[1]ITCM, Madero City Institute of Technology
[2]UPALT, Polytechnic University of Altamira
[3]UCA, University of Cadiz

Optimal Scheduling for Precedence-Constrained Applications on Heterogeneous Machines

Carlos Soto [1,3], Alejandro Santiago[2], Héctor Fraire-Huacuja[1], Bernabé Dorronsoro[3]

December 2018

Outline

Introduction: High performance computing

- High Performance Computing(HPC) systems are fundamental in a wide variety of scenarios as Wall Street or NASA [1], HPC systems requires high amount of power [2],
- HCP systems also known as super computers are formed by a set of machines interconnected over a bus or network platform.

Figure 1: Photo taken on June 16, 2013 shows the supercomputer Tianhe–2 developed by China's National University of Defense Technology. (tv.81.cn/He Shuyuan)

Task	m_1	m_2	m_3	$\overline{p_i}$
0	11	13	9	11.0
1	10	15	11	12.0
2	09	12	14	11.6
3	12	16	10	12.3
4	15	11	19	15.0
5	13	09	05	8.60
6	11	15	13	12.3
7	11	15	10	12.0

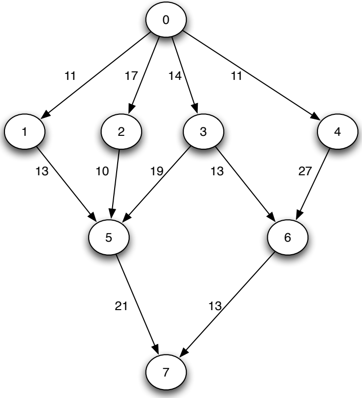

Figure 2: Example of a scheduling problem with a DAG.

Given a DAG, $G = (T, E)$, where:

- T is a set of n tasks $T = \{t_1, t_2, \ldots, t_n\}$.
- E is a set of edges, $(t_i, t_j) \in E$ are the precedence constraints.

A HCS is represented by:

- M is a set of machines $M = \{m_1, m_2, \ldots, m_k\}$.
- P a processing time for each task i in each machine k.

$$ts_i = \max\{\max\{tf_j\}\forall(t_j, t_i) \in E, Pw_{i,k} \in m_k\} \quad (1)$$

$$tf_i = ts_i + P_{i,k} + C_{i,k} \quad (2)$$

$$Makespan \quad \max_{i \in \{t_1, \ldots, t_n\}} tf_i \quad (3)$$

Methodology: Branch and bound

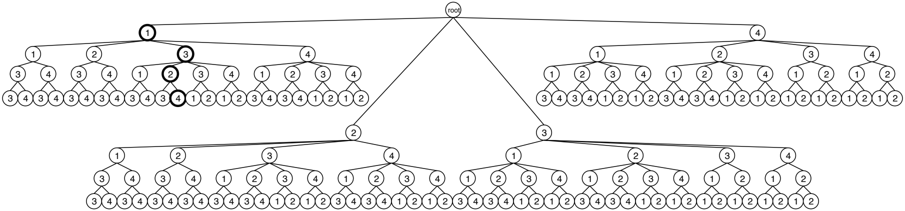

Figure 3: Example of a branch and bound tree.

Methodology: Branch and bound

Algorithm 1 Branch and bound scheduling.

```
function Scheduling(S_HEFT)
    S_best ← S_HEFT
    Bound ← evaluate S_best
    Initialize queue Q pushing {0, 1, ..., n − 1 }
    TopologialOrder(0)
    return S_best
```

Methodology: Branch and bound

Algorithm 2 Computation of the topological order.

function TopologialOrder($deep$)
    for i in $deep, \ldots, n$ **do**
        $e \leftarrow Q.pop()$
        $ord[level] \leftarrow e$
        Evaluate partial topological order $\{ord[0], ord[1], \ldots, ord[deep+1]\}$
        if partial order ord is feasible **then**
            if $level < n - 1$ **then**
                $TopologialOrder(deep + 1)$
            else ▷ The order is complete.
                $MachineAssignment(0)$
        $Q.push(e)$

Methodology: Branch and bound

Algorithm 3 Machine assignment.

function MachineAssignment($deep$)
  for j in $0, \ldots, k$ **do**
    $asig[ord[deep]] \leftarrow j$
    $S_{current} \leftarrow (ord, asig)$
    Evaluate partial solution $S_{current}$ in range $\{0, 1, \ldots, x+1\}$
    if partial evaluation $< Bound$ **then**
      if $prof < n - 1$ **then**
        $MachineAssignment(deep + 1)$
      else ▷ The solution is complete.
        $S_{best} \leftarrow S_{current}$
        $Bound \leftarrow$ partial evaluation

$$\min \quad z \tag{4}$$

Subject to:

$$\sum_{k=1}^{m} x_{ik} = 1, \quad i = 1, ..., n. \tag{5}$$

$$f_{jk} \leq s_{ik} + M(3 - x_{ik} - x_{jk} - q_{ij}), \quad \forall j \prec i; \\ i, j = 1, ..., n; k = 1, ..., m; i \neq j \tag{6}$$

$$f_{jk_2} \leq s_{ik_2} + M(3 - x_{ik_2} - x_{jk_1} - q_{ij}) + c_{ij}, \quad \forall j \prec i; \\ i, j = 1, ..., n; k_1, k_2 = 1, ..., m; i \neq j; k_1 \neq k_2 \tag{7}$$

$$s_{ik} + P_{ik} \leq f_{ik} + (M - x_{ik}), \quad i = 1, ..., n; k = 1, ..., m \tag{8}$$

$$f_{jk} - s_{ik} \leq M(1 - x_{jk}), \quad \forall j \prec i; i, j = 1, ..., n; k = 1, ..., m \tag{9}$$

Methodology: Mixed integer linear programming model

$$f_{jk_2} - s_{ik_1} \leq M(1 - x_{jk_2}) + c_{ij}, \quad \forall j \prec i; i,j = 1,...,n; k = 1,...,m \tag{10}$$

$$f_{ik} \leq s_{jk} + M * y, \quad i,j = 1,...,n; k = 1,...,m; i \neq j \tag{11}$$

$$f_{jk} \leq s_{ik} + M(1 - y), \quad i,j = 1,...,n; k = 1,...,m; i \neq j \tag{12}$$

$$z \geq f_{ik}, \quad i = 1,...,n; k = 1,...,m \tag{13}$$

$$f_{ik} \geq s_{ik}, \quad i = 1,...,n; k = 1,...,m \tag{14}$$

$$0 \leq x_{ik} \leq 1, \quad 0 \leq y \leq 1, \quad 0 \leq f_{ik}, \quad 0 \leq s_{ik} \tag{15}$$

$$M = \text{A big number} \tag{16}$$

Experimentation setup

- A time limit of 37,297 seconds.
- A small instance set from diverse authors (See Table 1).

Results

Table 1: Time results for synthetic applications in the literature

Instance	Optimal	Reference	BB	Time	MILP	Time
Ahmad_3_9_28	22	[3]	22	0.13	22	0.46
Bittencourt_3_9_184	184	[4]	184	0.41	184	0.73
Hsu_3_10_84	80	[5]	80	14.36	80	0.60
Demiroz_3_7_47	47	[6]	47	0.00	47	0.13
Eswari_2_11	100	[7]	100	5.61	100	0.71
Eswari_2_11_61	56	[7]	56	3.69	56	1.52
Hamid_3_10	100	[8]	100	5.59	100	0.70
Ilavarasan_3_10_77	73	[9]	73	6.84	73	0.64
Ilavarasan_3_15_114	121	[10]	121	37297.23	121	28.05
Kang_3_10_76	73	[11]	73	7.88	73	2.34
Kang_3_10_84	79	[12]	79	17.80	79	0.76
Kuan_3_10_28	25	[13]	26	1.89	26	0.75
Liang_3_10_80	73	[14]	73	6.83	73	0.65
Linshan_4_9_38	40	[15]	40	11.31	40	1.99
Mohammad_2_11_64	56	[16]	56	3.83	56	1.51
SahA_3_11_131	118	[17]	118	6.88	118	3.43
SahB_3_6_76	66	[17]	66	0.00	66	0.13
Samantha_5_11_31	32	[18]	32	3696.81	32	6.21
Sample_3_8_100	81	[19]	81	0.08	81	0.75
Liang_3_10_80	73	[14]	73	6.83	73	0.67
YCLee_3_8_80	66	[20]	66	0.11	66	0.28
Average time in seconds				1956.86		2.52

Conclusions

- Both strategies achieve the optimal values.
- The MILP gets a better computational time.
- A small instance set is proposed.

Referencias I

T. J. W. I. R. Center, D. Feitelson, Survey of scheduling in multiprogrammed parallel systems, Research report, IBM T.J. Watson Research Center.

W. C. Feng, The importance of being low power in high performance computing, CTWatch Quarterly 1 (3) (2005) 11–20. doi:10.1287/opre.8.1.101.

I. Ahmad, M. Dhodhi, R. Ul–Mustafa, DPS: dynamic priority scheduling heuristic for heterogeneous computing systems, IEE Proceedings - Computers and Digital Techniques 145 (6) (1998) 411, citation Key: Ahmad98. doi:10.1049/ip-cdt:19982345.

Referencias II

L. F. Bittencourt, R. Sakellariou, E. R. M. Madeira, DAG Scheduling Using a Lookahead Variant of the Heterogeneous Earliest Finish Time Algorithm, in: 2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing, IEEE, Pisa, Italy, 2010, pp. 27–34, series Title: PDP '10 Citation Key: Bittencourt2010. doi:10.1109/PDP.2010.56.

C.-H. Hsu, C.-W. Hsieh, C.-T. Yang, A Generalized Critical Task Anticipation Technique for DAG Scheduling, in: Algorithms and Architectures for Parallel Processing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 493–505, series Title: ICA3PP'07 Citation Key: Hsu2007. doi:10.1007/978-3-540-72905-1_44.

B. Demiroz, H. R. Topcuoglu, Static Task Scheduling with a Unified Objective on Time and Resource Domains, The Computer Journal 49 (6) (2006) 731–743, citation Key: Demiroz2006. doi:10.1093/comjnl/bxl030.

R. Eswari, S. Nickolas, Path-Based Heuristic Task Scheduling Algorithm for Heterogeneous Distributed Computing Systems, in: 2010 International Conference on Advances in Recent Technologies in Communication and Computing, IEEE, Washington, DC, USA, 2010, pp. 30–34, series Title: ARTCOM '10 Citation Key: Eswari2010. doi:10.1109/ARTCom.2010.19.

- H. Arabnejad, J. G. Barbosa, Performance Evaluation of List Based Scheduling on Heterogeneous Systems, in: Proceedings of the 2011 international conference on Parallel Processing, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 440–449, series Title: Euro-Par'11 Citation Key: Arabnejad2012b Medium: http://icl.eecs.utk.edu/heteropar2011/slides/heteropar_JorgeBarbosa.pdf. doi:10.1007/978-3-642-29737-3_49.

- E. Ilavarasan, P. Thambidurai, Performance Effective Task Scheduling Algorithm for Heterogeneous Computing System, in: The 4th International Symposium on Parallel and Distributed Computing (ISPDC'05), Vol. 3, IEEE, Washington, DC, USA, 2007, pp. 28–38, series Title: ISPDC '05 Citation Key: Ilavarasan2007a Issue: 2. doi:10.1109/ISPDC.2005.39.

E. Illvarasan, P. Thambidurai, Levelized Scheduling of Directed A-Cyclic Precedence Constrained Task Graphs onto Heterogeneous Computing System, in: First International Conference on Distributed Frameworks for Multimedia Applications, IEEE, Besançon, France, 2005, pp. 262–269, citation Key: Illvarasan2005. doi:10.1109/DFMA.2005.34.

Y. Kang, Z. Zhang, P. Chen, An activity-based genetic algorithm approach to multiprocessor scheduling, in: Y. Ding, H. Wang, N. Xiong, K. Hao, L. Wang (Eds.), 2011 Seventh International Conference on Natural Computation, Vol. 2, IEEE, Shanghai, 2011, pp. 1048–1052, citation Key: Kang2011b ISSN: 2157-9555. doi:10.1109/ICNC.2011.6022236.

Referencias VI

Y. Kang, Y. Lin, A recursive algorithm for scheduling of tasks in a heterogeneous distributed environment, in: Y. Ding, Y. Peng, R. Shi, K. Hao, L. Wang (Eds.), 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), Vol. 4, IEEE, Shanghai, 2011, pp. 2099–2103, citation Key: Kang2011a. doi:10.1109/BMEI.2011.6098753.

K.-C. Lai, C.-T. Yang, A dominant predecessor duplication scheduling algorithm for heterogeneous systems, The Journal of Supercomputing 44 (2) (2008) 126–145, publisher: Kluwer Academic Publishers Citation Key: Lai2008. doi:10.1007/s11227-007-0152-2.

L.-T. Lee, C.-W. Chen, H.-Y. Chang, C.-C. Tang, K.-C. Pan, A Non-critical Path Earliest-Finish Algorithm for Inter-dependent Tasks in Heterogeneous Computing Environments, in: 2009 11th IEEE International Conference on High Performance Computing and Communications, IEEE, Seoul, 2009, pp. 603–608, citation Key: Liang2009.
doi:10.1109/HPCC.2009.11.

L. Shen, T.-Y. Choe, Posterior Task Scheduling Algorithms for Heterogeneous Computing Systems, in: M. Daydé, J. M. L. M. Palma, A. L. G. A. Coutinho, E. Pacitti, J. C. Lopes (Eds.), High Performance Computing for Computational Science - VECPAR 2006, Vol. 4395, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 172–183, series Title: Lecture Notes in Computer Science Citation Key: Shen2006.
doi:10.1007/978-3-540-71351-7_14.

Referencias VIII

M. I. Daoud, N. Kharma, A hybrid heuristic–genetic algorithm for task scheduling in heterogeneous processor networks, Journal of Parallel and Distributed Computing 71 (11) (2011) 1518–1531, publisher: Academic Press, Inc. Citation Key: Daoud2011 Place: Orlando, FL, USA.
doi:10.1016/j.jpdc.2011.05.005.

S. K. Sah, R. S. Singh, Critical Path Based Scheduling of Multiple Applications in Heterogeneous Distributed Computing, in: 2009 IEEE International Advance Computing Conference, IEEE, 2009, pp. 99–104, citation Key: Sah2009.
doi:10.1109/IADCC.2009.4808988.

S. Ranaweera, D. Agrawal, A task duplication based scheduling algorithm for heterogeneous systems, in: Proceedings 14th International Parallel and Distributed Processing Symposium. IPDPS 2000, IEEE Comput. Soc, Washington, DC, USA, 2000, pp. 445–450, series Title: IPDPS '00 Citation Key: Ranaweera2000 Issue: February 2000.
doi:10.1109/IPDPS.2000.846020.

H. J. Fraire Huacuja, J. J. Gonzalez Barbosa, P. Bouvry, A. A. S. Pineda, J. E. Pecero, An iterative local search algorithm for scheduling precedence-constrained applications on heterogeneous machines, 6th Multidisciplinary International Conference on Scheduling : Theory and Applications (MISTA 2013) (2010) 47–485Citation Key: Pineda2010.

Young-Choon Lee, A. Zomaya, A Novel State Transition Method for Metaheuristic-Based Scheduling in Heterogeneous Computing Systems, IEEE Transactions on Parallel and Distributed Systems 19 (9) (2008) 1215–1223, citation Key: Lee2008. doi:10.1109/TPDS.2007.70815.