Al Sham Private University
Faculty of Informatics Engineering

جامعة الشام الخاصة
كلية الهندسة المعلوماتية

نظم معلومات موزعة
Distributed Information Systems

Lecture 1: Characterization of Distributed Systems

اعداد: أ. غاندي هسام

Introduction

- The Internet is one, as are the many networks of which it is composed.
- Mobile phone networks, corporate networks, factory networks, campus networks, home networks, – all of these, both separately and in combination, share the essential characteristics that make them relevant subjects for study under the heading distributed systems.

a distributed system as one in which hardware or software components located at networked computers communicate and coordinate their actions only by passing messages.

Definition

- A distributed system is a collection of independent computers that appears to its users as a single coherent system.
- This means that one way or the other the autonomous components need to collaborate.
- How to establish this collaboration lies at the heart of developing distributed systems.
- Note that no assumptions are made concerning the type of computers.
- Computers that are connected by a network may be spatially separated by any distance.

Consequences

❑Concurrency

- In a network of computers, concurrent program execution is standard.
- Sharing resources such as web pages or files when necessary.
- The capacity of the system to handle shared resources can be increased by adding more resources (for example: computers) to the network.
- The coordination of concurrently executing programs that share resources is also an important

❑No global clock

- When programs need to cooperate they coordinate their actions by exchanging messages.

- coordination often depends on a shared idea of the time at which the programs' actions occur.
- But it turns out that there are limits to the accuracy with which the computers in a network can synchronize their clocks – there is no single global notion of the correct time.

❑Independent failures

- All computer systems can fail, and it is the responsibility of system designers to plan for the consequences of possible failures.
- In fact, the programs on them may not be able to detect whether the network has failed or has become unusually slow.

The prime motivation

- Constructing and using distributed systems stems from a desire to share resources.
- The term ‘resource’ extends from hardware components such as disks and printers to software-defined entities such as files, databases and data objects of all kinds.
- The main goal of a distributed system is to make it easy for the users (and applications) to access remote resources, and to share them in a controlled and efficient way.
- However, as connectivity and sharing increase, security is becoming increasingly important.

Examples of distributed systems

Selected application domains and associated networked applications

Finance and commerce	The growth of eCommerce as exemplified by companies such as Amazon and eBay, and underlying payments technologies such as PayPal; the associated emergence of online banking and trading and also complex information dissemination systems for financial markets.
The information society	The growth of the World Wide Web as a repository of information and knowledge; the development of web search engines such as Google and Yahoo to search this vast repository; the emergence of digital libraries and the large-scale digitization of legacy information sources such as books (for example, Google Books); the increasing significance of user-generated content through sites such as YouTube, Wikipedia and Flickr; the emergence of social networking through services such as Facebook and MySpace.
Education	The emergence of e-learning through for example web-based tools such as virtual learning environments; associated support for distance learning; support for collaborative or community-based learning.

Healthcare	The growth of health informatics as a discipline with its emphasis on online electronic patient records and related issues of privacy; the increasing role of telemedicine in supporting remote diagnosis or more advanced services such as remote surgery (including collaborative working between healthcare teams); the increasing application of networking and embedded systems technology in assisted living, for example for monitoring the elderly in their own homes..
Transport and logistics	The use of location technologies such as GPS in route finding systems and more general traffic management systems; the modern car itself as an example of a complex distributed system (also applies to other forms of transport such as aircraft); the development of web-based map services such as MapQuest, Google Maps and Google Earth.
Environmental management	The use of (networked) sensor technology to both monitor and manage the natural environment, for example to provide early warning of natural disasters such as earthquakes, floods or tsunamis and to coordinate emergency response; the collation and analysis of global environmental parameters to better understand complex natural phenomena such as climate change.

Examples of distributed systems: Web search

- The task of a web search engine is to index the entire contents of the World Wide Web, surrounding a wide range of information styles including web pages, multimedia sources and (scanned) books.
- Given that most search engines analyze the entire web content and then carry out sophisticated processing on this enormous database, this task itself represents a major challenge for distributed systems design.
- **Google**, the market leader in web search technology, has put significant effort into the design of a complicated distributed system infrastructure to support search (and indeed other Google applications and services such as Google Earth).

- Highlights of Google infrastructure include:
  - an underlying physical infrastructure consisting of very large numbers of networked computers located at data-centres all around the world;
  - a distributed file system designed to support very large files and heavily optimized for the style of usage required by search and other Google applications;
  - an associated structured distributed storage system that offers fast access to very large datasets;
  - a lock service that offers distributed system functions such as distributed locking and agreement;
  - a programming model that supports the management of very large parallel and distributed computations across the underlying physical infrastructure.

Trends in distributed systems

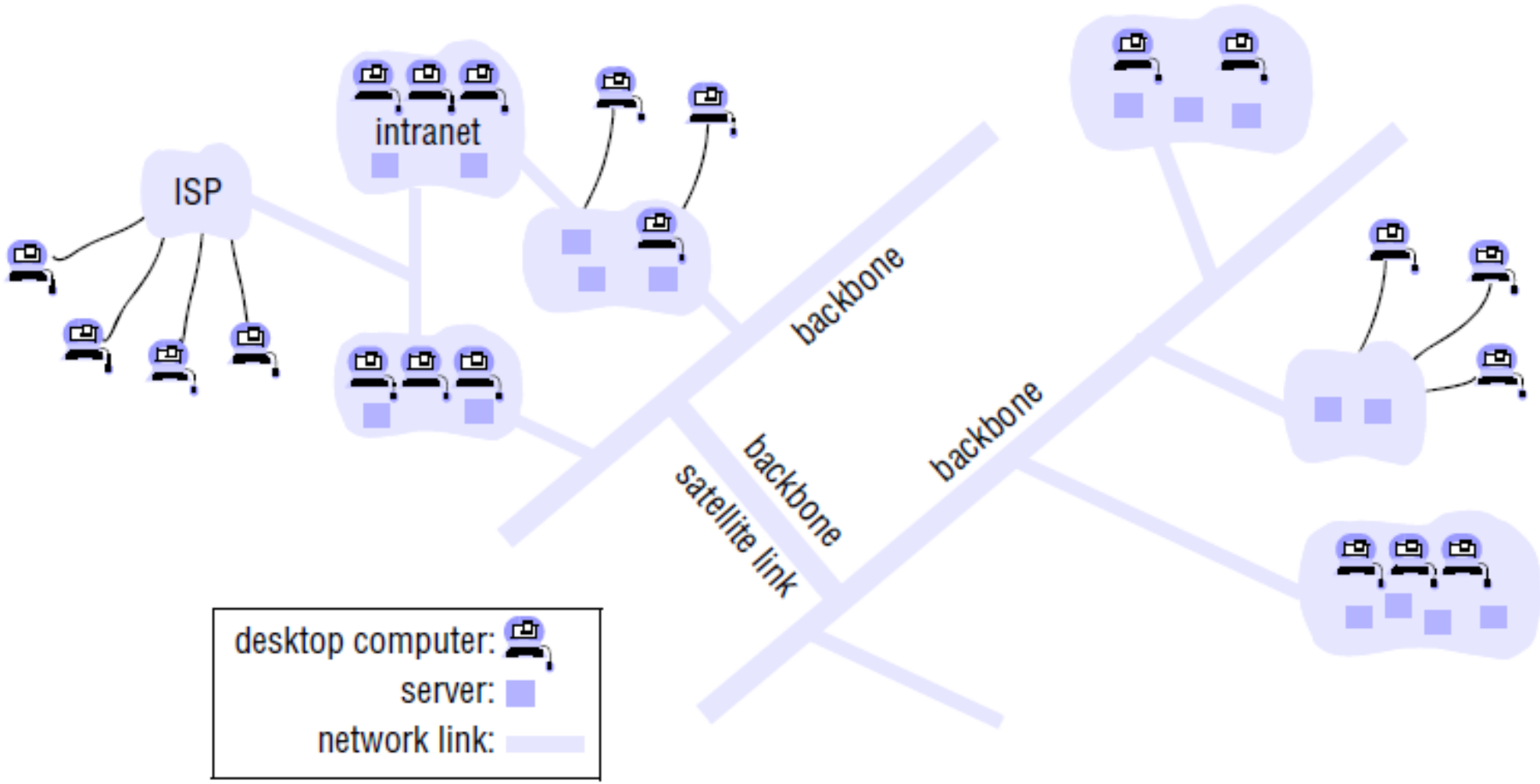
- Distributed systems are undergoing a period of significant change and this can be traced back to a number of powerful trends:
  - the appearance of spreading networking technology;
  - the appearance of ubiquitous computing coupled with the desire to support user mobility in distributed systems;
  - the increasing demand for multimedia services;
  - the view of distributed systems as a utility.

Spreading networking and the modern Internet

- The modern Internet is a vast interconnected collection of computer networks of many different types, with the range of types increasing all the time and now including, for example, a wide range of wireless communication technologies such as WiFi, WiMAX.
- Programs running on the computers connected to it interact by passing messages, employing a common means of communication.
- The design and construction of the Internet communication mechanisms (the Internet protocols) is a major technical achievement, enabling a program running anywhere to address messages to programs anywhere else

- The Internet is also a very large distributed system.
- It enables users, wherever they are, to make use of services such as the World Wide Web, email and file transfer.
- The set of services is open-ended – it can be extended by the addition of server computers and new types of service.
- Internet Service Providers (ISPs) are companies that provide broadband links and other types of connection to individual users and small organizations, enabling them to access services anywhere in the Internet as well as providing local services such as email and web hosting.

A typical portion of the Internet



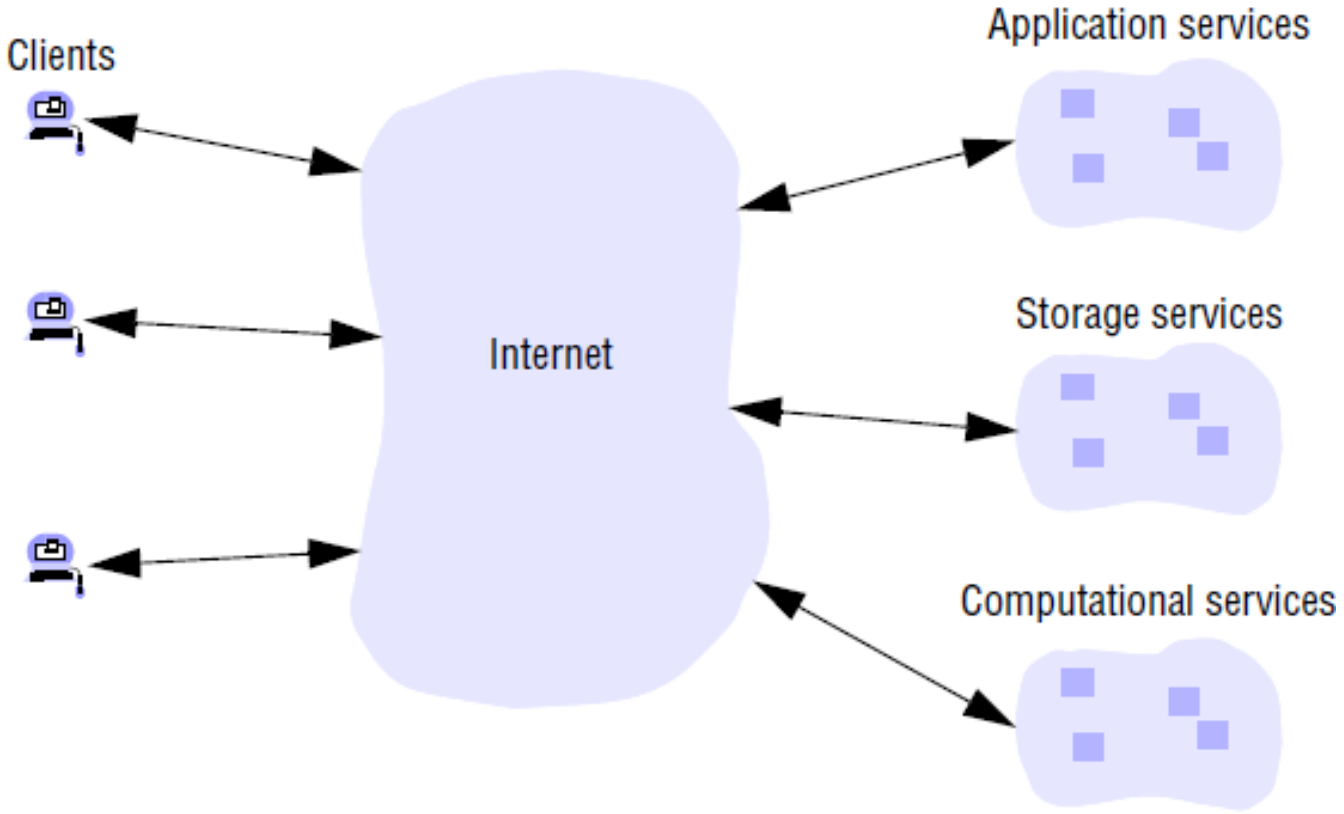
Distributed multimedia systems

- Multimedia support can usefully be defined as the ability to support a range of media types in an integrated manner.
- A distributed multimedia system should be able to perform the same functions for continuous media types such as audio and video; that is, it should be able to store and locate audio or video files, to transmit them across the network (possibly in real time as the streams emerge from a video camera).
- ***Webcasting*** is an application of distributed multimedia technology. Webcasting is the ability to broadcast continuous media, typically audio or video, over the Internet. It is now commonplace for major sporting or music events to be broadcast in this way, often attracting large numbers of viewers (for example, the Live8 concert in 2005 attracted around 170,000 simultaneous users at its peak).

- Distributed multimedia applications such as webcasting place considerable demands on the underlying distributed infrastructure in terms of:
  - providing support for an (extensible) range of encoding and encryption formats, such as the MPEG series of standards (including for example the popular MP3 standard otherwise known as MPEG-1, Audio Layer 3) and HDTV;
  - providing a range of mechanisms to ensure that the desired quality of service can be met;
  - providing associated resource management strategies, including appropriate scheduling policies to support the desired quality of service.

Cloud computing

- A cloud is defined as a set of Internet-based application, storage and computing services sufficient to support most users' needs, thus enabling them to largely or totally dispense with local data storage and application software.



- The term also promotes a view of everything as a service, from physical or virtual infrastructure through to software, often paid for on a per-usage basis rather than purchased.
- Note that cloud computing reduces requirements on users' devices, allowing very simple desktop or portable devices to access a potentially wide range of resources and services.
- Clouds are generally implemented on cluster computers to provide the necessary scale and performance required by such services.
- A ***cluster computer*** is a set of interconnected computers that cooperate closely to provide a single, integrated high-performance computing capability.

- Most clusters consist of commodity PCs running a standard (sometimes cut-down) version of an operating system such as Linux, interconnected by a local area network.
- Companies such as HP, Sun and IBM offer blade solutions.
- ***Blade servers*** are minimal computational elements containing for example processing and (main memory) storage capabilities.
  - A blade system consists of a potentially large number of blade servers contained within a blade enclosure.
  - Other elements such as power, cooling, persistent storage (disks), networking and displays, are provided either by the enclosure or through virtualized solutions.
  - Through this solution, individual blade servers can be much smaller and also cheaper

Blade servers

Video: Why Use Blade Server Systems?

End of Lecture 1