CS 370: OPERATING SYSTEMS
[INTER PROCESS COMMUNICATIONS]

Shrideep Pallickara
Computer Science
Colorado State University

COMPUTER SCIENCE DEPARTMENT

COLORADO STATE UNIVERSITY

1

Frequently asked questions from the previous class survey

- Can existing memory be marked "shared"
- How do processes agree/decide to open shared memory before anything else happens?
- Will the kernel ever return an error for requesting too much space in shared memory?
  - Yes: if the requested size exceeds available virtual memory and/or backing storage limits
- Does shared memory come with risks other than giving access to the wrong process?
  - Concurrency, stability, and debugging challenges
- Which is better?
  - Shared memory or message passing
- Which is better?
  - Direct or indirect communications
- How is it possible to have an "unbounded" buffer … in shared memory or otherwise?
- Wouldn't the producer and consumer operate at the same rate?

COLORADO STATE UNIVERSITY Professor: SHRIDEEP PALLICKARA COMPUTER SCIENCE DEPARTMENT INTER-PROCESS COMMUNICATIONS L6.2

2

Topics covered in this lecture

- Inter Process Communications
  - Messaging
  - Pipes
- Monolithic Kernels and Micro kernels

3

Just a castaway, an island lost at sea, oh
Another lonely day with no one here but me, oh
More loneliness than any man could bear
Rescue me before I fall into despair, oh
I'll send an S.O.S to the world
I'll send an S.O.S to the world

Message in a Bottle, The Police

MESSAGE PASSING IN WINDOWS XP

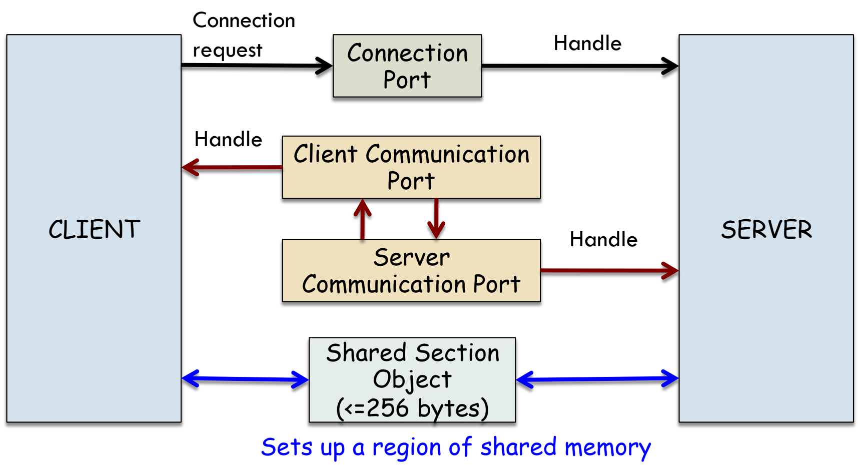
4

Message passing in Windows XP

- Called the local procedure call (LPC) facility
- Communications provided by **port** objects
  - Gives applications a way to set up communication channels
- Uses two types of message passing
  - Small messages (max 256 bytes)
  - Large messages

5

Connection ports are named objects visible to all processes [LPC in XP]



6

Windows XP message passing
Small messages

- Use port's internal message queue as intermediate storage
- Copy messages from one process to another

7

Windows XP message passing:
Large messages [1/2]

- Send message through **section object**
  - Sets up shared memory
- Section object info sent as a <u>small message</u>
  - Contains pointer + size information about section object

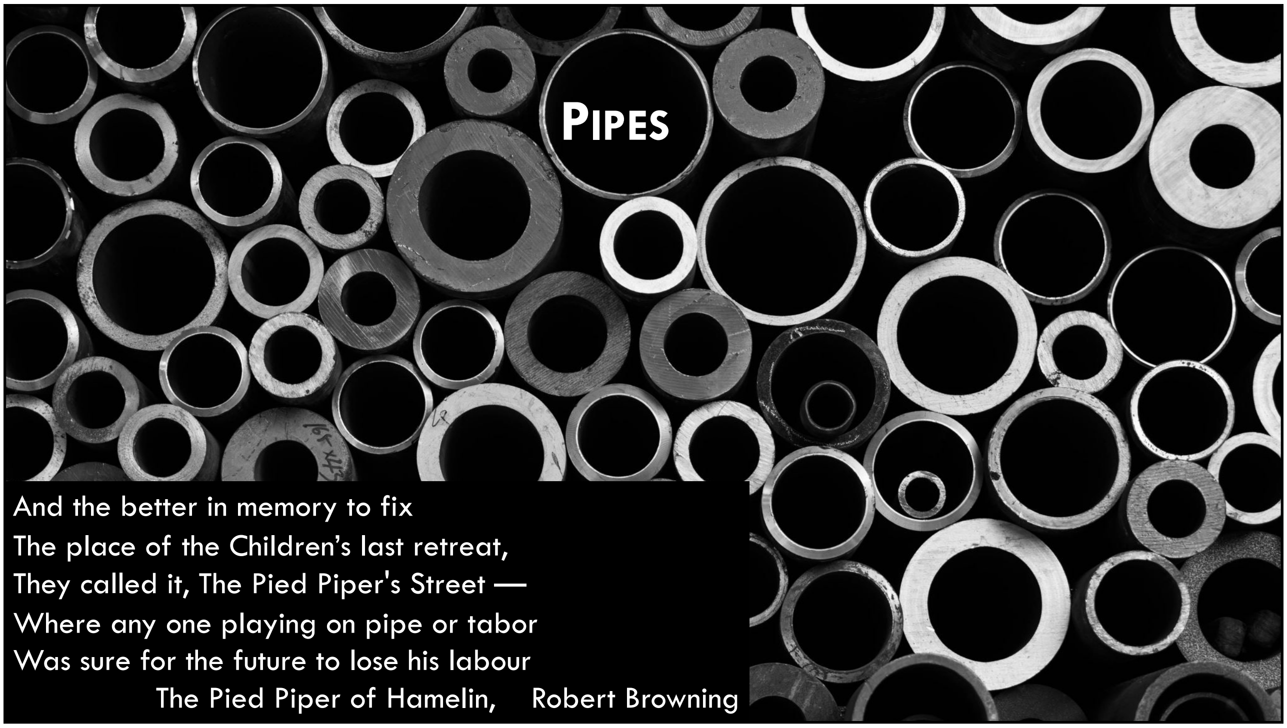
8

Windows XP message passing: Large messages [2/2]

- 2 ends of communications set up section objects if the request or reply is large
- Complicated, but **avoids data copying**
- **Callbacks** used if the endpoints are busy
  - Allows delayed responses
  - Allows asynchronous message handling
- So why did choose to cover this?
  - Shared memory and message passing are not ideological camps; real systems combine them!

But …

- Not every conversation between processes requires quite this much ceremony
- Unix offers a pipe!



11

Pipes

- Pipes serve as a **conduit** for communications between processes
- One of the first IPC implementation mechanisms

12

Issues to consider when implementing a pipe

- Unidirectional or bidirectional
- If it is bidirectional
  - **Half duplex**: Data can travel one way at a time
  - **Full duplex**: Data traversal in both directions *simultaneously*
- Must a relationship exist between the endpoints?
  - e.g., parent-child
- Range of communications
  - Intra-machine or Over the network

Pipes in practice

- Set up pipe between commands

```
ls | more
```

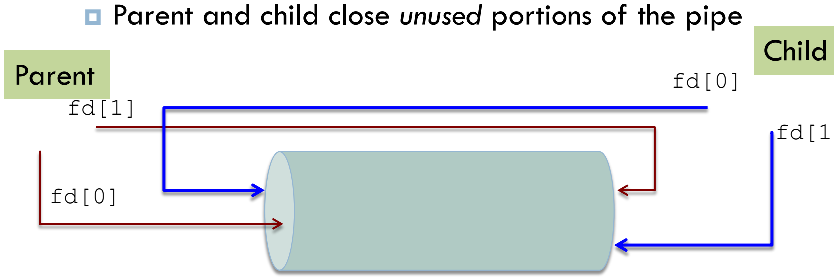
Output of **`ls`** delivered as input to **`more`**

Ordinary pipes

- Producer writes to one end of the pipe
- Consumer reads from the other end
- In UNIX: `pipe(int fd[])` to create pipe
  - `fd[0]` is the read-end
  - `fd[1]` is the write-end
  - Treats a pipe as a **special type of file**
    - Access with `read()` and `write()` system calls

A child inherits open files from its parent

- Since a pipe is a special type of file, the pipe is also inherited
  - Parent and child close *unused* portions of the pipe



`fd[0]` is the read-end
`fd[1]` is the write-end

Pipes: Example

```
if (pipe(fd) == -1) {
    /* creation failed */
}
pid = fork();

if (pid > 0) {
    close(fd[READ_END]);
   write(fd[WRITE_END], write_msg,…);
}

if (pid == 0) {
    close(fd[WRITE_END]);
   read(fd[READ_END], …);
}
```

Closing the unused ends makes each process's role explicit and, more importantly, allows EOF semantics to work correctly.

Windows Ordinary Pipes: These are unidirectional

- Anonymous Pipes
- Child **does not** automatically inherit pipe
  - Programmer *specifies* attributes a child will inherit
  - Initialize `SECURITY_ATTRIBUTES` to allow handles to be inherited
  - Redirect child's standard I/O handles to read/write handle of pipe
  - Pipes are half duplex

Some other things about ordinary pipes on UNIX and Windows

- Requires **parent-child** relationship
  - MUST be on same machine
- **Exists** only when processes communicate with one another
  - Upon termination, pipe ceases to exist
- But what if the two processes are unrelated?
  - There is no common parent from which they can inherit the pipe
  - Enter …

Named Pipes

- Can be bidirectional
- **NO** parent-child relationship needed
- Once named pipe is established
  - *Several* processes can use it for communications
- Continues to exist after communicating processes have finished

Named Pipes on UNIX/Windows

- Referred to as **FIFO** on UNIX systems
  - Created with `mkfifo()`
  - Manipulated with `open()`, `read()`, `write()` etc.
- FIFO: Normally used as a **one-way** (**half-duplex**) byte stream
  - If data must go both ways: use 2 FIFOs
  - Sockets used for inter-machine communications
- Windows: Full duplex communications

21

22

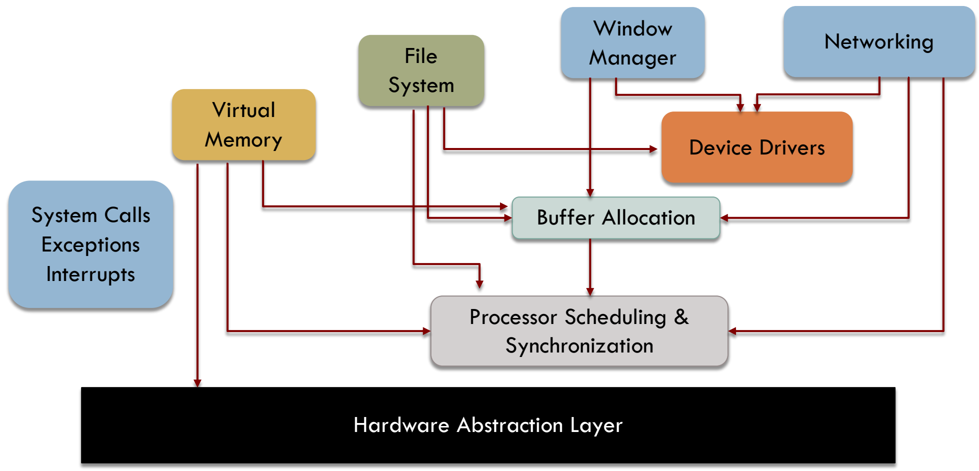
There are many dependencies (and interactions) among the modules inside the OS [1/2]

- Several modules depend on **synchronization primitives** for coordinating access to shared data structures with the kernel
- The virtual memory system depends on low-level hardware support for address translation
  - Support that is specific to a particular processor architecture

There are many dependencies (and interactions) among the modules inside the OS [2/2]

- Both the **file system** and the **virtual memory** system share a common pool of blocks of physical memory
  - They also both depend on the disk device driver
- The **file system** can depend on the **network protocol stack** if the disk is physically located on a different machine

OS Kernel Structure: A very crowded neighborhood



25

This has led operating system designers to wrestle with a fundamental tradeoff

- By **centralizing functionality** in the kernel, performance is improved
  - It makes it easier to arrange tight integration between kernel modules
- However, the resulting systems are less flexible, less easy to change, and less adaptive to user or application needs

? Prevalence of monolithic kernels?

26

Where is the monolithic kernel used?

- **Almost all** widely used commercial operating systems, such as Windows, MacOS, and Linux
- Monolithic is a bit of a misnomer, there are often large segments of what users consider the OS that runs outside the kernel
  - Either as utilities like the shell, or in system libraries, such as libraries to manage the user interface

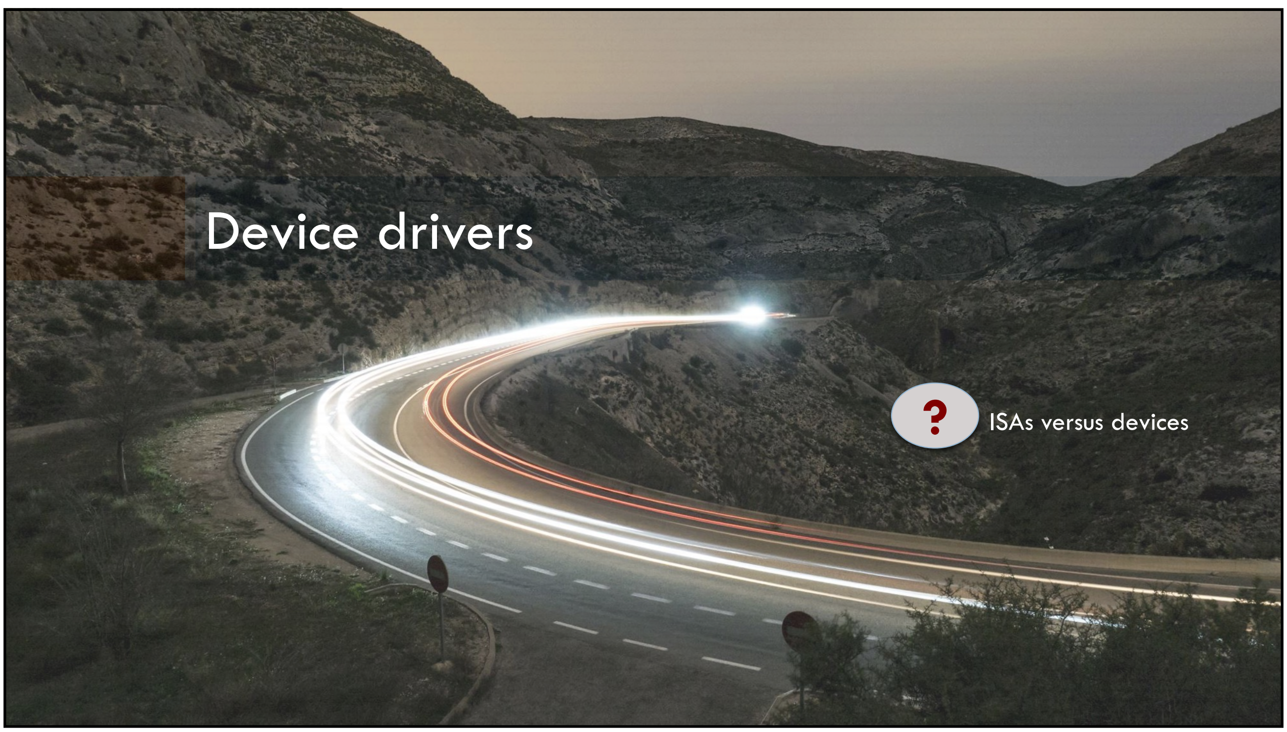
A key goal of operating systems is to be portable across a wide variety of hardware platforms [1/2]

- This requires careful design of the hardware abstraction layer
- **Portable interface** to machine configuration and processor-specific operations within the kernel

A key goal of operating systems is to be portable across a wide variety of hardware platforms [2/2]

- Within the same processor family, such as an Intel x86
  - Manufacturers use different machine-specific code to configure and manage interrupts and hardware timers
  - x86-based Mac vs x86-based Windows
- Across processor families, will need processor-specific code for process and thread context switches
  - Between an ARM and an x86 or
  - Between a 32-bit and a 64-bit x86

29



30

Operating Systems try to accommodate a wide variety of physical I/O devices

- There are **only a handful** of different instruction set architectures (ISA) in wide use today:
  - But there are a huge number of different types of physical I/O devices, manufactured by a large number of companies
  - There is diversity in the hardware interfaces to devices as well as in the hardware chip sets for managing the devices

What is a dynamically installed device driver?

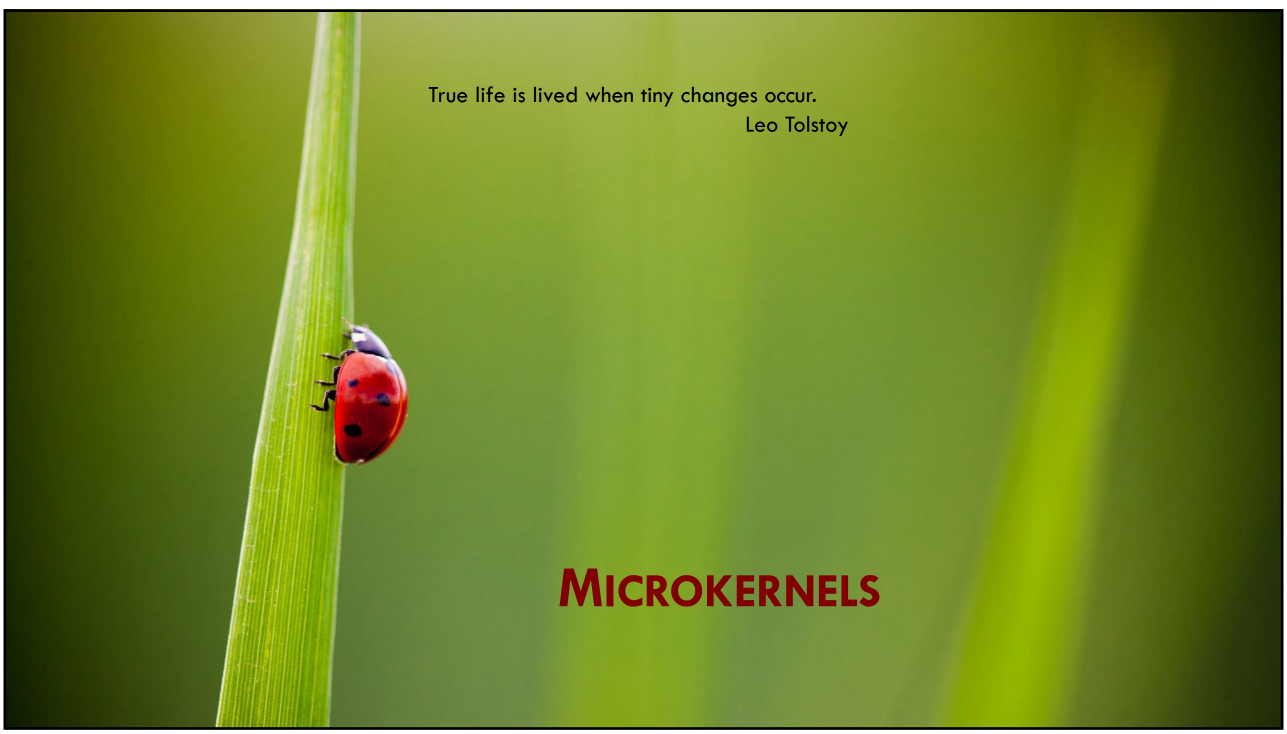
- Software to manage a specific device, interface, or chipset, that is added to the kernel **after it starts running**
  - To handle the devices that are present on a particular machine
- The device manufacturer typically provides the **driver code**, using a standard interface supported by the kernel
  - The kernel calls into the driver whenever it needs to read or write data to the device

A great deal of code; what could possibly go wrong?

- A recent survey found that approximately 60% of the code in the Linux kernel was in device-specific software
- However, 90% of all system crashes were due to bugs in device drivers, rather than in the operating system itself

COLORADO STATE UNIVERSITY Professor: SHRIDEEP PALLICKARA COMPUTER SCIENCE DEPARTMENT INTER-PROCESS COMMUNICATIONS L6.33

33



34

The Microkernel Approach [1/2]

- An alternative to the monolithic kernel approach is to run as much of the OS as possible in one or more user-level servers/services
- Structure OS by *removing non-essential components* from the kernel
  - Implement other things as system/user programs
  - **Mach**
- Provide minimal process and memory management
- Main function: Provide communication facility between client and services
  - **Message passing**

The Microkernel Approach [2/2]

- In monolithic kernels all the layers went in the kernel
  - But this is not really necessary
- In fact, it may be best to *put as little as possible* in the kernel
  - Bugs in the kernel can bring down the system instantly
- Contrast this with setting up user processes to have less power
  - A bug may not be fatal

Getting there …

- Achieve high reliability by **splitting OS into small, well-defined modules**
  - The microkernel runs in the kernel mode
  - The rest as relatively powerless ordinary user processes
- Running each device driver as a separate process?
  - Bugs cannot crash the entire system

Communications in the microkernel

- Client and service/server *never* interact directly
- Indirect communications by exchanging messages with the microkernel
- Advantages
  - Easier to port to different hardware
  - More security and reliability
    - Most services run as user, rather than kernel

Some examples

- The window manager on most operating systems works this way:
  - Individual applications draw items on their portion of the screen by sending requests to the window manager
  - The **window manager adjudicates** which application window is in front or in back for each pixel on the screen, and then renders the result
  - If the system has a hardware graphics accelerator present, the window manager can use it to render items more quickly
- Some systems have moved other parts of the operating system into user-level servers: the network stack, the file system, device drivers, and so forth

Another idea related to microkernels

- Put **mechanisms** for doing something in the *kernel*
  - But <u>not the policy</u>
- Example: Scheduling
  - Policy of assigning priorities to processes can be done in the user-mode
  - The mechanism to look for the highest priority process and to schedule it is in the kernel

The difference between monolithic/microkernel design is often transparent to the programmer

- The location of the service can be hidden in a user-level library
  - Calls go to the library, which casts the requests either as
    - System calls or
    - Reads and writes to the server through a pipe
- The location of the service can also be hidden inside the kernel
  - The application calls the kernel as if the kernel implements the service
    - But instead, the kernel reformats the request into a pipe that the service can read

Increased system function overhead can **degrade** microkernel performance

- A microkernel design offers considerable benefit to the operating system developer
  - Easier to modularize and debug user-level services than kernel code
- Aside from a potential reliability improvement, however, microkernels offer little in the way of visible benefit to end users and can **slow down overall performance considerably**
  - By inserting extra steps between the application and the services it needs

Isolation is lovely; it is also not free

- The very boundary that gives us fault isolation also gives us overhead
  - Crossing boundaries is not free!

A hybrid model

- Some operating system services are run at user-level and some are in the kernel
  - Depending on the specific tradeoff between code complexity and performance

Increased system function overhead can degrade microkernel performance

- Windows NT: First release, layered microkernel
  - Lower performance than Windows 95
- Windows NT 4.0 solution
  - Move layers from user space to kernel space
- By the time Windows XP came around
  - More monolithic than microkernel

COMMUNICATIONS IN CLIENT-SERVER SYSTEMS

Remote Procedure Calls

- Abstracts procedure call mechanisms for use with network endpoints
- Based on the **request/reply** model
- Message is addressed to the **RPC daemon** listening to a port for incoming traffic
  - Contains identifiers of function to execute
  - Parameters to pass to the function
  - TCP/UDP port number: 530
    - Other example ports: DNS(53), HTTP(80), NTP(123), etc.

Remote Procedure Calls

- Application makes CALL into a procedure
  - May be local or remote **and**
  - BLOCKS until call returns
- Origins:
  - **RFC 707** (1976)
  - First use by Xerox 1981 (Courier)
  - 1984 paper by Birell and Nelson

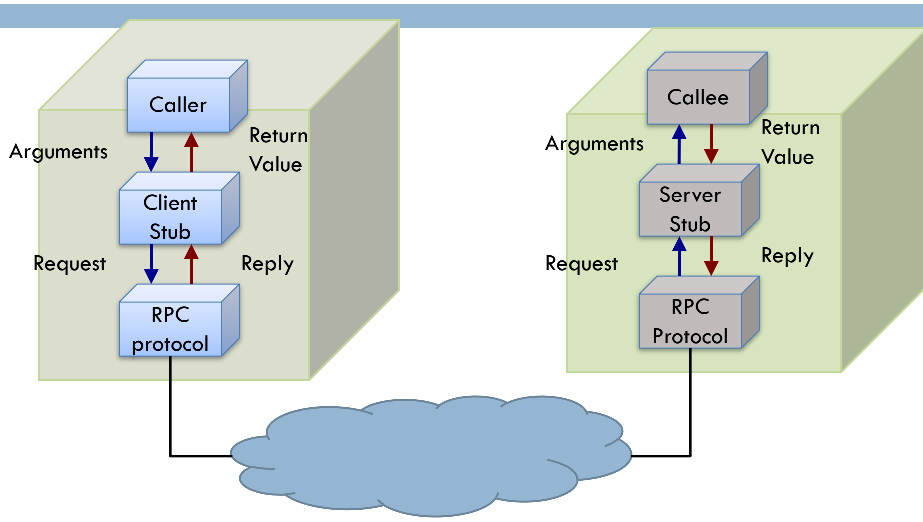
RPCs are slightly more complicated than local procedure calls

- Network between the *Calling* process and *Called* process can
  - **Limit** message sizes,
  - **Reorder** them or
  - **Lose** them
- Computers hosting processes may differ
  - Architectures and data representation formats

Resolving big-endian/little endian issues

- Big endian: Store **MSB** first
- Little endian: Store **LSB** first
- Machine independent data representation
  - XDR: e**X**ternal **D**ata **R**epresentation
  - Client side parameter marshalling
    - Convert machine-dependent data to XDR
  - Server side
    - Convert XDR data to machine dependent representation

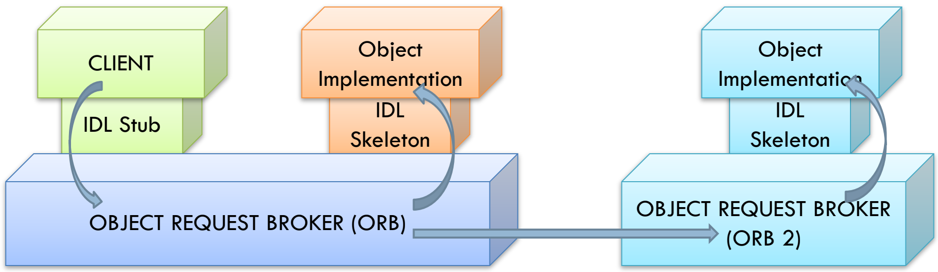
RPC mechanism



Distributed Objects

- RPC based on distributed objects with an **inheritance** mechanism
- *Create*, *invoke* or *destroy* remote objects, and interact as if they are local objects
- Data sent over network:
  - **References**: class, object and method
  - Method *arguments*
- CORBA early1990s, RMI mid-late 90s

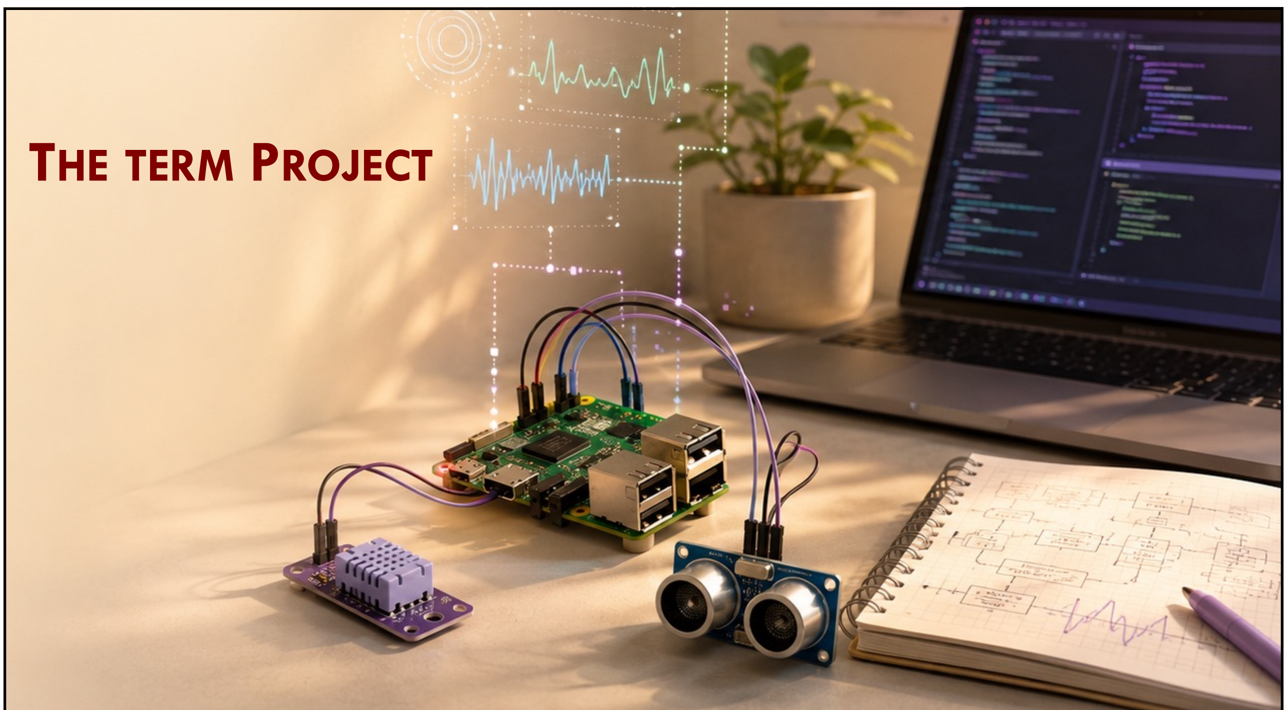
Distributed Objects in CORBA defined using the Interface Definition Language



GIOP/IIOP
General Inter-ORB Protocol/Internet Inter-Orb Protocol

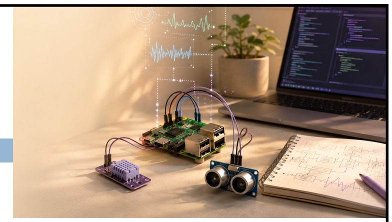
53

THE TERM PROJECT



54

Term project: Build a device someone actually needs



- Team size is 2
  - One Raspberry Pi, at least two *cooperating* sensors, and an honest user problem
- Claude Code is required in the workflow, but forbidden inside the shipped product
  - The agent helps you build the machine
  - It is not the machine … i.e., requests should not be funneled to Claude
- The question following you all semester: *Who wants this, and why does a phone app not already solve it?*

The contents of this slide-set are based on the following references

- *Avi Silberschatz, Peter Galvin, Greg Gagne. Operating Systems Concepts, 9th edition. John Wiley & Sons, Inc. ISBN-13: 978-1118063330.* [Chapter 3]
- *Thomas Anderson and Michael Dahlin. Operating Systems: Principles and Practice, 2nd Edition. Recursive Books. ISBN: 0985673524/978-0985673529. [Chapters 2-3]*
- *Andrew S Tanenbaum. Modern Operating Systems. 4th Edition, 2014. Prentice Hall. ISBN: 013359162X/ 978-0133591620.* [Chapter 2]
- *Kay Robbins & Steve Robbins. Unix Systems Programming, 2nd edition, Prentice Hall ISBN-13: 978-0-13-042411-2.* [Chapter 3, 4]