Processes & Threads

Dr. Padraig Corcoran

Process

- A **process** is a program in execution.
- Multiple processes may concurrently share same CPU by storing and later reloading a **process context.**
- Independent processes cannot affect the correctness of each other's behaviour (multiple **process contexts** cannot exist in memory at same time).
- This **concurrency transparency** has a performance cost (storing/reloading process context, switching processes, etc.).

Thread

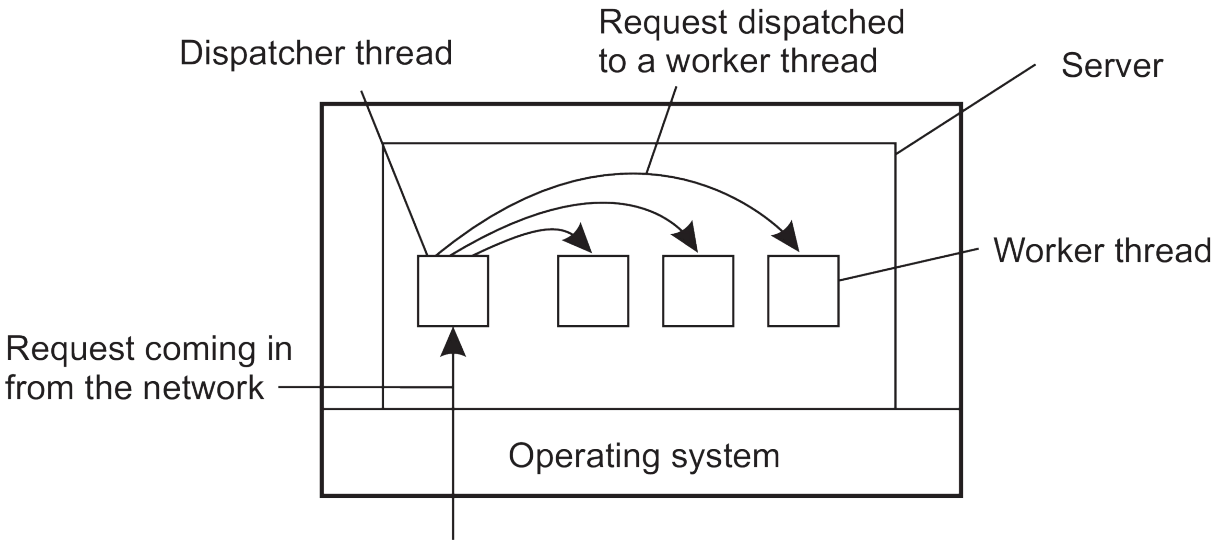
- A thread is also a program in execution.
- No attempt is made to achieve concurrency transparency if this results in performance degradation (multiple **thread contexts** can exist in memory at same time).
- The **developer** must ensure independent threads cannot affect the correctness of each other's behavior (e.g. protecting data against inappropriate access).
- For an example of why this is the case see corresponding lab session.

Threads in distributed systems - multithreaded clients

- Threads provide means of allowing blocking calls without blocking the entire process to which the thread belongs.
- To achieve distributed transparency it is necessary to conceal message propagation times.
- Communication latencies may be hidden by initiating communication and proceeding with something else.

Threads in distributed systems - multithreaded servers

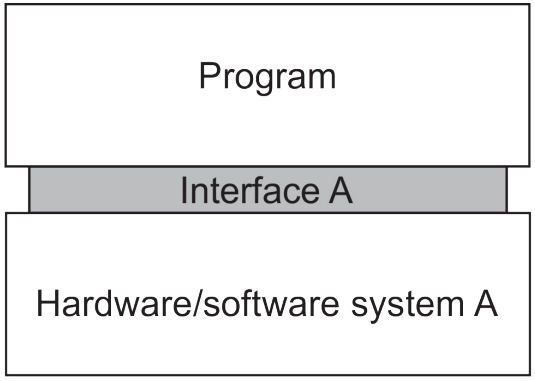
- Threads used extensively by servers in the client-server architecture.
- Allows sequential processes that make blocking system calls and while still achieving parallelism.



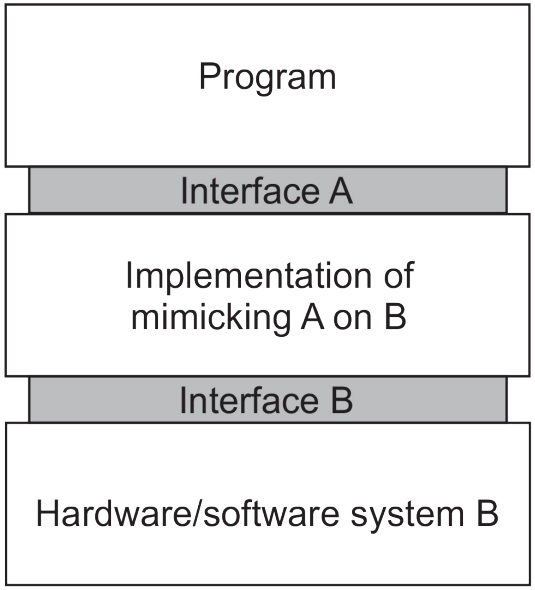
A multithreaded server organised in a dispatcher/worker model.

Virtualization

- Every (distributed) computer system offers a programming interface to higher-level software.
- Virtualization extends or replaces an existing interface so as to mimic the behavior of another system.



General organisation between a program, interface, and system.

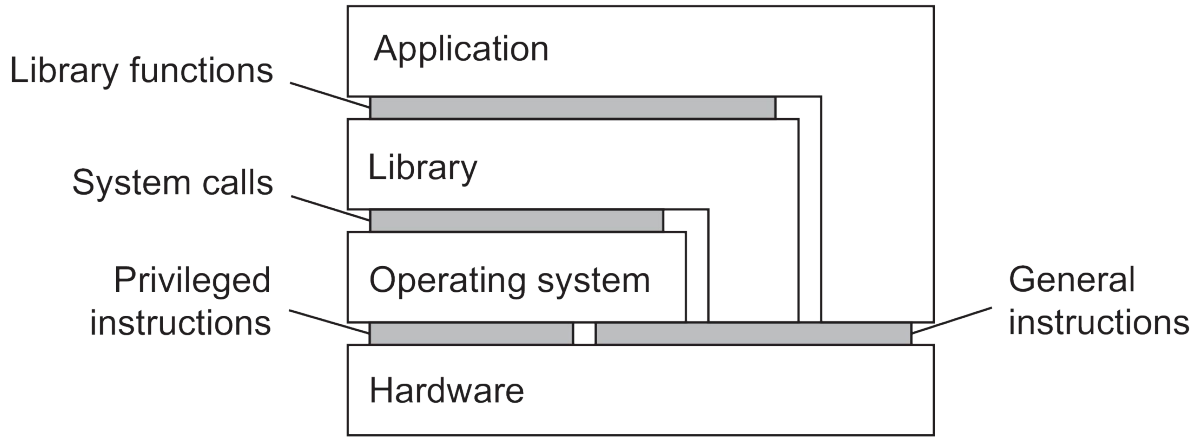


General organisation of virtualizing system A on top of B.

Types of virtualization

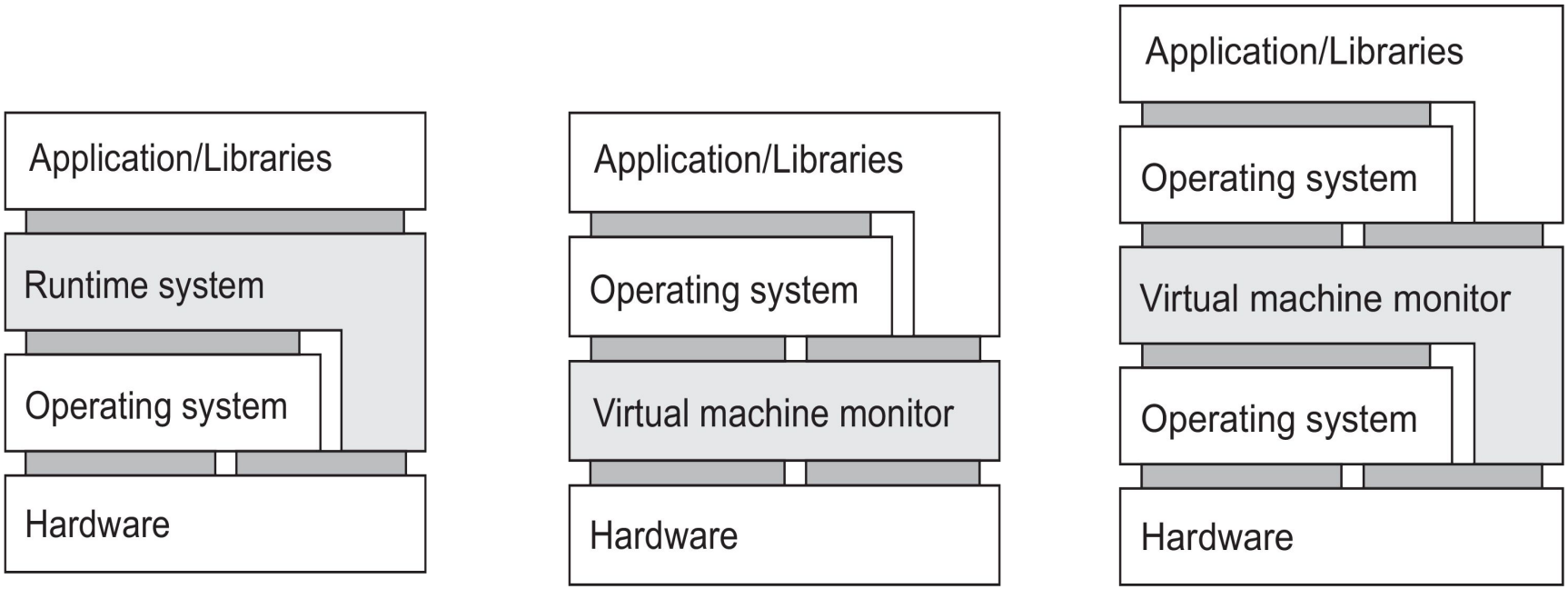
Computer systems offer interfaces at three levels:

- Interface between the hardware and software, referred to as the **instruction set architecture** (ISA).
- Interface consisting of **system calls** as offered by an operating system.
- Interface consisting of library calls, generally forming what is known as an **application programming interface** (API).



Various interfaces offered by computer systems.

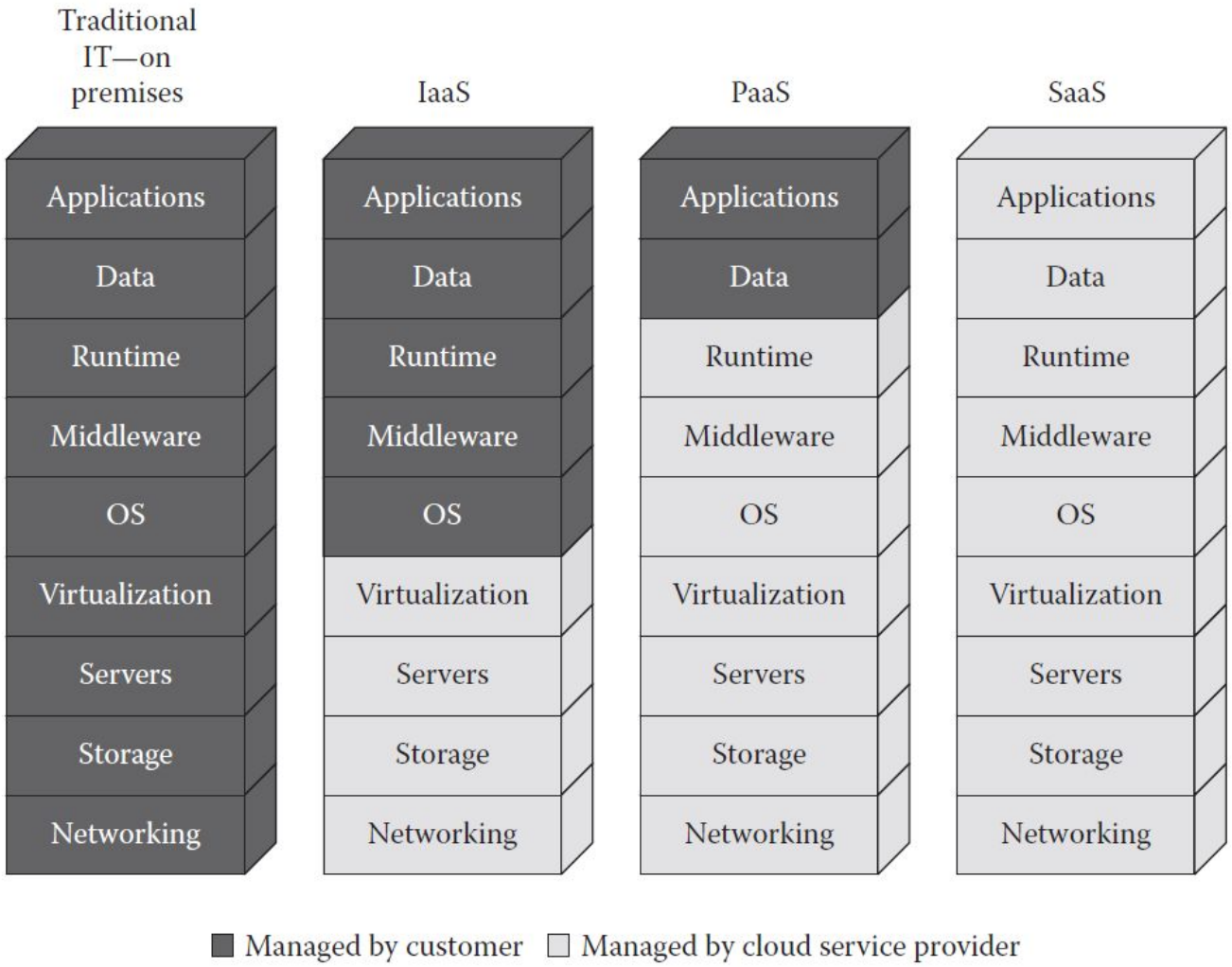
- Aim of virtualization is to mimic the behavior of these interfaces.



(a) A process virtual machine. (b) A native virtual machine monitor. (c) A hosted virtual machine monitor.

Application of virtual machines to distributed systems

- For distributed systems, an important application of virtualization lies in cloud computing.
- There are three main types of cloud computing services:
  - Infrastructure-as-a-Service (IaaS) - basic infrastructure.
  - Platform-as-a-Service (PaaS) - system-level services.
  - Software-as-a-Service (SaaS) - actual applications.
- Virtualization plays a key role in each of the above.
- For IaaS, instead of renting a physical machine, a client will rent a virtual machine (Amazon EC2).

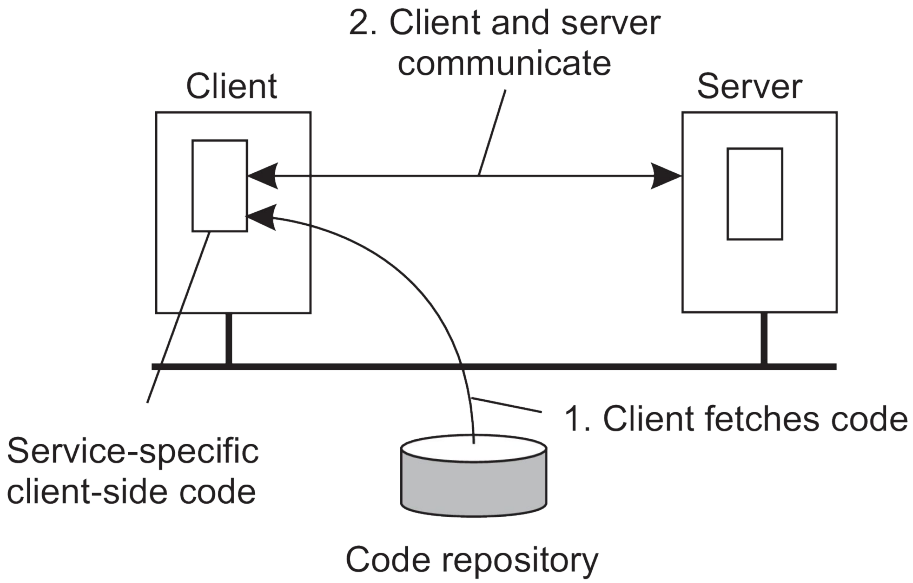


Separation of responsibilities in cloud operation (taken from Vacca 2016)

Code migration

- So far, we considered distributed systems in which communication between objects is limited to passing data.
- There are situations in which passing programs between objects is useful.
- Code migration can improve performance by:
  - Achieving better load-distribution.
  - Minimising communication/data transfer (code acting on the data is moved to the location of the data).
- Migrating code in heterogeneous systems is challenging.

- Besides improving performance, code migration allows greater flexibility.
- Code migration allows distributed systems to be dynamically configured.



The principle of dynamically configuring a client to communicate with a server.

Code migration challenges

- Migrating code in heterogeneous systems. A solution - use virtualization.
- Trusting downloaded code. A solution - execute code in a restricted environment called a sandbox.