Lecture 12: Monitors and Condition Variables

Instructor: Chengyu Song

Slide contributions from
Nael Abu-Ghazaleh, Harsha Madhyvasta and Zhiyun Qian
Higher-Level Synchronization

- Locks so far inefficient when critical sections are long
  - Spinlocks – inefficient
  - Disabling interrupts – can miss or delay important events

- Instead, we want synchronization mechanisms that
  - Block waiters
  - Leave interrupts enabled inside the critical section

- Plan:
  - Look at two common high-level mechanisms
    » Semaphores: binary (mutex) and counting
    » Monitors: mutexes and condition variables
  - Use them to solve common synchronization problems
Semaphores

- Semaphores are an abstract data type that provide mutual exclusion to critical sections
  - Block waiters, interrupts enabled within critical section

- Semaphores are integers that support two operations:
  - `wait(semaphore)`: decrement, block until semaphore is open
  - `signal(semaphore)`: increment, allow another thread to enter

- Semaphore safety property: the semaphore value is always greater than or equal to 0
Monitors

- A monitor is a programming language construct that controls access to shared data
  - Synchronization code added by compiler, enforced at runtime
  - Why is this an advantage?

- A monitor encapsulates
  - Shared data structures
  - Procedures that operate on the shared data structures
  - Synchronization between concurrent threads that invoke the procedures
Monitor Semantics

- A monitor guarantees mutual exclusion
  - Only one thread can execute any monitor procedure at any time (the thread is “in the monitor”)
  - If a second thread invokes a monitor procedure when a first thread is already executing one, it blocks
    - So the monitor has to have a wait queue...

- Monitors also support waiting on conditions
  - Situation: we enter a monitor and find that we need to wait
    - E.g., producer when the buffer is full
  - If we just wait, the monitor is blocked
    - So, monitors support waiting while releasing the monitor
Account Example

Monitor account {
  double balance;

  double withdraw(amount) {
    balance = balance – amount;
    return balance;
  }
}

Hey, that was easy!
But what if a thread wants to wait inside the monitor?
  » Such as “mutex(empty)” by reader in bounded buffer?

When first thread exits, another can enter. Which one is undefined.
Monitors, Monitor Invariants and Condition Variables

- A **monitor invariant** is a safety property associated with the monitor, expressed over the monitored variables. It holds whenever a thread enters or exits the monitor.
- A **condition variable** is associated with a condition needed for a thread to make progress once it is in the monitor.

```plaintext
Monitor M {
    ... monitored variables
    Condition c;

    void enter_mon (...) {
        if (extra property not true) wait(c);  \textit{waits outside of the monitor's mutex}
        do what you have to do
        if (extra property true) signal(c);    \textit{brings in one thread waiting on condition}
    }
```
Condition Variables

- Condition variables support three operations:
  - **Wait** – release monitor lock, wait for C/V to be signaled
    » So condition variables have wait queues, too
  - **Signal** – wakeup one waiting thread
  - **Broadcast** – wakeup all waiting threads

- Condition variables *are not* boolean objects
  - “if (condition_variable) then” … does not make sense
  - “if (num_resources == 0) then wait(resources_available)” does
  - An example will make this more clear
Monitor Bounded Buffer

Monitor \texttt{bounded\_buffer} \{ 
  Resource buffer[N];
  // Variables for indexing buffer
  // monitor invariant involves these vars
  Condition not_full; // space in buffer
  Condition not_empty; // value in buffer

  void \texttt{put\_resource} (Resource R) \{ 
    if (buffer array is full) 
      wait(not\_full);
    \textit{Add R to buffer array};
    signal(not\_empty);
  \}

  Resource \texttt{get\_resource}() \{ 
    if (buffer array is empty)
      wait(not\_empty);
    \textit{Get resource R from buffer array};
    signal(not\_full);
    return R;
  \}
\} // end monitor

\begin{itemize}
  \item What happens if no threads are waiting when signal is called?
\end{itemize}
Monitor Queues

Monitor `bounded_buffer`

```
Condition not_full;
...other variables...
Condition not_empty;

void put_resource () {
  ...wait(not_full)...
  ...signal(not_empty)...
}
Resource get_resource () {
  ...
}
```

Waiting to enter
Waiting on condition variables
Executing inside the monitor
Condition Vars != Semaphores

- Condition variables != semaphores
  - Although their operations have the same names, they have entirely different semantics
  - However, they each can be used to implement the other

- Access to the monitor is controlled by a lock
  - `wait()` blocks the calling thread, and gives up the lock
    » To call `wait`, the thread has to be in the monitor (hence has lock)
    » `Semaphore::wait` just blocks the thread on the queue
  - `signal()` causes a waiting thread to wake up
    » If there is no waiting thread, the signal is lost
    » `Semaphore::signal` increases the semaphore count, allowing future entry even if no thread is waiting
    » Condition variables have no history
Signal Semantics

- There are two flavors of monitors that differ in the scheduling semantics of signal()
  - **Hoare** monitors (original)
    - signal() immediately switches from the caller to a waiting thread
    - The condition that the waiter was anticipating is guaranteed to hold when waiter executes
    - Signaler must restore monitor invariants before signaling
  - **Mesa** monitors (Mesa, Java)
    - signal() places a waiter on the ready queue, but signaler continues inside monitor
    - Condition is not necessarily true when waiter runs again
      - Returning from wait() is only a hint that something changed
      - Must recheck conditional case
Hoare vs. Mesa Monitors

- **Hoare**
  
  ```
  if (empty)
  wait(condition);
  ```

- **Mesa**
  
  ```
  while (empty)
  wait(condition);
  ```

- **Tradeoffs**
  - Mesa monitors easier to use, more efficient
    - Fewer context switches, easy to support broadcast
  - Hoare monitors leave less to chance
    - Easier to reason about the program
Condition Vars & Locks

- Condition variables are also used without monitors in conjunction with **blocking** locks
- A monitor is “just like” a module whose state includes a condition variable and a lock
  - Difference is syntactic; with monitors, compiler adds the code
- It is “just as if” each procedure in the module calls acquire() on entry and release() on exit
  - But can be done anywhere in procedure, at finer granularity
- With condition variables, the module methods may wait and signal on independent conditions
Monitors and Java

- A lock and condition variable are in every Java object
  - No explicit classes for locks or condition variables
- Every object is/has a monitor
  - At most one thread can be inside an object’s monitor
  - A thread enters an object’s monitor by
    - Executing a method declared “synchronized”
      - Can mix synchronized/unsynchronized methods in same class
    - Executing the body of a “synchronized” statement
      - Supports finer-grained locking than an entire procedure
- Every object can be treated as a condition variable
  - Object::notify() has similar semantics as Condition::signal()
Summary

● Semaphores
  ◆ `wait()`/`signal()` implement blocking mutual exclusion
  ◆ Also used as atomic counters (counting semaphores)
  ◆ Can be inconvenient to use

● Monitors
  ◆ Synchronizes execution within procedures that manipulate encapsulated data shared among procedures
    » Only one thread can execute within a monitor at a time
  ◆ Relies upon high-level language support

● Condition variables
  ◆ Used by threads as a synchronization point to wait for events
  ◆ Inside monitors, or outside with locks
More synchronization practice

- Cars cross a narrow bridge. Bridge be open in one direction at a time. It can hold no more than three cars. Build threads to simulate this situation.

- Change the problem above to avoid starvation of one direction. Let's say we don’t let more than 10 cars cross in any direction if there are cars waiting on the other direction.
More synchronization practice

- You take a break from studying cs153 to play frisbee with your friends. We have one or more frisbees, and there are two or more of you. Each student waits until they have a frisbee and their neighbor doesn’t have one and then throws the frisbee.

- What happens if the number of frisbees is equal to the number of players?
More synchronization practice

- CS153 students are studying for the midterm over the national CS dish (Pizza). Each pizza pie has 8 slices. Each student eyes the pie, then grabs the next slice.
  - What race conditions can happen?
  - How can you resolve them?

- A student that grabs the last slice should order the next pie. Extend your implementation to do that
Next class

- Deadlock
- Preparation
  - Read Module 32