Portland State University
ECE 588/688

Interconnect and Routing

A Survey of Wormhole Routing

- Direct network architecture
  - Each node has a point-to-point or direct connection to a number of other nodes
  - Paper figures 1, 2
- Nodes communicate using messages of variable size
- A message is often transformed into packets before transmission
- A packet is the smallest unit of communication, containing:
  - A header with routing information
  - Data

Routers

- Communication can be between adjacent or non adjacent nodes
- Dedicated routers can be used within the nodes to handle communication
  - Dedicated routers allow overlap of computation and communication
- A typical router has multiple internal and external, input and output communication channels
- The network topology is defined by how the input channels of one router connects to the outputs of other routers
- Within a router, a cross bar typically connects all inputs to all outputs

Router Performance

- The time required to move messages between nodes is key to performance
  - Determines the granularity of parallelism that can be exploited
- Granularity refers to the size of the computation task
- Remember: fine-grain vs. coarse-grain parallelism
- **Communication Latency**: The performance metric used to characterize time to communicate data

Communication Latency

- Communication latency is the sum of
  - **Start-up latency**: Time needed to handle a packet at the source and destination nodes
  - **Network latency**: Elapsed time after the head enters the network until the tail leaves the network
  - **Blockage (congestion) time**: Measure of packet delays due to sharing of network communication resources with other packets
- Communication latency is determined by architectural characteristics
  - e.g., type of switching technology used in the router (Figure B)
    - Store and forward
    - Circuit switching
    - Wormhole

Characteristics of Direct Networks

- Topology defines how nodes are connected
  - Hypercube, torus, 3-D mesh, fully connected (paper figure A)
  - **Bisection width**: Minimum number of links that need to be cut to partition the network into two disjoint networks each containing half the nodes
  - Channel width
  - Channel rate
  - Channel bandwidth = width x rate
  - Bisection density = bisection width x channel width

Topology Tradeoffs

- Low-dimensional mesh networks have much lower bisection width than hypercubes
  - So they can offer wider channels
  - But they have larger average communication distance
- Before wormhole routing, hypercubes were popular due to shorter communication distance
- After wormhole routing, mesh networks became more popular
  - Wormhole routing makes communication latency almost independent of path length

Routing

- Routing determines the path a packet traverses from source to destination
- Source routing: the source node chooses the path
  - Each packet must carry the path information
  - Path cannot be changed after header leaves the source
- Distributed routing: each router determines after receiving a packet where to forward the message

Routing Classification

- **Deterministic routing**: path is determined statically by the source and destination addresses (paper figures 6,7)
- **Adaptive routing**: path is determined by source, destination and dynamic state of the network (paper figures 8, 9, 10)
- **Minimal routing**: selects the shortest possible path
- Non-minimal routing allows packet to traverse a longer path under some network conditions
  - Need to avoid continuously routing a packet without reaching destination

Flow Control

- A network consists of many channels and buffers
- Flow control deals with the allocation of buffers and channels to a packet
- A resource collision occurs when some packet cannot proceed because some resource is occupied by other packets

Collision Control Policy Options

- Drop packet
- Hold packet in place
- Buffer packet
- Reroute packet
- A good flow control policy avoids channel congestion while minimizing latency

Channel Allocation

- Routing algorithm selects which output channel to use for a packet arriving at a particular input
- Many arriving packets may select the same output channel
- An *input selection policy* determines which packet uses the output channel
  - Round robin
  - Fixed priority
  - First come, first served

Switching

- Mechanism that removes data from an input channel and places it on output channel
- Latency is highly dependent on the switching technique
- Switching techniques: (Figure B, C)
  - Store and forward
  - Circuit switching
  - Virtual cut-through
  - Wormhole

Store and Forward

- Also called packet switching
- Entire packet is stored in a buffer in an intermediate node
- The packet is then forwarded completely to an adjacent node
  - If channel and buffer are available
- Latency = L * D / B
  - L = Packet length (e.g., in bits)
  - D = Path length (between source and destination)
  - B = Channel Bandwidth (e.g., in bits/sec.)

Circuit Switching

- A physical circuit is established between source and destination
- Once established, the channels across the path cannot be shared
- Packet is then transmitted
- Circuit is torn down
- Latency = D x Lc/B + L/B
  - Lc = length of control packet transmitted to establish the circuit

Virtual Cut-Through

- Packet header is examined upon arrival at an intermediate node
- Packet is stored in intermediate node only if output channel is not available
- Latency = D x L_h/B + L/B
  - L_h = length of header field

Wormhole Routing

- Packet is divided into a number of “flits”
- The header flit governs the route
- The remaining flits follow in a pipelined fashion
- If output channel is blocked, flits remain in flit buffers along the route
- Supports broadcast and multicast
- Paper figure 3, 4
- Latency = path x L_f/B + L/B
  - L_f = length of each flit

Deadlocks

- When packets become blocked forever (example, paper figure 5)
- Can happen when packets are allowed to hold on resources while requesting other resources
- Deadlock avoidance
  - Preemption
    - Preempted packets are rerouted or discarded
    - Rerouting could lead to livelocks while discarding increases latency
  - By routing algorithm (deadlock-free routing)
    - Order network resources and allow usage only in strictly monotonic order
    - Use cycle-free dependence graphs

Virtual Channels

- Some adaptive routing algorithms require multiple pairs of channels between adjacent nodes
- Implementing each channel using wormhole routing is expensive and utilization may be low
- Virtual channels: each physical communication channel is time-multiplexed into multiple virtual channels
  - Each virtual channel has its own flit buffer, control, and data path
- Paper figure 12
- Disadvantages?

Reading Assignment

- Monday
  - Richard Russell, "The CRAY-1 Computer System," Communications of the ACM, 1978 (Read)
  - Steven Scott, "Synchronization and Communication in the T3E Multiprocessor," ASPLOS, 1996 (Read)
- No class next Wednesday
- Project progress report due Monday Nov 16