Synchronization: Implementing Barriers Promises + Futures

Chris Rossbach
CS378
Today

• Questions?
• Administrivia
  • Lab 2 due sooner than you’d like
• Material for the day
  • Barrier implementation
  • Promises & Futures

• Acknowledgements
  • Thanks to Gadi Taubenfield: I borrowed from some of his slides on barriers
Faux Quiz (answer any N, 5 min)

• How are promises and futures related? Since there is disagreement on the nomenclature, don’t worry about which is which—just describe what the different objects are and how they function.
Barriers
Prefix Sum

begin

a  b  c  d  e  f

end

a  a+b  a+b+c  a+b+c+d  a+b+c+d+e  a+b+c+d+e+f

time
Prefix Sum

begin

\[
\begin{array}{cccccc}
a & b & c & d & e & f \\
a & a+b & c & d & e & f \\
a & a+b & a+b+c & d & e & f \\
a & a+b & a+b+c & a+b+c+d & e & f \\
a & a+b & a+b+c & a+b+c+d & a+b+c+d+e & f \\
\end{array}
\]

time

end

\[
\begin{array}{cccccc}
a & a+b & a+b+c & a+b+c+d & a+b+c+d+e & a+b+c+d+e+f \\
\end{array}
\]
Parallel Prefix Sum

begin

\[ \begin{align*}
    a & \\
    b & \\
    c & \\
    d & \\
    e & \\
    f & \\
\end{align*} \]

\[ \begin{align*}
    a & \\
    a+b & \\
    b+c & \\
    c+d & \\
    d+e & \\
    e+f & \\
\end{align*} \]

time

\[ \begin{align*}
    a & \\
    a+b & \\
    a+b+c & \\
    a+b+c+d & \\
    b+c+d+e & \\
    c+d+e+f & \\
\end{align*} \]

time

end

\[ \begin{align*}
    a & \\
    a+b & \\
    a+b+c & \\
    a+b+c+d & \\
    a+b+c+d+e & \\
    a+b+c+d+e+f & \\
\end{align*} \]
Pthreads Parallel Prefix Sum

```c
int g_values[N] = { a, b, c, d, e, f }; 

void prefix_sum_thread(void * param) {

    int i;
    int id = *((int*)param);
    int stride = 0;

    for(stride=1; stride<=N/2; stride<<=1) {
        g_values[id+stride] += g_values[id];
    }
}

Will this work?
```
Pthreads Parallel Prefix Sum

```c
pthread_mutex_t g_locks[N] = { MUTEX_INITIALIZER, ...};
int g_values[N] = { a, b, c, d, e, f };;

void prefix_sum_thread(void * param) {
    int i;
    int id = *((int*)param);
    int stride = 0;

    for(stride=1; stride<=N/2; stride<<=1) {
        pthread_mutex_lock(&g_locks[id]);
        pthread_mutex_lock(&g_locks[id+stride]);
        g_values[id+stride] += g_values[id];
        pthread_mutex_unlock(&g_locks[id]);
        pthread_mutex_unlock(&g_locks[id+stride]);
    }
}
```
Parallel Prefix Sum

begin

barrier

barrier

end
Barrier Basics

- Coordination mechanism
  - Participants wait until all reach same point.
  - Once all reach it, all can pass.
  - Workhorse of BSP programming models

Fundamental primitive in many parallel models

Can you make a lock with a barrier?
Barriers: Goals

Desirable barrier properties:

• Low shared memory space complexity
• Low contention on shared objects
• Low shared memory references per process
• No need for shared memory initialization
• Symmetric: same amount of work for all
• Algorithm simplicity
• Simple basic primitive
• Minimal propagation time
• Reusability of the barrier (must!)
Barrier Building Blocks

• Conditions
• Semaphores
• Atomic Bit
• Atomic Register
• Fetch-and-increment register
• Test and set bits
• Read-Modify-Write register
Barrier with Semaphores
Barrier using Semaphores
Algorithm for N threads

```c
shared sem_t arrival = 1;    // sem_init(&arrival, NULL, 1)
sem_t departure = 0;         // sem_init(&departure, NULL, 0)
atomic int counter = 0;      // (gcc intrinsics are verbose)

sem_wait(arrival);
if (++counter < N)
  sem_post(arrival);
else
  sem_post(departure);
sem_wait(departure);
if (--counter > 0)
  sem_post(departure)
else
  sem_post(arrival)
```

**Phase I**
First N-1 threads post on arrival, wait on departure
Nth thread post on departure, releasing threads into phase II
(what is value of arrival?)

**Phase II**
First N-1 threads post on departure, last posts arrival
Semaphore Barrier Action Zone
N == 3

shared

sem_t arrival = 1;
sem_t departure = 0;
atomic int counter = 0;

sem_wait(arrival);
if(++counter < N)
    sem_post(arrival);
else
    sem_post(departure);
sem_wait(departure);
if(--counter > 0)
    sem_post(departure);
else
    sem_post(arrival)

Do we need two phases?

Still correct if counter is not atomic?
Barrier using Semaphores

Properties

• **Pros:**
  • Very Simple
  • Space complexity $O(1)$
  • Symmetric

• **Cons:**
  • Required a strong object
    • Requires some central manager
    • High contention on the semaphores
  • Propagation delay $O(n)$
Barriers based on counters
Counter Barrier Ingredients

Fetch-and-Increment register
- A shared register that supports a F&I operation:
- Input: register $r$
- Atomic operation:
  - $r$ is incremented by 1
  - the old value of $r$ is returned

```plaintext
function fetch-and-increment (r : register)
  orig_r := r;
  r:= r + 1;
  return (orig_r);
end-function
```

Await
- For brevity, we use the `await` macro
- Not an operation of an object
- This is just “spinning”

```plaintext
macro await (condition : boolean condition)
  repeat
    cond = eval(condition);
  until (cond)
end-macro
```
Simple Barrier Using an Atomic Counter

| shared     | counter: fetch and increment reg. – \( \{0, \ldots n\} \), initially = 0 |
| local      | local.go: a bit, initial value does not matter |
|            | local.counter: register |

1. local.go := go
2. local.counter := fetch-and-increment (counter)
3. if local.counter + 1 = n then
4.   counter := 0
5.   go := 1 - go
6. else await(local.go ≠ go)

Pros/Cons?

- There is high memory contention on go bit
- Reducing the contention:
  - Replace the go bit with \( n \) bits: \( go[1], \ldots, go[n] \)
  - Process \( p_i \) may spin only on the bit \( go[i] \)
A Local Spinning Counter Barrier
Program of a Thread $i$

| shared | counter: fetch and increment reg. – {0,..n}, initially = 0 |
|        | go[1..n]: array of atomic bits, initial values are immaterial |
| local  | local.go: a bit, initial value is immaterial |
|        | local.counter: register |

1. local.go := go[$i$]
2. local.counter := fetch-and-increment (counter)
3. if local.counter + 1 = $n$ then
   4. counter := 0
   5. for $j=1$ to $n$ { go[$j$] := 1 – go[$j$] }
6. else await(local.go ≠ go[$i$])
A Local Spinning Counter Barrier
Example Run for n=3 Threads

```
1 local.go := go[i]
2 local.counter := fetch-and-increment
3 if local.counter + 1 = n then
4     counter := 0
5     for j=1 to n { go[j] := 1 - go[j] }  # P1,P2 Busy wait
6 else await(local.go ≠ go[i])
```

Pros/Cons?
Does this actually reduce contention?
Comparison of counter-based Barriers

**Simple Barrier**

- Pros: 
- Cons: 

**Simple Barrier with go array**

- Pros: 
- Cons: 

Comparison of counter-based Barriers

<table>
<thead>
<tr>
<th>Simple Barrier</th>
<th>Simple Barrier with go array</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros:</strong></td>
<td><strong>Pros:</strong></td>
</tr>
<tr>
<td>• Very Simple</td>
<td>• Low contention on the go array</td>
</tr>
<tr>
<td>• Shared memory: O(log n) \textit{bits}</td>
<td>• In some models:</td>
</tr>
<tr>
<td>• Takes O(1) until last waiting process is awaken</td>
<td>• spinning is done on local memory</td>
</tr>
<tr>
<td><strong>Cons:</strong></td>
<td>• remote mem. ref.: O(1)</td>
</tr>
<tr>
<td>• High contention on the go bit</td>
<td><strong>Cons:</strong></td>
</tr>
<tr>
<td>• Contention on the counter register (*)</td>
<td>• Shared memory: O(n)</td>
</tr>
<tr>
<td></td>
<td>• Still contention on the counter register (*)</td>
</tr>
<tr>
<td></td>
<td>• Takes O(n) until last waiting process is awaken</td>
</tr>
</tbody>
</table>
Tree Barriers
A Tree-based Barrier

- Threads are organized in a binary tree
- Each node is owned by a predetermined thread
- Each thread waits until its 2 children arrive
  - combines results
  - passes them on to its parent
- Root learns that its 2 children have arrived \(\rightarrow\) tells children they can go
- The signal propagates down the tree until all the threads get the message
Assume $n = i 2^k - 1$

A Tree-based Barrier: indexing

2

Step 1: label numerically with depth-first traversal

Indexing starts from 2
Root $\rightarrow$ 1, doesn’t need wait objects
A Tree-based Barrier program of thread i

| shared | arrive[2..n]: array of atomic bits, initial values = 0 |
| shared | go[2..n]: array of atomic bits, initial values = 0 |

1. **if** i=1 **then**
   - **// root**
   - 
   - 
   - 
   - 

5. **else if** i ≤ (n-1)/2 **then**
   - **// internal node**
   - 
   - 
   - 
6. **await** (arrive[2i] = 1); arrive[2i] := 0
7. **await** (arrive[2i+1] = 1); arrive[2i+1] := 0
8. arrive[i] := 1
9. **await** (go[i] = 1); go[i] := 0
10. go[2i] = 1; go[2i+1] := 1
11. **else**
12. **// leaf**
13. arrive[i] := 1
14. **fi**

- **Root:**
  - Wait for arriving children
  - Tell children to go
- **Internal:**
  - Wait for arriving children
  - Tell parent about it
  - Wait for parent go signal
  - Tell children to go
- **Child:**
  - arrive
  - Wait for parent go signal
A Tree-based Barrier
Example Run for n=7 threads

Waiting for p_3 to arrive

Waiting for p_4 to arrive

Waiting for go[3]

Waiting for go[6]

Waiting for go[7]

At this point all non-root threads in some await(go) case
Tree Barrier Tradeoffs

• Pros:

• Cons:
Butterfly Barrier

- When would this be preferable?
Hardware Supported Barriers

• When would this be useful?
Barriers Summary

**Seen:**
- Semaphore-based barrier
- Simple barrier
  - Based on atomic fetch-and-increment counter
- Local spinning barrier
  - Based on atomic fetch-and-increment counter and go array
- Tree-based barrier

**Not seen:**
- Test-and-Set barriers
  - Based on test-and-test-and-set objects
  - One version without memory initialization
- See-Saw barrier
- Book has condition barriers
Asynchronous Programming
Events, Promises, and Futures
Programming Models for Concurrency

- Hardware execution model:
  - CPU(s) execute instructions sequentially

- Programming model dimensions:
  - How to specify computation
  - How to specify communication
  - How to specify coordination/control transfer

- Techniques/primitives
  - Message passing vs shared memory
  - Preemption vs Non-preemption

- Dimensions/techniques not always orthogonal

Futures & Promises touch all three dimension
Futures & Promises

• Values *that will eventually become available*

• Time-dependent states:
  • *Completed/determined*
    • Computation complete, value concrete
  • *Incomplete/undetermined*
    • Computation not complete yet

• Construct ( future X )
  • immediately returns value
  • concurrently executes X
Java Example

```java
static void runAsyncExample() {
    CompletableFuture cf = CompletableFuture.runAsync(() -> {
        assertTrue(Thread.currentThread().isDaemon());
        randomSleep();
    });
    assertFalse(cf.isDone());
    sleepEnough();
    assertTrue(cf.isDone());
}
```

- CompletableFuture is a container for Future object type
- cf is an instance
- runAsync() accepts
  - Lambda expression
  - Anonymous function
  - Functor
- runAsync() immediately returns a waitable object (cf)
- Where (on what thread) does the lambda expression run?
Futures and Promises: Why two kinds of objects?

Promise: “thing to be done”

Future: encapsulation (something to give caller)

Promise to do something in the future

future<int> f1 = async(foo1);
...
int result = f1.get();
Futures vs Promises

• **Future**: read-only reference to uncompleted value
• **Promise**: single-assignment variable that the future refers to
• Promises *complete* the future with:
  • Result with success/failure
  • Exception

<table>
<thead>
<tr>
<th>Language</th>
<th>Promise</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algol</td>
<td>Thunk</td>
<td>Address of async result</td>
</tr>
<tr>
<td>Java</td>
<td>Future&lt;T&gt;</td>
<td>CompletableFuture&lt;T&gt;</td>
</tr>
<tr>
<td>C#/.NET</td>
<td>TaskCompletionSource&lt;T&gt;</td>
<td>Task&lt;T&gt;</td>
</tr>
<tr>
<td>JavaScript</td>
<td>Deferred</td>
<td>Promise</td>
</tr>
<tr>
<td>C++</td>
<td>std::promise</td>
<td>std::future</td>
</tr>
</tbody>
</table>

Mnemonic: Promise to *do* something
Make a promise *for* the future
Putting Futures in Context
My unvarnished opinion

Futures:

• abstraction for concurrent work supported by
  • Compiler: abstractions are language-level objects
  • Runtime: scheduler, task queues, thread-pools are transparent

• Programming remains mostly imperative/sequential
  • Threads of control peppered with asynchronous/concurrent tasks

Compromise Programming Model between:

• Event-based
• Thread-based

Events vs. Threads!
GUI Programming

do {
    WaitForSomething();
    RespondToThing();
} until(forever);
GUI Programming

```
switch (message)
{
    // case WM_COMM
        // handle me
        break;
    case WM_PAINT:
        // draw our
        break;
    case WM_DESTROY:
        PostQuitMessage(0);
        break;
    default:
        // We do not want to handle this message so pass back to Windows
        // to handle it in a default way
        return DefWindowProc(hWnd, message, wParam, lParam);
}

// Step 2: Creating the Window
hwnd = CreateWindowEx(
    WS_EX_CLIENTEDGE,
    g_szClassName,
    "The title of my window",
    WS_OVERLAPPEDWINDOW,
    CW_USEDEFAULT, CW_USEDEFAULT, 240, 120,
    NULL, NULL, hInstance, NULL);

// Step 3: The Message Loop
while(GetMessage(&Msg, NULL, 0, 0) > 0)
{
    TranslateMessage(&Msg);
    DispatchMessage(&Msg);
}
```
GUI programming

```cpp
switch (message) {
    case WM_COMMAND:
        // handle menu select
        break;
    case WM_PAINT:
        // draw our window
        break;
    case WM_DESTROY:
        PostQuitMessage(0);
        break;
    default:
        // We do not want to handle this message so pass back to Win32
        // to handle it in a default way
        return DefWindowProc(hWnd, message, wParam, lParam);
}
```

```c
void OnMove() { ... }
void OnSize() { ... }
void OnPaint() { ... }
```

Over 1000 last time I checked!
GUI Programming Distilled

Pros
- Simple imperative programming
- Good fit for uni-processor

Cons
- Awkward/verbose
- Obscures available parallelism
```c
winmain(...) {
    while(true) {
        message = GetMessage();
        switch(message) {
            case WM_THIS: DoThis(); break;
            case WM_THAT: DoThat(); break;
            case WM_OTHERTHING: DoOtherThing(); break;
            case WM_DONE: return;
        }
    }
}
```

How can we parallelize this?
```c
winmain(...)
{
    while(true)
    {
        message = GetMessage();
        switch(message)
        {
            case WM_THIS: DoThis(); break;
            case WM_THAT: DoThat(); break;
            case WM_OTHERTHING: DoOtherThing(); break;
            case WM_DONE: return;
        }
    }
}
```
Parallel GUI Implementation 1

```c
void winmain() {
    pthread_create(&tids[i++], DoThisProc);
    pthread_create(&tids[i++], DoThatProc);
    pthread_create(&tids[i++], DoOtherThingProc);
    for(j=0; j<i; j++)
        pthread_join(&tids[j]);
}
```

```c
void DoThisProc() {
    while(true){
        if(ThisHasHappened){
            DoThis();
        }
    }
}
```

Pros/

Pros:
- Encapsulates parallel work

Cons:
- Obliterates original code structure
- How to assign handlers → CPUs?
- Load balance?!?
- Utilization
Parallel GUI Implementation 2

Pros:
• Preserves programming model
• Can recover some parallelism

Cons:
• Workers still have the same problem
• How to load balance?
• Shared mutable state a problem

Extremely difficult to solve without changing the whole programming model... so change it
Event-based Programming: Motivation

• Threads have a *lot* of down-sides:
  • Tuning parallelism for different environments
  • Load balancing/assignment brittle
  • Shared state requires locks →
    • Priority inversion
    • Deadlock
    • Incorrect synchronization
  • ...

• Events: *restructure programming model to have no threads!*
Event Programming Model Basics

• Programmer *only writes events*

• Event: an object queued for a module (think future/promise)

• Basic primitives
  • `create_event_queue(handler) → event_q`
  • `enqueue_event(event_q, event-object)`
    • Invokes handler (eventually)

• Scheduler decides which event to execute next
  • E.g. based on priority, CPU usage, etc.
Event-based programming

```c
switch (message)
{
    case WM_COMMAND:
        // handle menu selections etc.
        //break;
    case WM_PAINT:
        // draw our window - note: you must paint something here or not trap it!
        //break;
    case WM_DESTROY:
        PostQuitMessage(0);
    default:
        // We do not want to handle this message so pass back to Windows
        // to handle it in a default way
        return DefWindowProc(hWnd, message, wParam, lParam);
}
```

Is the problem solved?
Another Event-based Program

```
PROGRAM MyProgram {
  OnOpenFile() {
    char szFileName[BUFSIZE]
    InitFileName(szFileName);
    FILE file = ReadFileEx(szFileName);
    LoadFile(file);
    RedrawScreen();
  }
  OnPaint();
}
```

- Uses Other Handlers! (call OnPaint?)
- Burns CPU!
- Blocks!
No problem!
Just use more events/handlers, right?

```c
PROGRAM MyProgram {
    TASK.ReadFileAsync(name, callback) {
        ReadFileSync(name);
        Call(callback);
    }
    CALLBACK FinishOpeningFile() {
        LoadFile(file);
        RedrawScreen();
    }
    OnOpenFile() {
        FILE file;
        char szName[BUFSIZE]
        InitFileName(szName);
        EnqueueTask(ReadFileAsync(szName, FinishOpeningFile));
    }
    OnPaint();
}
```
Continuations, BTW

```cpp
PROGRAM MyProgram {
  OnOpenFile() {
    ReadFile(file, FinishOpeningFile);
  }
  OnFinishOpeningFile() {
    LoadFile(file, OnFinishLoadingFile);
  }
  OnFinishLoadingFile() {
    RedrawScreen();
  }
  OnPaint();
}
```
Stack-Ripping

PROGRAM MyProgram {
    TASK ReadFileAsync(name, callback) {
        ReadFileSync(name);
        Call(callback);
    }
    CALLBACK FinishOpeningFile() {
        LoadFile(file);
        RedrawScreen();
    }
    OnOpenFile() {
        FILE file;
        char szName[BUFSIZE];
        InitFileName(szName);
        EnqueueTask(ReadFileAsync(szName, FinishOpeningFile));
    }
    OnPaint();
}
Threads vs Events

• Thread Pros
  • Overlap I/O and computation
  • While looking sequential
  • Intermediate state on stack
  • Control flow naturally expressed

• Thread Cons
  • Synchronization required
  • Overflowable stack
  • Stack memory pressure

• Event Pros
  • Easier to create well-conditioned system
  • Easier to express dynamic change in level of parallelism

• Event Cons
  • Difficult to program
  • Control flow between callbacks obscure
  • When to deallocate memory
  • Incomplete language/tool/debugger support
  • Difficult to exploit concurrent hardware

Language-level Futures: the sweet spot?
Thread Pool Implementation

```c
void ThreadPool::StartThreads(
  __in UINT uiThreads,
  __in BOOL bWaitAllThreadsAlive
) {

  Lock();
  if(uiThreads != 0 && m_vhThreadDescs.size() < m_uiTargetSize)
    ResetEvent(m_hAllThreadsAlive);
  while(m_vhThreadDescs.size() < m_uiTargetSize) {
    for(UINT i=0; i<uiThreads; i++) {
      THREADDESC* pDesc = new THREADDESC(this);
      HANDLE * phThread = &pDesc->hThread;
      *phThread = CreateThread(NULL, 0, ThreadPoolProc, pDesc, 0, NULL);
      m_vhAvailable.push_back(*phThread);
      m_vhThreadDescs[*phThread] = pDesc;
    }
  }
  m_uiThreads = (UINT)m_vhThreadDescs.size();
  Unlock();
  if(bWaitAllThreadsAlive)
    WaitThreadsAlive();
```

Cool project idea: build a thread pool!
Thread Pool Implementation

```c
DWORD ThreadPool::ThreadPoolProc(
    __in THREADDESC * pDesc
)
{
    HANDLE hThread = pDesc->hThread;
    HANDLE hStartEvent = pDesc->hStartEvent;
    HANDLE hRuntimeTerminate = pTask::Runtime::GetRuntimeTerminateEvent();
    HANDLE vEvents[] = { hStartEvent, hRuntimeTerminate };;

    NotifyThreadAlive(hThread);
    while(!(pDesc->bTerminate)) {
        DWORD dwWait = WaitForMultipleObjects(dwEvents, vEvents, FALSE, INFINITE);
        pDesc->Lock();
        pDesc->bTerminate = bTerminate;
        if(pDesc->bRoutineValid && !pDesc->bTerminate) {
            LPTHREAD_START_ROUTINE lpRoutine = pDesc->lpRoutine;
            LPVOID lpParameter = pDesc->lpParameter;
            pDesc->bActive = TRUE;
            pDesc->Unlock();
            dwResult = (*(lpRoutine))(lpParameter);
            pDesc->Lock();
            pDesc->bActive = FALSE;
            pDesc->bRoutineValid = FALSE;
        }
        pDesc->Unlock();
        Lock();
        m_vhInFlight.erase(pDesc->hThread);
        if(!pDesc->bTerminate)
            m_vhAvailable.push_back(pDesc->hThread);
        Unlock();
    }
    NotifyThreadExit(hThread);
    return dwResult;
}```
ThreadPool Implementation

/// <summary>
/// Starts a thread: if a previous call to RequestThread was made with
/// the bStartThread parameter set to false, this API signals the thread
/// to begin. Otherwise, the call has no effect (returns FALSE). </summary>
///
/// <remarks>
/// crossbac, 8/29/2013. </remarks>
///
/// <param name="hThread"> The thread. </param>
///
/// <returns> true if it succeeds, false if it fails. </returns>
///
BOOL ThreadPool::SignalThread(
    in HANDLE hThread
)
{
    Lock();
    BOOL bResult = FALSE;
    std::set<HANDLE>::iterator si = m_vhWaitingStartSignal.find(hThread);
    if (si == m_vhWaitingStartSignal.end()) {
        m_vhWaitingStartSignal.erase(hThread);
        THREADDESC * pDesc = m_vhThreadDescs[hThread];
        HANDLE hEvent = pDesc->hStartEvent;
        SetEvent(hEvent);
        bResult = TRUE;
    }
    Unlock();
    return bResult;
}
Redux: Futures in Context

Futures:
• *abstraction* for concurrent work supported by
  • Compiler: abstractions are *language-level objects*
  • Runtime: scheduler, task queues, thread-pools are *transparent*
• Programming remains *mostly* imperative
  • Threads of control peppered with asynchronous/concurrent tasks

Compromise Model:
• Event-based programming
• Thread-based programming
Currently: 2nd renaissance IMHO

```java
static void runAsyncExample() {
    CompletableFuture cf = CompletableFuture.runAsync(() -> {
        assertTrue(Thread.currentThread().isDaemon());
        randomSleep();
    });
    assertFalse(cf.isDone());
    sleepEnough();
    assertTrue(cf.isDone());
}
```
Questions?