Distributed Systems Project, Spring 2014

Jussi Kangasharju
Course Outline

3 exercises to look at distributed systems in practice

Exercises mostly programming

2 individual exercises, 1 group exercise

Groups of up to 3 people allowed

Group work not mandatory, but recommended
Course Schedule

14.1. Start of first and second exercises (individual)
16.1., 21.1., and 23.1. Q&A for first two exercises

28.1. Deadline for first and second exercise

28.1. Start of third exercise (group)
30.1., 4.2., 6.2., 11.2., 13.2., 18.2., and 20.2. Q&A for third exercise

7.3. Deadline for third exercise
People

Jussi Kangasharju
  Office hour: Tue 13-14 or ask for appointment by email

Liang Wang
  Office hour: During meetings or ask appointment by email

Twitter: #tktl_dsp (also visible on course page)
Assignments

Distributed algorithms
   Individual assignments about algorithms

Hadoop
   Use Hadoop to analyze a data set

Overlay networks
   Design, analyze, and implement an overlay network

Details for assignment 3 presented later
Grading

Each assignment graded on scale 1-5
Must get at least 1 in every assignment

Same grade for all members of group

Overall grade is weighted average of assignment grades

  Assignments 1 and 2: Weight 1
  Assignment 3: Weight 2
Assignment 1: Algorithms

Link to assignment will be posted to course website
Individual Assignments on Distributed Algorithms

1. Lamport clocks
2. Vector clocks
3. Bully election algorithm
4. Gossiping
5. Token passing in ring

Simple programs communicating over the network
Select assignment: (student ID % 5) + 1
General Idea

Multiple programs on different machines

Everybody knows everybody

Programs communicate to implement a given algorithm

Key points: Network communication, correct algorithm
House Rules

Configuration file for nodes and ports
Format:

<ID> <IP/HOST> <PORT>
Command line argument indicates what is client’s ID
File has an arbitrary number of lines

Must conform to specified output format
Deviation results in a reduced grade
Programs must be runnable on Ukko cluster
Assignment 2: Hadoop

Link to assignment will be posted to course website
Analyzing Large Data Sets

Recall MapReduce from Distributed Systems course

Hadoop = Open source implementation of MapReduce (and several other things)

Assignment goal: Get familiar with Hadoop and MapReduce

Task: Analyze a large data set
Practical Matters

Provide your user ID to us

Either today or by email to Liang

We create work directories for everyone
Next steps

Q&A session on 16.1., 21.1., and 23.1.

Deadline for returning January 28th at 10:00
Same deadline for both assignments

Return to Liang.Wang@cs.helsinki.fi
See assignment sheet for instructions