

What is an Operating System?

OS is a collection of software programs. It is the first program to run on the system.

What are the main functions?

Process management

Memory management

I/O management

File System Management

What is a kernel?

Core of OS which manages core features of OS.

Handles communication between hardware and software.

Services are used through system calls.

Layer of shell wraps around the kernel.

First program loaded on start up.

What are the main functions of the kernel?

Process management

Device management

Memory management

Interrupt Handling.

What are the different types of kernel?

Monolithic

Micro kernel.

What is micro kernel?

User services and Kernel services are implemented in a different address space.

What is a command interpreter?

Interrupt the command input from user through keyboard. It raises an interrupt.

What is a process?

A process is a program that is running instance of a task or job.

A program under execution.

Functions of a process?

Creation and deletion of system processes.

CPU scheduling

Process communication

Synchronization

What is an interrupt?

Interrupt a signal from a device causing context switch.

What is a daemon?

Process that runs in the background.

Daemon processes are indicated with a d appended → httpd

What is a pipe?

A communication mechanism used between two processes.

What is semaphore?

Hardware or Software tag variable which is used as lock to a common resource.

In multi tasking the actions are synchronized by using semaphores.

What kind of operations are possible?

Wait

Signal

What is context switching?

An instance when the CPU switches execution of one process to another.

Critical Section?

Part of code which can be executed only by 1 process at a time.

Mutex?

Lock which protects access to shared data resources.

Before entering a critical section the mutex is locked.

What is Synchronization?

Controlling access to a shared resource which should be available for 2 or more threads or process.

Different types of synchronization?

Mutex

Semaphore

Condition variable

Read write locks

Critical region

What are condition variables?

Access to a shared resource only after a condition is satisfied.

What are read write locks?

Controlling access to a shared resource.

It makes easy when a data has very less write operations and has high read operations.

What is a deadlock?

A process waiting for resources used by other processes of the same group and it never happens.

What are page frames?

Same sized memory divided in the physical memory by the virtual memory.

What is trashing?

Under allocation of pages for a process.

What is a thread?

It is a independent flow of control by a process.

Path of execution within a process.

What is multithreading?

The idea to achieve parallelism by dividing processes into multiple threads.

Process vs Thread

Process	Thread
Run in separate memory space	Threads share same memory space
Process don't share resources with other processes	Threads share resources with other threads, code section, data section and OS resources.

Disadvantages of threads?

Cannot be re used

They correct the address space of process.

Synchronization to concurrent read write access to memory.

What are the different types of threads?

User Level	Kernel Level
User threads are implemented by users.	Kernel threads are implemented by OS.
Java Thread and POSIX	Windows Solaris

What is a compiler?

Source code into object code.

What is a driver?

Software interface of a hardware.

What does bit size of cpu denote?

No of bytes of info a cpu can access from ram.

RAM?

Temporary storage stores program data and volatile.

Static Ram: Bit of data stored by a combination of 6 transistors.

DRAM: A pair of transistor and capacitor. Data processing on edges of the pulse.

ROM?

Static data info is stored permanently access speed is slow.

BIOS info or Bootstrap

How is Inter-process communication handled?

Approaches

1. Shared Memory
2. Message Passing
3. Pipe

Consider 2 process 1 and 2 want to communicate.

Shared Memory: A new memory shared region is defined.

When process 1 wants to communicate to process 2. It writes to shared memory.

Message Passing: Process 1 wants to communicate to 2. It creates a message and passes on to the kernel. The kernel then sends to process 2

Pipe: Data written by a process 1 is buffered in the pipe until process 2 reads it.

Process Synchronization?

2 types of processes

Independent Processes: Execution of one process does not affect the execution of other processes.

Cooperative Processes: Execution of one process affects the execution of other.

Process synchronization occurs in the case of co-operative process.

Semaphore?

Code logic use to resolve the critical section problem.

Steps:

1. Initially S is set to 1.
2. A process ready to execute sets S value to zero, thereby not allowing other processes to execute.
3. After execution of critical section, it re initializes S value to 1.

```

P( Semaphore S)
{
    while( S==0 );           // makes other processes wait if
                             // a certain process is executing
    S = S - 1;               // asserts S value to zero so that
                             //no other process is allowed to execute
}
/*
critical section
*/
V( Semaphore S )
{
    S = S + 1;               // release the lock
}

```

Mutex Lock

Peterson's Solution for Mutual Exclusion

Considering two processes with thread id 0 and 1

```

void lock_init()
{
    flag[0] = flag[1] = 0;    // Initialize lock by resetting the
                             // desire of both the threads to acquire the locks.
    turn = 0;
}

void lock(int self)
{
    flag[self] = 1;           // Set flag[self] = 1 saying you want to acquire lock

    turn = 1-self;           // But, first give the other thread the chance to
                             // acquire lock

    while (flag[1-self]==1 && turn==1-self) ;    // Wait until the other thread loses the desire
                                                // to acquire lock or it is your turn to get the lock.
}

void unlock(int self)
{
    flag[self] = 0;           // You do not desire to acquire lock in future.
                             // This will allow the other thread to acquire
                             // the lock.
}

```

How to create a child process?

Parent process creates the child process by calling the fork().

Fork() will return

Child process pid → 0

Parent process → Child Pid

fork() → Create child process

getpid() → Get the process id of currently executing task.

getppid() → Gets the parent pid of the child process

wait(NULL) → Parent process waits for the execution of child process

execl() → executes the commands in the argument list

```
int main()
{
    pid_t return_id = fork();
    if( return_id < 0 )
    {
        printf("fork() error");
        exit();
    }
    if(return_id != 0)
    {
        printf(" I am parent %d, My child is %d", getpid(), return_id);
        wait(NULL);    //wait for child process to join with the parent.
    }
    if(return_id == 0)
    {
        printf(" I am child %d, My parent is %d", getpid(), getppid());
        execl("/bin/echo","echo","Hello World", NULL);
    }
}

/* Sample Ouput
 I am parent 20863, My child is 20864
 I am child 20864, My parent is 20863
Hello World
*/
```

Allocation of program variables on memory stack

4 parts of memory from top to bottom

Heap

Stack

Static/global

Code section

The memory allocated for heap does not extend after the end address that is why it decrements the counter.

The memory for stack is limited and fixed and program cannot request above its allocated space.

```
#include <stdio.h>           // entire instructions will be stored in the code section
int MAX = 3;                 // Will be stored on the global block of memory
int main () {

    int a=20;                // Will be stored on stack

    int *p;                  // p will be stored on stack
                             // but the content pointing to the location on heap

    p=(int*)malloc(sizeof(int)); // malloc will allocate a memory on heap and
                             // *p will be holding the address of the dynamic variable

    free(p);                 // deallocates the memory on heap, delete in c++

}
/*
c

malloc
calloc

free

c++

new

delete
*/
```

Memory

Memory is usually organized in 3 main ways

1. Contiguous
2. Paging
3. Segmentation

When u need two processes 1 and 2 to execute.

Contiguous:

Load process 1 in to memory 0 to x. and process 2 to x+1 to y. execute 1 and 2.

Paging:

Divide whole memory in to pieces of fixed size.

Load process 1 into pages 1,3,5,6,7

Load process 2 into 2,4,6,8

Execute 1 and 2

Segmentation:

Divide memory into pieces of different sizes.

Load process1 to a piece of suitable size.

Load process 2 to a piece of its size.

Execute 1 and 2.

Virtual or Logical address: Generated by the CPU.

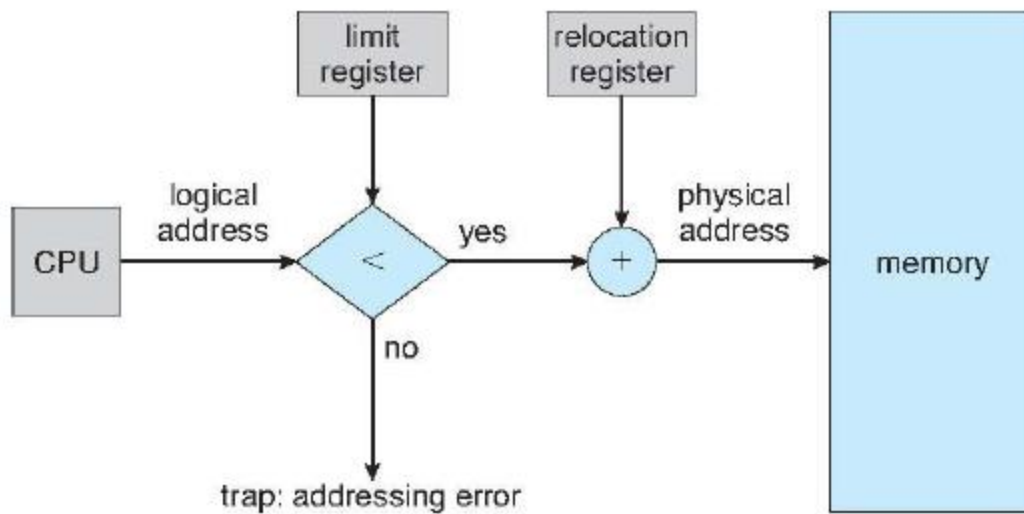
Physical address: address as seen by memory unit.

MMU : Memory management Unit

Contiguous:

The value needed to map the virtual address in to physical is stored in the **relocation register**.

The limit of the address space is maintained in the **limit register**.

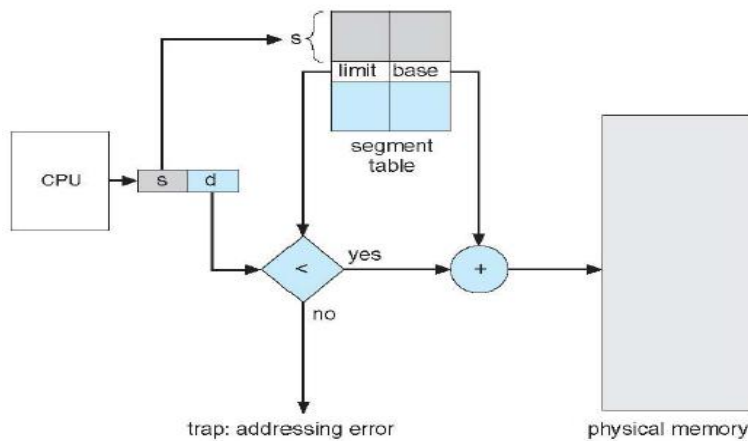


Segmentation:

Logical address consists of **<segment number, offset>**

Segment table base register: point to the segment table's location in the memory.

Segment table length register: Number of segments used by the program.



Static Linking: system libraries and program code combined by loader.

Dynamic Linking: Postponed until execution time.

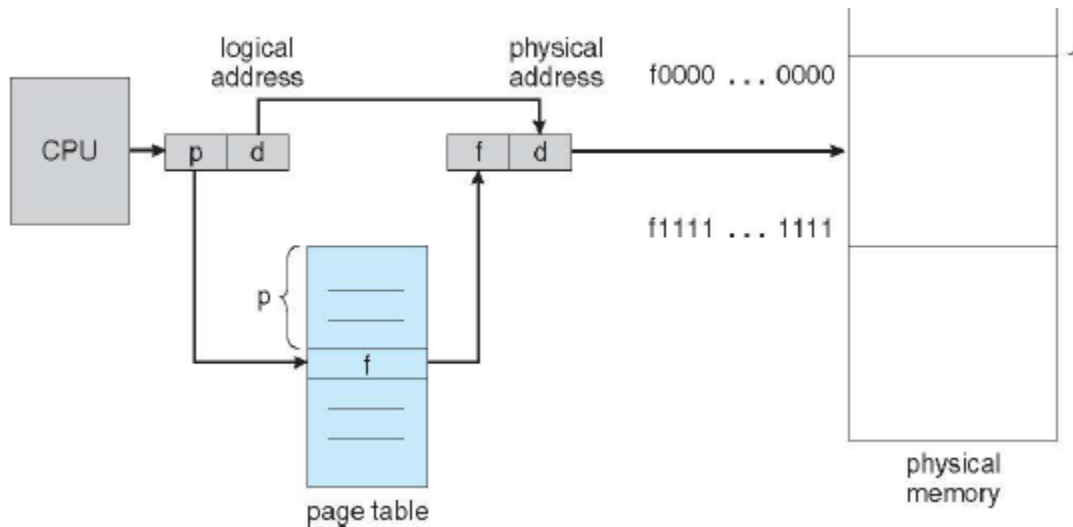
Stub: Used to locate the appropriate memory resident library routine.

Paging

Fixed size memory

Address translation scheme consists of <Page number, Offset>

Page number is then used as the line index on the page table to find the physical memory.



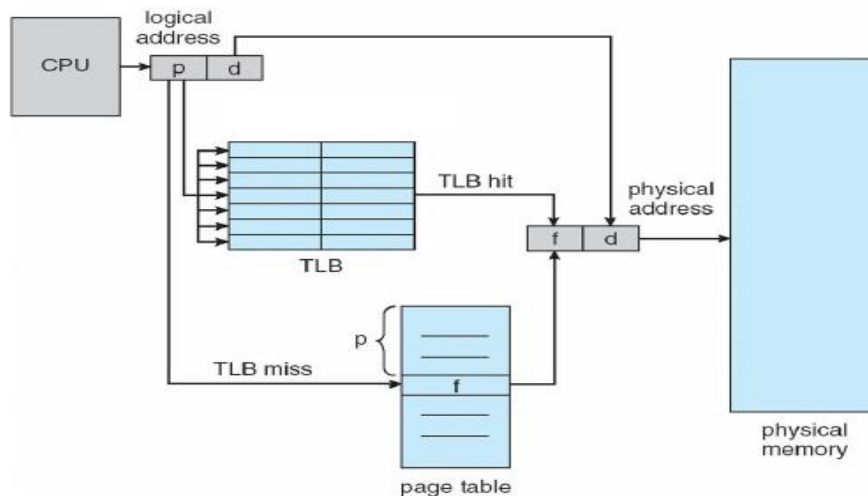
Page table and segmentation tables are stored in main memory.

Page table will have entries for all memory locations in the memory.

Why we need TLB?

Every fetch needs 2 memory access.

To speed up a Translation Look-aside Buffer (TLB) hardware cache is used.



Virtual Memory

Why we need virtual memory?

1st use.

Suppose consider a program has an address space of 4 GB. Ram Capacity is 1 GB.

For such a case a map is used which loads only the address space required for the execution in Ram.

It loads first 1 GB of memory.

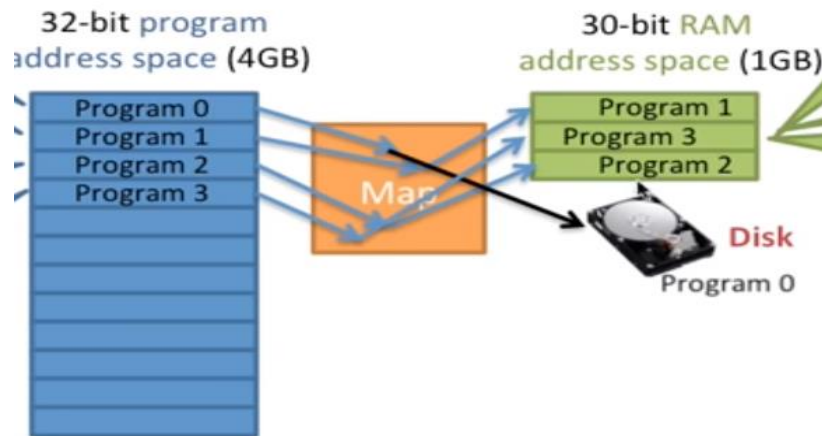
Once the processor needs to access a memory local above 1 GB. the map indicates the need for replacement.

Then the needed block of memory is replaced to a certain block in main memory. This is known as **Demand Paging**.

The mapping unit is also present in the memory itself.

Common page replacement algorithms used.

1. **First In First Out**
2. **Optimal Algorithm** – replace with the page that wont be used in future or used lately among the ones present.
3. **Least Recently Used**

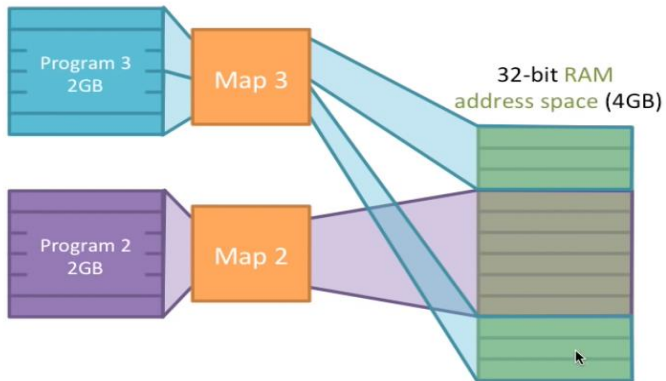


2nd Use

If a program is allocated a address space in the middle of the memory.

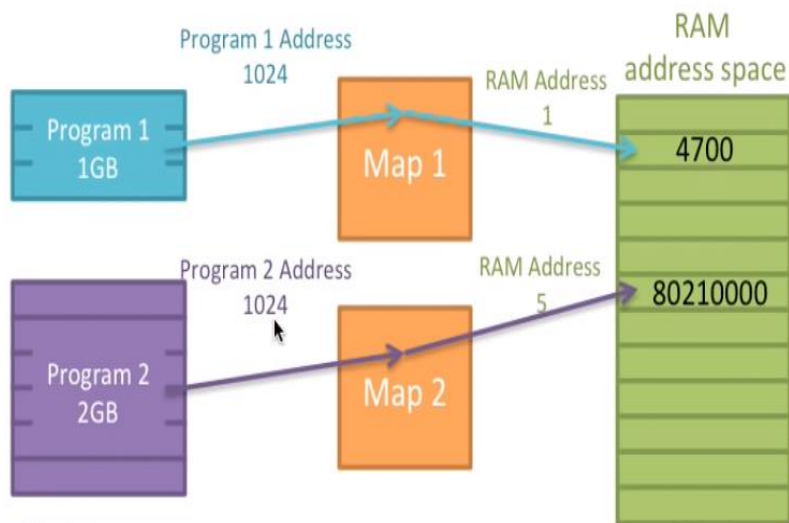
Another program has enough memory available but on different sections.

Then map is used to locate the sections where the program memory is used.

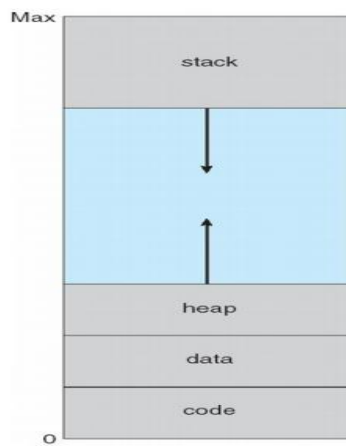


3rd use

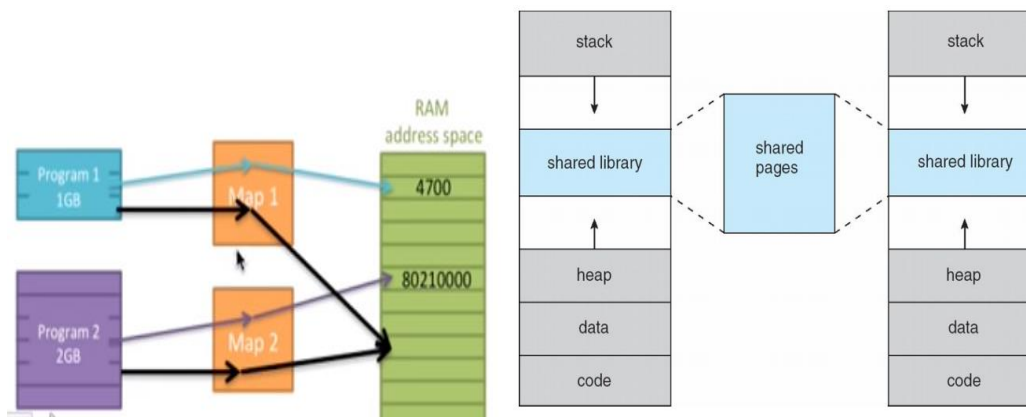
When 2 process needs to access the same address space.



Virtual Memory further maps data to particular sections in the memory



When 2 programs need to share the same memory.



Handling Page Fault?

1. When a page table has an entry which shows a page is on disk.
2. Hardware would generate a page fault exception.
3. Os then runs the page fault handler.
4. The OS chooses a page to be evicted on RAM if it is full.
5. If the page is dirty(which means data had to be modified) then it writes that to the disk first.
6. Then the page that needs to be copied from disk to memory is replaced to the location.
7. Os then updates the page table map.
8. Then jumps back and run the instruction that generated the page fault.

This is very long process.

Best way to handle is adding a larger RAM.

RISC (Reduced Instruction Set Computer):

Fewer cycles per instruction.

Smaller set of instructions.

Example: ARM, MIPS, RISC-V, Raspberry Pi, Android based systems, Iphone and IPad.

MIPS – Microprocessor without Interlocked Pipeline stages.

CISC (Complex Instruction Set Computer)

Example: x86, 8051.

Segmentation Fault in programs?

When programs try accessing memory out of its allocated space.

Assigning a pointer with an address out of address space wont throw an error.

Throws error only when process accesses the value.

Caches

Temporal Locality: If an item is reference it will be referenced again soon.

Spatial locality: If an item is reference items adjacent to it will be referenced.

Why we need virtual memory?

<https://www.youtube.com/watch?v=qlH4-oHnBb8>

<https://www.geeksforgeeks.org/operating-system-user-level-thread-vs-kernel-level-thread/>