Chapter 5

Distributed Memory Systems: Part II

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies

Assumptions

- All network Links are bidirectional
- All-Port-Communication: each node can send out messages on all outgoing links simultaneously
- The same holds for receiving messages
- A messages consists of several bytes sens uninterruptedly

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies

Assumptions

- The time for transmission of a message of m bytes size is

$$T(m) = t_s + m t_b,$$

  where t_s is a startup time and t_b is the time for sending a single byte.
- The communication is such that the length of the path from source node to destination node in the corresponding network graph determines the number of time steps required.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies

Landau Θ-notation

The $\Theta(g(x))$ notation describes a class of functions f for which roughly speaking we have that *"f is growing essentially as fast as g."*
More precisely we have,

$$\begin{aligned}\Theta(g(x)) = \{f(x) \mid \exists c_1, c_2 > 0 \text{ and } x_0, \text{ such that} \\ \forall x \geq x_0 \;\; c_1|g(x)| \leq f(x) \leq c_2|g(x)|\}\end{aligned}$$

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies

Critical operations are the collective communication operations, since they produce a notable load on the entire range of links in the network. We will investigate the following in more detail:

1. broadcast
2. scatter
3. multi-broadcast (each node broadcasts)
4. total exchange (each node scatters)

In the following p specifies the number of nodes in the network.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Complete Graph

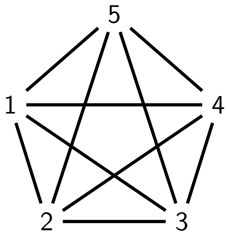


Figure: A complete graph network broadcast example

- all nodes connected, i.e., path length is one,
- by the assumptions all messages in all types of point to point and collective communication operations can be sent simultaneously,
- the operations can be performed in $\Theta(1)$.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Linear Array

Figure: A linear array network example

Single Broadcast

- The root node sends messages to its left and right neighbors starting with the most distant recipients,
- in all other steps each node forwards the message received from one neighbor in the previous step to its other neighbor.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Linear Array

1 —— 2 —— 3 —— 4 —— 5

Figure: A linear array network example

Single Broadcast

- The root node sends messages to its left and right neighbors starting with the most distant recipients,
- in all other steps each node forwards the message received from one neighbor in the previous step to its other neighbor.
- The minimal runtime is $\lfloor \frac{p}{2} \rfloor$ (root is the center node)
- The maximal runtime is $p-1$ (root is an end node)

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Linear Array

1 —— 2 —— 3 —— 4 —— 5

Figure: A linear array network example

Single Broadcast

- The root node sends messages to its left and right neighbors starting with the most distant recipients,
- in all other steps each node forwards the message received from one neighbor in the previous step to its other neighbor.
- The minimal runtime is $\lfloor \frac{p}{2} \rfloor$ (root is the center node)
- The maximal runtime is $p - 1$ (root is an end node)
- Thus the runtime class is $\Theta(p)$.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Linear Array

Multi Broadcast

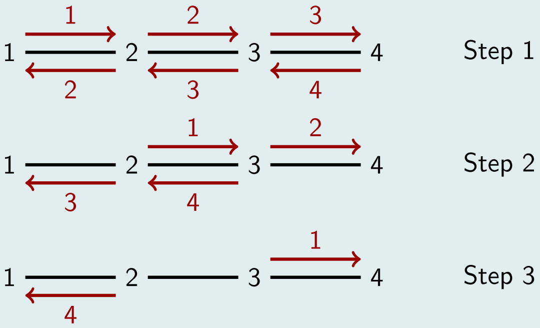


Figure: A linear array network multi broadcast example

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Linear Array

Scatter

The basic idea is that of the single broadcast, only the content of the messages need to be treated more carefully. Therefore, the complexity is $\Theta(p)$ as well.

Total exchange

An upper bound to the runtime is given by p scatter operations, resulting in basically p^2 communication steps. Intheir book Rauber and Rünger present an algorithm that can do it in $\frac{p^2}{4}$. Anyway the complexity is $\Theta(p^2)$.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Ring

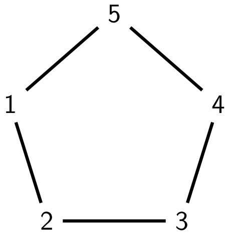


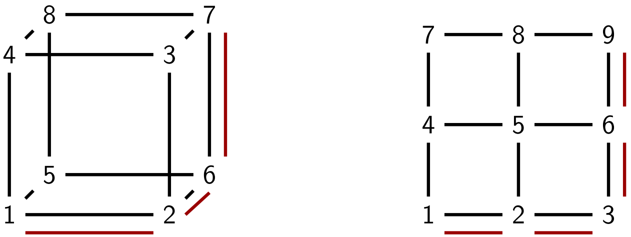
Figure: A ring network example

The ring is a prototype for the linear array where the root node is always in the center. Thus, we get the same complexities as in the best case for the linear array.
Note, however, that we need to cut the transmission at half way around the ring.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Mesh

We consider the d-dimensional mesh with $\sqrt[d]{p}$ nodes per direction, such that we have p nodes in total as before. The diameter then is $d(\sqrt[d]{p}-1)$.



(a) A 3d cubic mesh network with diameter $3 \cdot 1$

(b) A 2d square mesh network with diameter $2 \cdot 2$

Figure: Two mesh network examples with diameter indications.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Mesh

Single Broadcast

The single broadcast time is obviously proportional to the diameter of the network. This itself is proportional to the number of nodes in each direction. Therefore, the complexity class is $\Theta(\sqrt[d]{p})$.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Mesh

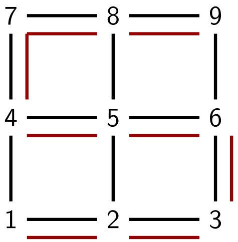


Figure: The linear array embedded in a 2d mesh.

Scatter

- The picture shows clearly that the communication time is limited by that for the linear array from above.
- On the other hand, each node has d to $2d$ outgoing connections and $p-1$ messages need to be sent, i.e., $\lfloor \frac{p-1}{d} \rfloor$ is a lower limit.
- a scatter is possible in $\Theta(p)$.

Communication Networks (revisited)

Asymptotic Message runtimes in some Standard Network Topologies: Mesh

Multi Broadcast

The multi broadcast time is observed similarly to be part of the complexity class $\Theta(\sqrt[d]{p})$.

Total Exchange

Rauber and Rünger in their book show a method that provides $\Theta(p^{\frac{d+1}{d}})$.