1. Concepts of Distributed Systems

Sistemes Distribuïts en Xarxa (SDX)
Facultat d'Informàtica de Barcelona (FIB)
Universitat Politècnica de Catalunya (UPC)
2018/2019 Q2

Contents

- **Definition of a distributed system**
- Challenges of distributed systems
- Architectures for distributed systems

Definition of a distributed system

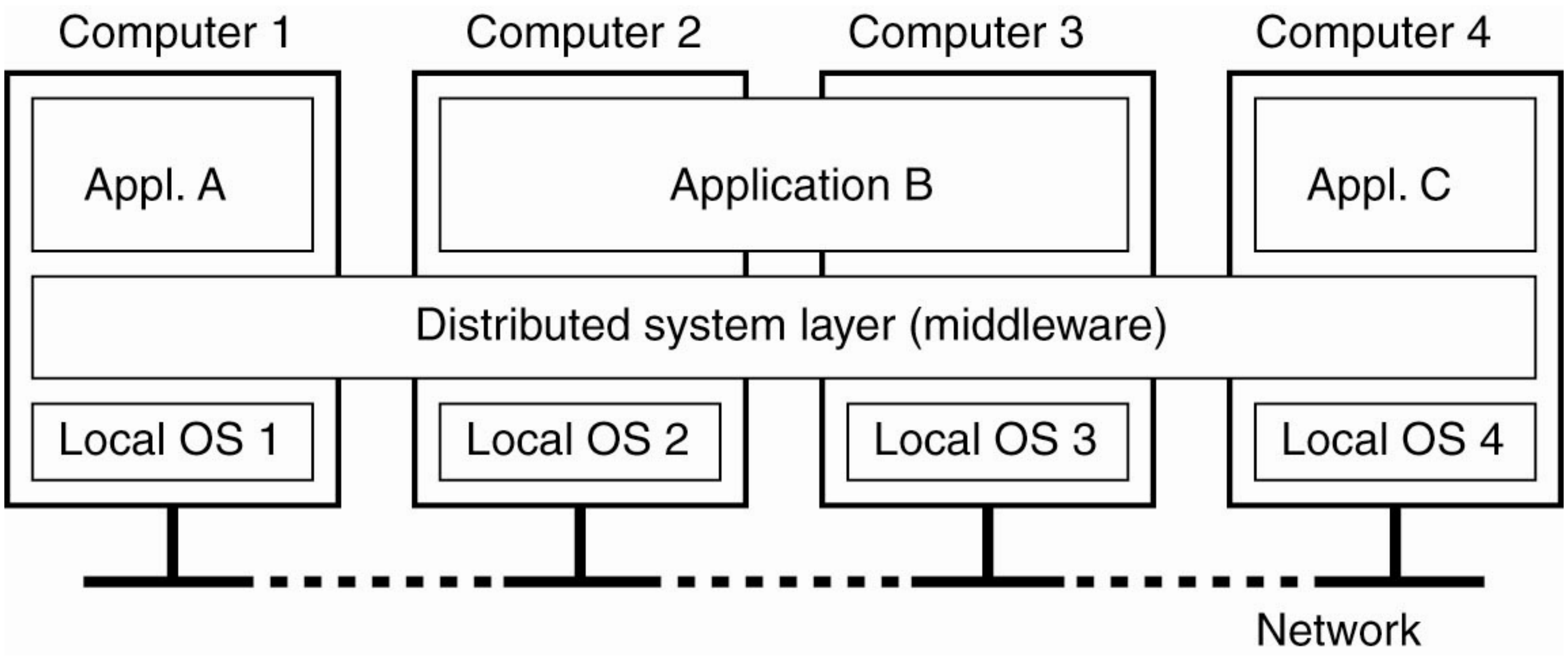
- It is one in which the failure of a computer you didn't even know existed can render your own computer unusable
  - [Lamport87]
- This apparently humorous definition entails the basic aspects of a distributed system:
  - The transparency of distribution ("a computer you didn't even know existed")
  - The interdependency of components ("the failure ... can render your own computer unusable")

Definition of a distributed system

- A collection of autonomous computing elements that appears to its users as a single coherent system
  - [Tanenbaum]
- A system where components located at networked computers communicate and coordinate their actions only by passing messages
  - [Coulouris]

Definition of a distributed system

- (1) independent networked computers but
- (2) single-system view: **middleware**



Definition of a distributed system

- Consequences of the definition:

1. Multiple heterogeneous computers
2. Multiple autonomous processes
3. Processes execute concurrently
4. Knowledge on each computer is local
5. Process address space is not shared
6. Only communication is by message passing
7. Network delays in process communication

Definition of a distributed system

- Consequences of the definition:

8. Independent (imperfect) clocks (no global time and perfect synchronization unfeasible)
9. Processes can fail independently
10. Processes may not always be available (network partitions, process disconnections)
11. High probability of failures (it grows with number of components)

READING REPORT

[Hebert13] Hebert, F., *Fallacies of Distributed Computing*, Learn You Some Erlang for Great Good! Chapter 26: Distribunomicon, No Starch Press, pp. 445-451, January 2013

Definition of a distributed system

- Why do we build distributed systems?
  - Functional: computers have different functionality
    - e.g. clients vs. servers
  - Load distribution: work divided among computers
    - e.g. P2P, distributed computing systems (SETI@HOME)
  - Fault tolerance: replication
  - Accommodate a natural geographic distribution
    - e.g. weather stations, remote resources
  - Resource sharing
    - e.g. printer in a LAN
  - Economic: many computers cheaper than one big

Contents

- Definition of a distributed system
- **Challenges of distributed systems**
- Architectures for distributed systems

Challenges of distributed systems

1. Heterogeneity
   - Need to accommodate different computers, programming languages, operating systems, …
2. Security
   - Need for confidentiality, integrity, authentication across administrative domains, …
   - Out of the scope of this course
3. Global view
   - Need to assemble meaningful global views of the system
   - Some problems that require a global view:

Challenges of distributed systems

a) Global time (we will cover this in Lesson 3)
   - Only low accuracy system clock on each node
   - Results in clock skew (two clocks, two times) and clock drift (two clocks, two count rates)

   i. Synchronize machines within a given bound with a master (with a UTC receiver) or with one another

   ii. Agree on the **order** in which events occur rather than the **time** at which they occurred

b) Global state (out of the scope of this course)
   - Each process is independent
   - Perfect clock synchronization is not feasible

   i. We need to assemble a meaningful global state from local states recorded at different real times

Challenges of distributed systems

4. Coordination
   - Need to coordinate actions in different machines, as they are running processes <u>concurrently</u> and <u>autonomously</u> (no single point of control)

   - Some problems that require coordination:

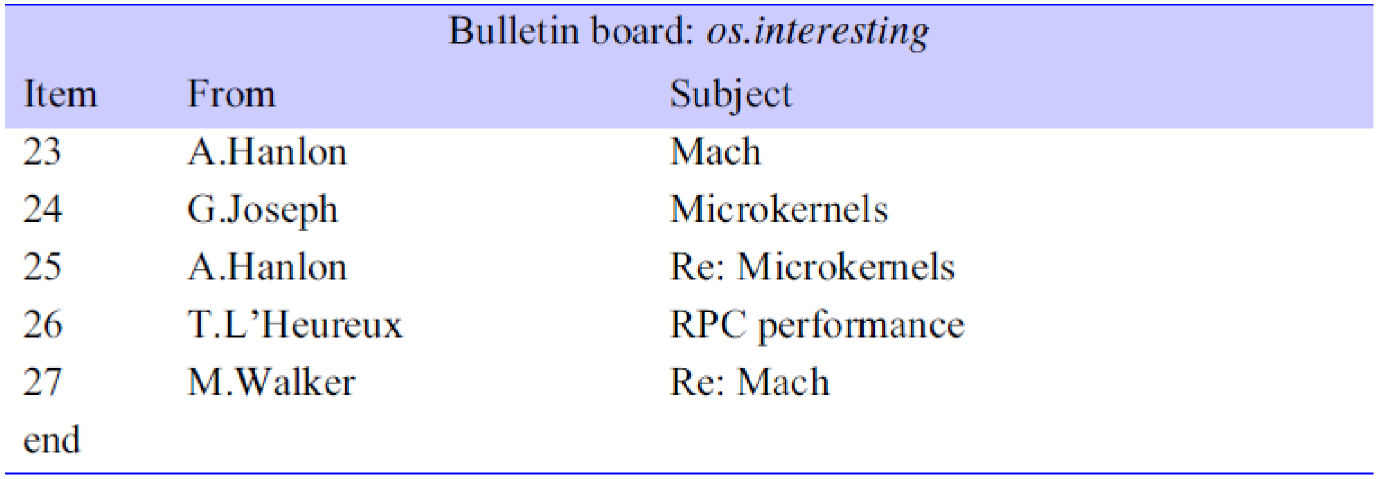
   a) Mutual exclusion (we will cover this in Lesson 4)
      - Processes coordinate their accesses to some shared resource (critical section) using solely message passing

   b) Leader election (we will cover this in Lesson 4)
      - Election of a unique coordinator of a group of processes (many algorithms need a process to act as coordinator)

Challenges of distributed systems

c) Atomic multicast (we will cover this in Lesson 4)

- Agree on which messages a group of processes receive and in which order. For instance, in a bulletin board:



Bulletin board: *os.interesting*		
Item	From	Subject
23	A.Hanlon	Mach
24	G.Joseph	Microkernels
25	A.Hanlon	Re: Microkernels
26	T.L'Heureux	RPC performance
27	M.Walker	Re: Mach
end		

d) Consensus (out of the scope of this course)

- A set of processes agree on some value after one or more of them have proposed what that value should be

Challenges of distributed systems

5. Asynchrony
   - Messages take (variable) time to be delivered
   - This results on two types of distributed systems:

   A. Synchronous distributed systems
      - Process execution time, message delay and clock drift rate have known bounds
      - **Timeouts** can be used to detect reliably process failures, lost messages, …

   B. Asynchronous distributed systems
      - Process execution time, message delay and clock drift rate are not bounded
      - No assumptions can be made about time intervals on any execution

Challenges of distributed systems

6. Openness
   - Interoperability
     - Components from different manufacturers can work together by merely relying on each other's services
   - Portability
     - Component made for a given system can run unmodified on another one implementing the same interfaces
   - Extensibility
     - Ability to add new components or replace existing ones without affecting those that stay in place

$\Rightarrow$ Offer services according to standard interfaces describing their semantics and syntax

Challenges of distributed systems

7. Transparency
   - Ability of a distributed system of presenting itself as a single computer system

Transparency	Description
Access	Enable local and remote objects to be accessed using identical operations by hiding differences in data representation and invocation mechanisms
Location	Enable objects to be accessed without knowledge of their physical location
Mobility	Allow the movement of objects and users within a system without affecting their operation
Replication	Allow multiple instances of an object to exist without knowledge of the replicas by users
Concurrency	Enable several users to operate concurrently on shared objects without interference between them
Failure	Mask from an object the failure and possible recovery of other objects

Challenges of distributed systems

- Transparency may be set as a goal, but fully achieving it is a different story
  A. There are communication latencies (due to distant locations) that cannot be hidden
  B. Completely hiding failures is impossible
     - In asynchronous systems, you cannot distinguish a slow node from a failing one
  C. Trade-off between transparency & performance
     - e.g. replication transparency requires keeping replicas up-to-date with the master copy
     - e.g. failure transparency tries repeatedly to contact a server before trying another one

Challenges of distributed systems

- Transparency may be set as a goal, but fully achieving it is a different story
  D. Exposing distribution could be good, especially when location is important
     - e.g. use of location-based services on mobile phones (finding your nearby friends or the nearest restaurant)
     - e.g. use a busy nearby printer rather than an idle one in another building
     - e.g. deal with users in different time zones

Challenges of distributed systems

8. Fault tolerance
   - A failure is any deviation of the observed behavior of a system from the specified behavior
   - Failures on a single machine often affect the **entire** system
   - Failures on distributed systems can be **partial** (only some nodes fail)
   - A **fault tolerant system** is able to meet its specifications even with partial failures
   - Important ability since the probability of failure grows with number of nodes in the system

Fault tolerance

- Classification of failure models

Type of failure	Description
Crash failure *Fail-stop crash* *Fail-silent crash*	Process halts and remains halted Other processes can reliably detect the failure Other processes may not be able to detect the failure
Omission failure *Receive omission* *Send omission*	Sent message never arrives at the other end Process fails to receive incoming messages Process fails to send outgoing messages
Timing failure	Process response lies outside a specified time interval
Response failure *Value failure* *State transition failure*	Process response is incorrect The value of the response is wrong The process deviates from the correct flow of control
Arbitrary (Byzantine) failure	Process may produce arbitrary messages at arbitrary times, commit omissions, stop, or take incorrect steps

Fault tolerance

- Use **failure detectors** to know about crash failures of processes
  - They are generally based on the use of timeouts
  - Setting timeouts properly is not easy
- Two implementation options:
  A. Heartbeats: I am alive!
  B. Pinging: Are you alive?: Yes!
- How reliable are the detectors?
  A. Unreliable
  B. Reliable

Fault tolerance

A. Unreliable: a process can be …

- Unsuspected: Recent evidence that the process has not failed
  - e.g. a message has recently received from it
- Suspected: Hint that the process may have failed
  - e.g. no messages from the process from a nominal maximum time

B. Reliable: a process can be Unsuspected or

- Failed: the detector has determined that the process has failed
- These are only possible in synchronous systems

Fault tolerance

- Hide the occurrence of failure from other processes using **redundancy**. Three types:
  1. Information redundancy
     - Add extra bits to allow for error detection/recovery
       - e.g. parity bits, cyclic redundancy checks (CRC)
  2. Physical redundancy (more details in Lesson 5)
     - **Replicate** hardware and/or software
       - e.g. process replication
  3. Time redundancy (more details in Lesson 2)
     - Retry an operation until a reply has been received
       - e.g. remote method invocations

Challenges of distributed systems

9. Scalability
   - Ability to support the growing in the number resources/users, the distance between nodes, or the number of administrative domains

A. Number of users and/or resources
   - **Size scalability**

B. Maximum distance between nodes
   - **Geographical scalability**

C. Number of administrative domains
   - **Administrative scalability**

Scaling problems

A. Size scalability

1. Centralized services and/or data
   - e.g. a single server providing a service becomes a bottleneck when the requests increase due to:
     - Limits on its computational capacity
     - Limits on its storage capacity and the I/O transfer rate between CPUs and disks
     - Limits on the network bandwidth with its users
2. Centralized algorithms
   - e.g. optimal routing needs complete information about the load of all machines and lines. This information would overload the network

Scaling problems

B. Geographical scalability
   1. Latency may easily prohibit synchronous client-server interactions in wide-area networks (WAN)
   2. Communication in WAN is less reliable and offers limited bandwidth
   3. WAN lack support for multipoint communication

C. Administrative scalability
   1. The various domains have to protect themselves against malicious attacks from each other
      - Components within a single domain can often be trusted by users within that domain

Scaling techniques

1. Hide communication latencies
   - Use <u>asynchronous communication</u>
   - Helpful to achieve geographical scalability

   ↓ Not every application can use this model
     - e.g. interactive applications
2. Partitioning and distribution
   - <u>Split</u> data and/or computations and <u>spread</u> them across multiple computers
     - e.g. move computations to clients (Java applets)
     - e.g. decentralized naming services (DNS)
     - e.g. decentralized information systems (WWW)

Scaling techniques

3. Replication/caching
   - **Replicate** components across the system
     - e.g. replicated web servers and databases
     - e.g. mirrored web sites
     - e.g. file and web caches (in browsers and proxies)
   - Caching: decision is taken by client of a resource
   - Having multiple copies leads to inconsistencies
     - Strong consistency requires global synchronization on each modification, which precludes large-scale solutions
     - Need for synchronization is reduced if we can tolerate some inconsistencies, but this is application dependent
   - We will cover this in Lesson 5

Scaling techniques

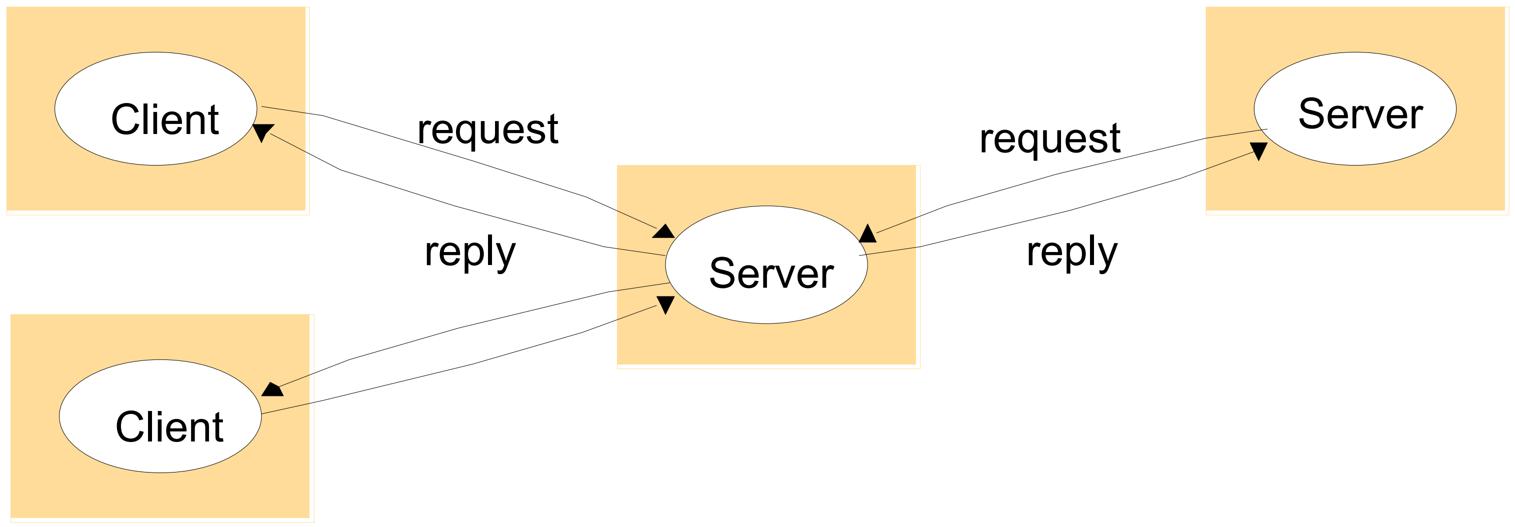
4. Use decentralized algorithms
   - No machine has complete information about the system state
   - Decisions based only on machine local information
   - Failure of one machine does not ruin the algorithm
   - No implicit assumption that a global clock exists
   - e.g. routing in P2P systems
     - We will see routing examples in Lesson 9

Contents

- Definition of a distributed system
- Challenges of distributed systems
- **Architectures for distributed systems**

System architectures

A. Client-server architectures
- Servers offer services and clients use them
- Based in the request-reply behavior
- Traditional (and successful) model until now
- Some scalability and robustness problems



Client-server architectures

- Clients
  - Software combines user interfaces and solutions for achieving distribution transparency, e.g.:
    - Client-side stubs for remote method invocations
    - Invoke several replicas, pass a single response to client
    - Mask communication failures by retrying connections
- Servers
  - <u>Sequential</u> server serves one request at a time
    - Can service multiple requests by employing events and asynchronous communication
  - <u>Concurrent</u> server spawns a process or thread for each request (can also use a pre-spawned pool)

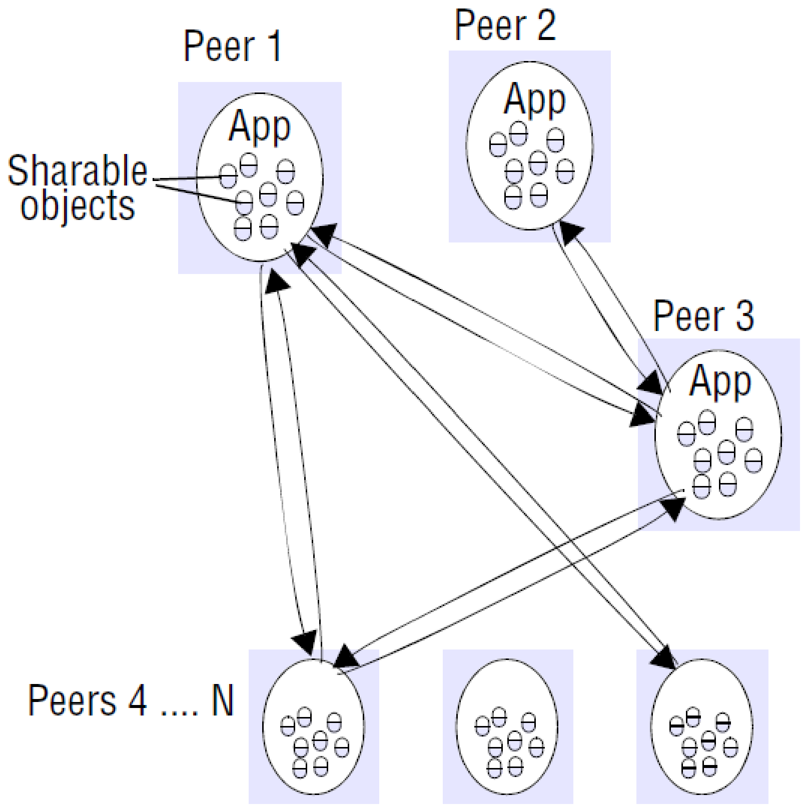
Client-server architectures

- Servers
  - Stateless server does not keep data of its clients, and can change its state without informing clients
    - e.g. web server (but cookies can store client’s info)
  - Stateful server keeps track of status of its clients
    - e.g. file server allowing client caching
      - Better client-perceived performance but complicates server recovery from a crash
  - Server can maintain a soft state for a limited time
- Servers can organize in tiered architectures
  - Servers may in turn be clients of other servers

System architectures

B. Peer-to-peer architectures

- Removes distinction between clients and servers
  - Nodes symmetric in function
- Overlay network
  - Data is routed over connections set up between the nodes
- No centralized control
- High availability & scalability
- Resilience to failures



Peer-to-peer architectures

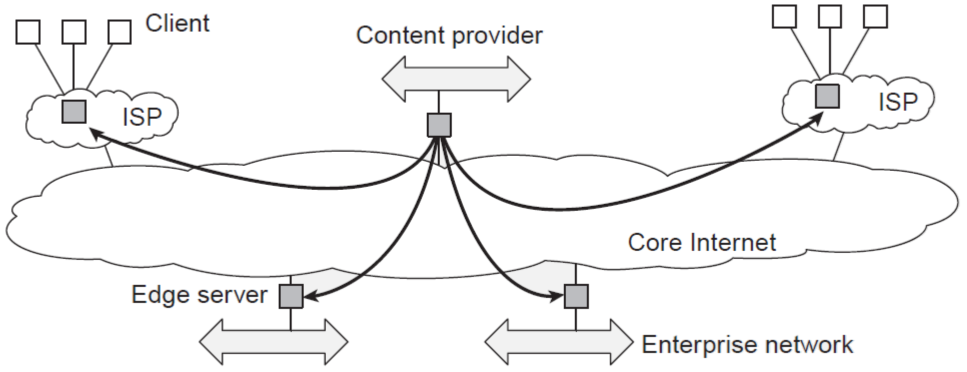
1. Structured Peer-to-Peer architectures
   - Topology constructed deterministically
     - Nodes are organized following a specific distributed data structure and they are responsible for data items based only on their ID
     - Typically Distributed Hash Tables (DHT) (e.g. Chord)
2. Unstructured Peer-to-Peer architectures
   - Topology based on randomized algorithms
     - Each node picks a random set of nodes and becomes their neighbor and data items are assumed to be randomly placed on nodes (e.g. Gnutella)

➢ We will cover this in Lesson 9

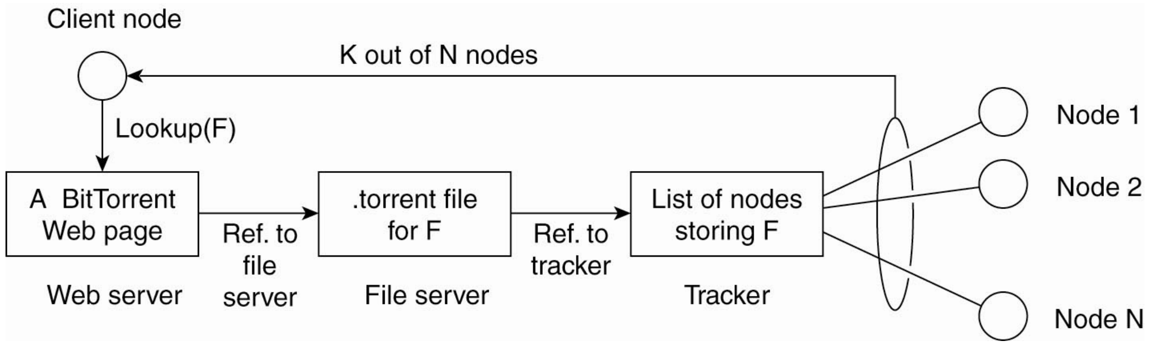
System architectures

C. Hybrid architectures

- Combination of client-server architectures with peer-to-peer solutions



e.g. edge-server systems (Content Distribution Networks)
we will cover this in Lesson 8



e.g. collaborative systems (BitTorrent)
we will cover this in Lesson 9

Summary

- Definition/motivation for distributed systems
- Challenges: heterogeneity, no global view, security, coordination, asynchrony, openness, transparency, fault tolerance, scalability
- System architectures: client-server, peer-to-peer, hybrid
- Further details:
  - [Tanenbaum]: chapters 1, 2.2, 3.3, and 3.4
  - [Coulouris]: chapters 1 and 2