Hwajung Lee

# ITEC452 Distributed Computing

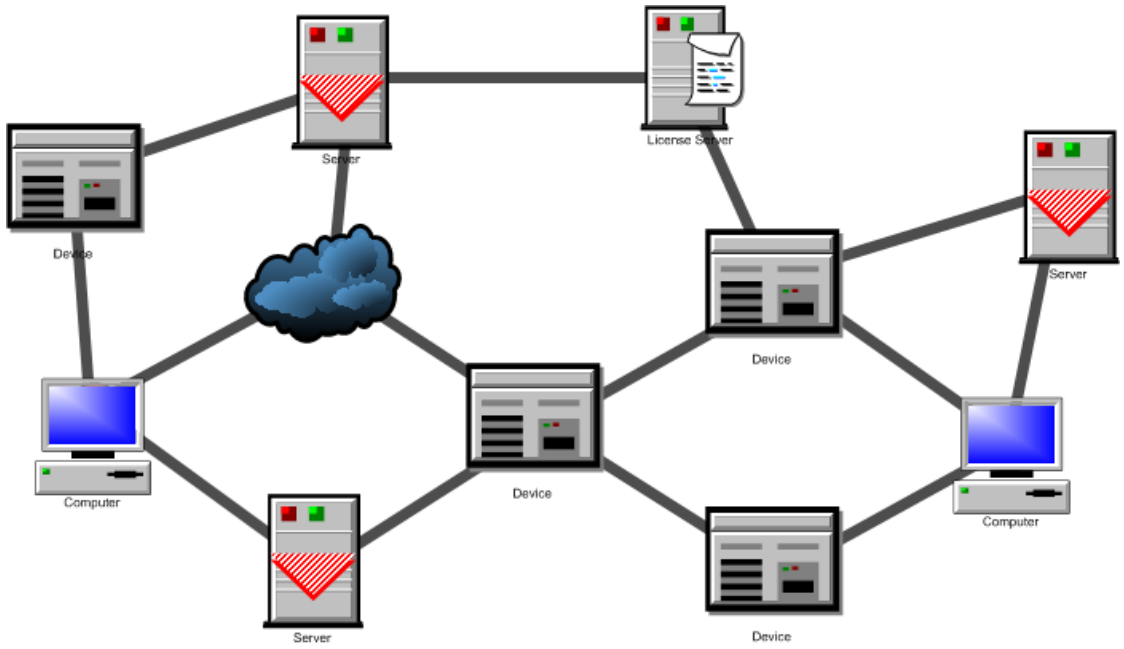
## Lecture 6
## Introduction to Distributed Systems

Maximilien Brice, © CERN

# Why distributed systems ?

- Fact: Processor population is exploding. Technology has dramatically reduced the price of processors.
- Geographic distribution of processes
- Resource sharing as used in P2P networks
- Computation speed up (as in a grid)
- **Fault tolerance**

# What is a distributed system? (1)




- A network of processes/resources.
- The nodes are processes/resources, and the edges are communication channels.

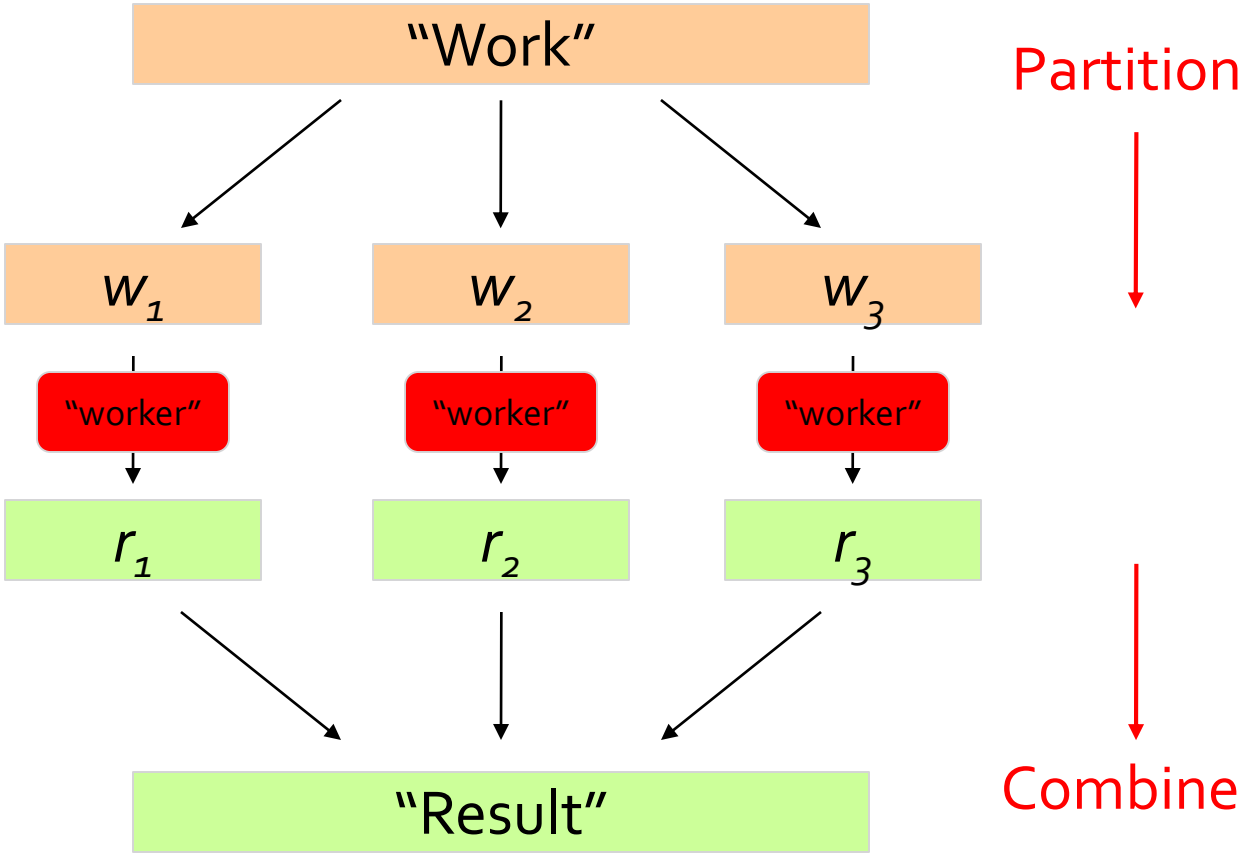
# What is a distributed system? (2)

- The logical distribution of functional capabilities
  - Multiple processes
  - Interprocess communication
  - Disjoint address space
  - Collective goal

# What is a distributed system? (2) Collective Goal?

- Don't hold your breath:
  - Biocomputing
    - Using biologically derived molecules
  - Nanocomputing:
    - the manipulation of matter on an atomic and molecular scale
  - Quantum computing
    - use of quantum-mechanical phenomena, such as superposition and entanglement
  - …
- It all boils down to…
  - Divide-and-conquer
  - Throwing more hardware at the problem

# What is a distributed system? (2) Divide and Conquer




Reference: Lecture Note by Jimmy Lin, Univ. of Maryland

# What is a distributed system? (2) Different Workers

- Different threads in the same core
- Different cores in the same CPU
- Different CPUs in a multi-processor system
- Different machines in a distributed system

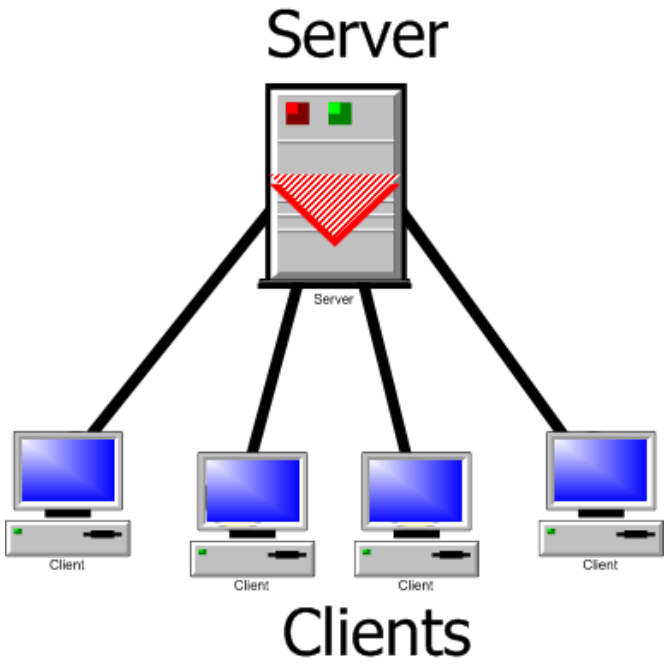
Reference: Lecture Note by Jimmy Lin, Univ. of Maryland

# What is a distributed system? (2) Choices, Choices, Choices

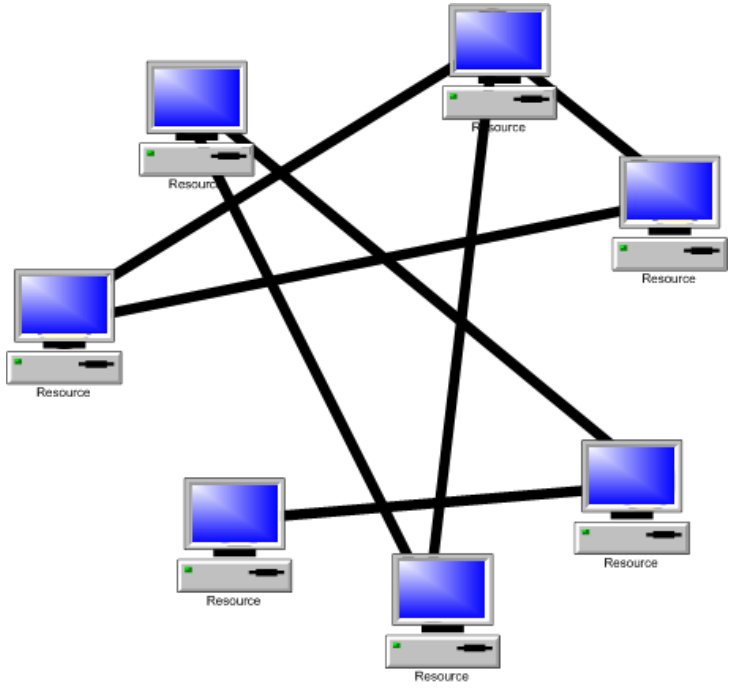
- Commodity vs. "exotic" hardware
- Number of machines vs. processor vs. cores
- Memory vs. disk vs. network bandwidth
- Different programming models

Reference: Lecture Note by Jimmy Lin, Univ. of Maryland

# A classification




Client-server model




Peer-to-peer model

# Parallel vs. Distributed

- In both parallel and distributed systems, the events are ***partially ordered***.

- In **parallel** systems, the primarily issue is speed-up
- In **distributed** systems the primary issues are fault-tolerance and availability of services

# Important services

- Internet banking
- Web search
- Net meeting
- Distance education
- Internet auction
- Google earth
- Google sky
- And so on…

# Examples

- Large networks are very commonplace these days. Think of the world wide web. Other examples are:
  - Ubiquitous Computing
  - Cloud computing
  - Grid computing, Grid computing networks
    - Ex. Computational grids (OSG, Teragrid, SETI@home)
  - Sensor networks
  - Network of mobile robots
  - And so on…

# Sensor Network

Checking the structural integrity

# Mobile robots

# Cloud Computing

Image Source: www.vemurivenkatrao.com/nature/cloud/

# Important issues

- Knowledge is local
- Clocks are not synchronized
- No shared address space
- Topology and routing
- Scalability
- Fault tolerance

# Some common sub-problems

- Leader election
- Mutual exclusion
- Time synchronization
- Distributed snapshot
- Replica management