Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicompurers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections

MCS 572 Lecture 2
Introduction to Supercomputing
Jan Verschelde, 24 August 2016
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
the classification of Flynn

In 1966, Flynn introduced the following classification:

**SISD** = Single Instruction Single Data stream
one single processor handles data sequentially
use pipelining (e.g.: car assembly) to achieve parallelism

**MISD** = Multiple Instruction Single Data stream
called systolic arrays, has been of little interest

**SIMD** = Single Instruction Multiple Data stream
graphics computing, issue same command for pixel matrix
vector and arrays processors for regular data structures

**MIMD** = Multiple Instruction Multiple Data stream
this is the general purpose multiprocessor computer
Single Program Multiple Data Stream

One model is SPMD: Single Program Multiple Data stream:

1. All processors execute the same program.
2. Branching in the code depends on the identification number of the processing node.

Manager worker paradigm:

- manager (also called root) has identification zero,
- workers are labeled $1, 2, \ldots, p - 1$. 
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
types of parallel computers

shared memory

\[
\begin{array}{cccc}
  m_1 & m_2 & m_3 & m_4 \\
  p_1 & p_2 & p_3 & p_4 \\
\end{array}
\]

network

distributed memory

\[
\begin{array}{cccc}
  p_1 & p_2 & p_3 & p_4 \\
  m_1 & m_2 & m_3 & m_4 \\
\end{array}
\]

network

One crude distinction concerns memory, shared or distributed:

- A shared memory multicomputer has one single address space, accessible to every processor.
- In a distributed memory multicomputer, every processor has its own memory accessible via messages through that processor.

Many parallel computers consist of multicore nodes.
clusters

**Definition**

A cluster is an independent set of computers combined into a unified system through software and networking.

Beowulf clusters are scalable performance clusters based on commodity hardware, on a private network, with open source software.

What drove the clustering revolution in computing?

1. **commodity hardware**: choice of many vendors for processors, memory, hard drives, etc...

2. **networking**: Ethernet is dominating commodity networking technology, supercomputers have specialized networks;

3. **open source software infrastructure**: Linux and MPI.
Message Passing and Scalability

\[
\text{total time} = \text{computation time} + \underbrace{\text{communication time}}_{\text{overhead}}
\]

Because we want to reduce the overhead, the

\[
\text{computation/communication ratio} = \frac{\text{computation time}}{\text{communication time}}
\]

determines the \textit{scalability} of a problem:

\textit{How well can we increase the problem size \( n \), keeping \( p \), the number of processors fixed?}

Desired: order of overhead \( \ll \) order of computation, so ratio \( \to \infty \),
examples: \( O(\log_2(n)) < O(n) < O(n^2) \).

Remedy: overlapping communication with computation.
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
In a distributed shared memory computer:
- memory is physically distributed with each processor, and
- each processor has access to all memory in single address space.

Benefits:
1. message passing often not attractive to programmers,
2. shared memory computers allow limited number of processors, whereas distributed memory computers scale well

Disadvantage: access to remote memory location causes delays and the programmer does not have control to remedy the delays.
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
some terminology

- **bandwidth**: number of bits transmitted per second

- on latency, we distinguish tree types:
  - **message latency**: time to send zero length message (or startup time),
  - **network latency**: time to make a message transfer the network,
  - **communication latency**: total time to send a message including software overhead and interface delays.

- **diameter of network**: minimum number of links between nodes that are farthest apart

- on bisecting the network:
  - **bisection width**: number of links needed to cut network in two equal parts,
  - **bisection bandwidth**: number of bits per second which can be sent from one half of network to the other half.
arrays, rings, meshes, and tori

Connecting $p$ nodes in complete graph is too expensive.

An array and ring of 4 nodes:

A matrix and torus of 16 nodes:
hypercube network

Two nodes are connected ⇔ their labels differ in exactly one bit.

e-cube or left-to-right routing: flip bits from left to right, e.g.: going from node 000 to 101 passes through 100.

In a hypercube network with \( p \) nodes:
- maximum number of flips is \( \log_2(p) \),
- number of connections is \( \ldots \)?
a tree network

Consider a binary tree:
- The leaves in the tree are processors.
- The interior nodes in the tree are switches.

Often the tree is *fat*: with an increasing number of links towards the root of the tree.
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
In a shared memory multicomputer, processors are usually connected to memory modules by a crossbar switch.

For example, for $p = 4$:

A $p$-processor shared memory computer requires $p^2$ switches.
2-by-2 switches

- a switch
- pass through
- cross over

Changing from *pass through* to *cross over* configuration:

![Diagram showing the change from pass through to cross over configuration.](image)
A multistage network

Rules in the routing algorithm:

1. bit is zero: select upper output of switch
2. bit is one: select lower output of switch

The first bit in the input determines the output of the first switch, the second bit in the input determines the output of the second switch.

<table>
<thead>
<tr>
<th>inputs</th>
<th>outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>10</td>
<td>01</td>
</tr>
<tr>
<td>01</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

The communication between 2 nodes using 2-by-2 switches causes blocking: other nodes are prevented from communicating.
a 3-stage Omega interconnection network

circuit switching:

number of switches for $p$ processors: $\log_2(p) \times \frac{p}{2}$
circuit and packet switching

If all circuits are occupied, communication is blocked.

Alternative solution: packet switching: message is broken in packets and sent through network.

Problems to avoid:

- **deadlock:** Packets are blocked by other packets waiting to be forwarded. This occurs when the buffers are full with packets. Solution: avoid cycles using e-cube routing algorithm.

- **livelock:** a packet keeps circling the network and fails to find its destination.
Scalability and Classifications

1. Types of Parallel Computers
   - MIMD and SIMD classifications
   - shared and distributed memory multicomputers
   - distributed shared memory computers

2. Network Topologies
   - static connections
   - dynamic network topologies by switches
   - ethernet connections
Computers in a typical cluster are connected via ethernet.

![Diagram of computer nodes connected via ethernet network]

- **compute nodes**
- **ethernet network**
- **manager node**
personal supercomputing

HP workstation Z800 RedHat Linux:
- two 6-core Intel Xeon at 3.47Ghz,
- 24GB of internal memory,
- 2 NVDIA Tesla C2050 general purpose graphic processing units.

Microway whisperstation RedHat Linux:
- two 8-core Intel Xeon at 2.60Ghz,
- 128GB of internal memory,
- 2 NVDIA Tesla K20C general purpose graphic processing units.
The UIC Condo Cluster

The Hardware Specs at http://rc.uic.edu/hardware-specs:

- Two login nodes are for managing jobs and file system access.
- 160 nodes, each node has 16 cores, running at 2.60GHz, 20MB cache, 128GB RAM, 1TB storage.
- 40 nodes, each node has 20 cores, running at 2.50GHz, 20MB cache, 128GB RAM, 1TB storage.
- 3 large memory compute nodes, each with 32 cores having 1TB RAM giving 31.25GB per core. Total adds upto 96 cores and 3TB of RAM.
- Total adds up to 3,456 cores, 28TB RAM, and 203TB storage.
- 288TB fast scratch communicating with nodes over QDR infiniband.
- 1.14PB of raw persistent storage.
summary and recommended reading

We classified computers and introduced network terminology, we have covered chapter 1 of Wilkinson-Allen almost entirely.

Available to UIC via the ACM digital library:

- Michael J. Flynn and Kevin W. Rudd: **Parallel Architectures**. 

Available to UIC via IEEE *Xplore*:

- George K. Thiruvathukal: **Cluster Computing**. 
  Guest Editor’s Introduction. 

Exercises

Homework will be collected at a to be announced date.

Exercises:

1. Derive a formula for the number of links in a hypercube with \( p = 2^k \) processors for some positive number \( k \).

2. Consider a network of 16 nodes, organized in a 4-by-4 mesh with connecting loops to give it the topology of a torus (or doughnut). Can you find a mapping of the nodes which give it the topology of a hypercube? If so, use 4 bits to assign labels to the nodes. If not, explain why.

3. We derived an Omega network for eight processors Give an example of a configuration of the switches which is blocking, i.e.: a case for which the switch configurations prevent some nodes from communicating with each other.

4. Draw a multistage Omega interconnection network for \( p = 16 \).