

6.911

Networks

A Great Place to Learn About
Contention, Collision, and Congestive Collapse

Where the bits are going, and why.

Two Broad Categories:

Static Networks:

- Network nodes are very close together (same box, board, or chip).
- Tight control of network design allows extreme specialization.
- Wide variety of topologies and routing schemes possible.
- Useful for very tightly-coupled multiprocessors, especially when using shared memory.

Dynamic Networks:

- Nodes may be separated by any distance.
- Characterized by looser control of topology, timing, and physical conditions.
- A more general approach. Active switching elements take the place of exotic topologies.

What to do with \$100 million.

The Specialization Approach

Make the Network Fit the Computation:

- The ideal static network is a complete graph. Unfortunately, it is not physically practical. Real computers settle for a subset of this.
- Choosing the best subset depends on the application. Different fixed topologies offer different balances between bandwidth, latency, and complexity.
- The smaller the subset, the more a dynamic a static network must be.
- Despite their specialization, most static networks are still subject to deadlock and contention issues.

Examples:

Illiac IV: 64 processing nodes in an Illiac mesh configuration. Each node is adjacent to four others, and no two nodes are more than seven links apart.

CM-2: 4096 processing nodes (16 processors each) in a 12-D hypercube. Good node degree and diameter, but not as scalable as Hillis claims.

The General Approach

Share the Pipes:

- Dynamic networks reduce cost and complexity by sharing communication channels. This is accomplished using switches and access control.
- Routing and arbitration systems assume no specific topology, allowing for extreme scalability.
- Generality comes with increased latency and lower reliability, making use in shared memory systems awkward.
- Switched networks typically shift the burden of receipt acknowledgement and deadlock avoidance onto higher-level software.

Examples:

Ethernet: As evidenced by the popularity of Beowulf clusters, off-the-shelf interconnect has a place in supercomputing.

Myrinet: This high-performance alternative to gigabit ethernet is specifically designed for cluster computing.

A look at the numbers that shape your network.

Performance Metrics:

- Link Bandwidth: The bandwidth of the channel connecting a single host to the network. This does not have to be the same for every link. It can also be variable.
- Bisection Bandwidth: The minimum bandwidth across any line that bisects the network graph.
- “Full” Bisection Bandwidth: In practical terms, a network has full BBW if it does not saturate before every node interface saturates.
- Latency: The time required for a message to reach its destination. Low latency is very important for tightly-coupled computations.
- Blocking policy: How a network deals with link contention affects both the latency and the need for higher-level software.

To Memphis and back again.

Routing:

- Static networks typically use special routing algorithms and hardware to fit the topology in question. Often, a fixed routing function applies.
- Dynamic networks need to support less structured topologies, and thus routing is often more complex. Switched networks must either map the network or maintain routing tables at some level.
- Network switches use either store-and-forward, wormhole routing, or virtual cut-through.
- Most dynamic networks use a low level addressing system in addition to medium-independent protocols such as TCP.

Always another problem.

Conflict:

- All networks (except complete graphs) must deal with resource contention on some level. This is often done with packet blocking or redirection.
- Packet blocking is problematic because it stops a packet in the middle of the network, creating more blocking conditions elsewhere and inviting deadlock.
- Packet redirection avoids blocking problems by sending packets to the wrong switch outputs when necessary. This prevents deadlock, but introduces the possibility of livelock, and decreases network efficiency.
- Deadlock may be avoided by means of virtual channels.
- Slow-start and backoff mechanisms allow efficient use of bandwidth in unpredictable conditions.

Questions:

What is Karn's clamped retransmit backoff? It almost sounds like a stunt done at the circus.

What does a combining switch do and why is it cool?

How does an Ethernet switch work? What kind of latency does it incur? How do they achieve such high aggregate bandwidth?

How robust can a network be made against denial-of-service attacks? This also relates to multiuser processing on large machines -- to what extent is it possible for one user's task to take down a machine?

Have any computers employed heterogenous network topologies?

What's the bisection bandwidth of the Internet?

Where is the opportunity for improvement in the fundamental characteristics of networks in the next 20 years? Just fatter and fatter pipes?