# Measurement of The Effects of Parallelism on Response Time and The Overhead Delay Due To Job Splitting

**V.S. Dixit**
Assistant Professor,
Dept. of Computer Science,
ARSD College,
DhaulaKuan, New Delhi-21
Delhi, India

**Abstract**

In this paper, a parallel processing System is described which consists of Central queue and job splitting as a bulk arrival M/M/C queuing system, where customer corresponds to tasks and bulks correspond to jobs respectively. The time that a job spends in the system is evaluated which includes synchronization delay and Bulk response time is computed.

**Keywords:** Central Queue, Response Time, Synchronous Delay.

## Introduction

In this paper, a parallel processing System is modelled with a central queue and job splitting as a bulk arrival M/M/C Queuing System [1][2][4][7] and Mean Response Time of a random customer expression is obtained. Here our interest is the time that a job spends in the System with Synchronisation delay. The main things common to all the systems that we consider is a flow of customers requiring service there being some restrictions on the service that can be provided. For example the customers may be patients arriving at an out patients clinic to see a doctor. The restriction on the services is again that only one customer could be served at a time. This is called single server queue. An example of multi -server queue is a queue for buying stamps at a main post office or for having goods checked at a super market. Here the restriction is not more than one customer could be served at a time. Parallel processing is the use of multiple processors to execute different parts of the same program simultaneously. One of the main concerns in any form of computation is to ensure that required items of data reach at right place in the right time to interact with each other. A synchronous parallel computation progress with a global synchronization mechanism which means that the operation on different items of data take exactly the same time, for the same operation [3][4][9][10].While doing performance modelling of parallel processing System, considerations of job structure and system structure is very important. In job structure relationship between jobs are shown by a task graph model. In Fork/join structure job consists of set independent tasks.Job is considered complete when all of its tasks are completed while System structure always modelled using queuing system consists of a system of queues and set of servers. A central queue is accessible by all the processors and distributed queue where each processor serves its own queue .A queue may consists of jobs or tasks depending upon whether jobs are split into tasks before entering into the queue or not. Tasks are indexed composing of jobs by their order of departure. The job response time with synchronous delay of the concerned job is hard to obtain even for simple system structure [1][3][5].The analysis is much easier in the non-splitting case since all the tasks from the same job are executed sequentially on the same processor.

## Aim of the study

In this paper, a parallel processing System is modelled with a central queue and job splitting as a bulk arrival M/M/C Queuing System [1][2][4][7] and Mean Response Time of a random customer expression is obtained. Here our interest is the time that a job spends in the System with Synchronisation delay.

## Formulation

In this paper, the system is modelled as a continuous time, discrete state Markov process. Statistics for the response time of a job is obtained by decomposing the problem into two parts each of which can be handled individually. Response Time Q of a random job could be expressed as the sum of two terms i.e. Q=P+R, where P is the job waiting time and corresponds to the time that the job waits in the queue before the first of its task is scheduled. When the jobs arrives to a system where one or more servers are idle then P=0, otherwise P>=0.R is the job service time and corresponds to the time required to process all the tasks attached to a job after scheduling of first job.

S[Q] = S[P] + S[R] (A)

P is obtained by analysing the bulk arrival M/M/C queue that underlines the parallel processing system. P corresponds to the time that a bulk of customers in the M/M/C must wait before the first customer begins services.

Let Nto be a random variable denoting the steady state number of tasks in the system. N has distribution $\propto_I$= P[N=i], i=0,1,……

The probability density function is,

$\emptyset(Z) = \sum_{i=0}^{\infty} \propto_I Z^I$

With the following constraints:

$\lambda\mu_{0= \mu\alpha_i}$ (B)

$(\lambda + i\mu)\alpha_i$= $\lambda\sum_{k=0}^{i-1} \alpha_k\, a_{i-k}$+(i+1)$\mu\alpha_{i+1}$ (C)

$(\lambda + i\mu)\alpha_i$= $\lambda\sum_{k=0}^{i-1} \alpha_k\, a_{i-k}$+cμ $\propto_{i+1}$, i>=c (D)

By multiplying both sides of each of te above equation by $z^i$ and substituting T(Z)=$\sum_{I=0}^{c-i}(c-i)\,\alpha_i z^i$ we get,

$$\emptyset(Z) = T(z)\pi(z-1) \Big/ \left\lfloor \lambda z\big(1-x(z)\big) + c\mu(z-1) \right\rfloor$$ (E)

Lets suppose Z tends to 1 we get

$\emptyset(0)= T(1)\mu \Big/ [c\mu - \lambda S(x)]$ if $\emptyset(1)$=1 then

T(1)=$[c\mu - \lambda S(x)]$/μ.

Now to calculate expected no of task the moment

S(N)= d/dz ($\emptyset(Z)$)

If we take a random variable M then

S(M)=S(N)-C+$\sum_{I=0}^{c-i}(c-i) \propto_i$ S(N)—c+T(1)

Where M is the no. of tasks in the queue awaiting services

S[P]=S[M]+S[N>=c]/cμ (F)

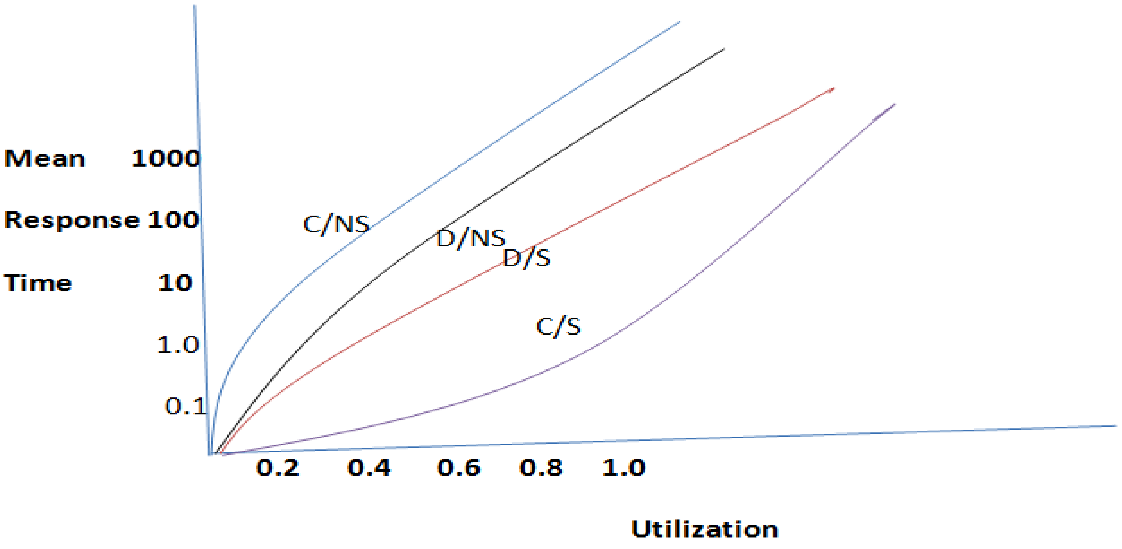
Now to calculate mean job service time for random job

S[R]=$\sum_{I=0}^{C-1}\sum_{n=1}^{\infty} R_{min}$ {n,c-i}n$g_0$+$\big(1-$
$i=0c-1\propto in=1\infty R1$n$g0$ (G)

Now Mean Job Response Time is calculated by substituting equations (F) AND (G) into (A).

**Analysis and Results**

We have assumed and considered four scenarios in Parallel processing system, i.e. distributed/splitting (D/S), distributed/nonsplitting (D/NS), centralized/ splitting (CS) and centralized/nonsplitting (C/NS) [9]. C is the no. of processors are taken, tasks are independent to each other and have exponentially distributed service requirements, arrival rate of job is lambda following poisson distribution. The systems differ in the way jobs queue for the processors and in the way jobs are scheduled on the processors. If each process has its own queue then queuing of jobs for processors are distributed. It would be centralized if there is a common queue for all the processors.The scheduling of job is no splitting if tasks are scheduled to run sequentially on the same processor once the job is scheduled. In the case of splitting tasks are scheduled such that they could run independently in parallel on different processors. A job is completed only when all of its tasks have finished execution. Mean Job response time is compared for all the scenarios to determine their relative performance over utilization.



**Comparison of four scenarios of four cases in parallel systemFor c=10 and lambda=1.**

It is evident that for all utilizations C/S has least Mean Response Time and D/NS has the greatest. DS has lower response time than D/NS.Mean response time of these systems depends on the utilization of the system. Evidently parallelism found in the splitting up jobs into tasks which reduced the mean response time in D/S system than C/NS system.

## Conclusion

In this paper, an expression for the mean job response time is obtained for different types of systems with the fact that there are some overheads associated with splitting jobs.

## References

1. *F Basseli and A.M. Makowaski, simple computable bounds for the fork/join queue," John Hopking Univ., 1985.*
2. *J. Hawng, "On the behaviour of algorithms in multiprocessing environment", PhD Dissertation , Computer Science, Dep- VCIA, 1988.*
3. *SA Nazaki and SM Ross, " Approximation in finite capacity multi server queues with poisson arrivals," J. Appl. Probavility, vol 15, no. 4, Dec 1978.*
4. *Y Yakahoshi," An approximation formula for the mean waiting time of M/G/M queue" J Oper res, vol 20, pp. 150-163,1977,*
5. *L Green, "Aqueuing System in which customers require a random numbers of servers, "oper. Res, vol 28, no. 6, page no. 1335-1346,1980.*
6. *AM Law and WD Kalton, Simulation modelling and analysis, second edition, New York Mc Graw-11[1],1991.*
7. *DD Yao, " Some results for the queues M/M/X and GI/G/C, oper research Lett, vol. 4, pp 79-83, july 1985.*
8. *"Referencing and diffusion approximation for m/g/m queue, oper. Res. Vol.33, pp. 1266-1277, Nov. 1985.*
9. *Joseph c. Jacob and Soo-Faury Lee, "Task spreading and shrinking on multi processor systems and network of work stations", IEEE Transaction on Parallel and Distributed Systems, vol. 10, October 1999;*
10. *Gh. Dodescu, B. Oancea, M. Raceanu, Parallel Processing, Ed. Economica, Bucharest, 2002.*