

LOAD BALANCING FOR MULTIPROCESSOR: A NOVEL ARCHITECTURE

Amit Sharma¹, Sushil Kumar²

^{1,2} Assistant Professor, IIMT College of Engineering, Greater Noida (India)

ABSTRACT

The goal of this paper is to represent the use of multiprocessor as a novel architecture for the parallel computing. This was not only for multiprocessing but it also work to support the computation of mathematical induction and also describe beyond the von Neumann architecture.

Keywords: Parallel, Computer, Multiprocessor, Load Balancing.

I. INTRODUCTION

One of the most important issues is how to effectively utilized parallel computers that has become increasingly complex to improve the performance. Such systems are constructed by different processor connected with communication link to operate in parallel with relatively low cost known as multi processor system. The parallel computer is one of the remarkable developments of methodology and technology in computer science in recent years. Due to multiprocessor structure of the computer architecture, this computer has a capability to execute multiple instructions or multiple data simultaneously. The parallel computer not only provides support for efficient computation of mathematical, economical, industrial, and ecological problems but also aims new computer architecture beyond the traditional von Neumann type. Multiprocessor system be very efficient at solving problems that can be partitioned into tasks with uniform computation and communication patterns. However, there exists a large class of non uniform problems with uneven and unpredictable computation and communication requirements. Therefore Dynamic load balancing (DLB) schemes are needed to efficiently solve non-uniform problems on multiprocessor systems.

In this paper , we can represent and mostly focus on the designing of a novel multiprocessor architecture and schedule the arriving load on the same to increase the performance.

II. LITERATURE REVIEW

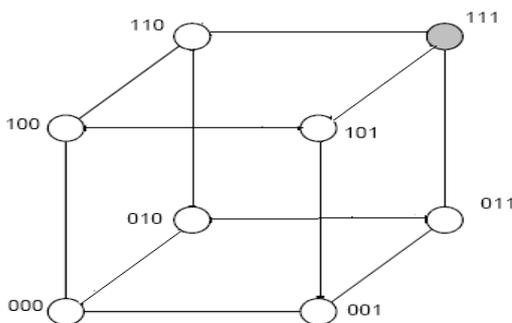
Achieving Parallelism is now a necessity to improve the performance of computer system. One of the main Issues is how to effectively utilized parallel computer that have become increasingly complex. It is estimated that many modern supercomputers and parallel processor deliver only 10 percent or less of their peak performance potential in a variety of applications. Yet high performance degradation are many. Performance losses occur because of mismatches among applications software and hardware. Research is active in the direction of developing new multiprocessor architectures and schedule the partitioned program on to it to achieve higher performance.[1,3,4].

Effort concentrated on the Study of all factors for developing a novel multiprocessor Architecture and after developing we will evaluate the performance of a multiprocessor architecture for server and networking with on study and to schedule the arriving load on to it in order to achieve higher performance. The other important issue are accessing delay and downloading the information while using multiprocessor technique. Comparable with similar architecture. In addition to designing an appropriate network, the efficient management of parallelism on the network involves optimizing performance needs, like the minimization of communication and scheduling over head. A simulation studies are carried out to compare the performance of different multiprocessor architecture (such as LEC, LET, Hypercube, Debruijn etc) with the various standard dynamic scheduling algorithms like Sender initiated diffusion (SID), receiver initiated diffusion (RID) etc .The simulation result will show that our Multiprocessor architecture (linearly extensible triangle) gives better performance as compare to other existing multiprocessor architecture with low cost and reduce the load balancing time and scheduling over head. Exploiting parallelism is now a necessity to improve the performance of computer systems One of the most important issues is how to effectively utilized parallel computers that have become increasingly complex to improve the performance. Such systems are constructed by different processor connected with communication link to operate in parallel with relatively low cost known as multi processor system [3,5].

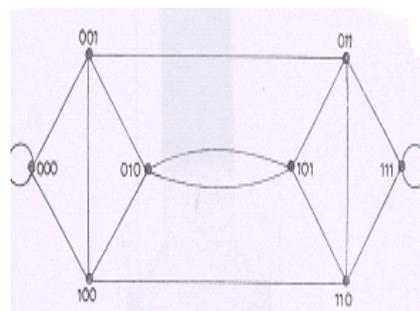
Multiprocessor system be very efficient at solving problems that can be partitioned into tasks with uniform computation and communication patterns. However, there exists a large class of non uniform problems with uneven and unpredictable computation and communication requirements. Therefore Dynamic load balancing (DLB) schemes [1,2,7] are needed to efficiently solve non-uniform problems on multiprocessor systems. Over the year, much different multiprocessor architecture, having different topological structure and different interconnection, have been used in different commercially available parallel systems. An enormous amount of research has centred on the design and their interconnection. Major architecture are found in ring network, hypercube, debruijn network, LET and LEC network.[3,4,8]. A number of multiprocessor architecture have been reported in literature.[6,7,8]

III. EXISTING MULTIPROCESSOR ARCHITECTURE

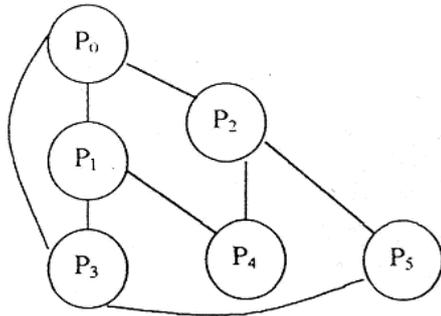
A processor organization can be represented by a graph in which nodes (Vertices) represent processors and the edge represent communication paths between pairs of processors.



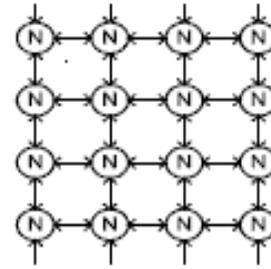
Hypercube Architecture



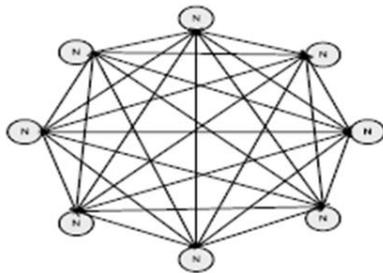
Debruijn Architecture



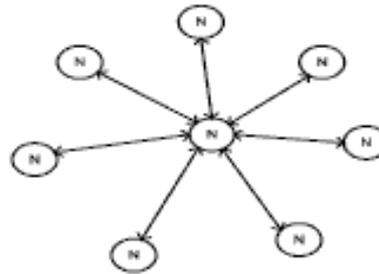
Linearly Extensible Tree



Two dimensional Mesh Architecture



Completely Connected Architecture



Star Connected Architecture

3.1 Linearly Extensible Cube

Characteristics of multiprocessor architecture and standard parameter:

- **Number of Node-**

It specifies total No. of processing element in the multiprocessor architecture. If no. of node increases the complexity as well as cost also increases whereas performance is decreases. So no of node should be minimum so that our system is less complex.

Note: Standard existing multiprocessor architecture has maximum $N=8$ nodes(processors).

3.2 Degree of Node

It specifies number of connections in each node .Higher the degree more complex the system. It is best if the degree or connectivity on each node is constant and independent of the network size implies that the processor organization scales more easily to systems with large no. of nodes.

Note: Degree should not be greater than four. [Degree 4]

3.3 Diameter

It is the measure of Maximum Inter-node distance in a network means the largest distance between two nodes.It specifies the communication delay between nodes as the size increases the diameter will also be increases.

Low diameter is better because low the diameter the communication delay between nodes is less.

- **Expandability-**

It is the property which facilitates constructing the large size system out of small one system with minimum change in the configuration of the architecture. It should be linear not exponential



- **Scalability-**

It specifies the ability of network to be modularly expandable with minimum change in the configuration.

- **Bisection Width-**

It specifies minimum No. edge that must be removed to divide the network into two halves. High bisection width is better because

In algorithm requiring large amount of data movement, the size of the data set divided by the bisection width puts a lower bound on the complexity of the parallel system.

- **Cost-**

Cost is directly proportional to no. of node in the network. cost should be minimum as possible as.

- **Complexity-**

Complexity is directly proportional to no of node and deg. of node in the network. Less complex system gives the higher performance. So less no. of node , lesser the complexity of the network.

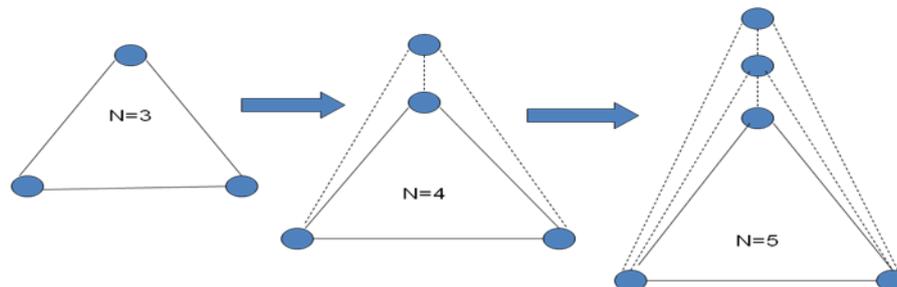
IV. CHARACTERISTICS OF EXISTING ARCHITECTURE (COMPARATIVE STUDY)

Network	No. of Link	Node degree.	Diameter	Bisection Width
Linear Array	N-1	2	N-1	1
Star	N-1	N-1	2	N-1
Completely Connected	$N(N-1)/2$	N-1	1	$(N/2)^2 N$
Binary Tree	N-1	3	$2(\log_2 N - 1)$	1
Illiac Mesh	2N	4	N - 1	2 N
Hypercube	$N \log_2 N / 2$	N	$\log_2 N$	N/2
De-bruijn	2N-1	4	$\log_2 N$	$N / \log_2 N$
LET	-	4	N	2

V. PROPOSED ARCHITECTURE

The proposed architecture named as “Linearly Extensible Triangle” triangular multi- processor network is simple extensible multiprocessor architecture having different topological properties comparable with existing architecture.

- **Linearly Extensible Triangle**



Characteristics of Proposed architecture

- No of Node- $N=5$
- Bisection Width- 3
- Degree of node- $N-1,3$
- Diameter- 2
- Complexity- Less in comparison to other existing Multi-processor architecture.
- Extensibility- Constantly Extensible on each level.
Linearly extensible by Layer of one processor.

VI. PROPOSED PERFORMANCE PARAMETER

To achieve high performance with low cost and minimization of the communication and scheduling overhead and uniform distribution of load among the processor Standard dynamic scheduling schemes are used namely:

- Hierarchical Balancing method
- Gradient model
- Sender Initiated Diffusion (SID)

A simulation study are carried out to compare the performance of proposed arch. And these scheduling schemes with the standard Hypercube multiprocessor architecture and other multiprocessor architecture.

VII. CONCLUSIONS

In this work , We have intended to evaluate the performance of our proposed architecture by comparing other existing multiprocessor architecture. So we have use two performance parameter (LIF & LBT). And “Ideal Load” for identifying processor is overloaded or underloaded. And one scheduling strategy SID (Source Initiated Diffusion) for scheduling or balancing the arriving the load on processor. On the basis of above two parameter we have seen that performance of our architecture is better than the other existing multiprocessor architecture in term of Load Balancing Time and Load Imbalance Factor. Our architecture has less complexity in terms of interconnection and less processing element in comparison to other existing multi processor architecture. So we can say that our proposed architecture(Linearly Extensible Triangle) give better performance in comparison to other existing multiprocessor architecture.



REFERENCES

- [1] A.Samad, M.Q. Rafiq and Omar Farooq "A novel algorithm for fast retrieval of information from a multiprocessor server" Proc. Intl. on parallel and distributed system, U.K. Feb 20 to feb 22, 2008.
- [2] Marc H. Willebeek-LeMair, Member, IEEE, and Anthony P. Reeves "Strategies for Dynamic Load Balancing on Highly Parallel Computers", Senior Member, IEEE, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 4, NO. 9, SEPTEMBER 1993
- [3] Z. Zeng and B.Veeravalli "Design and Performance of Queue and Rate Adjustment Dynamic load Balancing Polices for Distributed Network". IEEE trans. On computer, vol 55 ,no. 11, pp. 1410-1422,nov 2006.
- [4] Abdel A E and Khaled d," The Hyperstar interconnection Network"journal of Parallel and distributed computing no. 48,pp 175-199,1998
- [5] Razi-Azad H., "Constraint-based performance comparison of multi-dimensional interconnection networks with deterministic and adaptive routing strategies", Journal of Computers and Electrical Engineering, vol. 30, pp.167-82,2004.
- [6] Meraji S., Sarbazi-Azad H., Nayebi A., "Message routing and performance issues in necklace hypercubes", Technical Report, School of Computer Science, IPM, Tehran,Iran, 2006.
- [7] Nizadeh M., and Sarbazi-Azad, H., The necklace hypercube: a well scalable hypercube-based interconnection network for multiprocessors, ACM SAC 2005, pp.729-733, 2005.
- [8] K. G. Shin and Y.-C. Chang, "Load sharing in distributed realtime systems with state-change broadcasts," IEEE Trans. Comput., pp. 1124-1142, Aug. 1989.
- [9] V. A. Saletore, "A distrubuted and adaptive dynamic load balancing scheme for parallel processing of medium-grain tasks," in Proc. Fifth Distributed Memory Comput. Conf, Apr. 1990, pp. 995-999.
- [10] W. Shu and L. V. Kale, "A dynamic scheduling strategy for the Charekernel system," in Proc. ACM Supercomput. Con\$, 1989, pp. 389-398.
- [11] M. Willebeek-LeMair and A. P. Reeves, "A general dynamic load balancing model for parallel computers," Tech. Rep. EE-CEG-89- 1,Cornell School of Electrical Engineering, 1989.
- [12] T. L. Casavant and J. G. Kuhl, "A taxonomy of scheduling in generalpurpose distributed computing systems," IEEE Trans. Software Eng.,vol. 14, no. 2, pp. 141-154, Feb. 1988.
- [13] G. Cybenko, "Dynamic load balancing for distributed memory multiprocessors," J. Parallel and Distributed Comput., vol. 7:279-301, October, 1989.
- [14] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods. Englewood Cliffs, NJ: Prentice-Hall, 1989.K. M. Dragon and J. L. Gustafson,
- [15] Aebi A., Sarbazi Azad, H., Shamaei, A., Meraji, S., XMulator: An Object Oriented XML-Based Simulator,accessible at <http://www.XMulator.org>, 2006.