Lecture 16
Nondeterministic Parallel Programming

Charles E. Leiserson
November 5, 2013
**Determinism**

**Definition.** A program is **deterministic** on a given input if every memory location is updated with the same sequence of values in every execution.

- The program always behaves the same way.
- Two different memory locations may be updated in different orders, but each location always sees the same sequence of updates.

**Advantage:** debugging!
Rule of Thumb

Always write deterministic programs.
Rule of Thumb

Always write deterministic programs, unless you can’t!
Mutual Exclusion
Insert \( x \) into table:

1. \( \text{slot} = \text{hash}(x->\text{key}); \)
2. \( x->\text{next} = \text{table}[\text{slot}]; \)
3. \( \text{table}[\text{slot}] = x; \)
Concurrent Hash Table

RACE BUG!

1. slot = hash(x->key);
2. x->next = table[slot];
3. slot = hash(y->key);
4. y->next = table[slot];
5. table[slot] = y;
6. table[slot] = x;

© 2013 Charles E. Leiserson and I-Ting Angelina Lee
**Definition.** A **critical section** is a piece of code that accesses a shared data structure that must not be accessed by two or more threads at the same time (**mutual exclusion**).
Mutexes

Definition. A **mutex** is an object with `lock` and `unlock` member functions. An attempt by a thread to lock an already locked mutex causes that thread to **block** (i.e., wait) until the mutex is unlocked.

Modified code: Each slot is a `struct` with a mutex `L` and a pointer `head` to the slot contents.

```c
slot = hash(x->key);
lock(&table[slot].L);
  x->next = table[slot].head;
  table[slot].head = x;
unlock(&table[slot].L);
```

Cilk Plus uses TBB mutex locks (**tbb::mutex**), but Cilkscreen also understands Pthread locks.
Recall: Determinacy Races

**Definition.** A *determinacy race* occurs when two logically parallel instructions access the same memory location and at least one of the instructions performs a write.

- A program execution with no determinacy races means that the program is deterministic on that input.
- The program always behaves the same on that input, no matter how it is scheduled and executed.
- If determinacy races exist in an ostensibly deterministic program (e.g., a program with no mutexes), Cilkscreen guarantees to find such a race.
Definition. A data race occurs when two logically parallel instructions holding no locks in common access the same memory location and at least one of the instructions performs a write.

Cilk screen understands locks and will not report a determinacy race unless it is also a data race.

**WARNING:** Codes that use locks are nondeterministic by intention, and they weaken Cilk screen’s guarantee unless critical sections “commute.”
No Data Races $\neq$ No Bugs

Example

```c
slot = hash(x->key);
lock(&table[slot].L);
  x->next = table[slot].head;
unlock(&table[slot].L);
lock(&table[slot].L);
  table[slot].head = x;
unlock(&table[slot].L);
```

Nevertheless, the presence of mutexes and the absence of data races at least means that the programmer thought about the issue.
Benign Races

Example: Identify the set of digits in an array.
A: 4, 1, 0, 4, 3, 3, 4, 6, 1, 9, 1, 9, 6, 6, 6, 3, 4

```c
for (int i=0; i<10; ++i) {
    digits[i] = 0;
}
cilk_for (int i=0; i<N; ++i) {
    digits[A[i]] = 1;  // benign race
}
```

digits: 1 1 0 1 1 1 1 1 0 0 1

Caution: This code only works correctly if the hardware writes the array elements atomically — e.g., it races for byte values on some architectures.
Benign Races

Example: Identify the set of digits in an array.
A: 4, 1, 0, 4, 3, 3, 4, 6, 1, 9, 1, 9, 6, 6, 6, 3, 4

```c
for (int i=0; i<10; ++i) {
    digits[i] = 0;
}
cilk_for (int i=0; i<N; ++i) {
    digits[A[i]] = 1;  // benign race
}
```

digits: 1 1 0 1 1 1 1 1 0 0 1

Fake locks allow you to communicate to Cilkscreen that a race is intentional. (See Intel Cilk Plus Tools User's Guide.)
Implementation of Mutexes
Properties of Mutexes

- **Yielding/spinning**
  A yielding mutex returns control to the operating system when it blocks. A spinning mutex consumes processor cycles while blocked.

- **Reentrant/nonreentrant**
  A reentrant mutex allows a thread that is already holding a lock to acquire it again. A nonreentrant mutex deadlocks if the thread attempts to reacquire a mutex it already holds.

- **Fair/unfair**
  A fair mutex puts blocked threads on a FIFO queue, and the unlock operation unblocks the thread that has been waiting the longest. An unfair mutex lets any blocked thread go next.
Simple Spinning Mutex

**Spin.Mutex:**
```
  cmp 0, mutex ; Check if mutex is free
  je Get_Mutex
  pause ; x86 hack to unconfuse pipeline
  jmp Spin_MUTEX
```

**Get_Mutex:**
```
  mov 1, %eax
  xchg mutex, %eax ; Try to get mutex
  cmp 0, eax ; Test if successful
  jne Spin_MUTEX
```

**Critical_Section:**
```
  <critical-section code>
  mov 0, mutex ; Release mutex
```

**Key property:** xchg is an atomic exchange.
**Simple Yielding Mutex**

```assembly
Spin_Mutex:
  cmp 0, mutex ; Check if mutex is free
  je Get_Mutex
  call pthread_yield ; Yield quantum
  jmp Spin_Mutex
Get_Mutex:
  mov 1, %eax
  xchg mutex, %eax ; Try to get mutex
  cmp 0, eax ; Test if successful
  jne Spin_Mutex
Critical_Section:
  <critical-section code>
  mov 0, mutex ; Release mutex
```
Competitive Mutex

Competing goals:
- To claim mutex soon after it is released.
- To behave nicely and waste few cycles.

**Idea:** Spin for a while, and then yield.

**How long to spin?**
As long as a context switch takes. Then, you never wait longer than twice the optimal time.
- If the mutex is released while spinning, optimal.
- If the mutex is released after yield, $\leq 2 \times$ optimal.

**Randomized algorithm:** $e/(e-1)$-competitive.
Locking Anomalies

- Deadlock
- Convoying
- Contention
Deadlock

Holding more than one lock at a time can be dangerous:

Thread 1

1
lock(&A);
lock(&B);
   \textit{<critical section>}
unlock(&B);
unlock(&A);

Thread 2

2
lock(&B);
lock(&A);
   \textit{<critical section>}
unlock(&A);
unlock(&B);

The ultimate loss of performance!
Conditions for Deadlock

1. **Mutual exclusion** — Each thread claims exclusive control over the resources it holds.

2. **Nonpreemption** — Each thread does not release the resources it holds until it completes its use of them.

3. **Circular waiting** — A cycle of threads exists in which each thread is blocked waiting for resources held by the next thread in the cycle.
Dining Philosophers

C.A.R. Hoare

Edsger Dijkstra

Illustrative story of deadlock told by Charles Antony Richard Hoare based on an examination question by Edsger Dijkstra. The story has been embellished over the years by many retellers.
Dining Philosophers

Each of \( n \) philosophers needs the two chopsticks on either side of his/her plate to eat his/her noodles.

**Philosopher** \( i \)

```c
while (1) {
    think();
    lock(&chopstick[i].L);
    lock(&chopstick[(i+1)%n].L);
    eat();
    unlock(&chopstick[i].L);
    unlock(&chopstick[(i+1)%n].L);
}
```
Dining Philosophers

Starving

Each of $n$ philosophers needs the two chopsticks on either side of his/her plate to eat his/her noodles.

```c
while (1) {
    think();
    lock(&chopstick[i].L);
    lock(&chopstick[(i+1)%n].L);
    eat();
    unlock(&chopstick[i].L);
    unlock(&chopstick[(i+1)%n].L);
}
```

One day they all pick up their left chopsticks simultaneously.
Theorem. Suppose that we can linearly order the mutexes \( L_1 \preceq L_2 \preceq \cdots \preceq L_n \) so that whenever a thread holds a mutex \( L_i \) and attempts to lock another mutex \( L_j \), we have \( L_i \preceq L_j \). Then, no deadlock can occur.

Proof. Suppose that a cycle of waiting exists. Consider the thread in the cycle that holds the “largest” mutex \( L_{\text{max}} \) in the ordering, and suppose that it is waiting on a mutex \( L \) held by the next thread in the cycle. Then, we must have \( L_{\text{max}} \preceq L \). Contradiction. □
Philosopher $i$

```c
while (1) {
    think();
    lock(&chopstick[min(i,(i+1)%n)].L);
    lock(&chopstick[max(i,(i+1)%n)].L);
    eat();
    unlock(&chopstick[i].L);
    unlock(&chopstick[(i+1)%n].L);
}
```
void main() {
    cilk_spawn foo();
    lock(&L);
    cilk_sync;
    unlock(&L);
}

void foo() {
    lock(&L);
    unlock(&L);
}

- Don’t hold mutexes across `cilk_sync`’s!
- Hold mutexes only within strands.
- As always, try to avoid using mutexes (but that’s not always possible).
Locking Anomalies

• Deadlock
• Convoying
• Contention
Convoying

A lock **convoy** occurs when multiple threads of equal priority contend repeatedly for the same lock.

**Example: Performance bug in MIT–Cilk**
When random work–stealing, each thief grabs a mutex on its victim’s deque:

- If the victim’s deque is empty, the thief releases the mutex and tries again at random.
- If the victim’s deque contains work, the thief steals the topmost frame and then releases the mutex.

**Problem:** At start–up, most thieves quickly converge on the worker $P_1$ containing the initial strand, creating a convoy.
Performance Bug in MIT-Cilk

- Blue circle: busy worker
- Green circle: idle worker
- Blue square: successful steal in progress
- Blue arc: dependency from onto the lock on 's deque
Performance Bug in MIT–Cilk

: busy worker
: idle worker
: successful steal in progress
: dependency from onto the lock on ‘s deque
Performance Bug in MIT-Cilk

1: busy worker
2: idle worker
3: successful steal in progress
4: dependency from onto the lock on s's deque
5: dependency from onto the lock on s's deque
6: dependency from onto the lock on s's deque
Performance Bug in MIT–Cilk

1: busy worker
2: idle worker
3: successful steal in progress
4: dependency from
5: dependency on
6: 's deque

© 2013 Charles E. Leiserson and I–Ting Angelina Lee
Performance Bug in MIT–Cilk

1: busy worker  3: successful steal in progress
2: idle worker  4: dependency from onto
              the lock on ‘s deque
The work now gets distributed slowly as each thief serially obtains $P_1$’s mutex.
Solving the Convoying Problem

Use the nonblocking function \texttt{try\_lock()}, rather than \texttt{lock()}:

- \texttt{try\_lock()} attempts to acquire the mutex and returns a flag indicating whether it was successful, but it does not block on an unsuccessful attempt.

In Cilk Plus, when a thief fails to acquire a mutex, it simply tries to steal again at random, rather than blocking.
Locking Anomalies

• Deadlock
• Convoying
• Contention
int compute(const X& v);
int main() {
    const size_t n = 1000000;
    extern X myArray[n];
    // ...

    int result = 0;
    for (size_t i = 0; i < n; ++i) {
        result += compute(myArray[i]);
    }
    printf("The result is: %d\n", result);

    return 0;
}
```c
int compute(const X& v);
int main()
{
    const size_t n = 1000000;
    extern X myArray[n];
    // ...

    int result = 0;
    cilk_for (size_t i = 0; i < n; ++i) {
        result += compute(myArray[i]);
    }
    printf("The result is: %d\n", result);

    return 0;
}
```

**Work/span theory**
- $T_1(n) = \Theta(n)$
- $T_\infty(n) = \Theta(\lg n)$
- $T_P(n) = O(n/P + \lg n)$
Mutex Solution

#include <pthread.h>
int compute(const X& v);
int main() {
    const size_t n = 1000000;
    extern X myArray[n];
    // ...
    int result = 0;
    pthread_spinlock_t slock;
    pthread_spin_init(&slock, 0);
    cilk_for (size_t i = 0; i < n; ++i) {
        pthread_spin_lock(&slock);
        result += compute(myArray[i]);
        pthread_spin_unlock(&slock);
    }
    printf("The result is: \%d\n", result);
    return 0;
}
Scheduling with Mutexes

Greedy scheduler:

\[ T_p \leq T_1/P + T_\infty + B, \]

where \( B \) is the bondage — the total time of all critical sections.

This upper bound is weak, especially if many small mutexes each protect different critical regions. Little is known theoretically about lock contention.
Always write deterministic programs.
Always write deterministic programs, unless you can’t!