Advanced Computer Architecture II:
Multiprocessor Design

Distributed Memory Multiprocessors II: Communication Architecture Design Space

Professor Russ Joseph
Department of Electrical and Computer Engineering
Northwestern University

February 1, 2005

Administrative Stuff

Announcements
Exam 1 will be on Thursday, 2/3/05
  • Covers all lectures up to and including today
  • Bring your calculator

Projects
I will email you about your proposals,
Don't forget weekly progress reports,

Distributed Memory Multiprocessors I:
Building Scalable Systems

Last Time: Introduced scalable design techniques for massively parallel, large scale machines,

Today:
  Talk about communication architecture design space.
  Review for Thursday's exam.

Review: Network Transactions

The fundamental interaction in distributed machines is the network transaction.
Analogous to the bus transaction in SMP machines.
Key properties:
  • One-way transfer of information from source output buffer to destination input buffer.
  • Causes some action at destination.
  • Occurrence is not directly visible at source (and in general is not globally visible).

Communication Network

serialized_msg

Source Node

Destination Node

output buffer

input buffer
**Network Transaction Processing**

Key design issues:
- Amount of message interpretation,
- Amount of dedicated processing in the communication assist.

**Design Spectrum**

There are many tradeoffs between the amount of hardware support for communication assist, degree of specialization, intrusiveness, and performance.

Today we will look at three classes of machines:
- **Physical DMA**
- **User-Level Access**
- **Dedicated Message Passing**

**Physical DMA**

Communication assist makes no interpretation of information within network transaction, hardware can be rather simple, but processing overheads are large.

Node-to-Network Interface:
- OS initiates transfers via physical memory and registers,
- On sender-side: construct system “envelope” around user data in kernel area,
- On receiver-side: receive into system area and ready for user process

**Physical DMA: nCUBE Case Study**

Known as direct network machine since data is forwarded from source to destination through intermediate nodes.

Each input/output DMA channel maps to own port.

No count registers on input channels, so kernel must safeguard against buffer overruns.
Physical DMA: LAN Interface

Most LAN controllers use queue of transmit and receive descriptors.

User-level Access: CM-5 Case Study

The network interface (NI) chip acts as communication assist,
Provides input/output FIFOs for each network.

User-level Access

Communication assist distinguishes between user-level and system messages.
Allows for low-overhead user-level transfers (no system intervention).

Node-to-Network interface:
Region of address space mapped to network input/output ports and status register.
Communication assist performs protection and destination node number into address or route.
May also insert/check error checking information.

Dedicated Message Passing

Allow sophisticated processing of network transaction:

Using dedicated hardware resources
Without binding interpretation to hardware design
Communication assist performed by communication processor (CP) which operates on the network interface directly.

• Network-to-Network interface:
• Provides clean abstraction to compute processor.
• Communication processor handles all the dirty work.
Dedicated Message Passing: Intel Paragon

Each node is shared memory multiprocessor with two or more I860XP processors.

Summary

Very broad range of choices in communication assist design.

Key Issues:
- Amount of dedicated hardware
- Degree of message interpretation