University of British Columbia

CPSC 416
Introduction to Distributed Systems

Instructor: Dr. Alan Wagner

Email: wagner@cs.ubc.ca

My Research Interests

❑ Parallel Computing

❑ Interconnection Networks

❑ Distributed Computing

❑ Networking

Information and Organization

❑ Instructor: Dr. Alan Wagner

    ❑ Email: wagner at cs.ubc.ca

    ❑ Office Hours: M, W: 3:00-4:00pm

   ❑ Office RM 321 ICICS-CS building

    ❑ Class Website

          piazza.com/ubc.ca/winterterm22013/cpsc416

    ❑ TAs
- Imran Ahmed
- Nodir Kodirov

Resources

❑ <u>Reference Material:</u> (a) The Art of Multiprocessor Programming Maurice Herlihy Nir Shavit. (b) A.D. Kshemkalyani, M. Singhal, Distributed Computing: Principles, Algorithms, and Systems, (c) Concurrency J. Magee, J. Kramer, (d) Distributed Systems: Concpets and Design Coulouris, Dollimore, Kindberg, Blair

❑ Papers and Online Material
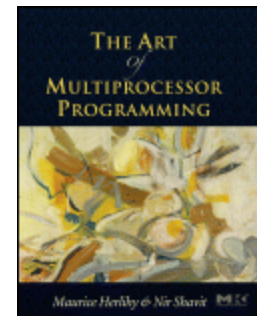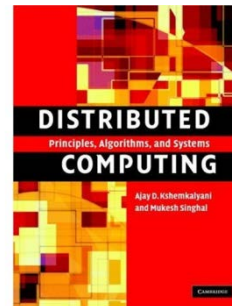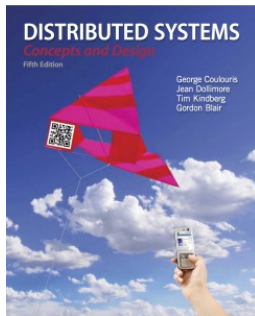
Evaluation

- ❑ Midterm   (20%)

- ❑ Final exam   (60%)

- ❑ Assignments   (20%)

Collaboration and Plagiarism

❑ http://www.cs.ubc.ca/about/policies/collaboration.shtml

❑ Serious!

❑ Professional Ethics – Integrity

❑ Giving others proper credit

Syllabus

Outcome

- ❑ Experience with Distributed Systems

- ❑ Understanding the Issues

- ❑ Understanding of Approaches, what is difficult and what is easy.

- ❑ Terminology and state-of-the-art

What is distributed computing?

Distributed computing is a field of computer science that studies distributed systems. A **distributed system** is a software system in which components located on networked computers communicate and coordinate their actions by passing messages.[1] The components interact with each other in order to achieve a common goal.

What is distributed Computing?

❑ Tannenbaum

"A distributed system is a collection of computers that appear to its users as a single coherent system"

❑ Similar to the Wikipedia one

"A distributed system is a collection of independent computers that are used to jointly perform a single task or to provide a single service"

What is distributed Computing?

❑ A distributed system is characterized by:

- multiple machines
- connected by some network
- co-operating on some task

  - viewed as a single system (Tannenbaum)
  - distributed by nature (email, web)
  - Provide a service (google)

Alternative Definition

...A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable...

Leslie Lamport

Challenges

Three significant characteristics of distributed systems are: concurrency of components, lack of a global clock, and independent failure of components.[1] An important goal and challenge of distributed systems is location transparency

Challenges

- ☐ Transparency

- ☐ Scalability

- ☐ Dependability

- ☐ Performance

- ☐ Flexibility

Challenges

- ❑ Availability
- ❑ Performance
- ❑ Scalability

- ❑ Concurrency
- ❑ Communication
- ❑ Synchronization

- ❑ Security
- ❑ Mobility

Transparency

- ❑ Access (identical access to remote and local resources -- dropbox)
- ❑ Location (content distribution networks)
- ❑ Migration (mobile phones)
- ❑ Replication (backup services)
- ❑ Failure (RAIDed disks)
- ❑ Concurrency (unaware that resources are shared)

The Eight Fallacies of Distributed Computing

Peter Deutsch

Essentially everyone, when they first build a distributed application, makes the following eight assumptions. All prove to be false in the long run and all cause *big* trouble and *painful* learning experiences.

1. The network is reliable
2. Latency is zero
3. Bandwidth is infinite
4. The network is secure
5. Topology doesn't change
6. There is one administrator
7. Transport cost is zero
8. The network is homogeneous
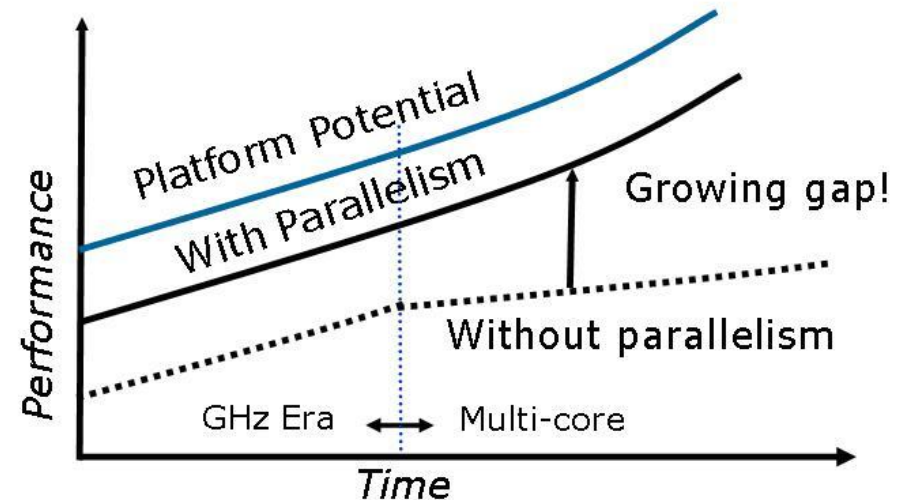
Current Technologies
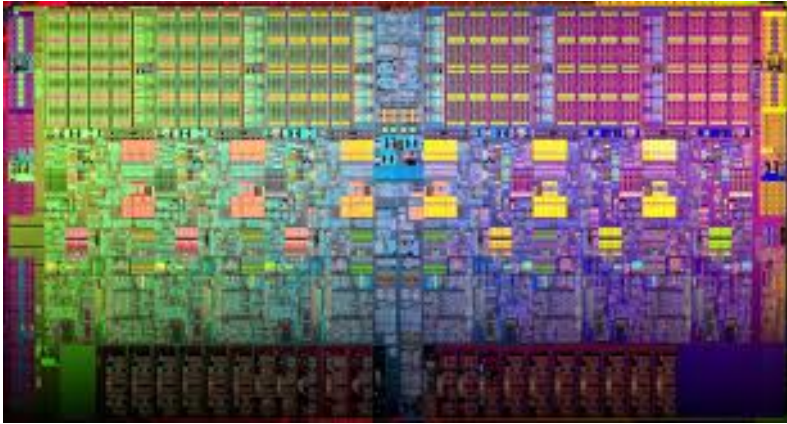
multicore

cloud/cluster computing

Paradigm Shift in Computing – 2003

❑ It's becoming harder to speed up processors
→ **"Multi-core" is the solution**

❑ Software must be built to utilize multiple cores in a machine and across machines

     → **Need for parallelism**

Need for Parallelism

Intel Technical Report

Distributed vs Parallel

Distributed systems are groups of networked computers, which have the same goal for their work. The terms "concurrent computing", "parallel computing", and "distributed computing" have a lot of overlap, and no clear distinction exists between them.[14]
In parallel computing, all processors may have access to a shared memory to exchange information between processors.[17]