# CS405
# Computer System Architecture

## MODULE 4

## Syllabus

**Message Passing Mechanisms**-Message Routing schemes, Flow control Strategies, Multicast Routing Algorithms.

**Pipelining and Superscalar techniques** – Linear Pipeline processors and Nonlinear pipeline processors

# COMPUTER SYSTEMS ARCHITECTURE

Message-Routing Schemes-
Store and Forward
Wormhole Routing
(Malayalam)

MODULE 4- PART1

## MESSAGE-PASSING MECHANISMS

- Message passing in a multicomputer network demands
- special hardware and software support.

- Message Routing schemes
- Flow control Strategies
- Multicast Routing Algorithms

## Message-Routing Schemes

- **Message Formats**
  - –A message is the logical unit for inter-node communication.
  - –It is often assembled from an arbitrary number of fixed-length packets, thus it may have a variable length.

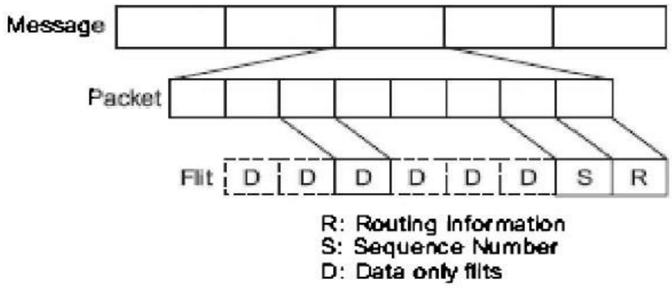



**Fig. 7.26** The format of message, packets, and flits (control flow digits) used as information units of communication in a message-passing network



- A packet is the basic unit containing the destination address for routing purposes.
- Because different packets may arrive at the destination asynchronously, a sequence number is needed in each packet to allow reassembly of the message transmitted.
- A packet can be further divided into a number of fixed-length flits (flow control digits).
- Routing information (destination) and sequence number occupy the header flits.
- The remaining flits are the data elements of a packet.

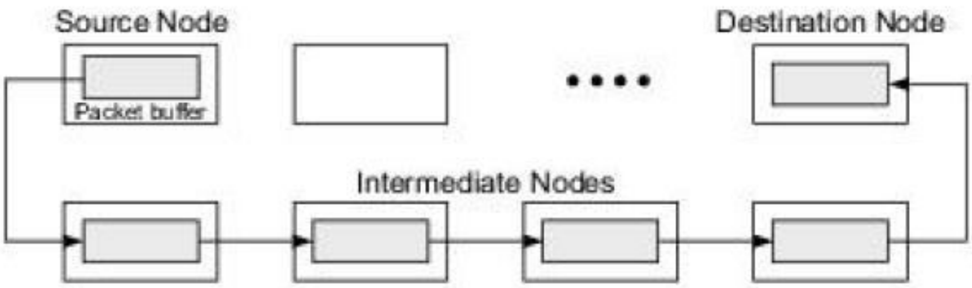
- In multicomputers with store-and-forward routing,
  - packets are the smallest unit of information transmission.
- In wormhole routed networks,
  - packets are further subdivided into flits (flow control digits).
- The flit length is often affected by the network size.
- The packet length is determined by the routing scheme and network implementation.
- Typical packet lengths range from 64 to 512 bits.
- The sequence number may occupy one to two flits depending on the message length.
- Other factors affecting the choice of packet and flit sizes include channel bandwidth, router design, network traffic intensity, etc.

7

## Store-and-Forward Routing

- Packets are the basic unit of information flow in a store-and-forward network.
- Each node is required to use a packet buffer.
- A packet is transmitted from a source node to a destination node through a sequence of intermediate nodes.
- When a packet reaches an intermediate node,
  - it is first stored in the buffer.
  - Then it is forwarded to the next node if the desired output channel and a packet buffer in the receiving node are both available.
- The latency in store-and-forward networks is directly proportional to the distance (the number of hops) between the source and the destination.
- This routing scheme was implemented in the first generation of multicomputers.

8




(a) Store-and-forward routing using packet buffers in successive nodes

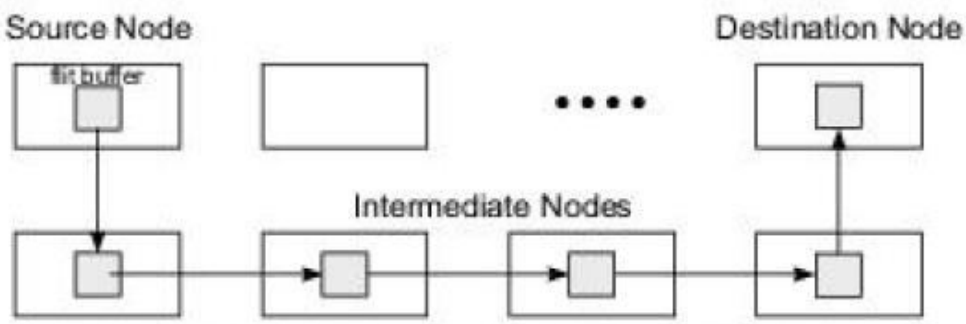


## Wormhole Routing

- By subdividing the packet into smaller flits, latter generations of multicomputers implement the wormhole routing scheme.
- Flit buffers are used in the hardware routers attached to nodes.
- The transmission from the source node to the destination node is done through a sequence of routers.
- All the flits in the same packet are transmitted in order as inseparable companions in a pipelined fashion.

- The packet can be visualized as a railroad train with an engine car (the header flit) towing a long sequence of box cars ( data flits).
- Only the header flit knows where the train (packet) is going.
- All the data flits (box cars) must follow the header flit.
- Different packets can be interleaved during transmission.
- However, the flits from different packets cannot be mixed up.
- Otherwise they may be towed to the wrong destinations.






(b) Wormhole routing using flit buffers in successive routers

# COMPUTER SYSTEMS ARCHITECTURE

## Asynchronous Pipelining Wormhole Routing (Malayalam)

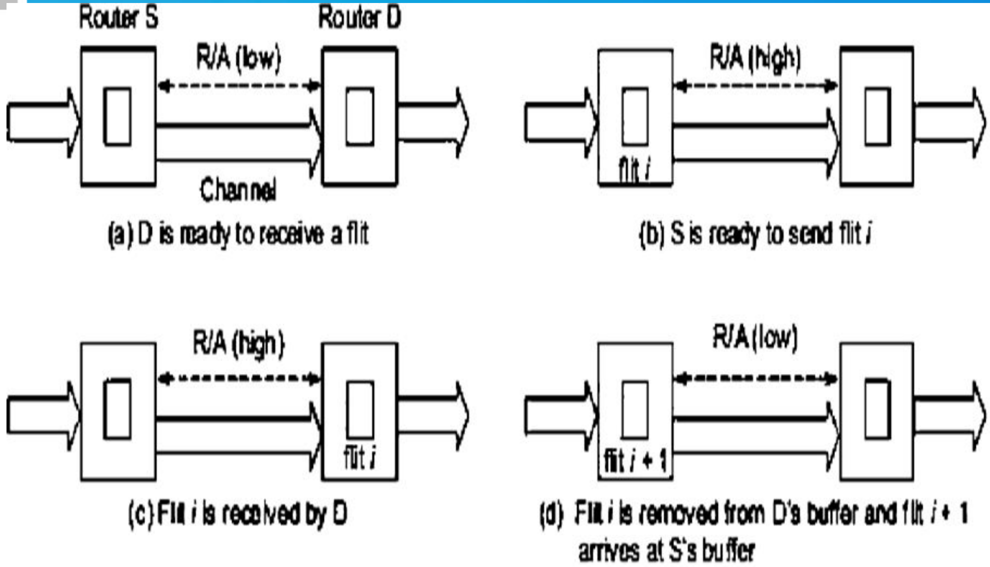
MODULE 4- PART 2

## Asynchronous Pipelining –WH routing

- The pipelining of successive flits in a packet is done asynchronously using a handshaking protocol.
- Along the path, a 1-bit ready/request (R/A) line is used between adjacent routers.
- When the receiving router (D) is ready to receive a flit
  - –the flit buffer is available
  - – it pulls the R/A line low.
- When the sending router (S) is ready,
  - –it raises the line high and
  - –transmits flit i through the channel.

- While the flit is being received by D, the R/A line is kept high.
- After flit i is removed from D's buffer (i.e. is transmitted to the next node),
  - the cycle repeats itself for the transmission of the next flit i+1 until the entire packet is transmitted.





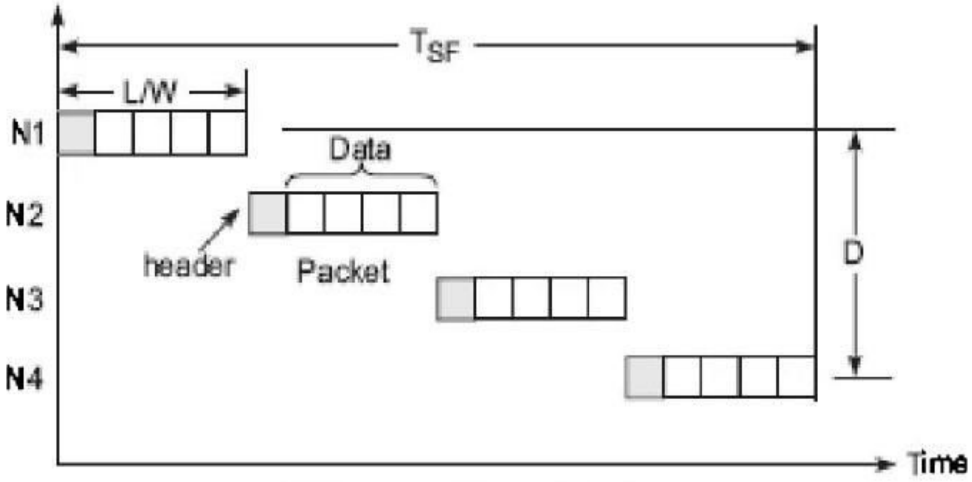

Handshaking protocol between two wormhole routers

- Asynchronous pipelining can be very efficient, and the clock used can be faster than that used in a synchronous pipeline.
- However, the pipeline can be stalled if flit buffers or successive channels along the path are not available during certain cycles.



## Latency Analysis- SF routing

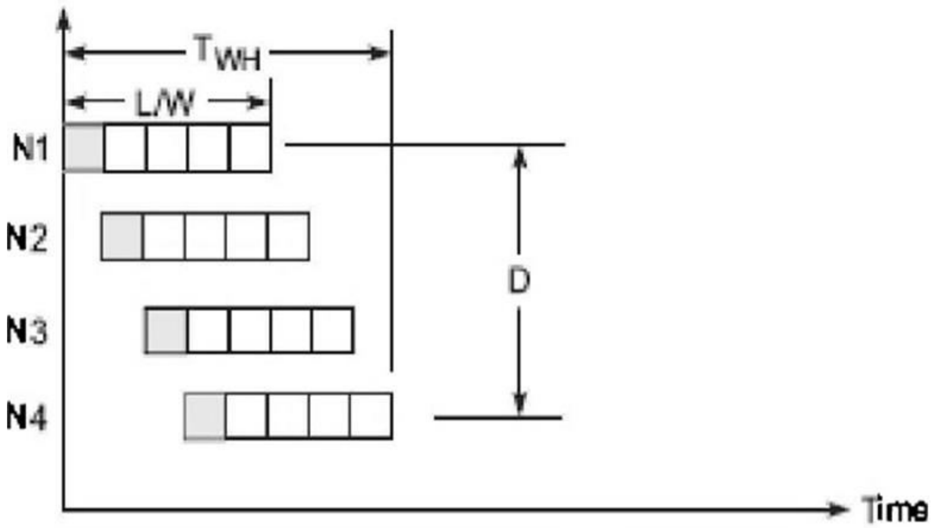
- Time comparison between the two routing techniques




(a) Store-and-forward routing



## Latency Analysis –WH routing




(a) Wormhole routing

- Let L be the packet length (in bits), W the channel bandwidth (in bits/s), D the distance (number of nodes traversed minus 1), and F the flit length (in bits).
- The communication latency $T_{SF}$ for a store-and-forward network is expressed by

$$T_{SF} = \frac{L}{W}(D+1)$$

- The latency $T_{WH}$ for a wormhole-routed network is expressed by

$$T_{WH} = \frac{L}{W} + \frac{F}{W} \times D$$

- $T_{SF}$ is directly proportional to D.
- $T_{WH} = L / W$ if $L >> F$. Thus the distance D has a negligible effect on the routing latency.



- Ignored the network startup latency and block time due to resource shortage (such as channels being busy or buffers being full, etc.)
- The channel propagation delay has also been ignored because it is much smaller than the terms in TSF or TWH·
- A typical first generation value of TSF is between 2000μs and 6000μs, while a typical value of TWH is 5μs or less.
- Current systems employ much faster processors, data links and routers.
- Both the latency figures above would therefore be smaller, but wormhole routing would still have much lower latency than packet store-and-forward routing.

# COMPUTER SYSTEMS ARCHITECTURE

## Flow Control Strategies (Malayalam)

MODULE 4- PART 4

## Flow Control Strategies

- Flow control strategies are used to control n/w traffic flow without causing congestion or deadlock situations
- When two or more packet collide at a node, policies must be set for resolving their conflict

- In order to move a flit between adjacent nodes in a pipeline of channels, three elements must be present:
  1. the source buffer holding the flit,
  2. the channel being allocated, and
  3. the receiver buffer accepting the flit.
- When two packets reach the same node, they may request the same receiver buffer or the same outgoing channel.
- Two arbitration decisions must be made
  1. Which packet will be allocated the channel? And
  2. What will be done with the packet being denied the channel?



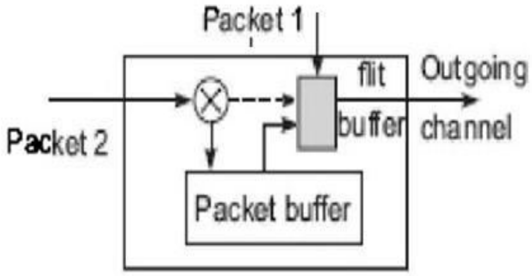
## Flow control Policies for Collision Resolution

- Buffering
- Blocking Policy
- Discard and retransmission
- Detour after being blocked

## Buffering Method

- This method is applied in virtual cut routing
- When packet 1and 2 collide at particular point
  - Packet 1 allocated to the channel
  - Packet 2 is denied
    - Packet 2 is temporarily stored in packet buffer
    - It will transmitted later, when the channel becomes available
- Advantage
  - Already allocated resources are not wasted
- Disadvantage
  - Require the use of large buffer to hold the entire packet
  - Cause significant storage delay

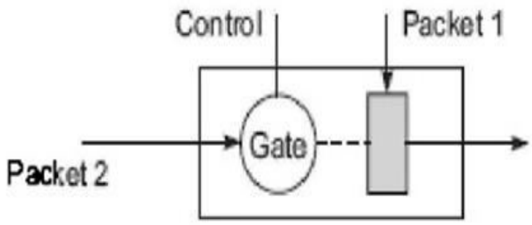





(a) Buffering in virtual cut-through routing

## Blocking Policy

- Pure wormhole routing use this scheme
- Second packet isblocked from advancing
- However it is not abandoned

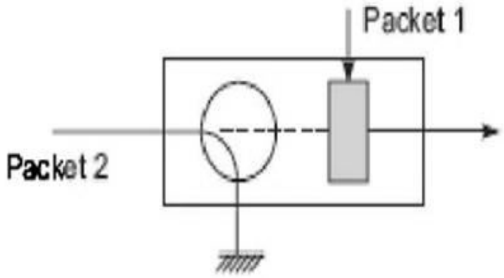



(b) Blocking flow control



## Discard Policy

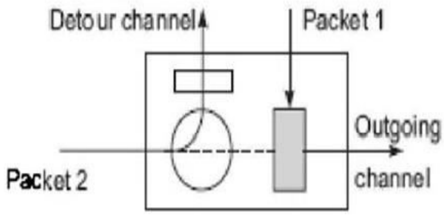
- It drops packet which is blocked
- This scheme results in severe wastage of resources
- It demands packet retransmission and acknowledgement
- Rarely used policy because of packet delivery rate




(c) Discard and retransmission

## Detour policy

- Blocked packet is routed through a detour channel
- It is economical to implement
- Offers more flexibility
- Disadvantage
  - Result in idling of resources allocated to blocked packet
  - Waste more channel recourses




(d) Detour after being blocked

31

# COMPUTER SYSTEMS ARCHITECTURE

Dimension-Order Routing
E-Cube and XY Routing
(Malayalam)

MODULE 4- PART 5

## Packet routing

- Packet routing can be conducted deterministically or adaptively.
- Deterministic routing
  - –the communication path is completely determined by the source and destination addresses.
  - –the routing path is uniquely predetermined in advance, independent of network condition.
- Adaptive routing
  - –may depend on network conditions, and alternate paths are possible.



- In both types of routing, deadlock free algorithms are desired.
- Dimension-order routing requires the selection of successive channels to follow a specific order based on the dimensions of a multidimensional network.
- Two such deterministic routing algorithms are given below, based on a concept called dimension order routing.
  - –X-Y routing
    - a routing path along the X-dimension is decided first before choosing a path along the Y-dimension
  - –E-cube routing.
    - For hypercube (or n-cube) networks

# E-cube Routing on Hypercube

- Consider an n-cube with N= $2^n$ nodes.
- Each node b is binary-coded as b = bn-1 bn-2 . . . b1 b0 .
- source node is S= Sn-1 . . . S1S0 and
- destination node is d = dn-1 . . . d1d0.
- We want to determine a route from s to d with a minimum number of steps.
- Denote the n dimensions as i = 1,2, . . . , n, where the ith dimension corresponds to the (i – 1)st bit in the node address.
- Let v = vn-1 . . . v1v0 be any node along the route.
- The route is uniquely determined as follows:

1. Compute the direction bit $r_i = s_{i-1} \oplus d_{i-1}$ for all $n$ dimensions ($i$ = 1, …, $n$). Start the following with dimension $i$ = 1 and $v = s$.
2. Route from the current node $v$ to the next node $v \oplus 2^{i-1}$ if $r_i = 1$. Skip this step if $r_i = 0$.
3. Move to dimension $i + 1$ (i.e. $i \leftarrow i + 1$). If $i \leq n$, go to step 2, else done.



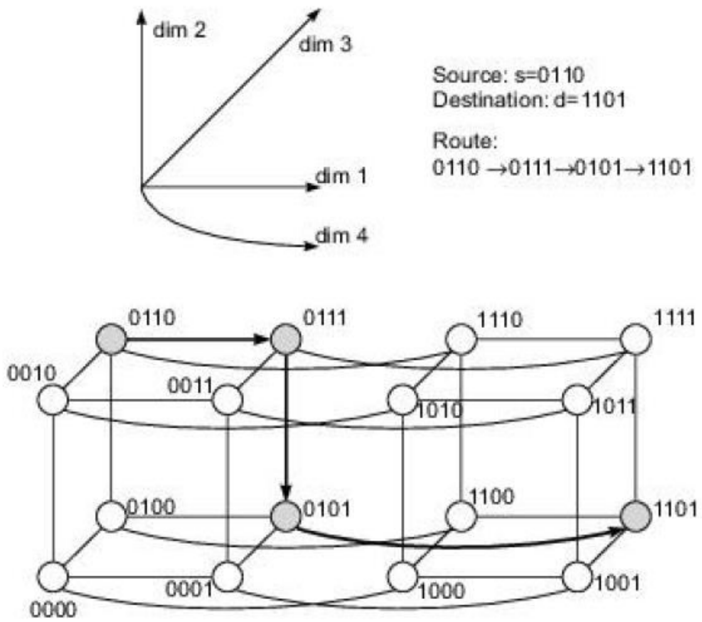



**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

- Here this is a 4-cube, N=2^4 nodes=16 nodes
- Source node S=S3S2S1S0=0110
- Destination node D=D3D2D1D0=1101
- Lets compute r in advance

$$r_i = s_{i-1} \oplus d_{i-1}$$

$$r_4 = s_3 \oplus d_3 = 1$$

$$r_3 = s_2 \oplus d_2 = 0$$

$$r_2 = s_1 \oplus d_1 = 1$$

$$r_1 = s_0 \oplus d_0 = 1$$

$$\begin{array}{rl} 0110 & - s \\ \oplus\ 1101 & - d \\ \hline 1011 & \\ r_4 r_3 r_2 r_1 & \end{array}$$



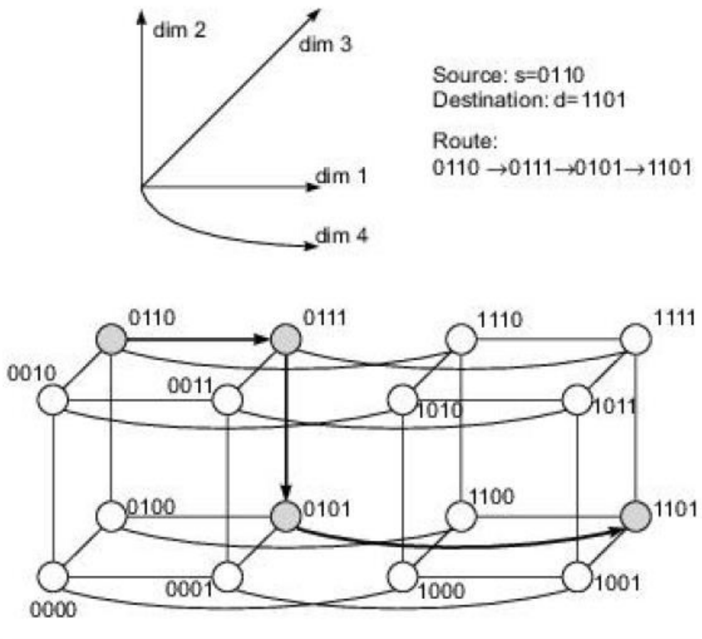



**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

- Stage 1: i=1, i<=n(4)
  1. v=s=0110
  2. .

$$v \oplus 2^{i-1} = 0110 \oplus (2^0 = 1)$$

$$\begin{array}{r} 0110 \\ 0001 \\ \hline 0111 \end{array}$$

    - Here we have to check status of ri also.
    - r1=1 so we can accept this v=0111 as next intermediate node
  3. Do i=i+1=2

  Now we move to next stage with i=2



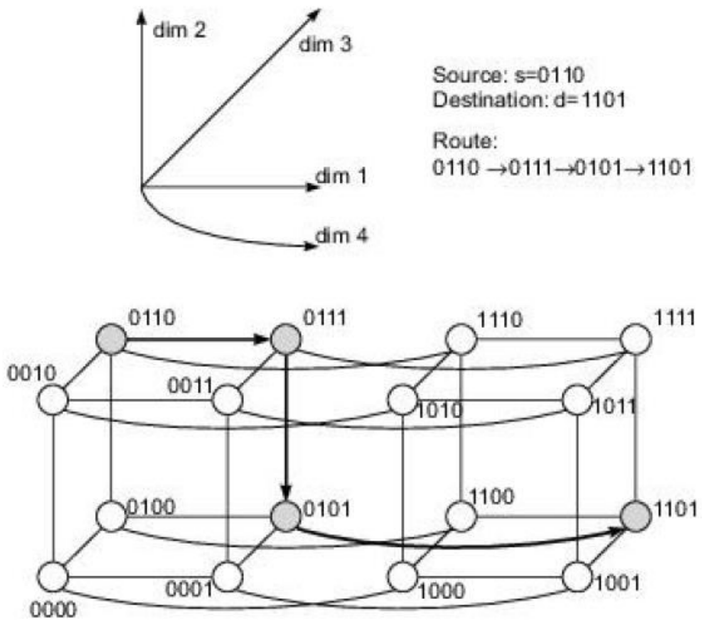



**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

- Stage 2: i=2, i<=n(4)
  1. v=0111
  2. .

$$v \oplus 2^{i-1} = 0111 \oplus 2^{2-1} \; (2^1=2)$$

$$\begin{array}{r} 0111 \\ \oplus\, 0010 \\ \hline 0101 \\ \hline r_4\, r_3\, r_2\, r_1 \end{array}$$

  - Here we have to check status of ri also.
  - r2=1 so we can accept this v=0101 as next intermediate node
  3. Do i=i+1=3

Now we move to next stage with i=3



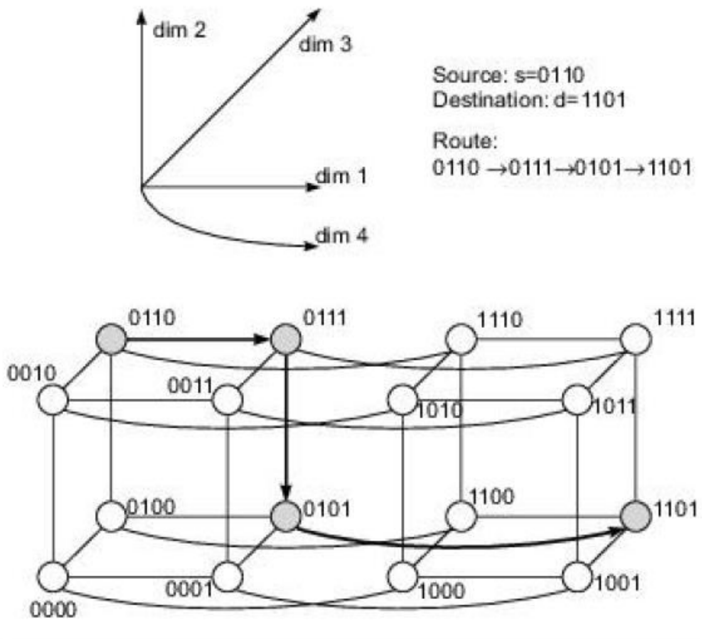



**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

- Stage 3: i=3, i<=n(4)
  1. v=0101
  2. Here when we check status of ri, r3=0
  3. so we can skip this and v remains old value v=0101
  4. Do i=i+1=4

  Now we move to next stage with i=4



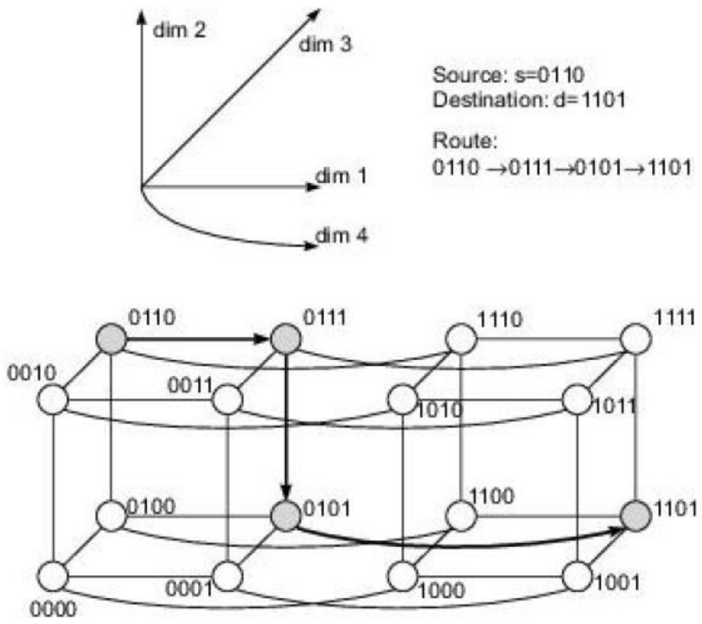



**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

- Stage 4: i=4, i<=n(4)
  1. v=0101
  2. .

$$v \oplus 2^{i-1} = 0101 \oplus 2^{4-1} \; (2^3=8)$$

$$\begin{array}{r} 1000 \\ \hline 1101 \\ r_4\, r_3\, r_2\, r_1 \end{array}$$

- Here we have to check status of ri also.
- r4=1

And we reached destination node d=1101

45

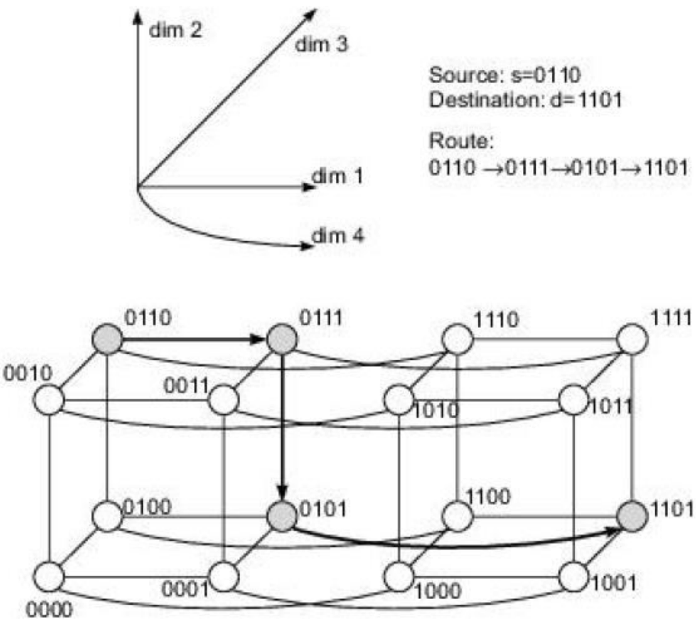


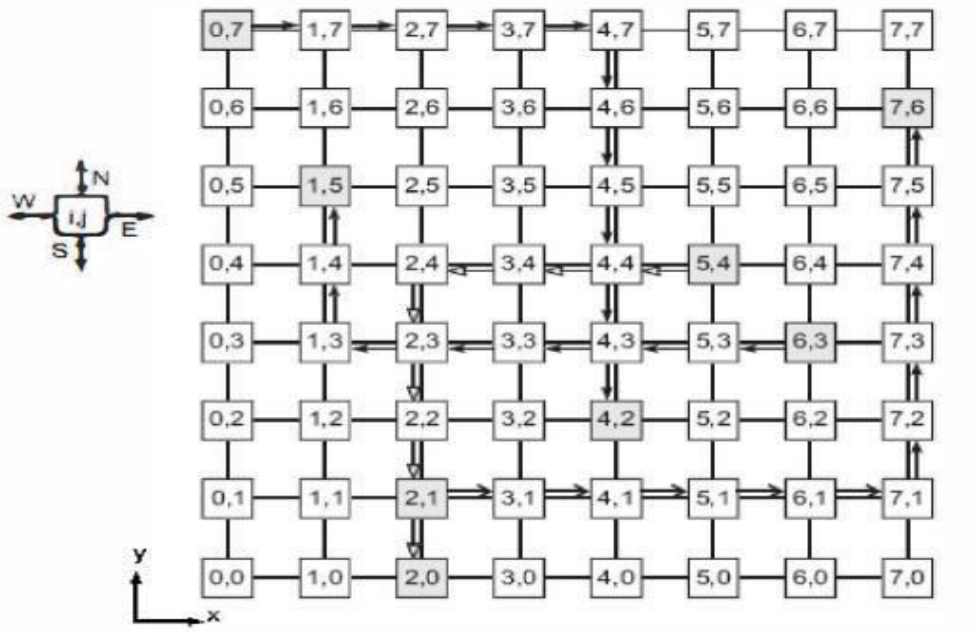

**Fig. 7.34** E-cube routing on a hypercube computer with 16 nodes

46

## X-Y Routing on a 2D Mesh

- From any source node s = (X1Y1) to any destination node d = (X2Y2),
- route from s along the X-axis first until it reaches the column Y2, where d is located.
- Then route to d along the Y-axis.
- There are four possible X-Y routing patterns
  - –east-north,
  - –east-south,
  - –west-north, and
  - –west-south paths chosen.






Four (source; destination) pairs: (2,1;7,6)⟶ (0,7;4,2)⟶ (5,4;2,0)⟶ (6,3;1,5)- - -→

**Fig. 7.35** X-Y routing on a 2D mesh computer with 8 × 8 = 64 nodes

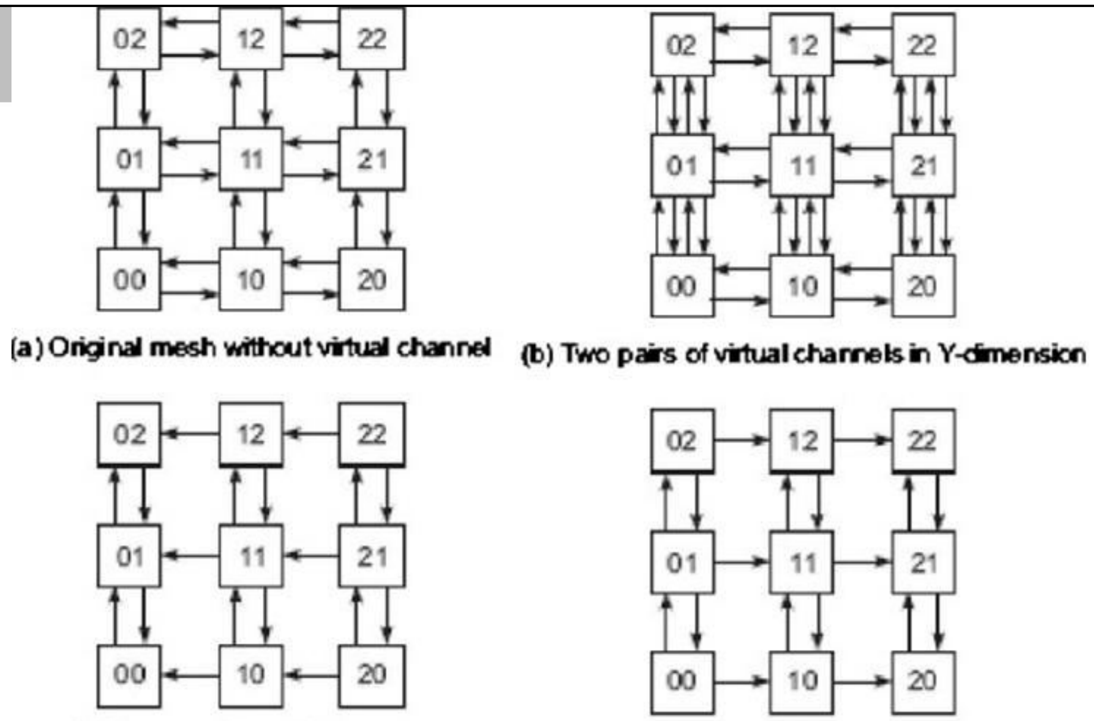
- Both E-cube and X-Y schemes can be applied in either store-and- forward or wormhole-routed networks, resulting in a minimal route with the shortest distance between source and destination.
- Non-minimal routing algorithms, producing deadlock-free routes,
  - –allow packets to traverse through longer paths, sometimes to reduce network traffic or for other reasons.



## Adaptive X-Y Routing cont.

- The main purpose of using adaptive routing is to achieve efficiency and avoid deadlock.
- The concept of virtual channels makes adaptive routing more economical and feasible to implement.
- The idea can be further extended by having virtual channels in all connections along the same dimension of a mesh-connected network
- Uses two pairs of virtual channels in the Y dimension of a mesh using X-Y routing.
- For westbound traffic, the virtual network can be used to avoid deadlock because all eastbound X-channels are not in use.
- Similarly, the virtual network supports only eastbound traffic using a different set of virtual Y-channels.
- The two virtual networks are used at different times; thus deadlock can be adaptively avoided.

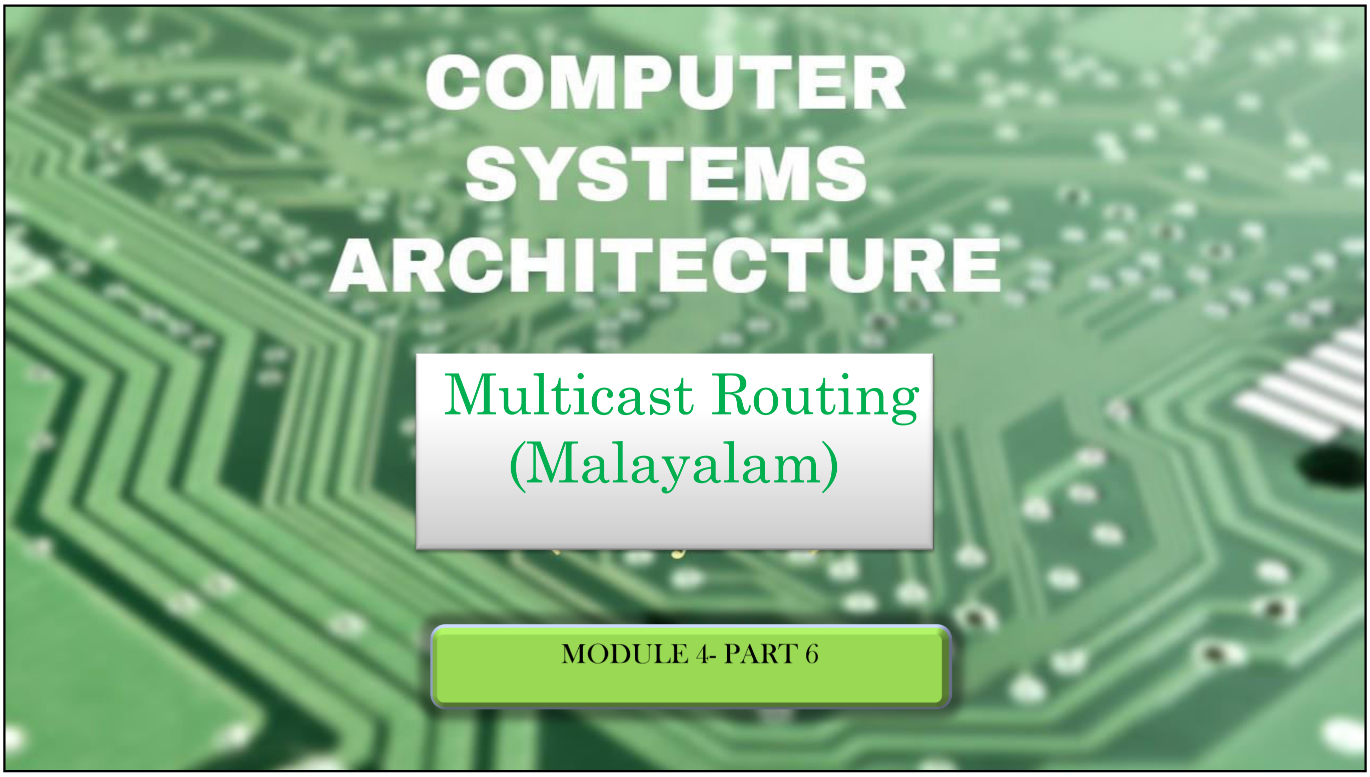
N THIS




(a) Original mesh without virtual channel (b) Two pairs of virtual channels in Y-dimension

(c) For a westbound message (d) For an eastbound message





# COMPUTER SYSTEMS ARCHITECTURE

## Multicast Routing (Malayalam)

MODULE 4- PART 6

## Multicast Routing Algorithms

- Various communication patterns are specified in this
- Routing efficiency is defined.
- The concept of virtual networks and network partitioning are applied to realize the complex communication patterns with efficiency.

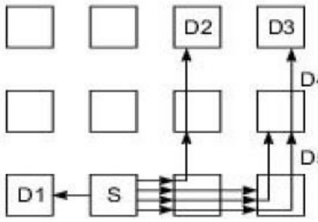


## Communication Patterns

- A multicast pattern
  - corresponds to one-to-many communication in which one source sends the same message to multiple destinations.
- A broadcast pattern
  - corresponds to the case of one-to-all communication.
- The most generalized pattern is the many-to-many conference communication.
- All patterns can be implemented with multiple unicasts sequentially, or even simultaneously if resource conflicts can be avoided.
- Special routing schemes must be used to implement these multi-destination patterns.

## Multicast and broadcast on a mesh connected network

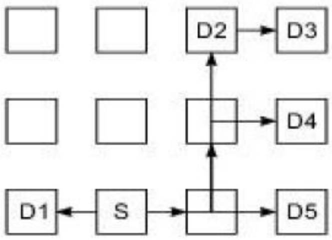
- The source nude is identified as S, which transmits a packet to five destinations labeled Di, for i = 1, 2, ..., 5.
- This five destination unicast can be implemented by five unicasts
- The X-Y routing traffic requires the use of 1 + 3 + 4 + 3 + 2 = 13 channels, and
- the latency is 4 for the longest path leading to D3.



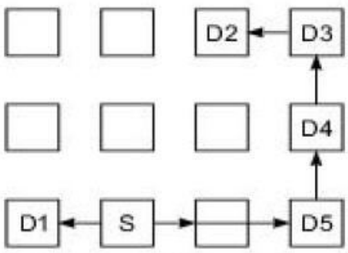

(a) Five unicasts with traffic = 13 and distance = 4



- A multicast can be implemented by replicating the packet at an intermediate node, and multiple copies of the packet reach their destinations with significantly reducec channel traffic.
- For a store-and-forward network, the route in Fig. 1 is better and has a shorter latency.
- On a wormhole-routed network, the multicast route in Fig. 2 is better.

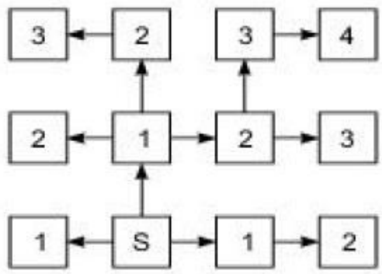



(b) A multicast pattern with traffic = 7 and distance = 4




(c) Another multicast pattern with traffic = 6 and distance = 5

- A four-level spanning tree is used from node S to broadcast packet to all the mesh nodes
- Nodes reached at level i of the tree have latency i.
- This broadcast tree should resu in minimum latency as well as i minimum traffic.




(d) Broadcast to all nodes via a tree (numbers in nodes correspond to levels of the tree)



## Routing Efficiency

- Two commonly used efficiency parameters are channel bandwidth and communication latency.
- The channel bandwidth
  - –at any time instant indicates the effective data transmission rate achieved to deliver the messages.
- The latency
  - –is indicated by the packet transmission delay involved.
- An optimally routed network
  - –should achieve both maximum bandwidth and minimum latency for the communication patterns involved.
- However, these two parameters are not totally independent.

- Achieving maximum bandwidth may not necessarily achieve minimum latency at the same time, and vice versa.
- Depending on the switching technology used, latency is the more important issue in a store-and-forward network, while in general the bandwidth affects efficiency more in a wormhole routed network.



- Extending the multicast tree, one should compare the reachability via all dimensions before selecting certain dimensions to obtain a minimum cover set for the remaining nodes.
- In case of a tie between two dimensions, select any one of them.
- Therefore, the tree may not be uniquely generated.
- The greedy multicast algorithm requires the least number of traffic channels compared with multiple unicasts or a broadcast tree.
- To implement multicast operations on wormhole-routed networks, the router in each node should be able to replicate the data in the flit buffer.
- In order to synchronize the growth of a multicast tree or a broadcast tree, all outgoing channels at the same level of the tree must be ready before transmission can be pushed one level down.
- Otherwise, additional buffering is needed at intermediate nodes.

## Virtual Networks

- Consider a mesh with dual virtual channels along both dimensions.
- These virtual channels can be used to generate four possible virtual networks.

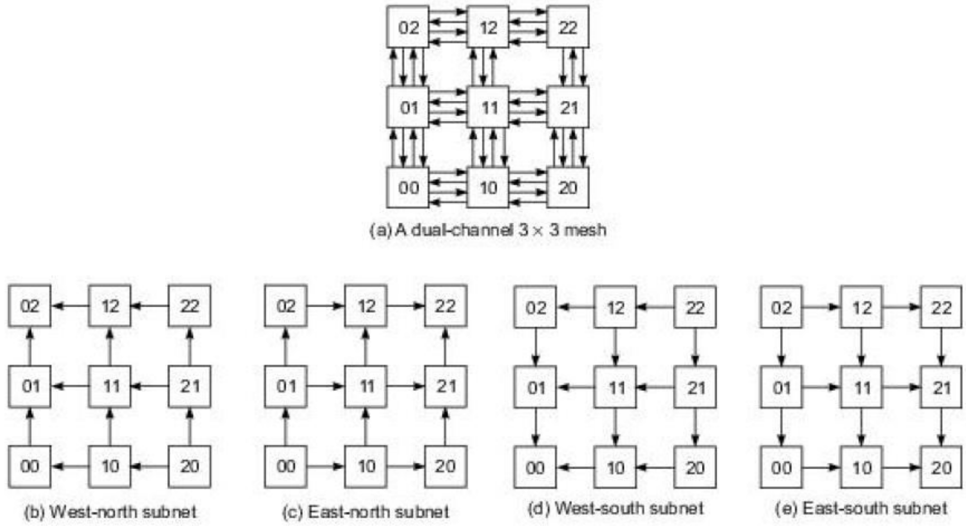





**Fig. 7.39** Four virtual networks implementable from a dual-channel mesh

- No cycle is possible on any of the virtual networks.
- Thus deadlock can be completely avoided when X-Y routing is implemented on these networks.
- If both pairs between adjacent nodes are physical channels, then any two of the four virtual networks can be simultaneously used without conflict.
- If only one pair of physical channels is shared by the dual virtual channels between adjacent nodes, then only (b) and (e) or (c) and (d) can be used simultaneously.
- Other combinations, such as (b) and (c), or (b) and (d), or (c) and (e), or (d) and (e), cannot coexist at the same time due to a shortage of channels.