Formalization and Investigation of Parallel Processes Dispatching

Radi Romansky
Technical University of Sofia (Bulgaria)

Contents

Abstract

The performance of computer systems in the joint multiple processing of applications is an important system parameter, as the main role is played by the dispatching strategy. In this reason, this paper deals with investigation of parallel processes dispatching based on mathematical formalization and analytical modelling. Two cases based on the characteristics of processes are discussed and some numerical assessments for parameters (execution time, workload, et.) are determined for comparison of possible plans.

Keywords

analytical modelling, formalization, inhomogeneous processes assessments, parallel processing, system dispatching.

I. Introduction

A key indicator of the good functionality of different computer environments is their performance, which depends on the organization of application service processes, including system dispatching. Different approaches related to monitoring and modelling are used to study and evaluate this basic system characteristic.

When planning parallel processes, one of requirements is to coordinate the time parameters of each task execution with the probabilistic characteristics (density, distribution law, interval, etc.) of the incoming information flow.

The paper discusses the problem of formalization and analytical model investigation organization of activity of a system dispatcher which regulates execution of processes with inhomogeneous character in a parallel structure based on multiple processors. A preliminary formalization of the system parameters in a multi-processors structure is made to support the planned investigation. The goal is to evaluate the level of load of the system components and to define a suitable plan of applications execution in concurrent mode.

II. Preliminary Analytical Formalization

Initial phase in the organization of deterministic analytical research of dispatching system is preliminary formalization of the processes by analytical description of system parameters and the level of load. The defined formal model is oriented towards dispatching strategies in parallel and pseudo-parallel systems, providing a methodology for deterministic study of homogeneous and inhomogeneous processes. It is described as an ordered structure $P = \langle S, A, G, T, F \rangle$ with the following components:

- $S=\{S_1, \ldots, S_m\}$ – set of system resources occupied by running processes for a certain time.
- $A=\{A_1, \ldots, A_n\}$ – finite discrete set of processes that run in environment S, occupying the free resource. In the general case, the processes are not independent of each other.
- $G=G(A, L)$ – directed graph presenting relations between the processes of A (nodes of the graph), defined by the arcs $l_{ij} \in L$ for each couple of nodes $\langle A_i, A_j \rangle$, defining the existence of information dependence of process A_j on process A_i ($l_{ij} : A_i \rightarrow A_j$). The presence of a directed path $pass(A_i \rightarrow A_j)$ from one node to another determines a fixed sequence in the execution of all processes involved in it.
- $T=\{t_i \,/\, i=1 \div n\}$ presents the execution times for all active processes from the set A. In a heterogeneous system resource, a given process A_i may request different devices and load them for different times $\{t_{iq} \,/\, q=1 \div k\}$, which can be presented by a matrix $T^*=\{t_{ij}\}$ whit $n=|A|$ row and $m=|S|$ columns.
- F presents a criterion for defining a plan based on the chosen dispatch strategy.

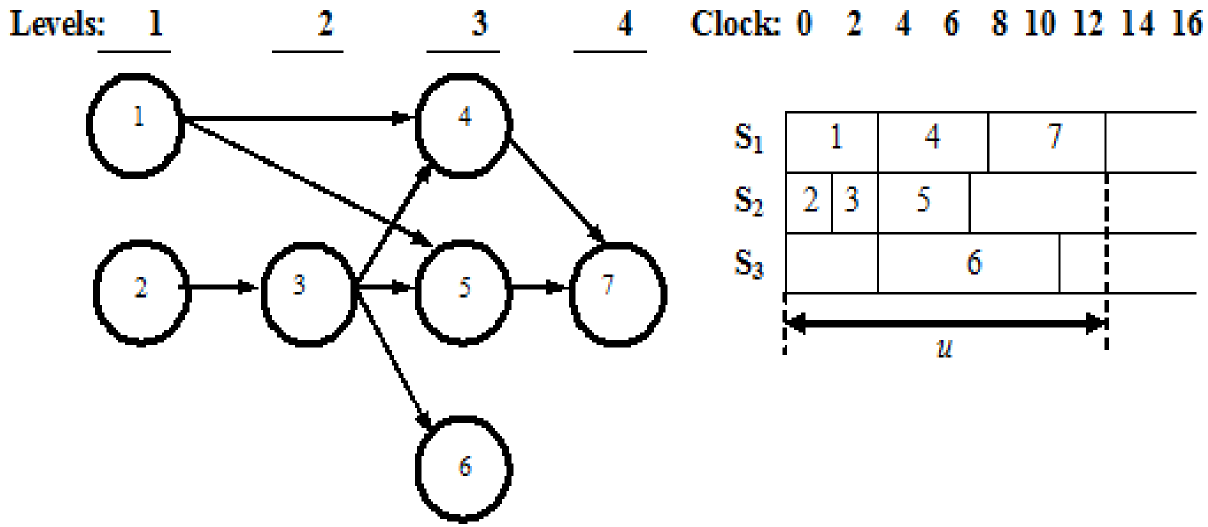
The dispatching plan is presented as a activities in the time $D(t)=\{d_i(t) \;/\; i=1 \div m\}$ in the interval $(0, \tau)$ with integer values of A, considered as numbers from 1 to n. In this case, the notation $d_i(t)=j$ $(1 \leq i \leq m;\ 1 \leq j \leq n)$ presents that the process A_j is executed in the resource S_i at time t.

In the general case, this defines a graph scheme of algorithm (GSA) and for the purpose of the discussed investigation it is necessary to form an ordered GSA (OGSA), which can be done by using program modules of TryAPL environment.

An example for 3 resources and 7 processes is the structure

***P* = <{1÷3}, {1÷7}, *G*, {4,2,2,5,4,8,5}, *FT*>**,

where *FT* shows minimization based on the total execution for all processes is illustrated in fig. 1, where there are two initial {1, 2} (in the level 1) and two final {6, 7} (in the level 4) processes executed in the environment *S*.



III. Dispatching Processes in a Parallel Environment

For the purposes of the investigation, it is accepted that the goal of the dispatcher is to define a plan with a minimum total time for the implementation of active processes of set A, analysing only the presence of information dependence between them. Thus, the weights of the nodes in OGSA set determine the times for the respective process realisation.

Two aspects of dispatching presented below are investigated by using the functions of the **APL2 program environment described in**

[3] Romansky, R. "Mathematical Modelling and Study of Stochastic Parameters of Computer Data Processing", Mathematics, vol. 9, no. 18, September 2021, art. 2240

A. Dispatching homogeneous processes

A homogeneous system resource $S=\{S_1, ..., S_m\}$ and a workload $A=\{A_1, ..., A_n\}$, formed by processes of equal volume and labor intensity, are considered. It is accepted that the processes have equal exaptation, i.e. $E[t(A_i)]\equiv E[t_i]$=const. This defines as acceptable the use of a binary direct graph for fixed times $t_i = \tau$ in the vector T ($i=1\div n$). The analysis was performed on the basis of fig. 1, as the determined model is $P=\{S, A, G, \tau, FT\}$ for $|S|=3$, $|A|=7$, and the criterion *FT* of planning requires determining the optimal configuration of *S* while providing a minimum total time *u* to implement the plan. This case have been discussed in [3] where to evaluate the dispatch plans D1(t)÷D5(t) the following parameters are introduced ($m=|S|=3$; τ is the clock whit two situations: τ' – effective and τ'' - passive):

- ✓ Relative average resource load factor:

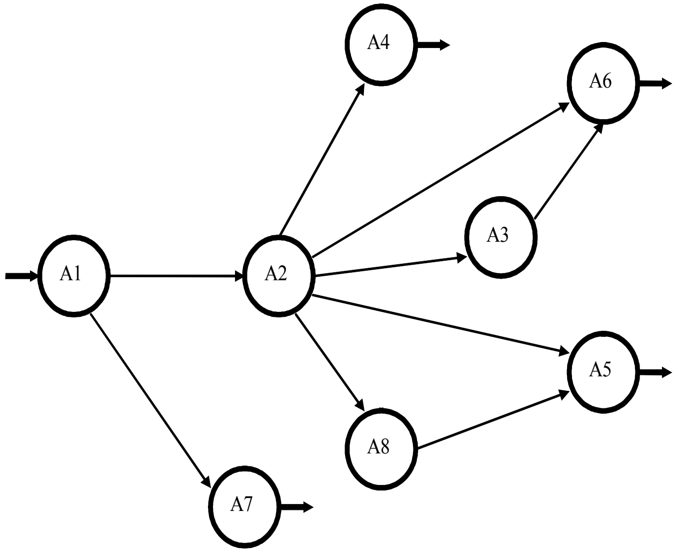
$$\eta = \frac{\sum \tau'}{\sum \tau} = \frac{\sum \tau'}{m.u}$$

- ✓ Relative weight of inefficient work:

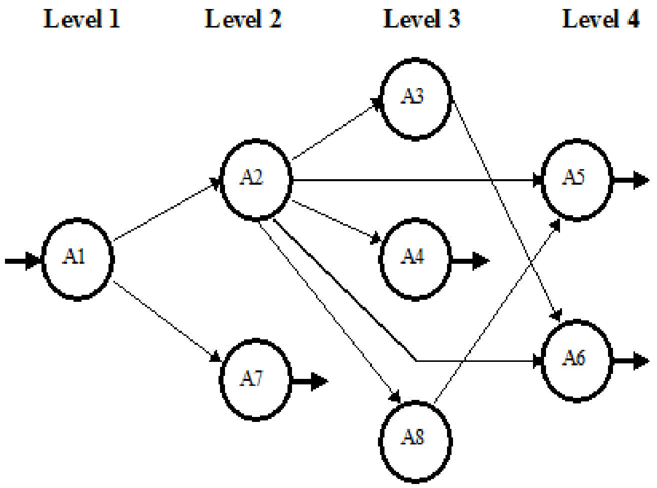
$$\chi = u\left(\frac{\sum \tau''}{m}\right)$$

B. Dispatching inhomogeneous processes

This case is an extension of the one presented above, as it is assumed that the planning processes have different execution times $t(A_i)$. The initial situation of the many active processes and the relations between them is shown in **fig. 2(a)**, at times T = {5, 10, 5, 6, 4, 3, 14, 4} for realization of the eight active processes.



The APL2 functions 'GSA' and 'PATH' (from [3]) are used to automate the experiments for determining communication between processes and formatting the ordered version of investigated case whit four layers – **Fig. 2(b)**. For the initial GSA, an additional determination of the scalar weights of the nodes is made, which does not change the functionality of the module "GSA" in the construction of OGSA and the module "PATH" used for forming all paths.



The generation of an optimal dispatching plan requires an adequate assessment of the weights of the processes in each of the defined layers. In the discussed case, four layers are defined, based on the elements of the set *T* (see fig. 2b) and with the help of the APL2-function ‘PATH’ the possible paths are formed. The obtained experimental results and calculated numerical estimates for their lengths are presented in **Table 1**.

Layers in OGSA	
I	$1^{(5)}$
II	$2^{(10)}$ $7^{(14)}$
III	$3^{(5)}$ $4^{(6)}$ $8^{(4)}$
IV	$5^{(4)}$ $6^{(3)}$

№	Weight path	Length
1	$1^{(5)}$ $2^{(10)}$ $4^{(6)}$	21
2	$1^{(5)}$ $2^{(10)}$ $5^{(4)}$	19
3	$1^{(5)}$ $2^{(10)}$ $8^{(4)}$ $5^{(4)}$	23
4	$1^{(5)}$ $2^{(10)}$ $6^{(3)}$	18
5	$1^{(5)}$ $2^{(10)}$ $3^{(5)}$ $6^{(3)}$	23
6	$1^{(5)}$ $7^{(14)}$	19

IV. Discussion

Table 2 presents several dispatch plans for different resource *S*, generated by the methodology with priority of the path with maximum length and priority of a process with a longer duration when choosing in a given layer.

<table>
<tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th></tr>
<tr><td colspan="30">Dispatch plan A</td></tr>
<tr><td>S_1</td><td colspan="5">1</td><td colspan="10">2</td><td colspan="5">3</td><td colspan="6">4</td><td colspan="3">6</td></tr>
<tr><td>S_2</td><td></td><td></td><td></td><td></td><td></td><td colspan="14">7</td><td colspan="4">8</td><td colspan="4">5</td><td></td><td></td></tr>
<tr><td colspan="30">Dispatch plan B</td></tr>
<tr><td>S_1</td><td colspan="5">1</td><td colspan="10">2</td><td colspan="5">3</td><td colspan="3">6</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>S_2</td><td></td><td></td><td></td><td></td><td></td><td colspan="14">7</td><td colspan="6">4</td><td></td><td></td><td></td><td></td></tr>
<tr><td>S_3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">8</td><td colspan="4">5</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="30">Dispatch plan C</td></tr>
<tr><td>S_1</td><td colspan="5">1</td><td colspan="10">2</td><td colspan="5">3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>S_2</td><td></td><td></td><td></td><td></td><td></td><td colspan="14">7</td><td colspan="4">5</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>S_3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">8</td><td colspan="3">6</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>S_4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="6">4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

The respective estimates are summarized in **Table 3**, as the best parameters were obtained for the plan with minimal system resource due to its significant usability. Such an analysis allows to choose the optimal configuration for the implementation of a specific graph scheme of parallel processing.

Dispatch plan	m	u	$\Sigma\tau'$	$\Sigma\tau''$	η	χ	σ1	σ3
A	2	29	51	7	0,879	101,5	2587,34	3348,69
B	3	25	51	24	0,68	200	3400	7352,94
C	4	23	51	41	0,554	235,75	3003,93	9787,45

V. Conclusion

An approach to mathematical formalization was applied in the organization of the research, as the experiments for obtaining analytical estimates were conducted with the help of developed program modules for the TryAPL2 operating environment. The choice of this software instrument is justified by the rich possibilities of language tools in describing discrete structures, including computer hardware and algorithmic structures. Built-in parallelism in language operators allows the creation of compact applications for the representation of systems and processes and their deterministic and stochastic study. These possibilities allowed to define a mathematical formal model for the description and organization of dispatching strategies, as a deterministic analytical study of homogeneous and inhomogeneous dispatch plans was conducted.

REFERENCES

15 publications are included in the list of references

Thank you for your attention

Radi Romansky

rrom@tu-sofia.bg