Path Fast Duplication) divides the DAG into three categories: CPN (Critical Path Node), IBN (In-Branch Node), OBN(Out Branch Node)based on these partitioning A CPN dominant sequence is made[23], LWB (Lower Bound)its first calculate lower bound start time and the find the critical edges and assign them to different processors[24].These are some heuristic based algorithm but they are not considered to give good results in some cased so some meta heuristic approaches such as GA[26] ,SA[27] is used. Research in Hybrid approaches [28] is still going on.

III. PROPOSED APPROACH

The main factor that effects performance of system in multiprocessor environment is make span .Various heuristic and meta heuristic methods are proposed till now, depending upon these researches hybrid approach is found to give better results over heuristic and meta heuristic approaches. Heuristic can easily stuck at local optima and only uses domain specific knowledge. To overcome these drawbacks Meta -Heuristic approaches were proposed that is a kind of general heuristic that will fit in many problems. Genetic algorithm is a Meta-heuristic method that imitates the natural evolution process for solving optimization problem. Recently, Genetic Algorithms (GAs) have been widely reckoned as a useful vehicle for obtaining high quality solutions or even optimal solutions for a broad range of combinatorial optimization problems including task scheduling problem ([29], [30]). There is large space for the multiprocessor task scheduling and it is expected that randomly generated initial solution provide the weak results [1]. So, here in hybrid approach Task Duplication heuristics method is used to produce initial solution and this solution is fed to Genetic Algorithm (GA) .The initial solution is obtained by Task Duplication based heuristic: LWB(Lower Bound)Algorithm and this solution is given as the input schedule to GA to provide optimal solution in considerable time.

A. *The LWB Algorithm:*

The LWB (Lower Bound) algorithm proposed by [Colin and Chretienne 1991][31] it first determines lower bound start time of the nodes and then find the critical edges and assign them to different processors. An edge is said to be a critical when parent's message-available time for the child is greater than the lower bound start-time of the child [31].The lower bound algorithm considers lower bound of entry node as zero. Lower bound for rest of the node depends on its parent finish time and message available time. The Algorithm is as follows:

1) For each node n_i, compute its lower bound start-time, denoted by $lwb(n_i)$, as follows:
   (a) For any entry node n_i, $lwb(n_i)$ is zero.
   (b) For any node n_i other than an entry node, consider the set of its parents. Let n_i be the parent such that $lwb(n_x)+ w(n_x)+ c(n_x, n_i)$ is the largest among all parents. Then, the lower bound of n_i, $lwb(n_i)$, is given by, with n_y != n_x,

MAX{$lwb\{ n_x \} + w(n_x)$, MAX{$lwb(n_y)+ w(n_y)+ c(n_y, n_i)$}}

2) Consider the set of edges in the task graph. An edge (n_x, ni) is labelled as "critical" if $lwb(n_x)+ w(n_x)+ c(n_x, n_i)> lwb(n_i)$.
3) Assign each path of critical edges to a distinct processor such that each node is scheduled to start at its lower bound start-time[29].

B. The Genetic Algorithm:

The father of original genetic algorithm was john holland[32].Genetic algorithm is stochastic search technique that can perform optimization based on natural selection and evolution theories[33].

There are five phases of genetic algorithm:

1. initial population
2. fitness function
3. selection
4. cross over
5. mutation

1) Initial population:
The first step of GA is initial population .The solution to the problem is encoded in each member of the population. It has been recognized that if the initial population to the GA is good, then the algorithm has a better possibility of finding a good solution [2].

2) Fitness Function: According to the fitness function the fitness value is assigned to the individual. For evaluating fitness of an individual, best optimal threshold should be found out (i.e. min value of the fitness function)

3) Selection Operator: Selection is process of selecting the individual from the initial population. After evaluating fitness values of the individual the process of normalization is done and the some individuals are then selected.

4) *Crossover:* Cross over is similar to biological crossover reproduction, in which genetic operator is used to produce chromosomes for next generation from the present generation.
5) *Mutation:* Mutation selects randomly any part of the chromosome and regenerates, Mutation is extractive, and it creates random small diversions, thereby introducing new information.
6)

Working Principle of Simple GA:

1) Begin
2) CREATION of new population
3) EVALUAE Fitness Function for each candidate
4) REPEAT UNTIL (TERMINATION CONDITION is satisfied)
5) DO
   a. SELECT parents;
   b. CROSSOVER pairs of parents to produce children;
   c. MUTATION of the Children Produced
   d. EVALUATE Fitness of new candidate;
   e. SELECT individuals for the next generation;
6) DO
7) END

C. The Hybrid Algorithm(LWB-GA):

1) Calculate the lower bound of each node and find the critical edges, assign them different processor, calculate the make span and Schedule.
2) Fed the schedule as an input, calculated at step 1 to the initial population of genetic algorithm.
3) Run the Genetic Algorithm to find the schedule.
4) Note the final make span and Schedule.

IV. CONCLUSION

In this paper a hybrid approach LWB-GA is presented for solving Multi Processor Task Scheduling problem. Various research works on heuristic approach is done. Some of the researchers concluded that duplication based algorithm is better than other heuristic method. Heuristic algorithms are concluded to be problem specific while Meta-Heuristic being general .Meta heuristic approach are beneficial but in some cases they provide weak results where there is large solution space. Some researchers have combined both heuristic and Meta heuristic method to produce better results and it is concluded that hybrid approach may give better near optimal solution. So, LWB and GA may be considered for hybridization.

References

1. Sunita Dhingra, Satinder Bal Gupta, Ranjit Biswas," Hybrid GASA for Bi-criteria multiprocessoer task scheduling with precedence constraints" Computer Applications: An International Journal (CAIJ), Vol.1, no.1,pp.11-21,August 2014.
2. E. K. Burke, S. Gustafson, and G. Kendall, "Diversity in genetic programming: An analysis of measures and correlation with fitness," IEEE Transactions on Evolutionary Computation, vol. 8, No. 1, pp. 47–62, 2004.
3. R. Hwang, M. Gen and H. Katayama, "A comparison of multiprocessor task scheduling algorithms with communication costs", Computers and Operations Research, Vol. 35, No. 3, pp. 976–993 ,2008.
4. EL-REWINI, H., LEWIS, T. G., AND ALI, H. H. 1994. "Task scheduling in parallel and distributed systems. Prentice-Hall series in innovative technology" Prentice-Hall, Inc. Upper Saddle River, NJ.
5. KWOK, Y.-K. AND AHMAD. "Efficient scheduling of arbitrary task graphs to multiprocessors using a parallel genetic algorithm". J. Parallel Distrib. Comput. 47,58–77, 1997.
6. SIH, G. C. AND LEE, E. A.."A compile-time scheduling heuristic for interconnection-constrained heterogeneous processor architectures". IEEE Trans. Parallel Distrib. Syst. 4, 2 , 75–87 feb 1993a..
7. T.L. Adam, K. Clhandy and J. Dickson, "A Comparison of List Scheduling for Parallel Processing Systems," Communications of the ACM, vol. 17, no. 12, pp. 685-690, Dec. 1974
8. KASAHARA, H. AND NARITA." Practical multiprocessor scheduling algorithms for efficient parallel processing". IEEE Trans. Comput. C-33, 11,pp. 1023–1029,Nov. 1984.
9. Hwang, J., Chow,Y.,Anger, A., Lee, C "Scheduling precedence graphs in systems with inter-processor communication times". SIAM J. Comput. 8(2), pp. 244–257 1989.
10. Sivanandam, S.N., Visalakshi, P., Bhuvaneswari. "A Multiprocessor scheduling using hybrid particle swarm optimization with dynamically varying inertia". Int. J. Comput. Sci. Appl. 4(3),pp. 95–106 (2007)
11. Kwok, Y.-K., Ahmad "Dynamic critical path scheduling: an effective technique for allocating task graphs to multiprocessors". IEEE Trans. Parallel Distrib. Syst. 7(5), pp. 506–521 , 1996.
12. Hwang R.K., Gen M "Multiprocessor scheduling using genetic algorithm with priority-based coding". Proceedings of IEEJ Conference on Electronics, Information and Systems, 2004
13. Kwok, Y.K Ahmad," Dynamic critical path scheduling: an effective technique for allocating task graphs to multiprocessors". IEEE Trans. Parallel Distrib. Syst. 7(5), pp. 506–521 (1996).
14. K. Kaur, A. Chhabra and G. Singh, "Modified Genetic Algorithm for Task Scheduling in Homogeneous Parallel System Using Heuristics", International Journal of Soft Computing, Vol. 5, No. 2, pp. 42-51, 2010.
15. I. Ahmad and M.K. Dhodhi, "Multiprocessor scheduling in a genetic paradigm", Parallel Computing, Vol. 22, No. 3, pp. 395 –406, 1996.
16. HU, T. C. "Parallel sequencing and assembly line problems". Oper. Res. 19, 6 (Nov.), 841–848 ,1961..
17. Sarkar, "Partitioning and Scheduling Parallel Programs for Multiprocessors". MIT Press, Cambridge (1989)
18. KIM, S. J. AND BROWNE, J. C.. A general approach to mapping of parallel computation upon multiprocessor architectures. In Proceedings of International Conference on Parallel Processing,pp. 1–8 Aug., 1988.
19. Yang, T., Gerasoulis, " List scheduling with and without communication delays". Parallel Comput. 19(12), 1321–1344 (1993).
20. PAPADIMITRIOU, C. H. AND YANNAKAKIS."Towards an architecture-independent analysis of parallel algorithms". SIAM J. Comput.19,pp. 322–328, Apr. 1990.
21. CHUNG, Y.-C. AND RANKA," Applications and performance analysis of a compile-time optimization approach for list

scheduling algorithms on distributed memory multiprocessors". In Proceedings of the 1992 Conference on Supercomputing (Supercomputing '92, Minneapolis, MN, Nov. 16–20), R. Werner, Ed. IEEE Computer Society Press, Los Alamitos, CA, pp.512–521,1992.

22. CHEN, H., SHIRAZI, B., AND MARQUIS. "Performance evaluation of a novel scheduling method: Linear clustering with task duplication". In Proceedings of the 2nd International Conference on Parallel and Distributed Systems, 270–275,dec 1993.
23. AHMAD, I. AND KWOK, Y.-K." On exploiting task duplication in parallel program scheduling". IEEE Trans. Parallel Distrib.Syst. 9, 9, pp.872–892,1998a.
24. COLIN, J. Y. AND CHRETIENNE. "C.P.M. scheduling with small computation delays and task duplication". Oper. Res. 39, 4, pp.680–684, 1991.
25. KRUATRACHUE, B. AND LEWIS." Grain size determination for parallel processing". IEEE Software 5, 23–32, Jan.1998.
26. Wu, A.S., Yu, H., Jin, S., Lin, K.-C., Schiavone,"An incremental genetic algorithm approach to multiprocessor scheduling". IEEE Trans. Parallel Distrib. Syst. 15(9), pp.824–834,2004.
27. Kirkpatrick S, Gelatt CD, Vecchi M.P., "Optimization by simulated annealing, Science", Vol.220, pp. 671–680,1983.
28. Mostafa R.Mohamed and Medhat H. A. Awadalla, "Hybrid Algorithm for multiprocessor Task Scheduling", International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, pp. 79-89, May 2011.
29. yu-kwong kwok and isfaq ahmad,"static scheduling algorithms for allocating directed task graphs to multiprocessors" ACM computing surveys ,vol.31,no.4 ,pp. 407-427 , December 1999.
30. A.S. Wu, H. Yu, S. Jin, K.-C. Lin, and G. Schiavone, "An incremental genetic algorithm approach to multiprocessor scheduling", IEEE Transactions on Parallel and Distributed Systems, Vol. 15, No. 9, pp. 824–834, 2004.
31. COLIN, J. Y. AND CHRETIENNE. "C.P.M. scheduling with small computation delays and task duplication". Oper. Res. 39, 4, pp. 680–684, 1991.
32. Paraskevas Evripdou And Jean-Luc Gaudiot " Input/Output Operations For Hybrid Data-Flow /Control-flow System"1991.
33. E.Hou, R.Hong And N Ansari,"Multiprocessor scheduling Based On Genetic Algorithms", Dept Of ECE, New Jersey Institute Of Technology ,Technical Report,Aug.1990.

