Modeling of a Queueing System with Tasks Requiring Multiprocessor Service

Vladimir Sahakyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
email: vladimir.sahakyan@sci.am

Artur Vardanyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
email: artur.vardanyan@iiap.sci.am

https://doi.org/10.62343/csit.2024.17

Abstract — This paper investigates a multiprocessor queueing system in which each incoming task requires a variable number of service nodes, reflecting the diverse resource demands found in modern parallel computing environments. By modelling the task arrival and service processes using Poisson and exponential distri-butions, the system's behaviour is analyzed through differential-difference equations that govern the steady-state probabilities. The research provides a comprehensive solution for determining the probability of different numbers of tasks being serviced, considering both the acceptance and rejection of tasks based on system capacity. The derived results offer valuable insights into optimizing system performance in terms of throughput, task completion, and resource utilization for multiprocessor systems.

Keywords — Queueing Theory, Multiprocessor System, Multi-server Queueing System, Multiprocessor Queueing System, Steady-State Probabilities.

I. INTRODUCTION

The rapid advancement of high-speed computing technologies has significantly transformed various scientific and industrial domains. Parallel computing systems, in particular, have become crucial for efficiently processing large-scale tasks that require substantial computational power. In recent years, multiprocessor systems, characterized by their ability to perform simultaneous operations across multiple processors, have gained considerable attention due to their potential to enhance performance and reduce execution times [1], [2]. In the context of queueing theory, multiprocessor systems can be modelled as multi-server service systems, where incoming tasks require the allocation of a specific number of service nodes for processing. Unlike traditional queueing models, which assume one service node per task [3], [4], this research introduces a custom number of service nodes for each task, depending on its resource requirements. This approach allows for a more realistic representation of modern computing envi-ronments, where task complexity and resource demands vary significantly.

The core objective of this study is to analyze the steady-state probabilities of the multiprocessor queueing system with tasks requiring multiple processors to service, considering exponential distributions for both task arrival and execution.

II. QUEUEING MODEL

This paper investigates a queueing system, with the follow-ing assumptions:

- A multiprocessor computing system comprising m processors (also referred to as cores or nodes, where $m \geq 1$) is considered a queueing system.
- Each task within the system is characterized by a random parameter, denoted as ν. Here, ν represents the number of computing resources required by the task for servicing, which could be processors, cores, cluster nodes, etc.
- Upon arrival in the system, tasks are subjected to either acceptance for servicing or rejection. The duration required for task servicing is partly contingent, it represents the maximum allowable time for task completion but is inherently random and may be shorter. Tasks face service denial if, upon entry into the system, it becomes evident that their specified parameters cannot be met. This occurs when the system lacks the requisite idle processors to initiate service.
- Tasks arrive at the system individually following a Poisson process characterized by a rate parameter α:

$$P(\alpha < t) = 1 - e^{-at},$$

  where a is the intensity of the incoming stream($a > 0$).
- Service times are modelled by an exponential distribution with a density function described as follows:

$$P(\beta < t) = 1 - e^{-bt},$$

  where β is a random value of the task execution time and b is the intensity of service($b > 0$).
- ν represents the random number of computational resources(service nodes) required for task execution. This parameter follows a probability distribution:

$$P(\nu = k) = \frac{1}{m},$$

  where $k = 1, 2, ..., m$.

Upon arrival, tasks undergo either acceptance for service or rejection. Once service begins, it continues uninterrupted until completion. These assumptions provide a framework for analyzing the dynamics of task arrival and service completion within the multiprocessor queueing system.

III. MATHEMATICAL MODEL FORMULATION

To analyze the queuing system, it is essential to identify the following notation: $P_k(t)$ represents the probability that k tasks are being serviced in the system at time t.

It is a well-established principle that the flow resulting from multiple elementary flows r emains e lementary. Furthermore, the probability of multiple events occurring within a short interval h is negligible typically denoted as $o(h)$. Leveraging this fact, and considering all possible scenarios concerning the system's states at time t, specifically, f ocusing o n instances where the system transitions during the time interval h to the state where k tasks are being serviced, the differential-difference equations of the system are given by:

$$\frac{dP_0(t)}{dt} = -aP_0(t) + bP_1(t),\ k = 0 \tag{1}$$

$$\begin{aligned}\frac{dP_k(t)}{dt} &= a\delta_{k-1}^{(1)}P_{k-1}(t) - [a(1-\delta_k^{(0)}) + kb]P_k(t) \\ &+ (k+1)bP_{k+1}(t),\ k \geq 1\end{aligned} \tag{2}$$

where the $\delta_k^{(0)}$ and $\delta_k^{(1)}$ probabilities are determined as follows:

$$\delta_k^{(0)} = P\left(\sum_{i=1}^{k+1}\nu_i > m \Big/ \sum_{i=1}^{k}\nu_i \leq m\right)$$

$$\delta_k^{(1)} = P\left(\sum_{i=1}^{k+1}\nu_i \leq m \Big/ \sum_{i=1}^{k}\nu_i \leq m\right)$$

It is evident that the following expressions express the conditional probability and by using the formula for calculating the conditional probability $\delta_k^{(0)}$ and $\delta_k^{(1)}$ probabilities can be calculated as follows:

$$\delta_k^{(0)} = \frac{P\left(\sum_{i=1}^{k}\nu_i \leq m < \sum_{i=1}^{k+1}\nu_i\right)}{P\left(\sum_{i=1}^{k}\nu_i \leq m\right)}$$

$$\delta_k^{(1)} = \frac{P\left(\sum_{i=1}^{k+1}\nu_i \leq m\right)}{P\left(\sum_{i=1}^{k}\nu_i \leq m\right)}$$

Using some sporting probability formulas from previous works[5] and simplifying the expressions, the following formulas are obtained:

$$\delta_k^{(0)} = \frac{k(m+1)}{m(k+1)} \tag{3}$$

$$\delta_k^{(1)} = \frac{m-k}{m(k+1)} \tag{4}$$

IV. STEADY-STATE PROBABILITIES

This section will outline the process of solving the recurrent equations derived from (1) and (2) when the queueing system goes into the steady state [6].

In steady state

$$\lim_{k\to\infty}\frac{dP_k(t)}{dt} = 0$$
$$\lim_{k\to\infty}P_k(t) = P_k$$

Therefore the equations (1) and (2) become:

$$-aP_0 + bP_1 = 0,\ k = 0$$

$$a\delta_{k-1}^{(1)}P_{k-1} - [a(1-\delta_k^{(0)}) + kb]P_k + (k+1)bP_{k+1} = 0,\ k \geq 1$$

Now, taking into account (3) and (4) formulas for the $\delta_k^{(0)}$ and $\delta_k^{(1)}$ probabilities and simplifying the expressions:

$$P_1 = \alpha P_0, \tag{5}$$

$$P_{k+1} = \frac{\alpha(m-k) + mk(k+1)}{m(k+1)^2}P_k - \frac{\alpha(m-k+1)}{mk(k+1)}P_{k-1}$$

where α is defined by $\alpha = a/b$.
By expanding the terms using the following definitions:

$$x_k = \frac{\alpha(m-k) + mk(k+1)}{m(k+1)^2}$$
$$y_k = \frac{\alpha(m-k+1)}{mk(k+1)}$$

The formula for P_{k+1} can be rewritten as:

$$P_{k+1} = x_kP_k - y_kP_{k-1} \tag{6}$$

The equations (5) and (6) are solved recurrently to obtain:

$$P_k = \frac{\alpha^k\prod_{i=1}^{k-1}(m-i)}{(k!)^2m^{k-1}}P_0,\ k \geq 1. \tag{7}$$

By substituting some binomial coefficient formulas and performing some straightforward mathematical transformations into equation (7), the probability P_k can be determined by the following formula:

$$P_k = \frac{\alpha^k}{k!\,m^k}\binom{m}{k}P_0,\ k \geq 1.$$

As the capacity of the system (i.e. the total number of tasks that can accommodate) is finite, say n, then using the condition of normality:

$$\sum_{k=0}^{n}P_k = 1$$

and by substituting the derived formulas for P_k into it, P_0 can be determined as the following:

$$P_0 = \frac{1}{1 + \sum_{k=1}^{n}\frac{\alpha^k}{k!\,m^k}\binom{m}{k}}$$

V. CONCLUSION

In tIn this paper, a queueing model has been developed to analyze the steady-state behaviour of a multiprocessor system, where each task requires a custom number of service nodes. The inclusion of varying task resource requirements significantly enhances the model's relevance to modern high-performance computing environments. By solving the differential-difference equations and deriving closed-form expressions for the steady-state probabilities. These results provide a foundation for further optimization of multiprocessor systems, enabling better management of computational resources in environments with varying task complexities. The findings contribute to the broader field of queueing theory by extending traditional models to more accurately reflect contemporary parallel computing challenges.

REFERENCES

1. Vladimir Sahakyan, Artur Vardanyan, ”The Queue State for Multipro cessor System with Waiting Time Restriction”, Computer Science and Information Technologies 2019, Conference Proceeding, Yerevan, pp. 116–119, 2019. DOI: https://doi.org/10.1109/CSITechnol.2019.8895093.
2. Jain Anamika, Madhu Jain, Dheeraj Bhardwaj, ”Controllable multipro cessor queueing system”, Applications of Mathematical Modeling, Ma chine Learning, and Intelligent Computing for Industrial Development, pp. 61-76, 2023. ISBN: 9781003386599
3. Bocharov P.P., D'Apice C., Pechinkin A.V., Salerno S., ”Queueing Theory”, VSP, Utrecht, Boston, pp. 94-98, 2004. ISBN: 90-6764-398-X
4. Shortle J. F., Thompson J. M., Gross D., Harris C. M., Fundamentals of Queueing Theory, John Wiley and Sons, New York, pp. 35-475, 2018. ISBN: 111894352X
5. Vladimir Sahakyan, Artur Vardanyan, ”A Computational Approach for Evaluating Steady-State Probabilities and Virtual Waiting Time of a Multiprocessor Queuing System”, Programming and Computer Software, Volume 49, pp. S16–S23, 2023. DOI: https://doi.org/10.1134/S0361768823090098.
6. Kleinrock L., Queueing Systems: Vol. — Theory, John Wiley and Sons, New York, 1975, ISBN: 0-471-49110-1.