Thinking a Bit About Decomposition

CIS 322
Feb 28, 2017

My time has somewhat blown up, so todays lecture will likely be a little thin.

From the Reports

- I'm having trouble starting assignments early, but when I do it really improves what I get from the lectures. I am motivated to try to start work on assignments as early as possible now
- Other Assign 8 questions: Can we get a demo or a link to an example on how to implement the functionality described in step 3? Specifically the part where the user can click on the row of a table? Step 5; Check the argument that identifies the request to be approved? How should we be storing these arguments in the db? Can you save a db query in the db? P.s. Can we get some mock-ups?

I would like to do the mockups as an exercise during class. I'm also hoping we can do a quick demo on passing some arguments.

Why have functions?

Why do we have functions? What is the value in this as a programmer or designer or architect?

Simplifies Complexity

- Who has taken 314? Do you want to write everything in assembly?
- Code reuse and sharing
- Developer doesn’t generally need to know the how
- Reasoning about larger units of computation
- Allows implementation to be delegated

When you use the python standard library, you likely choose functions to solve particular problems for your program…

We had a json parsing problem so we used functions from a json library. We had a database to interact with so we used functions from an sql library. How these tasks are actually done aren’t really what we care about, we care about getting the functionality into our system.

Why have programs?

- Do you want to write everything yourself?
- Code reuse and sharing
- User doesn’t generally need to know the how
- Reasoning about larger units of computation
- Allows implementation to be delegated

One way to think about programs is as another unit of computational segmentation.

With functions, the location of computation is known both physically and logically. Working at the level of programs brings in challenges of locality and how to move data between programs.

Conceptual Computation

- How do you think about how the computer executes your code?
- How do you think about your computer running multiple applications simultaneously?
- How do you think about network applications like MMORPGs and Facebook?

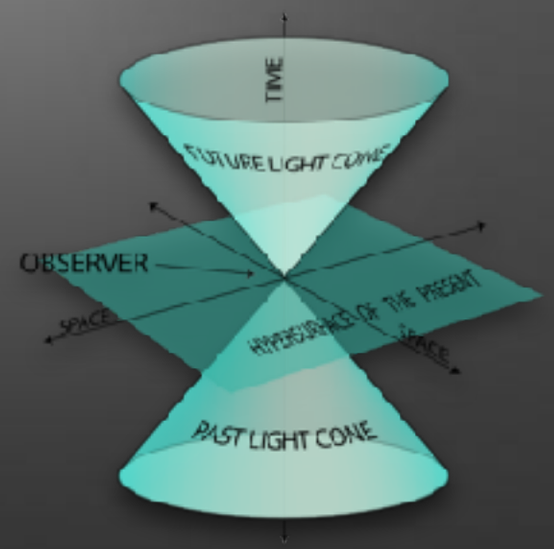
What is the model that you use to understand and think about what the computer does with your program?

What is the model that you use to understand and think about what the computer is doing when running multiple programs? I come from a time when a computer could only do things one at a time… no one uses a computer like that anymore.

What is the model you use to understand network applications? How do you think those work?

Computation in Physical Space and Time

- Computation is done physically via transistor state changes
- Transistors exist in both space and time
- At 3 GHz, light travels roughly 10 cm in a vacuum per clock tick.



https://commons.wikimedia.org/wiki/File:World_line.svg

The problem of physical causality leads to multi-layer caching and a bunch of other annoying problems for coordinating behavior across a computer. Causality and the light cone puts serious restrictions on what can be accounted for within a computation.

Due to the high clock speeds of modern processors, I really like to think about how my poor understanding of relativity can be useful in thinking about computation and system design.

Computation in Logical Space and Time

- Function - a sequence of instructions to execute
- Program - a sequence of instructions to execute that an operating system can start (contains at least one thread)
- Thread - a sequence of executing instructions
- Process - one or more threads sharing memory (usually one program)
- System - one or more collaborating processes
- Distributed system - one or more collaborating systems

Reasoning about execution within a single thread is relatively easy and based on our understanding of serial execution, branching, and loops. A thread guarantees that only one thing happens at a time from the threads perspective. My understanding is that our CS program sticks mostly to programs with a single thread of execution running on a single host.

Multithreading gets tricky quickly. When two threads share memory there is a chance that the threads will try to read/write to the same locations. Without coordination this can cause all kinds of trouble (data corruption, time of check time of use issues). On the other hand, shared memory is probably the easiest way to transfer data between threads.

Processes, under most operating systems, have strongly separated/protected memory. This is a really good thing and central to a lot of security assumptions we make about our computers. Internally, shared memory is often used to communicate between threads but IPC mechanisms could be used.

For systems, InterProcess Communication (IPC) is used to move data between processes (OS managed shared memory, pipes, sockets, files, signals, etc). Since memory is partitioned, moving data requires some kind of interface to move between boundaries.

Distributed systems refer to systems made of networked computers (systems). IPC in these settings requires traversing network links in addition to the OS enforced memory boundaries. Remote Procedure Calls (RPC) and other network based protocols are used to communicate across systems.

Shared Memory Problem

- Example is in JAVA since python GIL serializes all python programs
- Program counts to 2 million?

```
128-223-223-92-wireless:Desktop dellsmor$ javac T.java
128-223-223-92-wireless:Desktop dellsmor$ java T
1318878
128-223-223-92-wireless:Desktop dellsmor$ java T
1817207
128-223-223-92-wireless:Desktop dellsmor$ java T
1818469
128-223-223-92-wireless:Desktop dellsmor$ java T
1822139
128-223-223-92-wireless:Desktop dellsmor$ java T
1538599
128-223-223-92-wireless:Desktop dellsmor$ java T
1811568
```

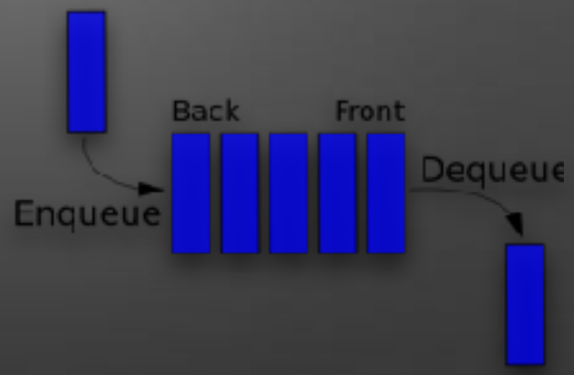
```
class T extends Thread {
    static int cntr=0;

    public void run() {
        for(int i=0;i<1000000;i++) {
            cntr++;
        }
    }

    public static void main(String[] argv) {
        Thread t1 = new T();
        Thread t2 = new T();
        t1.start();
        t2.start();
        try {
            t1.join();
            t2.join();
        } catch(Exception e) { }
        System.out.println(cntr);
    }
}
```

This is the fundamental problem with having concurrent execution… Each thread has a sensible behavior, but interactions between threads cause trouble. IPC mechanisms that use shared memory and files suffer some the same problem. There is no longer total order when threads execute concurrently.

In your operating systems class or in the parallel computing class you should learn about locks, mutexes, and other tools to enable the safe use of shared memory.

Sockets, Pipes, Queues

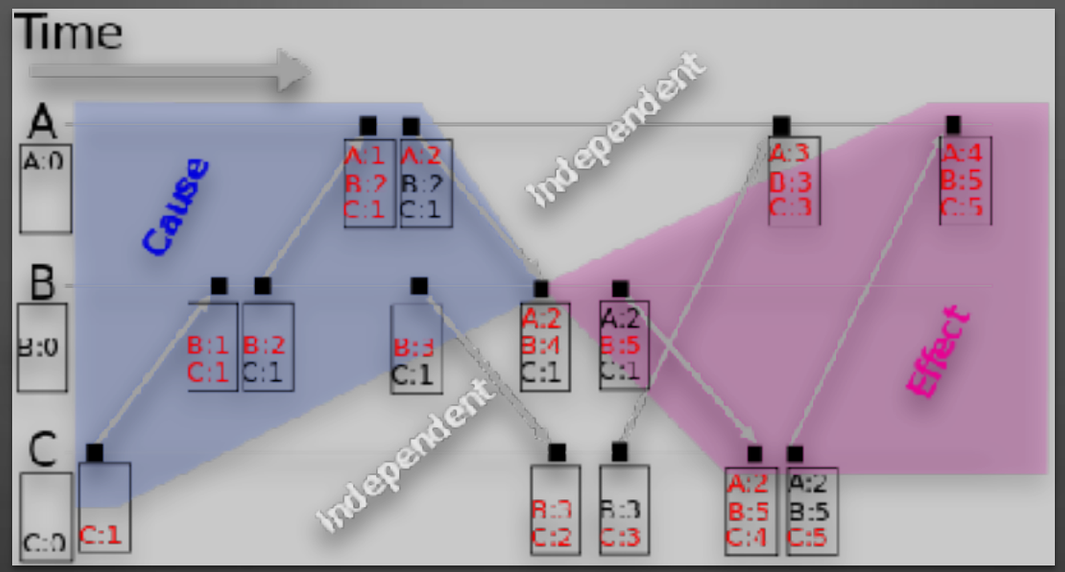
- Avoid memory corruption problems by making copies
- Provider must push data to the queue
- Consumer must pull data from the queue
- Queued data leads to inconsistency in state between provider and consumer



https://commons.wikimedia.org/wiki/File:Data_Queue.svg

Sockets and pipes are effectively data queues… without the packetization.

These all make copies of the data. This is safer than manipulating the data in another thread directly but is slow… and there is no way to know when/if the data will ever be read by the recipient. There are also changes with shared state, the provider likely thinks what was written happened while the consumer does not.



https://commons.wikimedia.org/wiki/File:Vector_Clock.svg

Vector Clocks

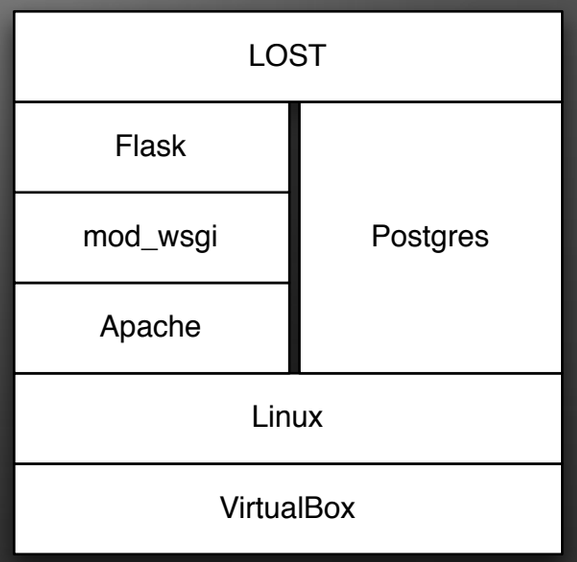
Causality based time…

One way to think about time in a distributed computation is by keeping track of the causality chain. This provides information on what past events caused the currently generated event; providing a partial order to the distributed computation.

Leslie Lamport gets credited for the work on logical clocks.

Our Tech Stack

- What problem does each part solve?
  - VM/Linux
  - Apache
  - mod_wsgi
  - Postgres
  - Flask
  - LOST
- What Mechanisms are used for communication?



VM/Linux - Operating system to coordinate processes, manage memory, handle low level memory
Apache - Web server to handle HTTP protocol issues
mod_wsgi - WSGI interface to Apache that allows currency and coupling Apache to python
Flask - session support and other HTTP helper capabilities
Postgres - data persistence, coordination between concurrent operations
LOST - Desired user functionality

Get Arguments…

Were going to drop out to look at some code for working with get args.

Assignment 8 Mockups

What are mockups? Why do we like them? Who wants to run the chalkboard?