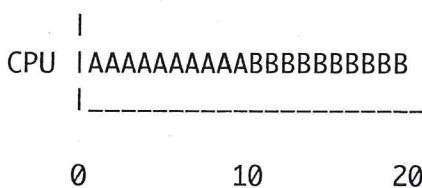


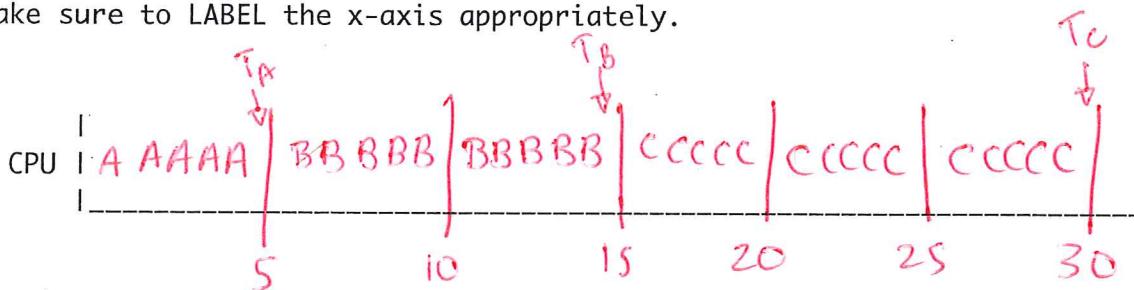
Scheduling policies can be easily depicted with some graphs. For example, let's say we run job A for 10 time units, and then run job B for 10 time units. Our graph of this policy might look like this:



In this question, you'll show your understanding of scheduling by drawing a few of these pictures.

- (a) Draw a picture of Shortest Job First (SJF) scheduling with three jobs, A, B, and C, with run times of 5, 10, and 15 time units, respectively.

Make sure to LABEL the x-axis appropriately.



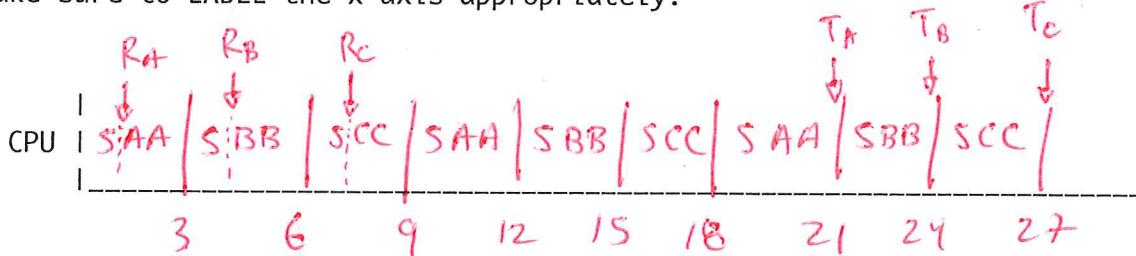
- (b) What is the average TURNAROUND TIME for jobs A, B, and C?

turnaround is $(\text{Time}_{\text{end}} - \text{Time}_{\text{submit}})$

$$T_A = (5-0) \quad T_B = (15-0) \quad T_C = (30-0) \quad T_{\text{avg}} = \frac{5+15+30}{3} = \frac{50}{3}$$

- (c) Draw a picture of ROUND-ROBIN SCHEDULING for jobs A, B, and C, which each run for 6 time units, assuming a 2-time-unit time slice; also assume that the scheduler (S) takes 1 time unit to make a scheduling decision.

Make sure to LABEL the x-axis appropriately.



- (d) What is the average RESPONSE TIME for round robin for jobs A, B, and C?

response time is $(\text{Time}_{\text{firstrun}} - \text{Time}_{\text{submit}})$

$$R_A = 1 \quad R_B = 4 \quad R_C = 7 \quad R_{\text{avg}} = \frac{1+4+7}{3} = \underline{\underline{4}}$$

- (e) What is the average TURNAROUND TIME for round robin for jobs A, B, and C?

$$T_A = 21 \quad T_B = 24 \quad T_C = 27$$

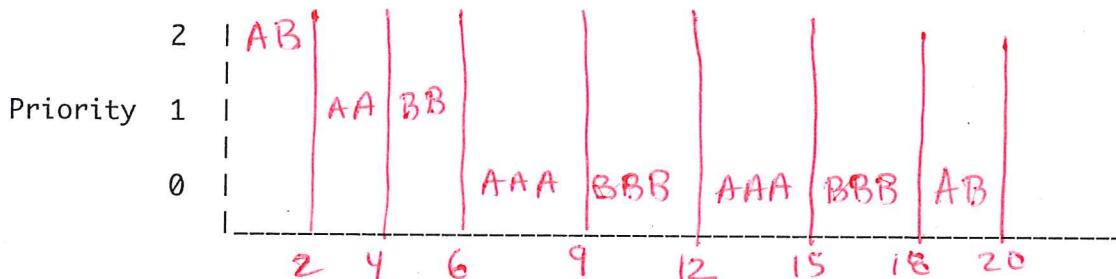
$$T_{\text{avg}} = \underline{\underline{24}}$$

Assume you have a multi-level feedback queue (MLFQ) scheduler.
In this question, we'll draw a picture of how it behaves over time.

Unlike the drawings in the previous problem (for SJF and RR), the y-axis will also be important for these pictures, as it will show the PRIORITY of the jobs over time.

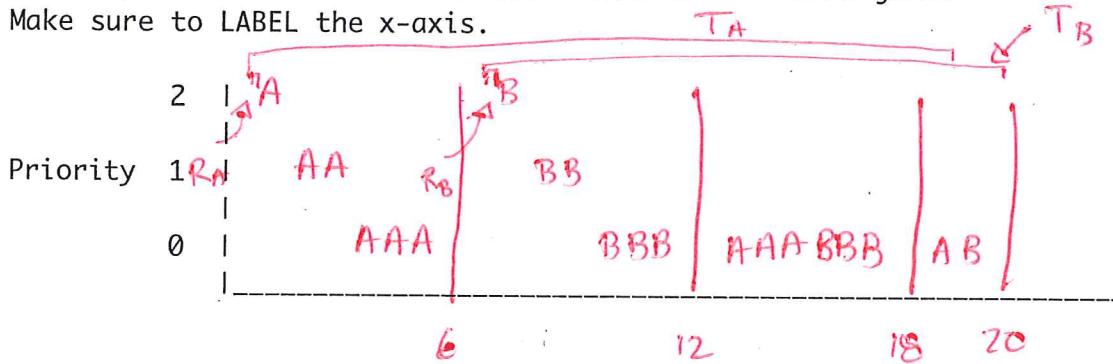
- (a) Assume a 3-level MLFQ (high priority is 2, low priority is 0).
Assume two jobs (A and B), both BATCH jobs (no I/O), each with a run-time of 10 time units, and both entering the system at T=0.
Assume the quantum length at the highest priority level is 1, then 2 at the middle, and 3 for the lowest priority.

Draw a picture of how the scheduler behaves for these jobs.
Make sure to LABEL the x-axis.



- (b) Assume the same scheduling parameters as above. Now the jobs are different; A and B both are BATCH jobs that each run for 10 time units (no I/O again), but this time A enters at T=0 whereas B enters the system at T=6.

Draw a picture of how the scheduler behaves for these jobs.
Make sure to LABEL the x-axis.



- (c) Calculate the RESPONSE TIME and TURNAROUND TIME (in part b) for Job A

$$R_A = 0 \quad T_A = 19$$

- (d) Calculate the RESPONSE TIME and TURNAROUND TIME (in part b) for Job B

$$R_B = 0 \quad T_B = 20 - 6 = 14$$

Assume you have a chunk of memory that you need to manage. When someone requests a chunk, you take the first available chunk and return it, starting at the lowest address in the space you are managing (i.e., a LOWEST-ADDRESS-FIRST policy, perhaps). The space is managed with a simple free list; when someone returns a chunk, you COALESCE the list, thus merging smaller allocated chunks back into a bigger free space.

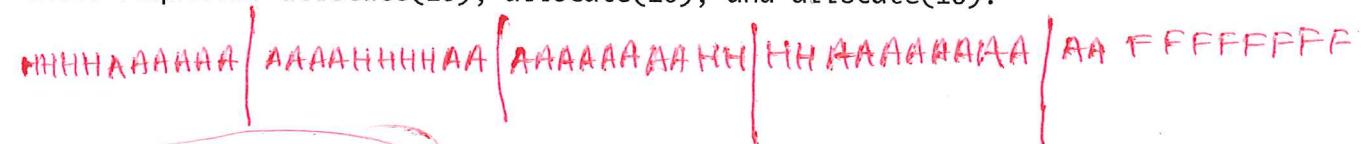
Assuming you have 50 bytes of memory to manage, and that exactly one allocation has taken place (for 10 bytes), here is what memory would look like (with spaces in-between every 10 bytes for readability):

HHHHAAAAAA AAAAFFFFFF FFFFFFFFFF FFFFFFFFFF FFFFFFFFFF

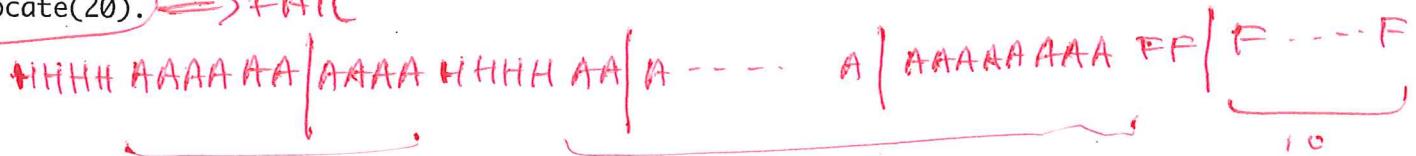
(low) Addresses of Managed Space (high)

In the picture, A means allocated, F means free, and H is a 4-byte header that is REQUIRED before every allocated chunk.

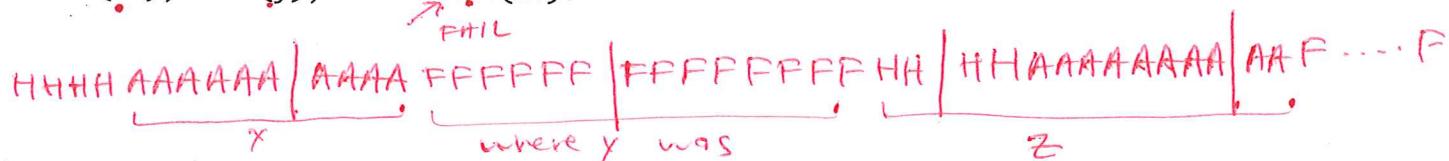
- (a) Assume a 50-byte free space. Draw what it would look like after these requests: `allocate(10)`, `allocate(10)`, and `allocate(10)`.



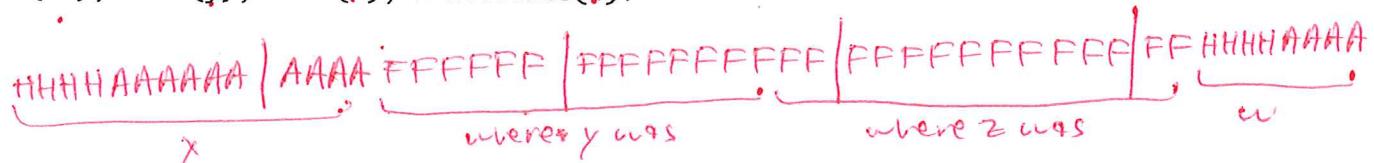
- (b) Assume a 50-byte free space. Draw what it would look like after allocation requests of `allocate(10)`, `allocate(20)`, and `allocate(20)`. FAIL



- (c) Assume a 50-byte free space. Draw what it would look like after the following requests: `x=allocate(10)`, `y=allocate(10)`, `z=allocate(10)`, `free(y)`, `w=allocate(24)`.



- (d) Assume now that there is NO COALESCING of free space. Also assume that instead of allocating via the policy of LOWEST-ADDRESS-FIRST, you instead use a BEST-FIT policy. Assume a 50-byte free space. Draw what it would look like after the following requests: `x=allocate(10)`, `y=allocate(10)`, `z=allocate(10)`, `free(y)`, `free(z)`, `w=allocate(4)`.



Assume virtual memory hardware that uses segmentation, and divides the address space in two by using the top bit of the virtual address. Each segment is thus relocated independently.

What we'll be drawing in this question is what physical memory looks given some different parameters. We'll also label where a particular memory reference ends up.

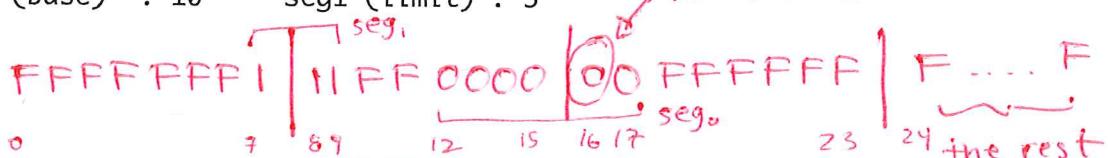
For all questions, assume a virtual address space of size 16 bytes (yes tiny!) and a physical memory of size 64 bytes. Thus, if we had a virtual address space placed in physical memory, it might look like this (with spaces between every 8 physical bytes):

0000FFFF FFFFFFFF FFFFFFFF FFFFFFFF FFFFFFFF FFFFFFFF FFFF1111

In this example, the segment 0 base register is 0, segment 1 base is 64 (it grows backwards), and both length registers are 4. 0's are used to record where segment 0 is in memory; 1's are for segment 1; F means free.

(a) What would physical memory look like if we had the following values instead? (draw a picture below)

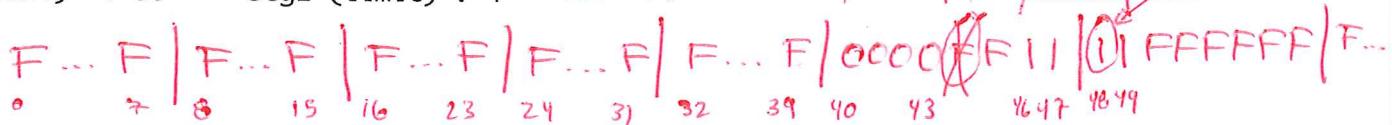
seg0 (base) : 12 seg0 (limit) : 6
seg1 (base) : 10 seg1 (limit) : 3



(b) In your picture above, CIRCLE which byte of memory is accessed when the process generates a byte load of virtual address 4 (or DRAW AN X on the physical-memory address if the access is illegal)

(c) What would physical memory look like if we had the following values instead? (draw a picture below)

seg0 (base) : 40 seg0 (limit) : 4 40 ... 43
seg1 (base) : 50 seg1 (limit) : 4 46 ... 49

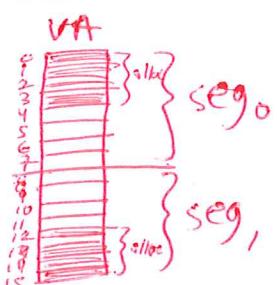


(d) In your picture above, CIRCLE which byte of memory is accessed when the process generates a byte load of virtual address 14 (or DRAW AN X on the physical-memory address if the access is illegal)

(e) In your picture above, CIRCLE which byte of memory is accessed when the process generates a byte load of virtual address 4 (or DRAW AN X on the physical-memory address if the access is illegal)

VA:4 =>
seg 0
seg 0 base: 12
+
4
16

VA:4 (fault)



Which memory is accessed during the execution of an instruction?
 For this question, assume a linear page table, with a 1-byte page-table entry. Assume an address space of size 128 bytes with 32-byte pages. Assume a physical memory of size 128 bytes. The page-table base register is set to physical address 16. The contents of the page table are:

VPN	PFN
0	1
1	Not valid
2	3
3	Not valid

Now, finally assume we have the following instruction, which loads a SINGLE BYTE from virtual address 70 into register R1:

10: LOAD 70, R1

This instruction resides at virtual address 10 within the address space of the process.

In the diagram of physical memory below:

(a) Put a BOX around each valid virtual page (and label them)

(b) Put a BOX around the page table (and label it)

(c) CIRCLE the memory addresses that get referenced during the execution of the instruction, including both instruction fetch and data access (there is no TLB).

(d) LABEL these addresses with a NUMBER that indicates the ORDER in which various physical addresses get referenced.

Physical Memory

0	1	2	3	4	5	6	7
8 ①	9	10 ③	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71
72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87
88	89	90	91	92	93	94 ④	95
96 ₆₄	97 ₆₅	98 ₆₆	99 ₆₇	100 ₆₈	101 ₆₉	102 ₇₀	103
104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119
120	121	122	123	124	125	126	127

instruction fetch:

VA : 10 \Rightarrow VPN₀

translate
by reading
first entry
of PT
 \Downarrow
42 \Rightarrow PA

data fetch :

VA : 70 \Rightarrow

read
third entry
of PT
to xlate
PA : 102

page table : base is 16, 4 entries each 1 byte
in size

VPN₀ \Rightarrow PFN₁

VPN₂ \Rightarrow PFN₃

In this question, you will examine virtual memory reference traces. An access can be a TLB hit or a TLB miss; if it is a TLB miss, the reference can be a page hit (present) or a page fault (not present).

Assume a TLB with 4 entries, and a memory that can hold 8 pages. Assume the TLB and memory both are empty initially. Finally, assume LRU replacement is used for both the TLB and memory.

- (a) What happens on each access in the following reference trace?
 a TLB hit, TLB miss/page hit, or TLB miss/page fault?
 (these can be abbreviated H, M, or PF)

0	PF
1	PF
2	PF
3	PF
0	H
1	H
2	H
3	H

} after faulting in,
 all are in mem
 and mapped in TLB

- (b) What happens on each access in the following reference trace?
 (write H, M, or PF)

0	PF
1	PF
2	PF
3	PF
4	PF
0	M
1	M
2	M
3	M
4	M

} this time, pages are in
 memory, but TLB
 not big enough \Rightarrow TLB misses

- (c) Now assume a memory that can only hold 3 pages.
 What happens on each access in the following reference trace? (H, M, PF)

0	PF	LRU \rightarrow 0 1 2		
1	PF			
2	PF			
0	H	1	2	0
1	H	2	0	1
3	PF	0	1	3
0	H	1	3	0
3	H	1	0	3
1	H	0	3	1
2	PF	3	1	2
1	H	3	2	1

TLB : big enough to hold
 all mappings \Rightarrow no TLB misses

In this question, we'll examine a multi-level page table, like that found in the (optional) homework. The parameters are the same:

- The page size is an unrealistically-small 32 bytes.
- The virtual address space for the process in question (assume there is only one) is 1024 pages, or 32 KB.
- Physical memory consists of 128 pages.

Thus, a virtual address needs 15 bits, 5 of which are the offset. A physical address requires 12 bits, also with 5 as the offset.

The system assumes a multi-level page table. Thus, the upper five bits of a virtual address are used to index into a page directory; the page directory entry (PDE), if valid, points to a page of the page table. Each page table page holds 32 page-table entries (PTEs). Each PTE, if valid, holds the desired translation (physical frame number, or PFN) of the virtual page in question.

The format of a PTE is thus:

VALID | ~~PFN7~~ PFN6 ... PFN0

and is thus 8 bits or 1 byte.

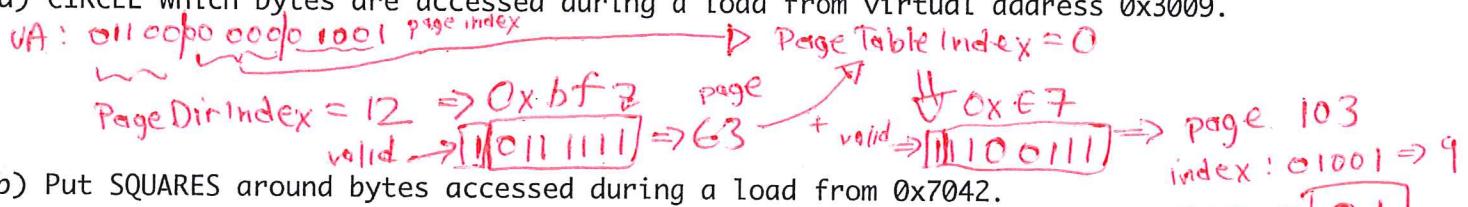
The format of a PDE is essentially identical:

VALID | ~~PT7~~ PT6 ... PT0

For this question, assume the PDBR (page-directory base register) is 73.

On the next page is the physical memory dump, where your answers will go.

(a) CIRCLE which bytes are accessed during a load from virtual address 0x3009.



(b) Put SQUARES around bytes accessed during a load from 0x7042.

11100 | 00010 | 00010
 PDindex PTindex offset
 on page

page of page table
for (a)

PD indexed entry for (d)

Ah, virtual machine monitors. You use them, and now (hopefully) you understand them (a little bit).

Assume in this question some hardware that has a software-managed TLB.

Assume we are running a virtual machine monitor (VMM), an operating system (OS) on top of the VMM, and a user process running on the OS.

Draw a picture of the control flow during a TLB miss generated by the user process. The picture should reflect a time-line of what happens during this miss, including when the user process, OS, and VMM run, and what they do when they run.

