P: ISSN NO.: 2394-0344 RNI No.UPBIL/2016/67980 VOL-2* ISSUE-9* December- 2017
E: ISSN NO.: 2455-0817 *Remarking An Analisation*

Derivation of Bounds on the Performance Gain in Parallel Processing

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

Abstract

In this paper, parallel processing models are described, in which a job submitted is divided into two or more tasks during its execution. In the described System, task executes without synchronization and independent to each other. The performance of the system is analysed with the help of queueing models and an approximate solution is developed. The aim of the study of this paper is to calculate bounds on the performance gain when none of the devices is fully utilised in parallel processing System.

Keywords: Multiprogramming, Multitasking, Utilisation, Throughput.

Introduction

In this paper, two applications of Multi programming system i.e. terminal oriented system and CPU-I/O overlap are discussed. The first one used to represent transactions that are split into two tasks. The terminal user will wait for completion of the first task before issuing the next one transaction. In second application, a job can issue I/O requests served either synchronously or asynchronously. When I/O is neither synchronous nor asynchronous then the job could continue CPU processing in the simultaneous way along with the execution of I/O. Job never waits for an asynchronous I/O request to complete. [1] describes a general model to split job into two or more synchronous tasks, must complete execution before the job may resume processing. In [2], CPU and I/O activity can be overlapped with tight synchronization between two concurrent tasks. [2] Explained that the performance gain due to this type of overlap is greatest for balanced systems and low level of multiprogramming, Performance model of various I/O buffering schemes are described in [3].There are fixed number of buffers per file. [4] Showthat performance improvements become insignificant as the size of the buffer pool increases. Other studies of multitasking include [4][5][8].

While modelling a system it consists of active resources. Workload consists of a set of similar jobs where each job has primary tasks.

Aim of the Study

The aim of the study of this paper is to calculate bounds on the performance gain when none of the devices is fully utilised in parallel processing System.

Derivation

Consider performancegain by multitasking on the benefits of overlapping I/O with computation in Central Server Model. Higher throughput is achieved when the bottlenecked server is completely utilized. When the utilization would be the highest then bottleneck situation would arise. Bottleneck situation would arise when there is highest utilization. Now suppose a job comprised of a CPU burst and one I/O service.

So definition is

$$z_{i1} = z_{i2} = 1 \quad \text{(A)}$$

z denotes here the arrival rate.

Per visit to the CPU made by the primary task if the probability of a secondary task is t,

The average CPU time required for the primary task would be $P_{i1} = P_p + (1-t)P_O$,where the secondary task is P_O.

Total CPU time after completing a job will be

$$W = z_{i1}P_{i1} + tz_{i2}P_{i2} = P_p + P_O \quad \text{(B)}$$

Average time needed per job completion on server i, i>1 will be

W_i=(1-t)$z_{i1}P_{i1}$+t$z_{i2}P_{i2}$=((1-t)O_{i1}+tO_{i2})P_i (C)

For the fixed value of W_i the maximum possible throughput is

Q_{max} =1/max(W_i) (D)

Branching probabilityg_i=(1-t)O_{i1}+tO_{i2} (E)

When jobs are processed sequentially, the minimum throughput will be

Q_{min} =1/$\sum_1^m W_i$ (F)

So maximum possible throughput [6] improvement factor will be

Q_{max} /Q_{min} (G)

To obtain a tighter upper bound on the gain due to overlapped I/O

Q_{un} (n)=A/$\sum_1^m W_i$ +g_{un} (O)

g_{un} is the queuing delay due to other jobs competing for same resource

Q_{ol} (n) <= n/P_{I1}+$\sum_{i=1}^{m}(1-t)\,P_i O_{i1}$+$g_{ol}$ (P)

$Q_{ol}(n)$<= 1/(1/max(W_i))=Q_{max} (Q)

So the gain due to overlapping

R(n)= (Q_{ol}(n)/Q_{un}(n))<= ($\sum_{i=1}^{m} W_i$+g_{un})/max{n/Q_{max} (P_{i1}+(1-t)$\sum_{i=2}^{m} P_i O_{i1} + g_{ol}$)} (R)

By using the homogeneity [7]and monotonicity [7] of the throughput function of a close queuing network, It is evident that if none of the devices is fully utilised, then gain is bounded above by 1/(1-t).

Conclusion

In this paper, the expression for calculating bounds on the performance gain in Parallel Processing is derived which shows that in parallel processing system if none of the devices is fully utilized, then gain is bounded above by 1/(1-t).

References

1. *[1] I. Green, "A queuing system in which customers require a random number of servers," oper. res., Vol.28, 1980.*
2. *A.M. Law and W.D. Kelton, "Simulation modelling and analysis", Second edition, New York: McGraw-Hill, 1991.*
3. *C.D. Cromelin, "Delay probability formula", POEbc Eng. J., Vol 26, 1984.*
4. *L. Flatto, "Two parallel queues created by arrivals with two demands", SIAM J. Appl. Math., Vol. 45, OCT. 1985.*
5. *S.I.Brumelle, "Some inequalities for parallel server queues," Oper. Rev. Vol 19,pp. 402-413,1971.*
6. *G.P. Cosmotator, "Some approximation equilibrium results for the multi server queues (M/G/r)," opnl. Res. Vol. 27, 1976.*
7. *Queuing system, Vol.2: Computer Applications, New York, Wiley 1976.*
8. *Gh. Dodescu, B. Oancea, M. Raceanu, Parallel Processing, Ed. Economica, Bucharest, 2002.*