01_Introduction

1. Definitions of Distributed Systems

1.1. Definition 1

A **Distributed System (DS)** is a system that is characterized by

- **autonomy** (distribution of **authority**)
- **cooperation** (distribution of **functionality**)
- **communication** (distribution of **information**)

1.2. Definition 2

A **Distributed Computer System**

- consists of multiple autonomous processors

- **do not share memory**,
- **cooperate by sending messages** over a communications network

1.3. Examples of Important Distribued Systems

Distributed Databases

e.g. WWW

Distributed Operating Systems

e.g. P2P

Distributed Applications

e.g. Multiplayer online games

2. Properties of Distributed Systems

2.1. Properties

- **not necessarily a regular structure** such as leader/identical processor
- **no directly accessible common state** such as shared variable in shared memory
- **no common clock**
- **non-determinism:** components make progress independently
- **independent failure modes**

2.2. (De-)centralized versus distributed

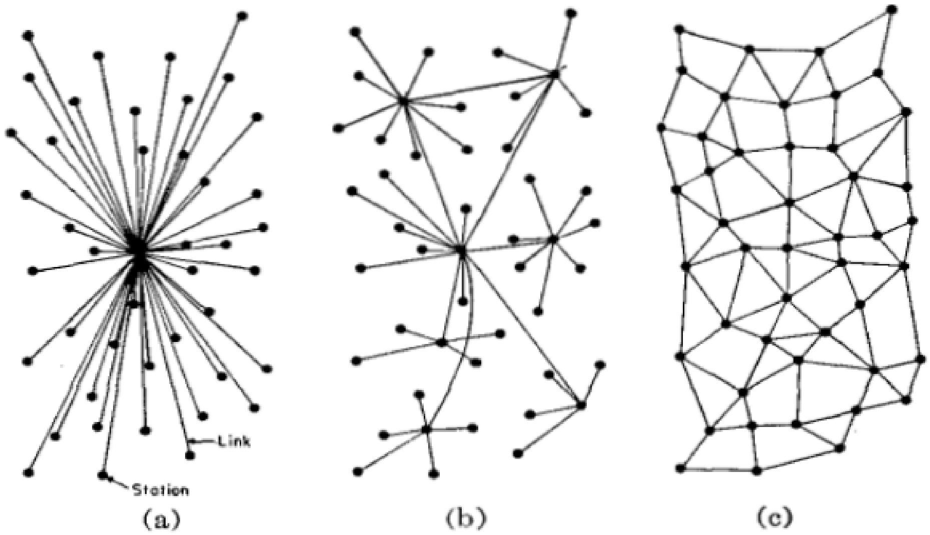
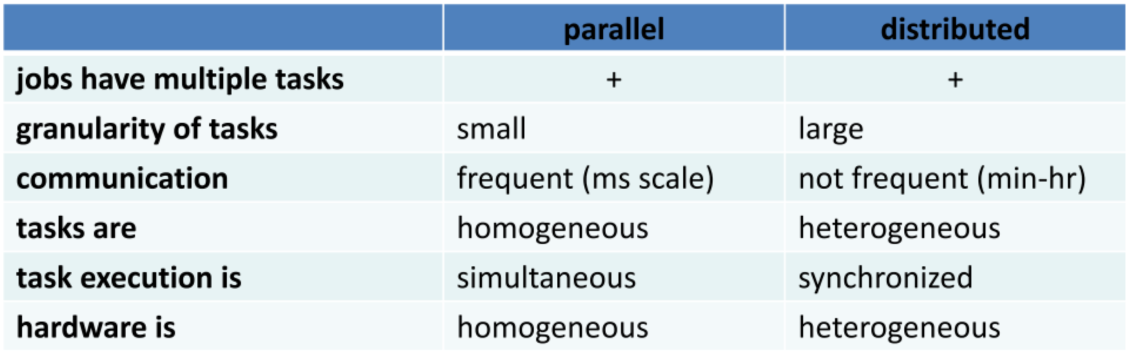


Fig. 1—(a) Centralized. (b) Decentralized. (c) Distributed networks.

2.3. Parallel versus distributed



	parallel	distributed
jobs have multiple tasks	+	+
granularity of tasks	small	large
communication	frequent (ms scale)	not frequent (min-hr)
tasks are	homogeneous	heterogeneous
task execution is	simultaneous	synchronized
hardware is	homogeneous	heterogeneous

3. Motivations for distributed systems

- organizational
- resource sharing
- extensibility
- availability
- reliablity
- security
- performance

4. Requirements for distributed systems

4.1. Transparency

a DS should present itself as a **single entity**

- Location transparency
- Access transparency
- Failure transparency
- Replication transparency

4.2. Scalability

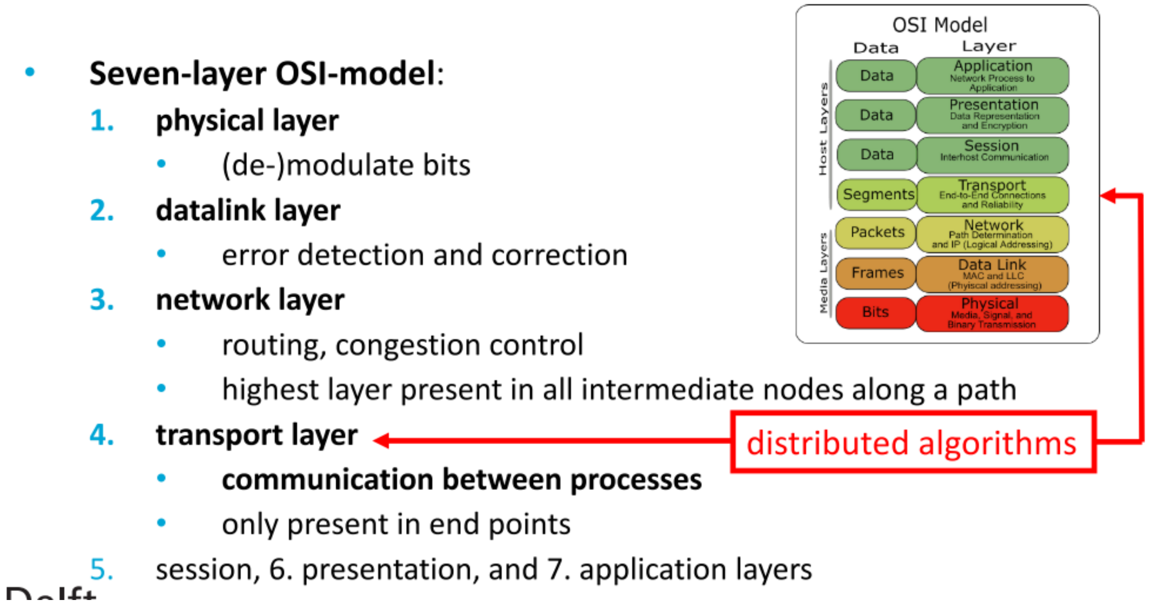
extending the system with the same structure yields a **proportional increase** in capacity (no potential bottleneck)

4.3. Consistency

in terms of

- performance
- user interface
- global state

5. Relation with networks



- **Seven-layer OSI-model**:
  1. **physical layer**
     - (de-)modulate bits
  2. **datalink layer**
     - error detection and correction
  3. **network layer**
     - routing, congestion control
     - highest layer present in all intermediate nodes along a path
  4. **transport layer**
     - **communication between processes**
     - only present in end points
  5. session, 6. presentation, and 7. application layers

OSI Model
Data
Layer
Host Layers
Media Layers
Data
Application
Network Process to Application
Data
Presentation
Data Representation and Encryption
Data
Session
Interhost Communication
Segments
Transport
End-to-End Connections and Reliability
Packets
Network
Path Determination and IP (Logical Addressing)
Frames
Data Link
MAC and LLC (Phyiscal addressing)
Bits
Physical
Media, Signal, and Binary Transmission
distributed algorithms

- DSs and DAs are based upon the **transport layer**
- But: **routing (layer 3)** and **link failures (layer 2)** do play a role in some distributed algorithms

6. Important techniques in DSs/DAs

6.1. Replication

multiple copies of the same entity for better availability and performance

6.2. Caching

keep results of **recent operations**

6.3. Locality

take into account the **location of objects**

6.4. Time stamps

assign logical or real times to events

6.5. Time outs

after some amount of time, try again

6.6. Randomization

flip a coin to **speed up**, or to have a solution in the first place