Scalable Internet Systems and Services - Course Overview

Kai Shen
Dept. of Computer Science, University of Rochester

Course Goal - Scalable Internet Systems and Services

- What are the problems? - Real-world examples/applications.
- What does it take? - Overview of fundamental technologies involved.
- What have people done? - Read some papers.
- How to make it better? - Discuss it and work on it.
- What differentiates this course? - performance measurements.

Topic #1 - Basic components

- Online information gathering or crawling
  - efficiency - how fast you can gather them?
  - control your pace - hits/minute and bandwidth.
  - do it at opportune time (at night).
- A review on important protocols: TCP, UDP, HTTP, etc.
  - understand their performance - TCP slow start, HTTP persistency
- Designs of concurrent servers
  - multi-processing
  - multi-threading
  - event-driven

Topic #2 - Performance measurements

- Basics:
  - measure the TCP/UDP roundtrip latency
  - clock precision
  - prediction of the future based on past measurements
  - measurement of non-repetitive events
- Performance measurements of concurrent servers
  - metrics - throughput, mean response time, or a combination.
  - workload design - choosing requests, request arrival interval (constant, Poison process, trace-driven)
  - efficiency of the testing client
**Topic #3 – Applications/Services**
- Read-only services
  - Web servers
  - Keyword search engines – Google
- Services that allow updates
  - Hashtable – query cache
  - Web email – hotmail
- Database-driven apps
  - Online discussion groups - slashdot
  - Auctions - eBay

**Topic #4 – Building Scalable Systems**
- Rely on supercomputers
- Cluster-based systems
  - using commodity servers connected with commodity network technologies
  - load balancing
  - fail-over
  - replication and consistency maintenance
  - advanced resource management and service differentiation

**Topic #5 – Some Wide-area Systems**
- Web caching
  - scalability and performance
- Content distribution networks
  - scalability and performance
- Performance monitoring and network management

**Topic #6 – Peer-to-peer Systems**
- Application domains
  - Internet overlay networks
  - Mobile ad hoc networks
- Key advantages:
  - scalability and robustness
- Applications/protocols
  - routing
  - keyword searching
  - storage system
  - load balancing
  - replication management
  - trust and reputation
  - anti-censorship (Mixmaster remailer)
Topic #7 - Network security

- Distributed DoS attack
  - performance and scalability

A few words on the meetings

- Required/recommended readings
- I will present some prepared materials
- Interaction, please!
- Meeting notes will be posted on the Web afterwards

A few words on assignments and the final project

- Some possible assignments
  - crawling and information gathering
  - performance measurement of a Web server
  - building a keyword search system
  - building a clustered system with load balancing
- Final project
  - you can choose from a number of options or do something of your own
- C/C++ networking and Perl
- No written homeworks/exams

Test your brain and have some fun!

- Give a simple C expression to test whether a number $N$ is a power of 2.
  - $((N-1) \& N) == 0$
- You have a balance and $N$ balls, one of them is a defect (lighter than others), how many times you will need to use the balance to find the defect.
  - $N = 9$
  - $N = 26$
  - $\log_3 N$
- What if you don't know if the defect is lighter or heavier? $N=13$