5118020-03 Operating System

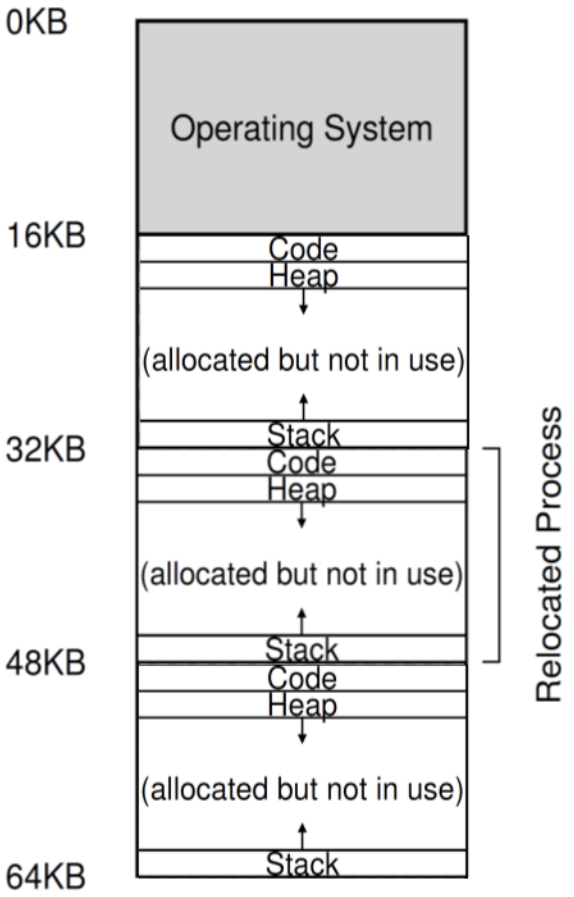
Segmentation

OSTEP Chapters 16 & 17

Shin Hong

Motivation

- **Internal fragmentation problem**
  - a process typically uses a small portion of an address space in a sparse manner
  - the rest of the address space remains unused and wasted
- **Redundant data**
  - The same piece of code may be loaded redundantly if multiple processes use it



Approach

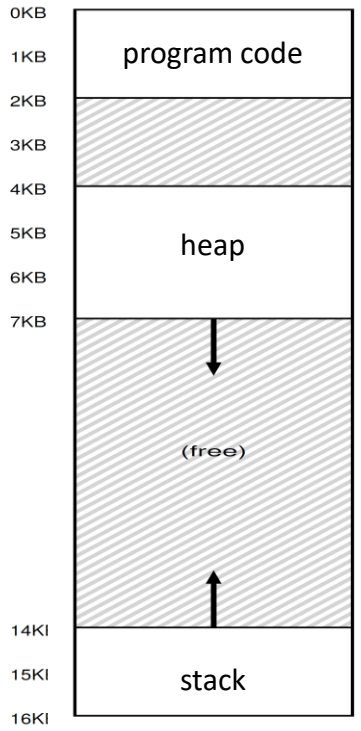
- Ideas
  1. split an address space into multiple pieces and manage each separately
  2. allocate the memory to each piece depending on the actual needs

- Define the address space of a process as a set of **segments**
  - a segment is a continuous memory region defined by a pair of base and bounds addresses
  - the available portion of the address space of a process can be represented as a set of segments
  - a segment is initially defined at a loading time
  - a segment can be relocated, extended, or shrunk over time

- Possible to make multiple processes share the same segment if they use the same code or data

Segmentation

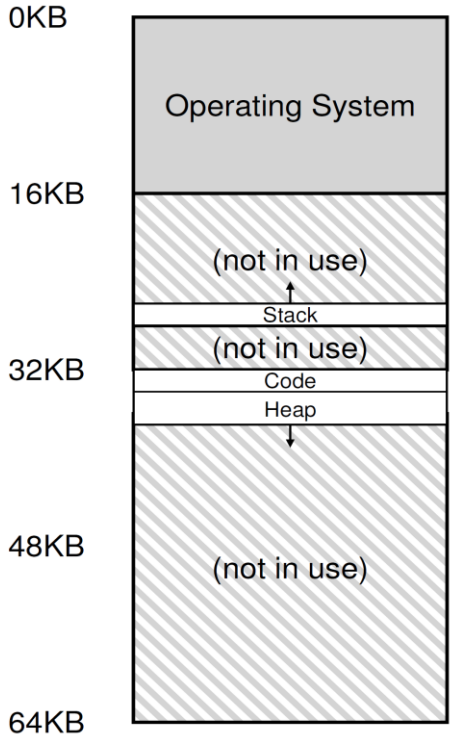
- A process has three segments, **code**, **heap** and **stack** in an address space
- MMU holds three pairs of base-bound values, and for each memory access, it identifies which segment the access is on
  - For a virtual address, MMU finds the base value, and add it to the offset part to obtain the corresponding physical address
- On demand, OS can change bounds to extend/shrink segments of a process



Segment	Base	Size
Code	32K	2K
Heap	34K	3K
Stack	28K	2K

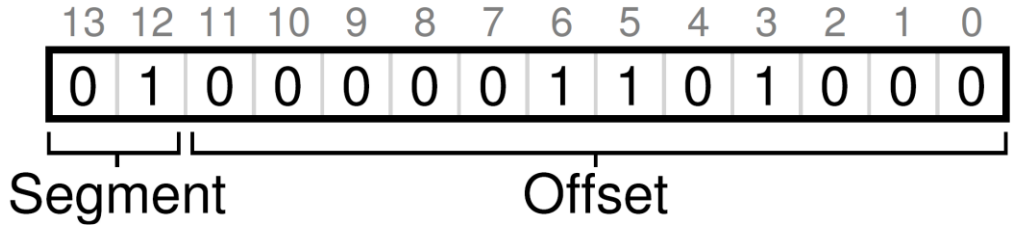
100 (0x0064) → 32K + 100 (0x8064)

4200 (0x1068) → 34K + 104 (0x8868)



Which Segment A Memory Access is On?

- By virtual address
  - use few most significant bits as a segmentation indicator
  - ex. virtual address 4200



00: code
01: heap
11: stack

- By instruction type
  - referring the code segment if the address is derived from PC
  - referring the stack segment if the address is derived from stack pointer
  - referring the heap segment otherwise

Segment Attributes

- A bit to indicate whether a segment grows forward or backward
  - ex.

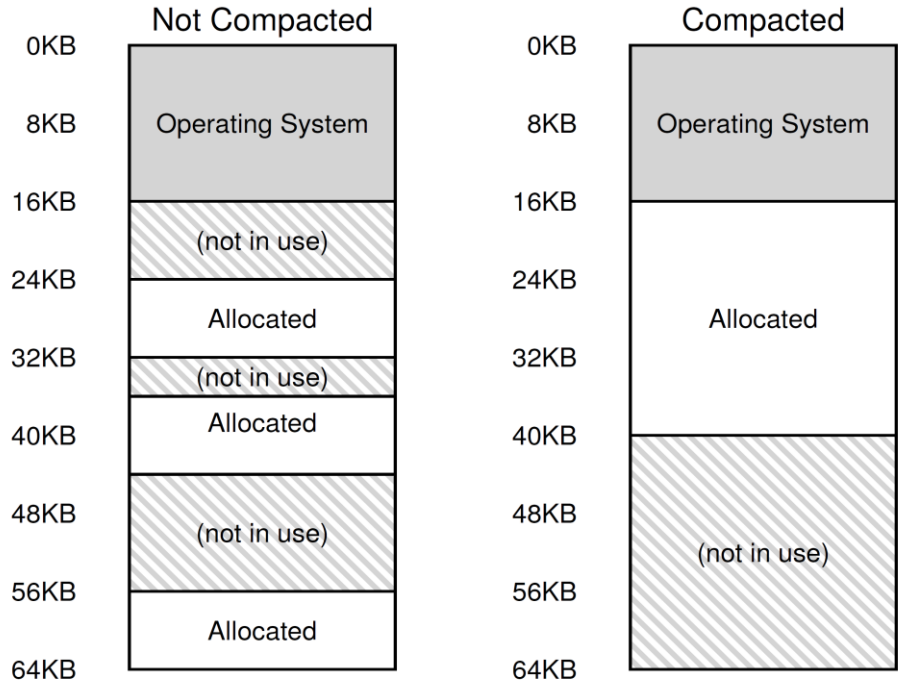
Segment	Base	Size (max 4K)	Grows Positive?
Code_{00}	32K	2K	1
Heap_{01}	34K	3K	1
Stack_{11}	28K	2K	0

- A bit to indicate whether a segment is for read-write or read-only
  - ex.

Segment	Base	Size (max 4K)	Grows Positive?	Protection
Code_{00}	32K	2K	1	Read-Execute
Heap_{01}	34K	3K	1	Read-Write
Stack_{11}	28K	2K	0	Read-Write

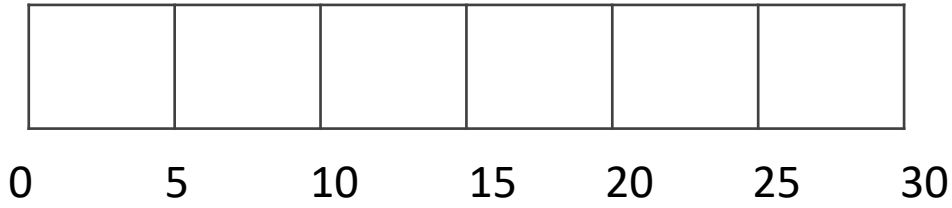
Operating System Supports

- context switching
- serve requests of growing/shrinking a segment
- manage free space to counter external fragmentation
  - compaction
  - free-list management



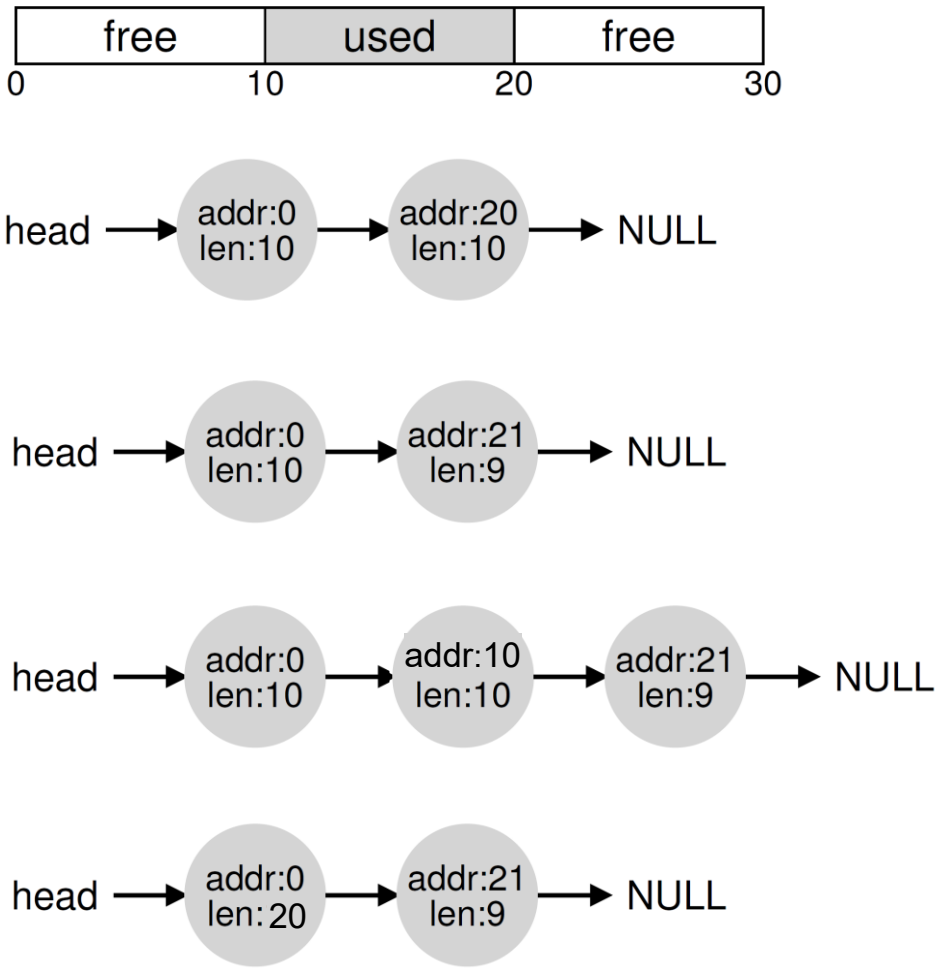
Free-Space Management

- External fragmentation hurts memory utilization
  - commonly happens for dynamic allocation of variable-length memory units
  - example. with a 30-bytes address space
    - alloc 15 bytes as A
    - alloc 10 bytes as B
    - free A (15 bytes)
    - alloc 5 bytes as C
    - alloc 15 bytes as D



Free List

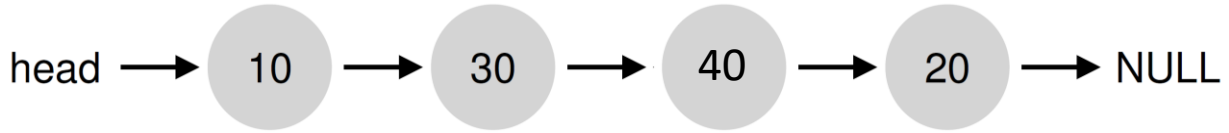
- maintains free space as a linked list of free chunks (i.e., continuous unused memory regions)
- to allocate a chunk of size *m*
  - find a node of a free chunk whose size is greater than or equal to *m*
  - split the free chunk if its size is greater than *m*
- to free an allocated memory chunk,
  - add a node of the newly freed chunk to the free list
  - merge adjacent chunks into a single larger chunk (coalescing)



Allocation Strategies

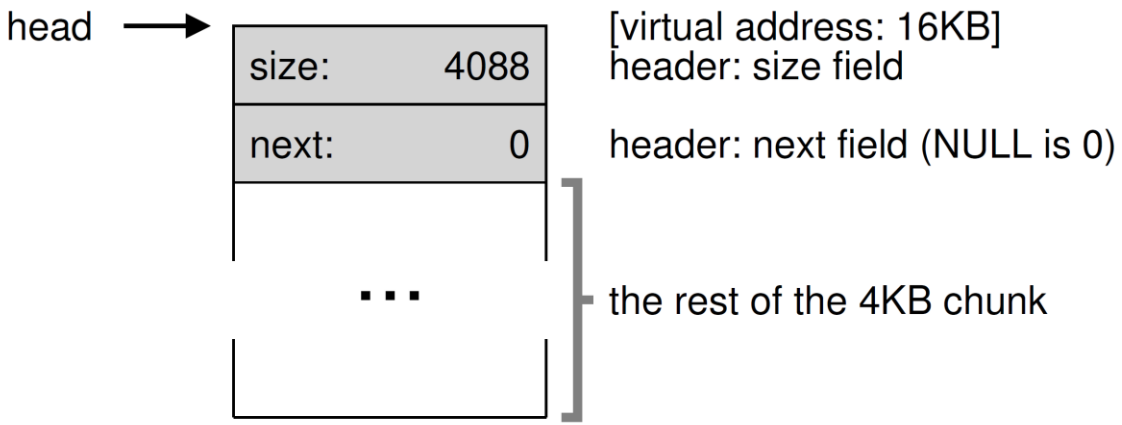
- The ideal allocator should be fast and minimize fragmentation.
- **Best fit**: search through the free list and find the smallest chunks that are large enough to afford the memory request (i.e., closest to what the user asks)
- **Worst fit**: find the largest chunk in order to keep large chunks remain in the free list
- **First fit**: find the first chunk that is large enough to afford the requested memory
- **Next fit**: conduct the first fit search from the node where the last search was stopped (not from the beginning of the free list)

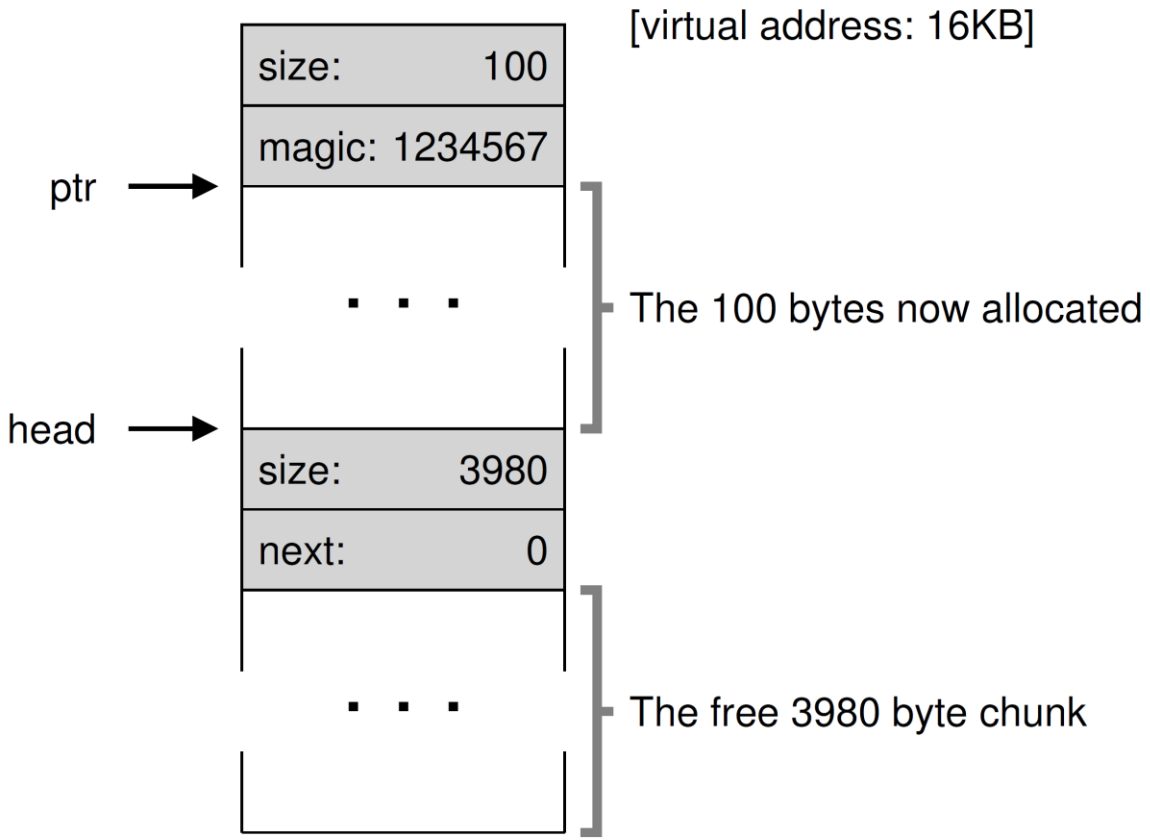
Ex. the user asks 18

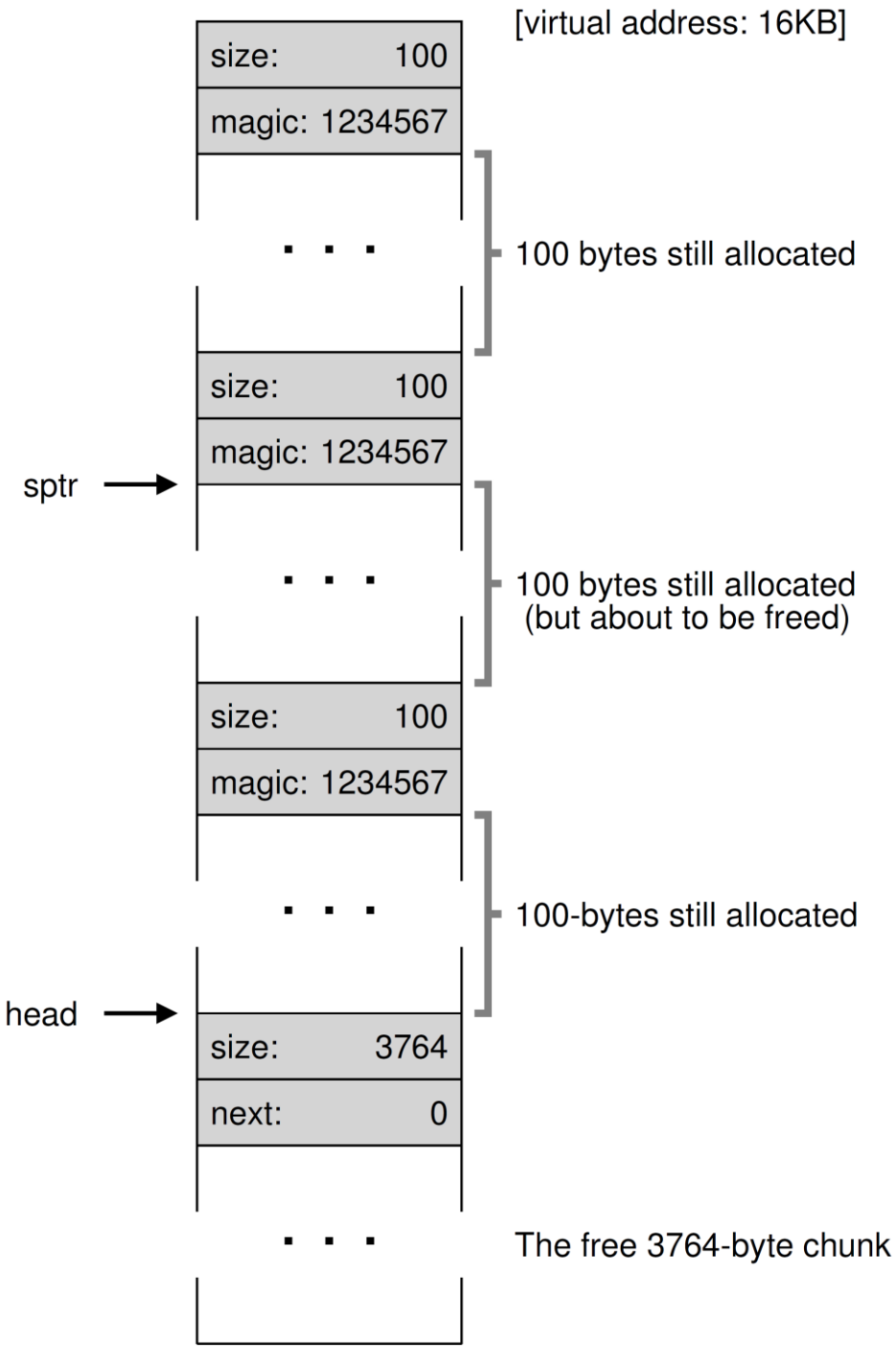


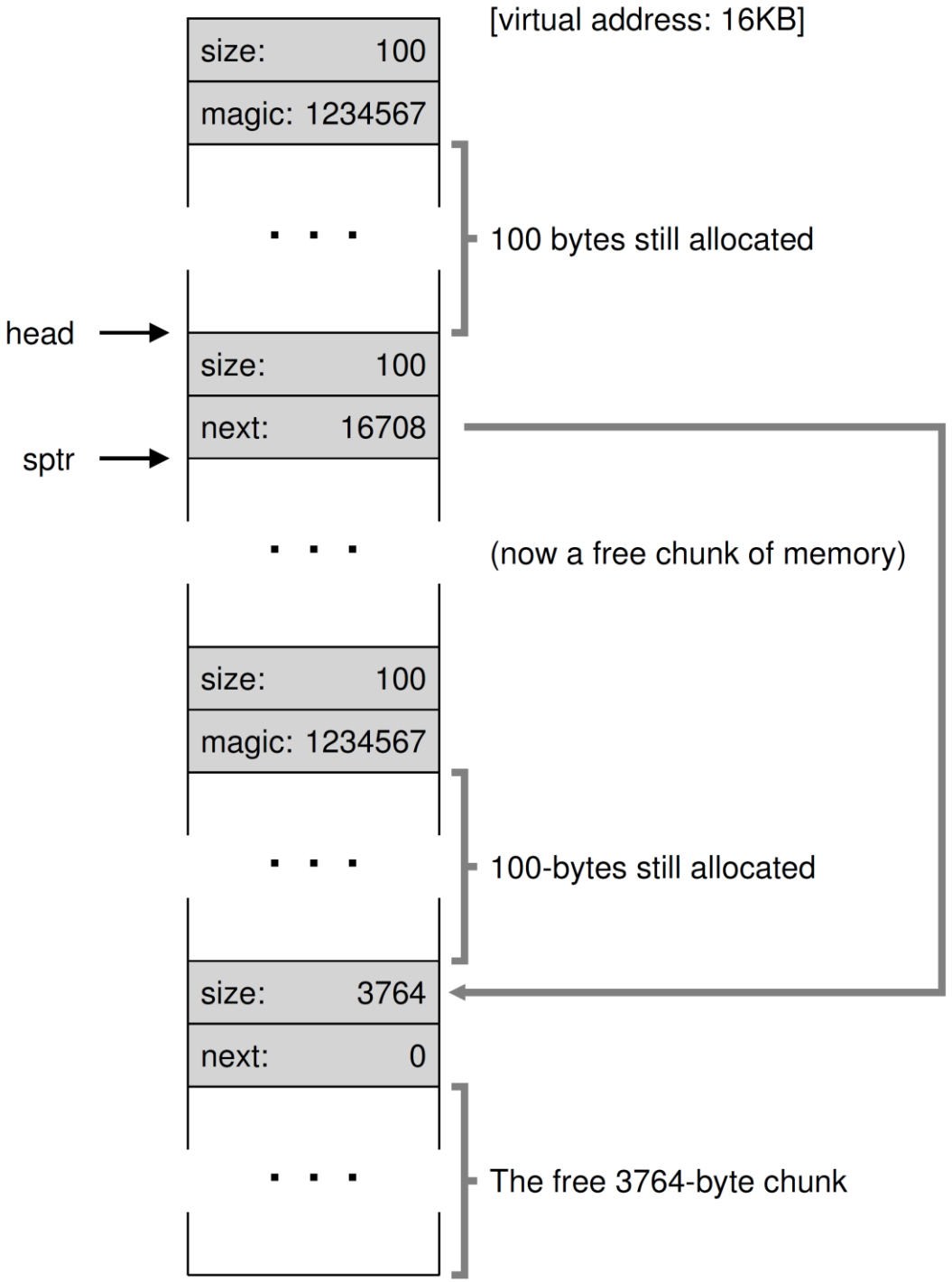
Embedding A Free List

- Example: 4096-byte memory






[virtual address: 16KB]
size: 100
magic: 1234567
100 bytes still allocated
size: 100
magic: 1234567
sptr
100 bytes still allocated
(but about to be freed)
size: 100
magic: 1234567
100-bytes still allocated
head
size: 3764
next: 0
The free 3764-byte chunk

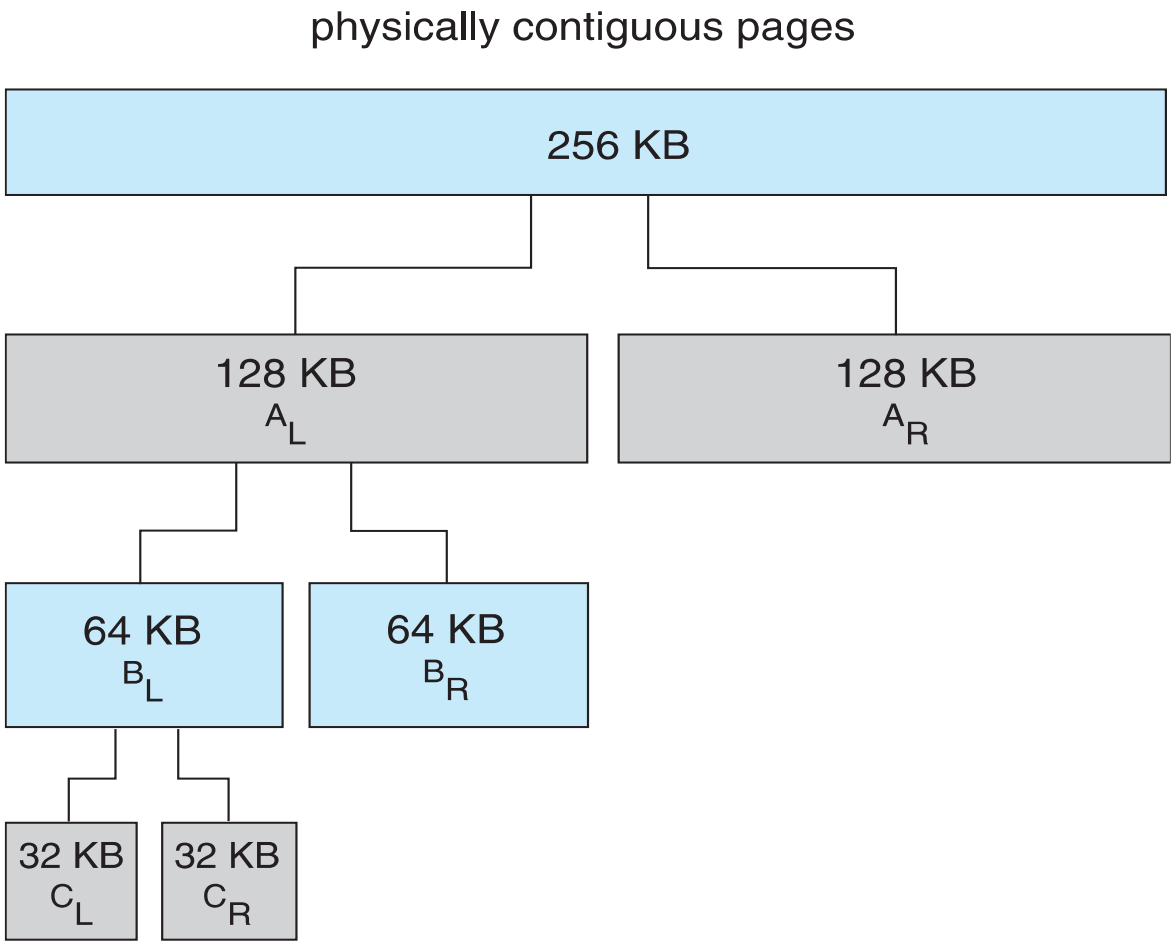

[virtual address: 16KB]
size: 100
magic: 1234567
100 bytes still allocated
head
size: 100
sptr
next: 16708
(now a free chunk of memory)
size: 100
magic: 1234567
100-bytes still allocated
size: 3764
next: 0
The free 3764-byte chunk

Buddy System (1/3)

- Use **power-of-2 allocator**
  - satisfies requests in units sized as a power of 2 (e.g., 16, 32, 64)
  - request rounded up to next smallest power of 2 (e.g., 32 for 20, 64 for 40)
  - split an existing chunk into two buddies when the smaller chunk is needed
    - continue until an appropriately-sized chunk is available

- E.g., assume one chunk of 256KB is available, there is a request of 21 KB
  - split the chunk into A_L and A_R of 128 KB each
  - split A_L into B_L and B_R, each of which has 64 KB
  - split B_L into C_L and C_R of 32 KB each – one used to satisfy request

Buddy System (2/3)



Buddy System (3/3)

- Advantage: quickly **coalesce** unused chunks into larger chunk
- Disadvantage: internal fragmentation