Implementation of Parallel Search using PVM

Presented by

- **Sudipto Chakraborty 022810**
- **Rahul Chakraborty 022823**
- **Rohan Datta 022829**

Performed Under Guidence of

- **Sri Utpal Ray**

Project Goal

- Implement Cluster Programming environment on a LAN.
- Perform Parallel Search in PC Clusters using PVM.

Definition of Parallel Processing

- **Parallel processing is a form of computing in which a number of activities are carried out concurrently so that the effective time required to solve the problem is reduced.**
- **In the simplest sense, parallel computing is the simultaneous use of multiple compute resources to solve a computational problem.**

•Why we need it?

- Save time
- Solve larger problems
- Taking advantage of non-local resources
- Fault tolerance
- Cost effective than a super computer
- Overcoming the problem of memory and pc resource limitations

What is Cluster programming?

- Group of loosely coupled computers that work together closely so that in many respects they can be viewed as though they are a single computer.
- Clusters are commonly, but not always, connected through fast local area networks.
- Clusters are usually deployed to improve speed and/or reliability over that provided by a single computer, while typically being much more cost-effective than single computers of comparable speed or reliability.

Modes and Categories of Clustering

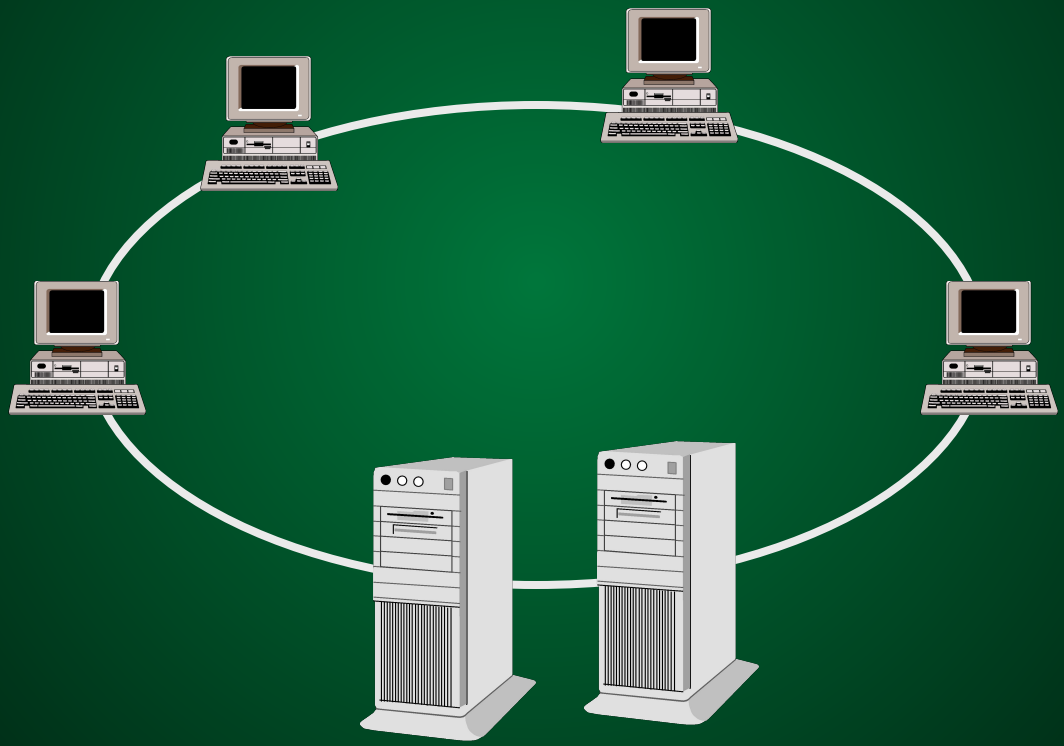
- Modes:
- Asymmetric Clustering
- Symmetric Clustering
- Categories:
- High-availability(HA) Clusters
- Load balancing Clusters
- High Performance Clusters
- Grid Computing

Need for Cluster Programming

- It can gather together multiple procesors to accomplish computational work.
- Clusters allow multiple hosts to access the same data on the shared storage.
- Capable of solving large problems in an efficient way.
- High parallelism can be achieved.
- Capable of fault tolerance.

Cluster examples

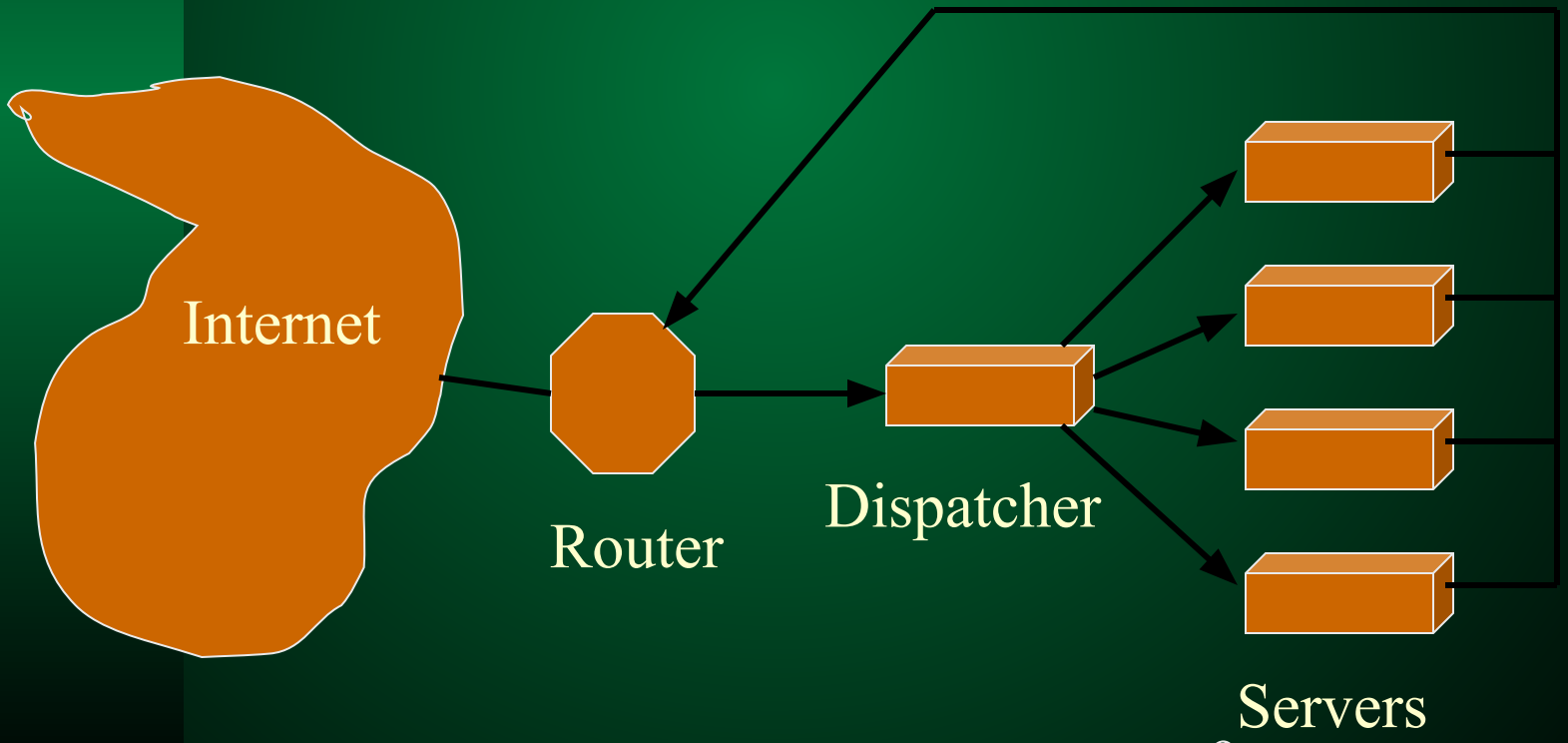
If the active server fail the work will be picked up by the standby server. The standby server's disk is consistent with the active server's disk at all times



- Cluster Example - Web-sever

Some popular Internet sites need more than one server in order to handle all incoming requests. In that case one can send all requests to a dispatcher and let the dispatcher distribute the load among a number of servers.

.



Clusters versus distributed systems

- The nodes in a distributed system have their own identity; from the outside the cluster nodes are anonymous.
- The computers in a distributed system often have dedicated roles, e.g. servers and clients; the computers in a cluster are usually equal.
- A cluster can be one node in a distributed system.

Problems with cluster systems

- Lack of “single system image” software. The important exception in parallel processing is the SMP. This is probably a major reason why SMPs have been relatively successful
- Limited exploitation. Only a limited number of software products currently support clusters

  Consequently, the problem with clusters is not hardware; it is software.

The standard reasons for using parallel and distributed systems in general

- **Availabilty (in most cases the most important reason for using clusters)**
- **Price/performance (clusters consist of standard computers, which generally have good price/performance ratio)**
- **Incremental Growth (one can incrementally extend the system by adding more computers)**
- **Scaling (there is no upper limit on the number of computers in cluster, as opposed to the maximum number of processors in an SMP)**
- **Performance (always important)**
- **Scavenging (turn the idle time on organization's computers into something useful)**

PARALLEL VIRTUAL MACHINE

- A software application that enables you to turn TCP/IP networked computers into a single virtual machine in order to run parallel programming.
- PVM system is the message-passing model .

- It makes a collection of computers appear as one large *virtual* machine.

appear as one large *virtual* machine

Why PVM ?

- Ability to establish a parallel machine using ordinary (low cost) workstations
- Ability to connect heterogeneous Machines
- Use of simple C/Fortran functions
- Takes care of data conversion and low-level communication issues
- Easily installable and configurable

Designing a PVM program

- There are a number of possible algorithm configurations that one could use when designing a PVM program. We will present one such configuration here:

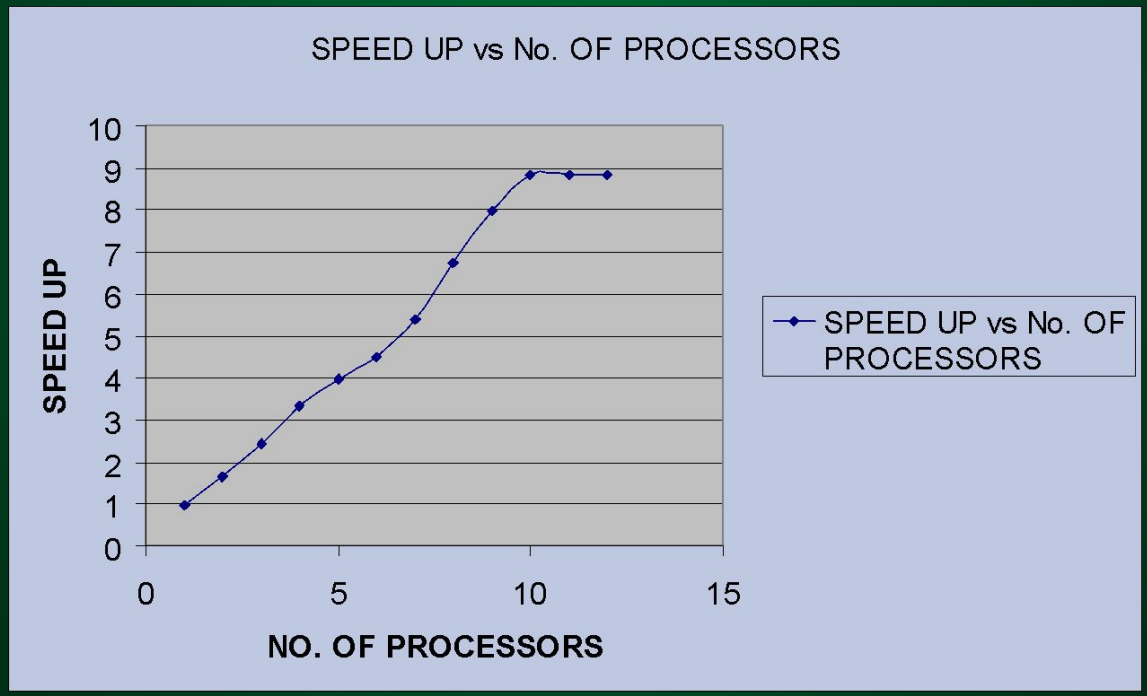
Master - Slave Approach:

- The master-slave approach is the most common method for a PVM computation. It utilizes the power of parallel programming more easily than the other methods because some computational task can be divided into sub-tasks, and then these smaller units can be worked on independently by the slave tasks, and the results sent back to the master when complete

Our Experiment

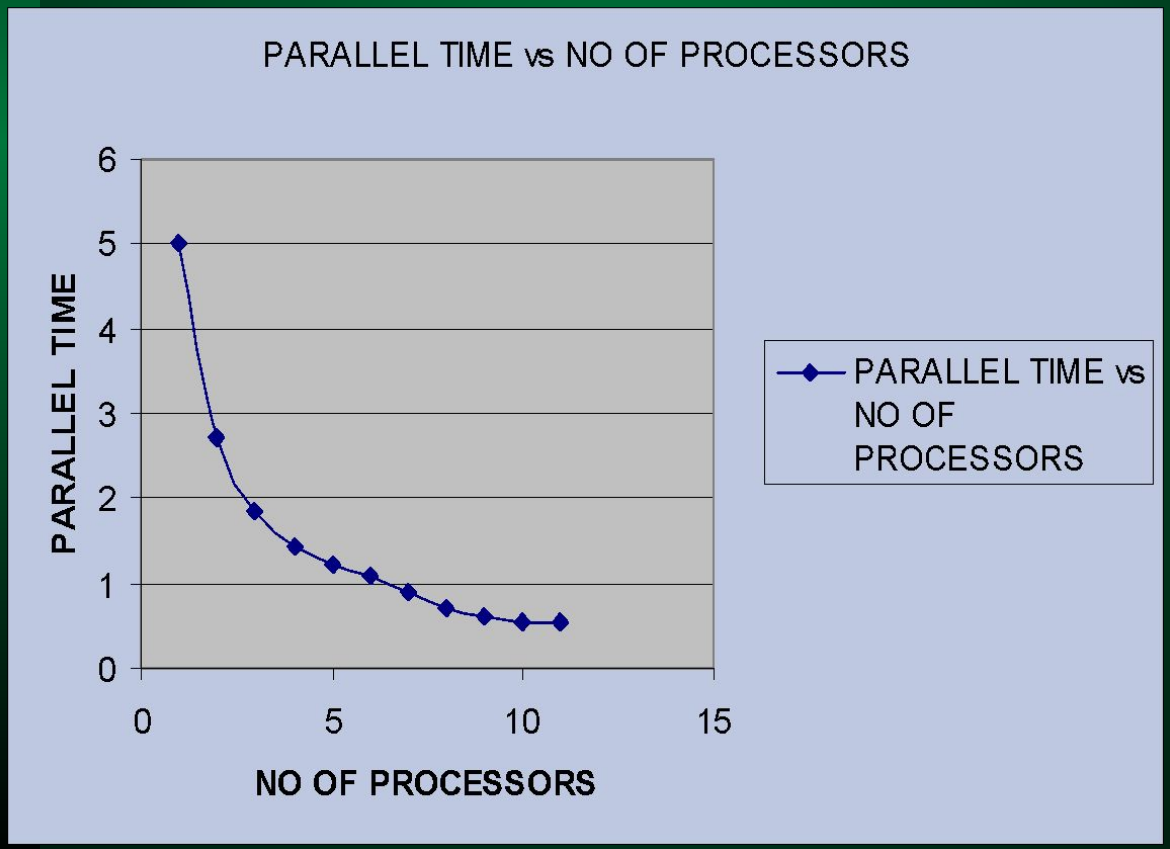
- A text file is divided equally among the available processors. The available processors act as slaves and record the number of occurrences of the particular text string under consideration and reports the number of occurrences to the master after completion of searching their domain space. The master adds the result from all of the processor and displays the result which also includes the parallel seraching time taken by the slave processors.After computing the sequential searching time in the master processor, performance testing is done by computing the speed up for varying number of slaves.

Performance Testing
Speed up vs No. of Processors



Performance testing

Parallel time vs No. of processors



Conclusion

By increasing the Processing power i.e.by adding more slaves to the master computer, speed-up increases almost linearly and computation time decreases, but to certain point which we can call the saturation point. After reaching that point speed up does not increase any more as the computation time remains constant, at, that point, there is no change in speed up and computation time even though any increase in processing power, i.e., speed up cannot be increased any more by adding more number of slaves.so, achieved speed up at that point is highest.

We have observed that efficient searching can be done by cluster programming model, and, it is more helpful if the search domain is too large for sequential searching.

THANK YOU