Network Model of the Processor System

VLADIMIR HOTTMAR

The University of Zilina, Department of Telecommunications, Slovak Republic
hottmar@fel.utc.sk

Keywords: processor, queuing theory, service centre, closed network, dual port memory, data request

The paper deals with modelling of a double-processor system by closed service network. The article presents the queuing system as a method of modelling. It analyses the performance and quantifies time characteristics of the processor systems.

1. Introduction

Questions connected with monitoring and valuation of processor systems (PS) are actual in many different levels of its utilization. One of many possible processor system applications appears from assumption of dividing the final number of processed tasks among fractional processor systems, which activity is mutually and relatively independent, and every processor system \( p_i \) of general system \( M \) process certain class of tasks \( r_i \) of total set \( R \), where \( p_i \in M \) and \( r_i \in R \), when \( i,j = 1, 2,... \)

Our task will be to understand a certain input set of data processed by given programme, and activities resulting from this processing will be the output set of data. Data transmission among processor systems is realised in the form of messages \( S \) that, on basis of routing and information content of message, is possible to consider as requests (answers) \( N \). Requests are in every processor system saved into memory where they create a family of requirements. Systems processors solve the operation of requests. Lets define data transmission (independently of its information content) as requests coming (outgoing) into (from) PS and lets define processing of messages as operation of requests by processor system. Assign to the set of data the set of requests \( S \rightarrow N \), where

\[
S = \{1, 2, ..., s\} \rightarrow N \{1, 2, ..., p\} \rightarrow \Sigma_i \text{ where } i = 1, 2,...
\]

Computing memories are specific. Function of these memories is to save results and intermediate results of operations with data. As the processor always has free access to all data in the memory, it is needful to define the strategy of saving and selection of requests into (from) the memory. Then it is possible to transform operation memories to the family of requests, and to define the strategy of its service. Service strategies are worked out in detail in [1,2,3,4].

Assignment of the set of data to the set of requests, and the set of processors to the set of service centers allow to model every local processor system by stochastic model and to solve this model by queuing theory.

2. Network model of double-processor system

Network model comes from the principle of closed service network. Characteristic feature of closed service network is the uniformity of total number of requests. According to this, the intensities of arrivals into particular service centres are non-constant.
The state of closed service network is at every instant of time defined as

\[ K = \sum_{i=1}^{N} k_i \]  

(1)

where

\( N \) – is a total number of network nodes (Figure 1), \( K \) – is a total number of requests in network, \( k_i \) – is number of requests in \( i \)-th node.

We define the total number of network states as

\[ Z = \binom{N+K-I}{N-I} \]  

(2)

The state of closed service network is at every instant of time defined as

\[ K = \sum_{i=1}^{N} k_i \]  

(1)

Local balance condition [1],[2] is the theoretical presumption for existence of closed service network solution by service centre. If every service centre in network fulfills the local balance condition, then even the whole network fulfill the balance condition. A solution is possible for service networks composed of local service centres of M/M/n/FIFO type [1],[2]. A precise solution of closed service networks was given by Gordon and G. F. Newell [1] and is defined as

\[ p(k_1, k_2, \ldots, k_N) = \frac{1}{G(K)} \prod_{i=1}^{N} x_i^{k_i} \beta(k_i) \]  

(3)

which is the probability that \( k_i \) requests are in \( i \)-th service centre, where \( x_i \) can be obtained from equation system solution

\[ \mu_i x_i = \sum_{j=1}^{N} \mu_j x_j p_{ji} \]  

(4)

where \( i = 1, 2, \ldots, N \) and \( G(K) \) is a normalization constant.

Figure 1. Block structure of computing system and its network model

If every service centre has only one service device, then \( \beta(k_i) = 1 \) and (3) changes to the form

\[ p(k_1, k_2, \ldots, k_N) = \frac{1}{G(K)} \prod_{i=1}^{N} x_i^{k_i} \]  

(5)

while, for the stationary state probability vectors, the following condition has to be fulfilled:

\[ \sum_{k} p(k_1, k_2, \ldots, k_N) = 1 \]  

(6)

Closed service network is characterized by input parameters:

- \( N \) – number of service centres (nodes) of model service network
- \( K \) – total number of requests in the model
- \( \mu_i \) – service intensity of requests processing in \( i \)-th service centre, \( i = 1, 2, \ldots, N \)
- \( p_{ji} \) – transmission probability, that means – the request for service in \( j \)-th service centre will also require service in \( i \)-th service centre.

If we put \( j = 1 \) and \( i = 2, 3, \ldots, N \) in equation system (4) then this will take the form

\[ x_i \mu_i = \mu_1 x_1 p_{1i}, \]  

(7)

or

\[ x_i = \frac{\mu_1}{\mu_i} x_1 p_{1i} \text{ for } i = 2, 3, \ldots, N \]  

(8)

For \( x_i \) we get an equation system of \((N-1)\) equations which are independent. This equation system enables to express unknown \( x_i \) through the parameter, for instance \( x_1 \).

If \( x_1 = 1 \), the normalization constant we define from equation

\[ G(K) = \sum_{i=1}^{N} \prod_{k=1}^{i} x_i^{k_i} \]  

(9)

The basic indicator is utilization (loading) \( \rho_1, \rho_2 \) of service centres \( \Sigma_1 \) and \( \Sigma_2 \) [1,3,7]. By analysing (9) we arrive to the following conclusion. Exponents \( k_i \) define possible number of requests in \( i \)-th service centre. Marginal attributes for \( k_i \) are zero (condition, when \( i \)-th service centre is empty) and \( K \) (all service centres are empty except \( i \)-th). Consequently, two following alternatives can occur for \( k_i \)

\[ (k_i = 0) \cup (k_i = K) \]  

(10)

If we want to define \( \rho_1 \), then we consider only those states, in which \( k_1 > 0 \), that means, \( k_1 = K \) for \( i = 2, 3, \ldots, N \) can never happen at the same time.

As the minimum of requests in first service centre \( (i=1) \) is \( k_1 = 1 \), then the maximum

\[ k_{i_{\text{max}}} = (K-1) \]  

(11)

for \( i = 2, 3, \ldots, N \) is divided into the rest of service centres [2],[9].
According to (9) we can write
\[
\sum_{k=0}^{N} x_i^k = G(k_{\text{in}}) = G(K-1) \tag{12}
\]

Utilization of \( i \)-th service centre is defined as
\[
\rho_i = \sum_{k=0}^{N} p(k_1,k_2,\ldots,k_N) = \sum_{k=0}^{N} \frac{1}{G(k)} \prod_{k=0}^{N} x_i^k \tag{13}
\]

Then the utilization of \( \rho_1 \) service centre \( i = 1 \) in terms of (12) and (13)
\[
\rho_1 = \sum_{k=0}^{N} p(k_1,k_2,\ldots,k_N) = \sum_{k=0}^{N} \frac{1}{G(k)} \prod_{k=0}^{N} x_i^k = \frac{G(K-1)}{G(K)} \tag{14}
\]

We can interpret the meaning of (14) as follows.
Utilization of \( \rho_1 \) service centre \( i = 1 \) is given by sum-\( k \)-ary of state probabilities in which \( k \)-values larger than zero \( k > 0 \) are gained. As an event, for \( k = 0 \), surely occurs once at least, for \( \rho_1 \) we can write as
\[
\rho_1 = \frac{G(K-1)}{G(K)} \tag{15}
\]

Utilization of \( \rho_i \) service centres \( i = 2, 3, \ldots, N \) we get from balance condition of input and output flow intensities of requests in steady-state regime in \( i \)-th service centre. Input flow intensity from service centre \( i \) is given by summary of service intensity \( \mu_i \) and probability, that \( i \)-th service centre is occupied, which is just \( \rho_i \).

Concerning (8) and (15) the following equation is valid
\[
\rho_i = x_i \rho_1 \quad \text{or} \quad \rho_i = \frac{G(k_{\text{in}})}{G(K)} \tag{16}
\]

Intensities of \( \lambda_i \) requests arriving in individual service centres are defined for steady state of input and output request flow equation of local service centre. For \( \lambda_i \) will be valid:
\[
\lambda_i = \rho_{i,1} \mu_{i,1} \quad \text{for} \quad i = 1
\]
\[
\lambda_i = \rho_{i,1} \mu_{i,1} \mu_{i,1} \quad \text{for} \quad i = 2, 3, \ldots, N \tag{17}
\]

Average amount of requests \( K_i \) in \( i \)-th service centre will be defined from the condition
\[
K_i = \sum_{e=0}^{K} k_e p(k_1,k_2,\ldots,k_N) \quad \text{for} \quad e = 1, 2, \ldots, K \tag{18}
\]

where \( k_{\text{in}} \) is number of requests in \( e \)-th service centre and \( p(k_1,k_2,\ldots,k_{\text{in}}) \) is the already known vector of state line probabilities. We perform the summation through lines, in which \( k_{\text{in}} = 0 \), and according to (1), condition of requests uniformity in closed network has to be fulfilled. We define the average length of \( L_i \) line in \( i \)-th service centre from deference of serviced \( \rho_i \) and from average number of requests waiting for service, provided that number of service devices (processors) in \( i \)-th service centre is equal to one.
\[
L_i = K_i - \rho_i \quad \text{for} \quad i = 1, 2, \ldots, N \tag{19}
\]

Time characteristics of service centres – the average stopping time of \( T_i \) requests in \( i \)-th service centre and average waiting period of \( T_wi \) requests in \( L_i \) line we can define from Little’s law:
\[
T_i = \frac{K_i}{\lambda_i} \tag{20}
\]
\[
T_{wi} = \frac{L_i}{\lambda_i} = T_i - \frac{1}{\mu_i} \tag{21}
\]

### 3. Model analysis

We analyse the model of Figure 1. We assume a real technical environment and we set a number of connected terminals to 9, as an example (T_1 till T_9). Implementation of \( \mu_i \) and \( p_{ji} \) parameters will help us to define all performance parameters and time characteristics of the system. The number of closed network states for \( K = 9 \) and \( N = 2 \) define (2), according to which \( Z = 10 \).

From system (4) we define the equations for parameters \( x_i \) calculation, for two service centres, then
\[
x_1 \mu_2 (p_{21} + p_{22}) = x_1 \mu_{12} \tag{22}
\]
\[
x_1 \mu_1 (p_{11} + p_{12}) = x_2 \mu_{21} (p_{21} + p_{22}) + x_1 \mu_{11} \tag{23}
\]

for complete probabilities it is valid that
\[
p_{12} + p_{11} = 1, \quad p_{21} + p_{22} = 1 \tag{23}
\]

By elimination of equations for \( x_{i=1} \) we get
\[
x_2 = \frac{\rho_{11}}{\mu_2} \tag{24}
\]

We see that parameter \( x_2 \) is dependent on ratio \( \mu_1/\mu_2 \) and on \( p_{12} \) probability. These three variables crucially affect performance parameters of the system model. Let’s analyse the model when the service intensity \( \mu_1, \mu_2 \) of both processors \( P_1, P_2 \) are the same, and requests for service in service centre \( \Sigma_1 \) demand the service in service centre \( \Sigma_2 \) with \( p_{12} = 1/2 \) probability. According to (24), the parameter \( x_2 \) will obtain value \( x_2 = 0.5 \).

Substitution of \( x_1 \) and \( x_2 \) into (9) is normalization constant
\[
G(K) = \frac{G(9)}{1.998046} \tag{20}
\]

For performance parameters definition we use numerical figures from Table 1. Table shows in the first column the number of \( S \) network states, in second and third column we can see the distribution of requests \( k_1 \) and \( k_2 \) in service centres \( i = 1, 2 \).

<table>
<thead>
<tr>
<th>( S )</th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( x_1^k \cdot x_2^k )</th>
<th>( p(k_1,k_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0.001953</td>
<td>0.000977</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>8</td>
<td>0.003906</td>
<td>0.001955</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>7</td>
<td>0.007812</td>
<td>0.00391</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>0.015625</td>
<td>0.00782</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>0.03125</td>
<td>0.01564</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0.0625</td>
<td>0.03128</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td>0.125</td>
<td>0.062561</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>2</td>
<td>0.25</td>
<td>0.125122</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1</td>
<td>0.5</td>
<td>0.250244</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.500488</td>
</tr>
<tr>
<td>( \sum )</td>
<td></td>
<td></td>
<td>1.998046</td>
<td>0.999997</td>
</tr>
</tbody>
</table>
The fourth column defines on the right side of the relation fractional conjunction of (9) and in the last column there are values of state line probabilities. The value \( G(K-1) \) is defined for \( i=1 \) from (12), or from Table 1 for \( G(8)=1.99414 \), then from (15) we define \( \rho_1=0.998 \), and from (16) we get by substitution of \( \rho_1 \) the value for \( \rho_2=0.499 \).

For the definition of \( \lambda_1, \lambda_2 \) we need to know the numerical value for service intensity \( \mu_1 \). Then by means of (17) the average number of requests in service centres will be \( K_1=8.00975 \) and \( K_2=0.990214 \), while the relation (1) has to be valid for total number of requests in network, that means \( K=8.999964 \).

The average length of lines \( L_1=7.01175 \) and \( L_2=0.491214 \) requests waiting for service. We will define time characteristics of \( T_1, T_2, T_W1, T_W2 \) from well known \( \lambda_1, \lambda_2 \), and from (20) and (21).

Now let’s analyze the model in the case when the service intensities are mutually deferent \( \mu_1 \neq \mu_2 \), and for probability of transition it will be again valid \( \rho_2=1/2 \). On the basis of (24) \( x_2=1 \), and \( x_1=1 \) we leave unchanged. After replacement \( x_1 \) and \( x_2 \) into (5) we obtain numerical data shown in Table 2.

<table>
<thead>
<tr>
<th>( S )</th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( x_1^<em>, x_2^</em> )</th>
<th>( p(k_1,k_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Achievement parameters:

\( G(9) = 10, G(8) = 9, p_1 = p_2 = 0.9, K_1 = K_2 = 4.5, K = 9, L_1 = L_2 = 3.6, \lambda_1 = 0.9, \mu_1, \lambda_2 = 0.45, \mu_1 \)

4. Discussion of the results

Our conclusions are as follows.

• Equal service intensities \( (\mu_1 = \mu_2) \) cause different coefficients of utilization of \( p_1 \) and \( p_2 \), while service centre \( \Sigma_1 \) (that is the processor \( P_1 \)) works on saturation limit and creates a bottleneck in the system [7], the processor \( P_2 \) works with approximately 50% utilization.

• Coefficient \( \rho \rightarrow 1 \), which means that we can expect a breach of local balance in model (the steady state stops to be valid), what will be at the real processor \( P_1 \) expressed by not accepting the request for service and by memory conflicts.

• Unequal dividing of the requests in lines \( L_1, L_2 \) represent the increase of memory claims at real memories.

• Change in service intensity ratio \( (\mu_1/\mu_2 = 2) \) at \( p_{12} = 1/2 \) result in equal dividing of requests \( K_1, K_2 \) in service centres \( \Sigma_1, \Sigma_2 \). utilization ratio \( \rho_1, \rho_2 \) gain equal values, and the processors \( P_1, P_2 \) work with approximately 90% utilization.

Acknowledgement

The author gratefully acknowledges the support by the VEGA project No. 1/2043/05 “Systems for the interactive television cabling distribution ICATV”

References