Slide 1

Administrivia

- (None.)

Slide 2

9/21 Minute Essay Followup

- “Seems like this competition for memory is a lot like competition for I/O, and generalized is just deadlock.” (Review definition of mutual exclusion.)
- “Issues with multiple critical regions may cause problems with these solutions.” (Review definition of critical region.)
- More examples of mutual exclusion (or similar problems): users changing same password, Web servers, concurrent access to database, buying a ticket online and having price change mid-transaction, print server.

9/23 Minute Essay Followup

Slide 3

- "Does Peterson's algorithm scale up to more than two processes?" It can be extended to n processes, but the result is complicated. A simpler algorithm for n processes that also doesn't need hardware support is Lamport's bakery algorithm.
- "If every process uses a register, wouldn't this be unscalable?" (Review "virtual CPU" idea.)

Semaphores — Recap

Slide 4

- Semaphore ADT:
  - Value — non-negative integer.
  - Two operations, up and down; both atomic.
- Last time — solution to mutual exclusion problem using semaphores.

Slide 5

Bounded Buffer Problem

- Idea — we have a buffer of fixed size (e.g., an array), with some processes ("producers") putting things in and others ("consumers") taking things out. Synchronization:
  - Only one process at a time can access buffer.
  - Producers wait if buffer is full.
  - Consumers wait if buffer is empty.
- Example of use: print spooling (producers are jobs that print, consumer is printer — actually could imagine having multiple printers/consumers).

Slide 6

Bounded Buffer Problem, Continued

- Shared variables:

```
    buffer B(N); // initially empty, can hold N things
```

Pseudocode for producer:

```
while (true) {
    item = generate();
    put(item, B);
}
```

Pseudocode for consumer:

```
while (true) {
    item = get(B);
    use(item);
}
```

- Synchronization requirements:
  1. At most one process at a time accessing buffer.
  2. Never try to `get` from an empty buffer or `put` to a full one.
  3. Processes only block if they "have to".

Slide 7

Bounded Buffer Problem, Continued

- We already know how to guarantee one-at-a-time access. Can we extend that?
- Three situations where we want a process to wait:
  - Only one get/put at a time.
  - If B is empty, consumers wait.
  - If B is full, producers wait.

Slide 8

Bounded Buffer Problem, Continued

- What about three semaphores?
  - One to guarantee one-at-a-time access.
  - One to make producers wait if B is full — so, it should be zero if B is full — "number of empty slots"?
  - One to make consumers wait if B is empty — so, it should be zero if B is empty — "number of slots in use"?

Slide 9

Bounded Buffer Problem — Solution

- Shared variables:

```
buffer B(N); // empty, capacity N
semaphore mutex(1);
semaphore empty(N);
semaphore full(0);
```

Pseudocode for producer:

```
while (true) {
    item = generate();
    down(empty);
    down(mutex);
    put(item, B);
    up(mutex);
    up(full);
}
```

Pseudocode for consumer:

```
while (true) {
    down(full);
    down(mutex);
    item = get(B);
    up(mutex);
    up(empty);
    use(item);
}
```

Slide 10

Implementing Semaphores

- We want to define:
  - Data structure to represent a semaphore.
  - Functions `up` and `down`.
- `up` and `down` should work the way we said, and we'd like to do as little busy-waiting as possible.

Implementing Semaphores, Continued

- Idea — represent semaphore as integer plus queue of waiting processes (represented as, e.g., process IDs).
- Then how should this work . . .

Slide 11

Implementing Semaphores, Continued

- Variables — integer `value`, queue of process IDs `queue`.

```
down() {
    bool zero;
    enter_cr();
    zero = (value == 0);
    if (!zero)
        value -= 1;
    else
        enqueue(current_process, queue);
    leave_cr();
    if (zero)
        block();    // mark current process blocked
}
```

```
up() {
    process p = null;
    enter_cr();
    if (empty(queue))
        value += 1;
    else
        p = dequeue(queue);
    leave_cr();
    if (p != null)
        unblock(p);     // mark p runnable
}
```

- `enter_cr()`, `leave_cr()` mostly like before; see p. 113.

Slide 12

Minute Essay

- Alleged joke (from some random Usenet person):

  A man's P should exceed his V else what's a sema for?

  Do you understand this? (Remember that P is "down" and V is "up".)

Slide 13

Minute Essay Answer

- It's a pun. The idea is roughly that if you never have a situation in which you've attempted more "down" operations than "up" operations, you didn't need a semaphore. (Or that's what I think it means. The author might have another idea!)

Slide 14