Parallel & Distributed System

Introduction to PDS

Md. Biplob Hosen

Lecturer, IIT-JU

Email: biplob.hosen@juniv.edu

Contents

Reference Books

- Distributed Systems: Principles and Paradigms, 3rd Edition by Andrew S. Tanenbaum & Maarten van Steen, Publisher: Pearson Prentice Hall.

- Distributed Systems: Concepts and Design, 5th Edition by George Coulouris, Jean Dollimore & Tim Kindberg, Publisher: Addison-Wesley.

Introduction to PDS

- **Distributed System** is a model where distributed applications are running on multiple computers linked by a communications network.
- Sometimes it is also called loosely coupled systems because in which each processor has its own local memory and processing units. LOCUS and MICROS are some examples of distributed operating systems.
- **Parallel Systems** are designed to speed up the execution of programs by dividing the programs into multiple fragments and processing these fragments at the same time.
- Flynn has classified computer systems into four types based on parallelism in the instructions and in the data streams.
  - Single Instruction stream, single data stream
  - Single Instruction stream, multiple data stream
  - Multiple Instruction stream, single data stream
  - Multiple Instruction stream, multiple data stream

Continue…

- A distributed system and a parallel system are two different types of computer systems, and the main difference between them is how they manage the processing and communication of tasks across multiple processors.
- A **distributed system** is a computer system that consists of multiple interconnected computers or nodes, that work together to perform a task or a set of tasks. The processing is distributed across multiple nodes, and each node is responsible for performing a part of the task.
- In a distributed system, the nodes communicate with each other using a network, and the system is designed to handle data and tasks that are geographically distributed.
- Examples of distributed systems include the internet, cloud computing, and peer-to-peer networks.

Continue…

- On the other hand, a **parallel system** is a computer system that consists of multiple processors that work together to perform a task. In a parallel system, the processing is divided into multiple tasks, and each processor performs a separate task simultaneously.
- The processors communicate with each other using shared memory or message passing, and the system is designed to handle data and tasks that require high computational power.
- Examples of parallel systems include supercomputers and clusters.
- In summary, the main difference between a distributed system and a parallel system is how they manage the processing and communication of tasks across multiple processors.
- In a distributed system, the processing is distributed across multiple nodes connected by a network, while in a parallel system, the processing is divided among multiple processors that work together on a single task.

Continue…

Advantages of Distributed Systems:

- Scalability: Distributed systems can be easily scaled by adding more computers to the network.
- Fault Tolerance: Distributed systems can recover from failures by redistributing work to other computers in the network.
- Geographical Distribution: Distributed systems can be geographically distributed, allowing for better performance and resilience.

Disadvantages of Distributed Systems:

- Complexity: Distributed systems are more complex to design and maintain compared to single computer systems.
- Communication Overhead: Communication between computers in a distributed system adds overhead and can impact performance.
- Security: Distributed systems are more vulnerable to security threats, as the communication between computers can be intercepted and compromised.

Continue…

Advantages of Parallel Systems:

- High Performance: Parallel systems can execute computationally intensive tasks more quickly compared to single processor systems.
- Cost Effective: Parallel systems can be more cost-effective compared to distributed systems, as they do not require additional hardware for communication.

Disadvantages of Parallel Systems:

- Limited Scalability: Parallel systems have limited scalability as the number of processors or cores in a single computer is finite.
- Complexity: Parallel systems are more complex to program and debug compared to single processor systems.
- Synchronization Overhead: Synchronization between processors in a parallel system can add overhead and impact performance.

Overview of DS

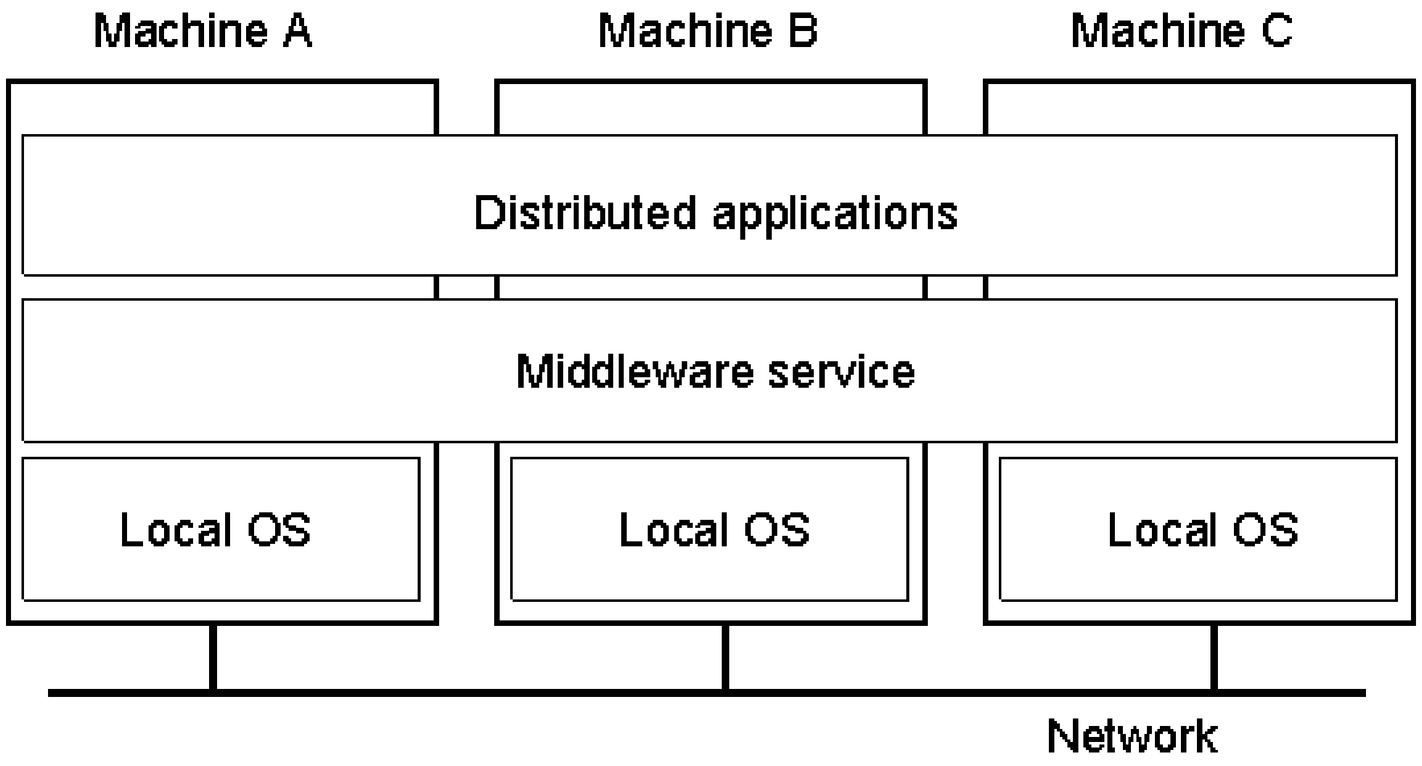
- A distributed system is a collection of independent computers that appears to its users as a single coherent system.
- This definition has two aspects:
  - The first one deals with hardware: the machines are autonomous.
  - The second one deals with software: the users think they are dealing with a single system.
- One important characteristic of a distributed system is that the differences between the various computers and the way in which they communicate are hidden from the users.
- Another important characteristic is that users and applications can interact with a distributed system in a consistent and uniform way, regardless of where and when interaction takes place.

Continue…

- Distributed systems are often organized by means of a layer of software that is logically placed between a higher level layer consisting of users and applications and a layer underneath consisting of operating system.
- For example, consider a network of workstations in a university.
- In addition to each user’s PC there might be a pool of processors in the machine room that are not assigned to specific users but are allocated dynamically as needed.
- Such a system might have a single file system, with all files accessible from all machines using the same path name.
- If the system as a whole looks and acts like a classical single-processor timesharing system, it qualifies as a distributed system

Continue…

- A distributed system organized as middleware.
- Note that the middleware layer extends over multiple machines.



Continue…

- Networks of computers are everywhere!
- Examples of networks of computers are:
  - Mobile phone networks
  - Corporate networks
  - Factory networks
  - Campus networks
  - Home networks
  - In-car networks
  - On board networks in aero planes and trains

Continue…

Properties of Distributed Systems

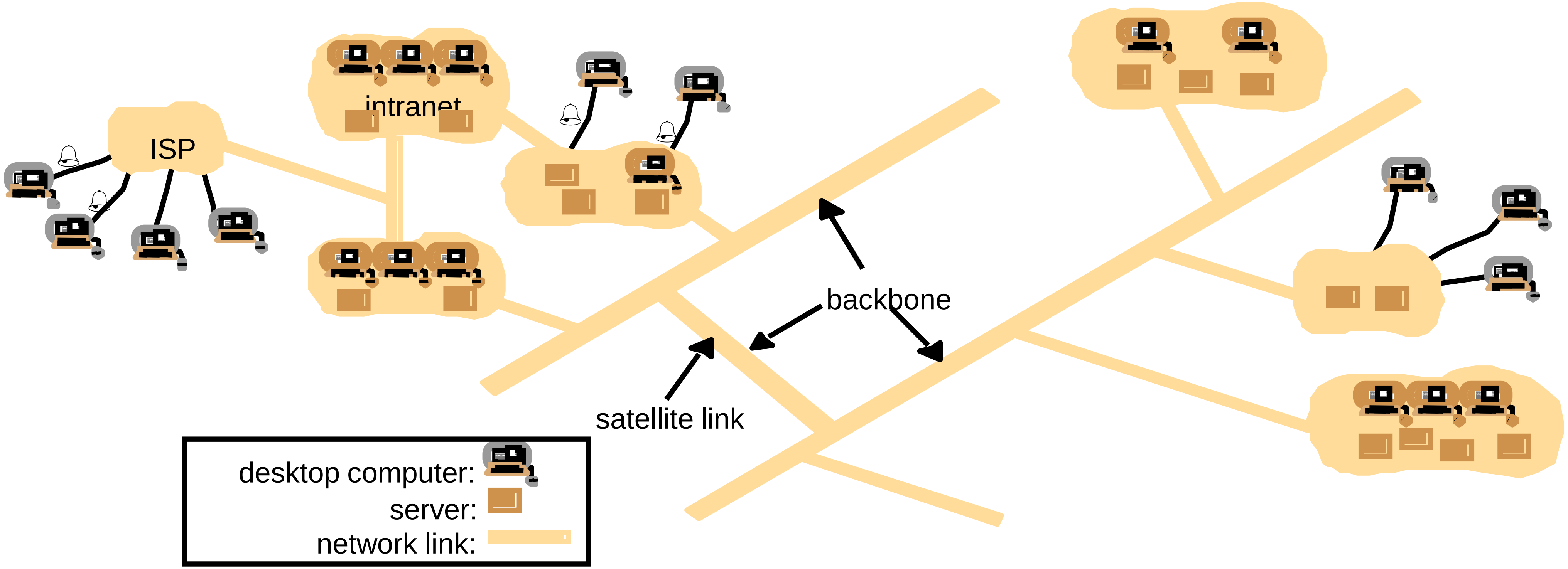
- Concurrency
  - Tasks carry out independently
- No global clock
  - Tasks coordinate their actions by exchanging messages
- Independent Failures
  - Faults in the network result in the isolation of the computers that are connected to it.
  - Each component of the system can fail independently, leaving the others still running.

Examples of Distributed Systems

- Internet
- Intranets
- Mobile networks
- Web search
- Massively multiplayer online games (MMOGs)
- Financial trading

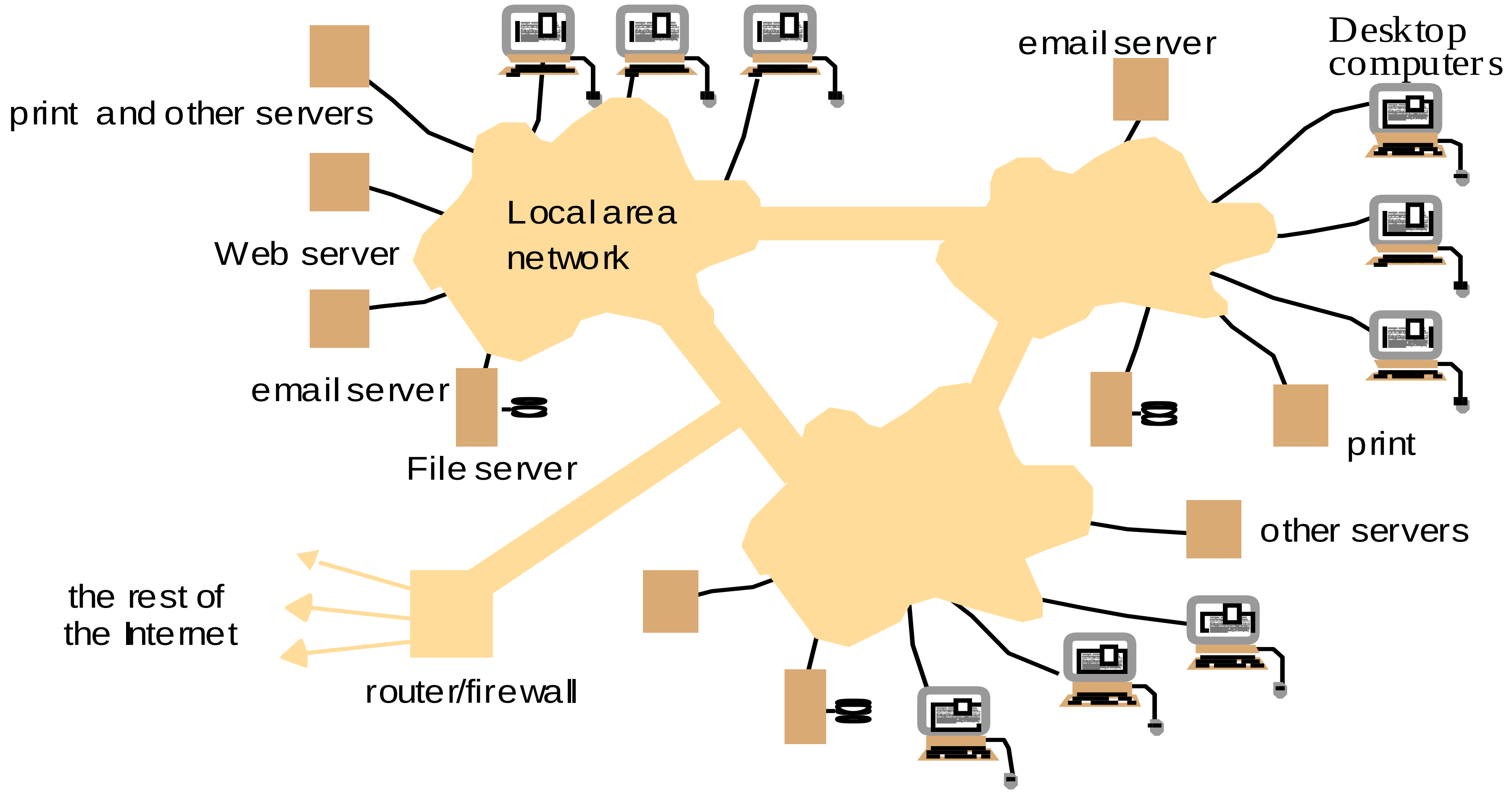
The Internet

- The Internet is a vast interconnected collection of computer networks of many different types.
- Multimedia services are available in the Internet enabling users to access audio and video data including music, radio, TV channels, phone, and video conferencing.



The Intranet

- An intranet is a portion of the Internet that is separately administered and has a boundary that can be configured to enforce local security policies.



Mobile Networks

- Technological advances in device miniaturization and wireless networking have led increasingly to the integration of small and portable computing devices into distributed systems.
- These devices include:
  - Laptop computers
  - Handheld devices
  - Personal digital assistants(PDAs)
  - Mobile phones
  - Video cameras
  - Digital cameras
- General Examples of Devices with Mobile Computing in Distributed Systems:
  - Wearable devices: Smart watches with functionality similar to a PDA
  - Devices embedded in appliances: Washing machines, Hi-fi systems, Cars, Refrigerators

Mobile Computing

- Mobile computing (also known as nomadic computing) is the performance of computing tasks while the user is on move, or visiting places other than their usual environment.
- In mobile computing, users who are away from their “home” intranet are still provided with access to resources via the devices they carry with them.
- They can continue to access the Internet; and there is increasing provision for users to utilize resources such as printers that are conveniently nearby as they move around. This is known as location aware or context-aware computing.
- For example, turning on AC in room by sitting in the car while coming back from the office through mobile phone.

Mobile Computing

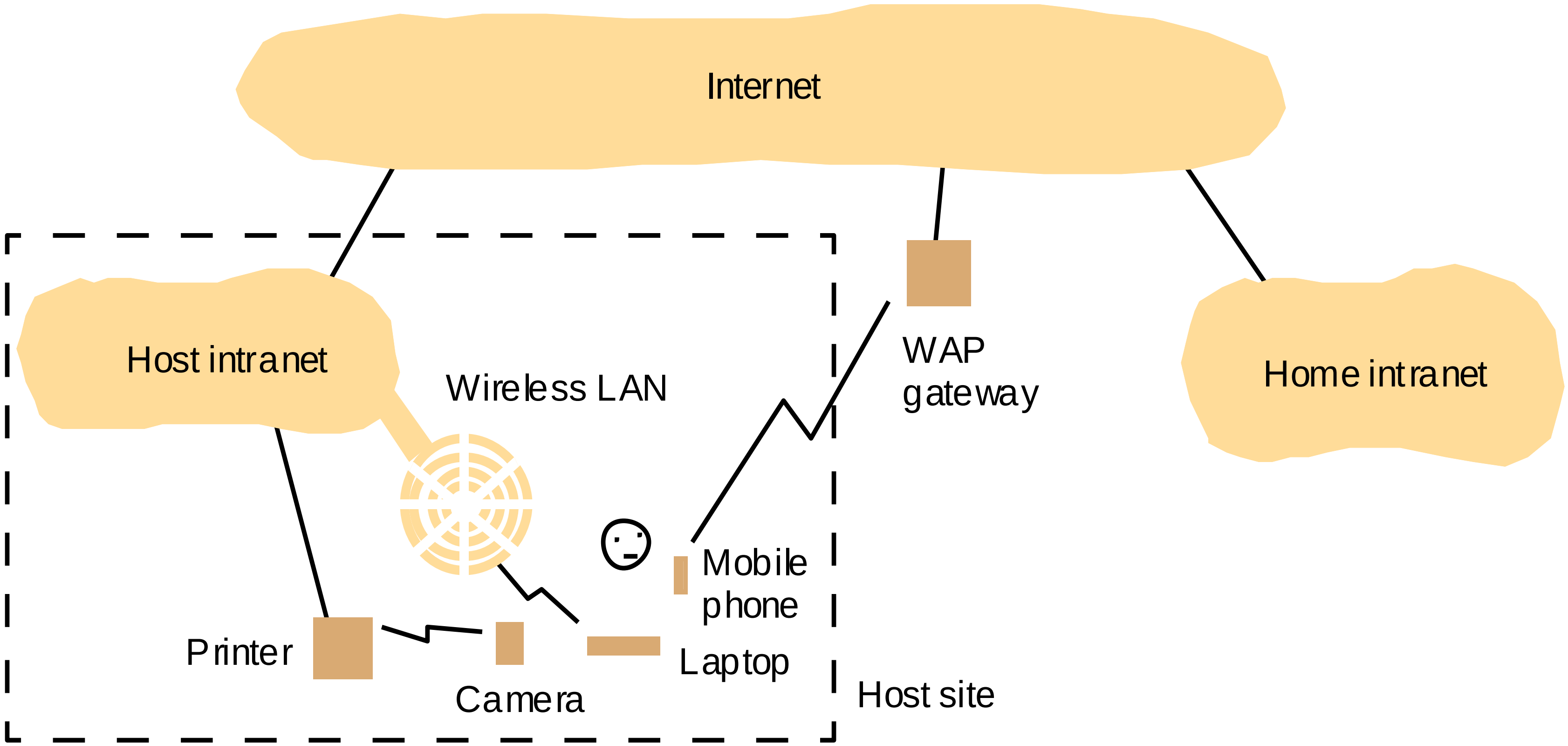


Figure- Portable and handheld devices in a distributed system.

Web Search

- The task of a web search engine is to index the entire contents of the World Wide Web, encompassing a wide range of information styles including web pages, multimedia sources and (scanned) books.
- This is a very complex task, as current estimates state that the Web consists of over 63 billion pages and one trillion unique web addresses.
- 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 sophisticated distributed system infrastructure to support search.
- This represents one of the largest and most complex distributed systems installations in the history of computing and hence demands close examination.

Web Search…

Highlights of this infrastructure include:

1. An underlying physical infrastructure consisting of very large numbers of networked computers located at data centers all around the world;
2. 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 (especially reading from files at high and sustained rates);
3. An associated structured distributed storage system that offers fast access to very large datasets;
4. A lock service that offers distributed system functions such as distributed locking and agreement;
5. A programming model that supports the management of very large parallel and distributed computations across the underlying physical infrastructure.

Massively Multiplayer Online Games (MMOGs)

- The engineering of MMOGs represents a major challenge for distributed systems technologies, particularly because of the need for fast response times to preserve the user experience of the game.
- Other challenges include the real-time propagation of events to the many players and maintaining a consistent view of the shared world.
- This therefore provides an excellent example of the challenges facing modern distributed systems designers. A number of solutions have been proposed for the design of massively multiplayer online games:

1**) EVE Online** utilizes a client-server architecture where a single copy of the state of the world is maintained on a centralized server and accessed by client programs running on players' consoles or other devices. To support large numbers of clients, the server is a complex entity in its own right consisting of a cluster architecture featuring hundreds of computer nodes. The centralized architecture helps significantly in terms of the management of the virtual world and the single copy also eases consistency concerns. The goal is then to ensure fast response through optimizing network protocols and ensuring a rapid response to incoming events. To support this, the load is partitioned by allocating individual 'star systems' to particular computers within the cluster, with highly loaded star systems having their own dedicated computer and others sharing a computer. Incoming events are directed to the right computers within the cluster by keeping track of movement of players between star systems.

MMOGs…

- Other MMOGs adopt more distributed architectures where the universe is partitioned across a number of servers that may also be geographically distributed. Users are then dynamically allocated a particular server based on current usage patterns and also the network delays to the server (based on geographical proximity for example). This style of architecture, which is adopted by EverQuest, is naturally extensible by adding new servers.
- Most commercial systems adopt one of the two models presented above, but researchers are also now looking at more radical architectures that are not based on client-server principles but rather adopt completely decentralized approaches based on peer-to-peer technology where every participant contributes resources (storage and processing) to accommodate the game.

Financial Trading

- The financial industry has long been at the cutting edge of distributed systems technology with its need, in particular, for real-time access to a wide range of information sources (for example, current share prices and trends, economic and political developments). The industry employs automated monitoring and trading applications.
- Note that the emphasis in such systems is on the communication and processing of items of interest, known as events in distributed systems, with the need also to deliver events reliably and in a timely manner to potentially very large numbers of clients who have a stated interest in such information items. Examples of such events include a drop in a share price, the release of the latest unemployment figures, and so on. This requires a very different style of underlying architecture from the styles mentioned above (for example client-server), and such systems typically employ what are known as distributed event-based systems.

Financial Trading…

- Figure1.2 illustrates a typical financial trading system. This shows a series of event feeds coming into a given financial institution. Such event feeds share the following characteristics:
- Firstly, the sources are typically in a variety of formats, such as Reuters market data events and FIX events (events following the specific format of the Financial Information eXchange protocol), and indeed from different event technologies, thus illustrating the problem of heterogeneity as encountered in most distributed systems. The figure shows the use of adapters which translate heterogeneous formats into a common internal format.
- Secondly, the trading system must deal with a variety of event streams, all arriving at rapid rates, and often requiring real-time processing to detect patterns that indicate trading opportunities. This used to be a manual process but competitive pressures have led to increasing automation in terms of what is known as Complex Event Processing (CEP), which offers a way of composing event occurrences together into logical, temporal or spatial patterns. This approach is primarily used to develop customized algorithmic trading strategies covering both buying and selling of stocks and shares, in particular looking for patterns that indicate a trading opportunity and then automatically responding by placing and managing orders.

Financial Trading…

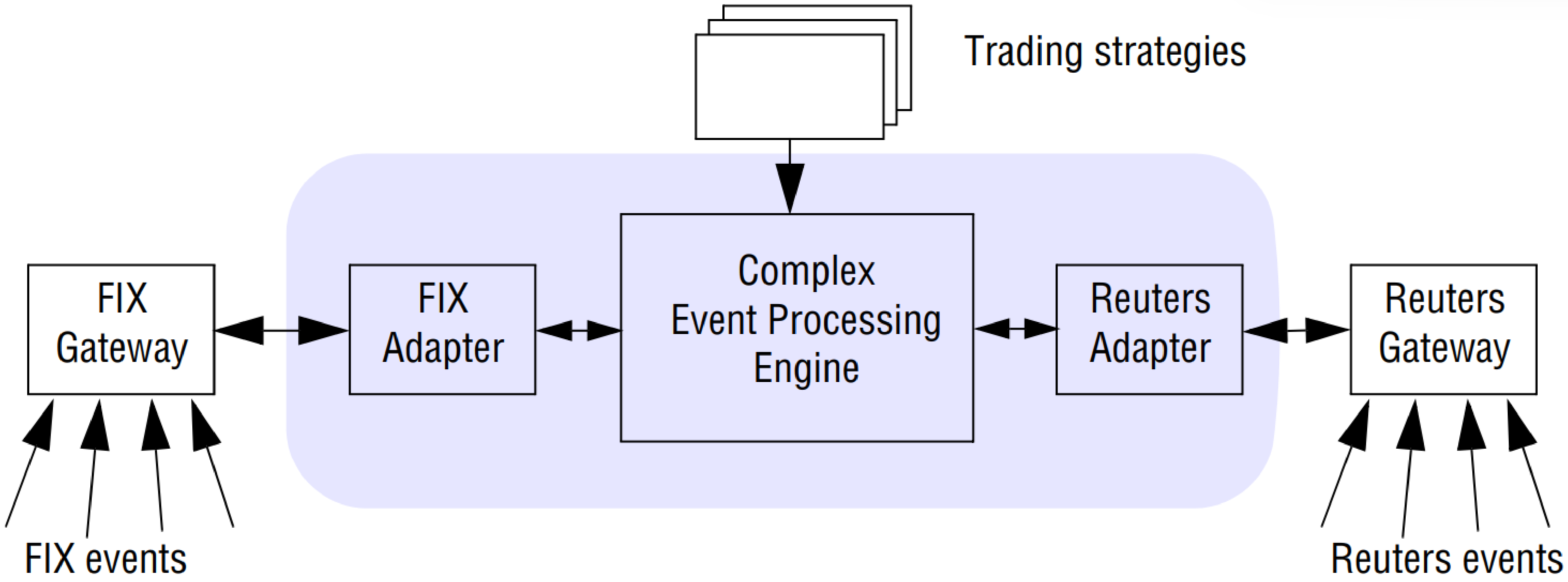


Figure 1.2 An example financial trading system

Financial Trading…

- As an example, consider the following script:
- This script is based on the functionality provided by Apama [www.progress.com], a commercial product in the financial world originally developed out of research carried out at the University of Cambridge.

```
WHEN
    MSFT price moves outside 2% of MSFT Moving Average
FOLLOWED-BY (
    MyBasket moves up by 0.5%
    AND
        HPQ’s price moves up by 5%
        OR
        MSFT’s price moves down by 2%
    )
)
ALL WITHIN
    any 2 minute time period
THEN
    BUY MSFT
    SELL HPQ
```

Financial Trading…

- The script detects a complex temporal sequence based on the share prices of Microsoft, HP and a basket of other share prices, resulting in decisions to buy or sell particular shares.
- This style of technology is increasingly being used in other areas of financial systems including the monitoring of trading activity to manage risk (in particular, tracking exposure), to ensure compliance with regulations and to monitor for patterns of activity that might indicate fraudulent transactions.
- In such systems, events are typically intercepted and passed through what is equivalent to a compliance and risk firewall before being processed

Thank You ☺