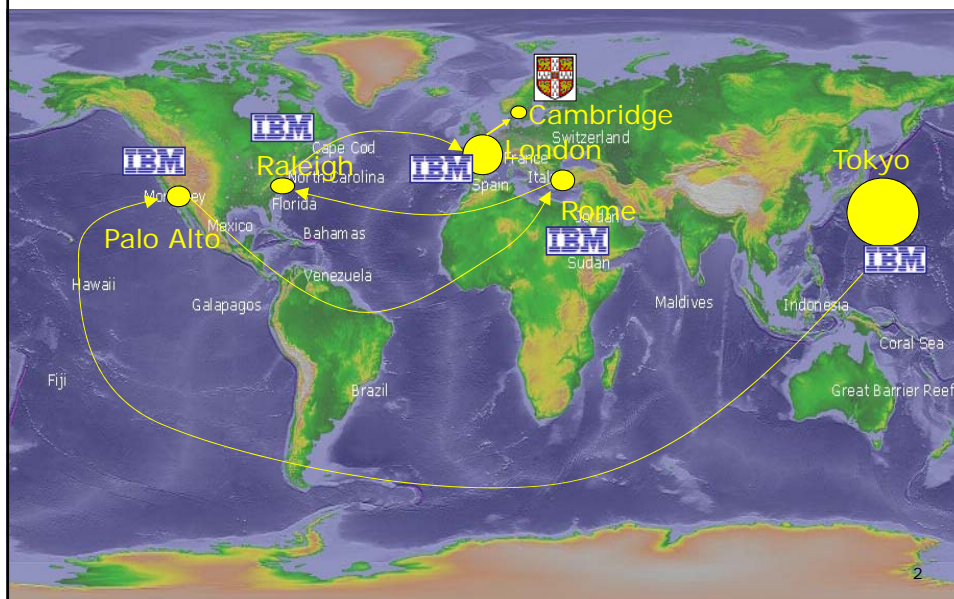# Data Centric Networking and Systems (DCNS)

Session 1: Introduction to R202

Eiko Yoneki

Systems Research Group
University of Cambridge Computer Laboratory

My Trajectory

My Research Interests

- **Distributed Systems, Networking and Database**
  - Delay Tolerant Networking
  - Data Driven Complex Networks (Epidemiology)
  - Declarative Networking

- **Large-Scale Graph Processing (e.g. 1 Billion Vertices)**
  - Graph Database
  - DSL for Graph Processing
  - Parallel Data-Flow Programming
  - Graph Algorithms/Parallelism

3

Introduction to R202

- **Welcome to R202**

- **First introduce yourselves**
  - Tell about yourself
    - Your name and where you studied before ACS
    - What modules have you taken in Michaelmas term
    - What is your research interests
    - What is your ACS project
    - Why are you interested in R202
    - Do you want to continue research career after ACS?

4

Data Centric Networking and Systems

- Shift of Communication Paradigm
  - From end-to-end to data centric
  - Data as communication token

- Integration of complex data processing with networking
  - A key vision for future computing

- Different aspects of data centric approaches

5

R202 Course Objectives

- Understand key concepts of data centric approaches

- Understand how to build distributed systems in data driven approach

- Research skills

  - Read systems/networking papers
  - Establish basic research domain knowledge in data centric networking and systems
  - Obtain your view of research area for thinking forward

http://www.cl.cam.ac.uk/~ey204/teaching/ACS/R202_2012_2013/

6

Course Structure

- Reading Club
    - ~3-4 Paper review presentations and discussion per session (~=25 minutes each)
    - Each of you will present about 2 reviews during the course
        - You can use your own laptop or USB key with your PowerPoint or PDF file
        - Revised (if necessary) presentation slides needs to be emailed on the following day
    - *Review_Log*: minimum 1 per session
        - Email me by noon on Monday
        - Prepare a couple of questions
    - Active participation to review discussion!

7

Review_Log

Paper Review Log: Session 2 (2013/01/29)

Name and (crsid):

Paper Title and Authors

1. Paper Summary (<100 words)
Describe a brief summary (extract essentials).

2. List other papers you read or skimmed for understanding this paper

3. Punch-line of the Paper (<250 words):
What is the significant contribution?
What is the difference from the existing work?

4. What didn't you understand?
Crystallise what you did not get from the paper and describe your potential questions to the presentation/discussion.

5. Any major criticism to the authors?
Any criticism and suggestions to the authors?

8

4

Review_Log

1. Paper summary (<100 words)
   - Describe a brief summary
   - Aim: you have read and extracted essentials

2. List other papers you read or skimmed for understanding this paper

3. Punch-line of the Paper (<250 words)
   - What is the significant contribution?
   - What is the difference from the existing works?

   - What is the novel idea?
   - What is required to complete the work?

9

Review_Log

5. What didn't you understand? (<100 words)
   - Crystallise what you did not get from the paper and describe your potential questions to the presentation/discussion

6. Any major criticism to the authors?

10

Course Work: Reports

- **Review report** on full length of paper (1800 words ~3 pages)
  - Describe the contribution of paper in depth with criticism
  - Crystallise the significant novelty in contrast to the other related work
  - Suggestion for future work
- **Survey report** on sub-topic in data centric networking (<2000 words)
  - Pick up to 5 papers as core papers in your survey scope
  - Read them and expand your reading through related work
  - Comprehend your view and finish as your survey paper
- **Hand in reports**
  - Report 1: February 19 noon
  - Report 2: March 11 noon
  - No particular order

11

Course Work: Reports

- **Report on project study** and exploration of a prototype (<2500 words)

  - What is the significance of the project in the research domain?
  - Compare with the similar and succeeding projects
  - Demonstrate the project by exploring its prototype
  - Project selection by February 4, 2012
    - Title and brief description (100 words) by email
  - Project presentation on March 12, 2012
  - Final report on the project study on March 26, 2012

12

Study of Open Source Project

- Open Source project normally comes with new proposal of system/networking architecture
- Understand the prototype of proposed architecture, algorithms, and systems through running an actual prototype
- Any additional work
  - Writing applications
  - Extending prototype to another platform
  - Benchmarking using online large dataset
- Present/explain how prototype runs
- Some projects are rather large and may require extensive environment and time; make sure you are able to complete this assignment

13

Candidates of Open Source Project

http://www.cl.cam.ac.uk/~ey204/teaching/ACS/R202_2012_2013/opensource_projects.html

- List is not exhausted and discuss with me if you find more interesting one for you
- Expectation of workload on open source project study is about intensive 3 full days work except writing up report
- One approach: pick one in the session topic, which you are interested in along your survey report

- Apache Giraph, Naiad, GraphLab, CIEL…

14

Important Dates

- **February 4 (Monday)**
  - Project selection

- **February 19 (Tuesday)**
  - Review report or Survey report

- **March 11 (Monday)**
  - Review report or Survey report

- **March 26 (Tuesday)**
  - Open source project study report

15

Assessment

- The final grade for the course will be provided as a letter grade or percentage and the assessment will consist of two parts:

- 25%: for a reading club (presentation, participation and *review_log*)

- 75%: for the three reports
  - 20%: Intensive review report
  - 25%: Survey report
  - 30%: Project study

16

Topic Areas

- *Session 1:* Introduction to Data Centric Networking and Systems
- *Session 2:* Content-Centric Networking (CCN) and Content Distribution Networks (CDN)
- *Session 3:* Programming in Data Centric Environment
- *Session 4:* Map/Reduce Hands-on Tutorial with CIEL
- *Session 5:* Stream Data Processing and Data/Query Model + Guest lecture
- *Session 6:* Graph Structured Data: Network, Storage, and Query Processing
- *Session 7:* Network holds Data in Delay Tolerant Networks (DTN)
- *Session 8:* Project study presentation

17

Reading Papers

- Scope of DCNS is wide

- …includes distributed systems, OS, networking, middleware, programming language, database…

- Understand where DCN functionality resides and how whole system works

- Type of papers
  - Building a real networking component and system
  - Proposing algorithm/mechanism on routing or architecture design
  - New idea (w/ or w/o simulation)

18

Critical Thinking

- Reading a research paper is not like reading a text book

- But the most important one is that the paper is not necessary the *truth*
  - there is no right and wrong, just good and bad
  - There are inherently subjective qualities...but you can't get away with just your opinion: must argue

- Critical thinking is the skill of marrying subjective and objective judgment of a piece of work

First Let's Argue for...

- What is the problem?

- What is important?

- Why isn't it solved in previous work?
  - Why CCN? Current Internet naming is not good enough?

- What is the approach?
  - MapReduce for Big data

- Why is this novel/innovative?
  - MapReduce can solve all big data?

And Now against...

- **Problem is overstated (or oversold)**
  - CCN – does flat name scale?

- **Problem does not exist**

- **Approach is broken**
  - Functional programming language too difficult for regular programmers?

- **Solution is insufficient**
  - Only works when data rate is lower than ...

- **Evaluation is unfair/biased**
  - ZebraNet only uses 5 nodes for evaluation...can it be applied on the general case?

S. Hand'10   21

So Which is RIGHT Answer?

- **There isn't one!**
  - Most of arguments are mostly correct...

- **Your judge on what is valuable on topic**

- **In this course, we'll be reviewing a selection of +16 papers (3-4 per week)**
  - Cover 5 different aspects of data centric networking
  - All of these papers were peer-reviewed and published
  - However you can pick your opinion on papers!

S. Hand'10   22

Reviewing Tips & Tricks

- Identify a core paper for the topic

- Read *related work and/or background* section and read key other papers on the topic

- Capture the author's claim of contribution in *introduction* section and judge if it is delivered

- Identify major idea from main section, normally described at beginning

- Understand the methodology to demonstrate paper's approach

- Capture what authors evaluate and judge if that is a good way to evaluate the proposed idea

- For theory/algorithm paper, capture what it produces as a result (rather than how)

23

Elements in Review Comments

- Paper Summary
  - Provide a brief summary of the paper
  - At this stage you should try to be objective
- Problem
  - What is the problem? Why is it important? Why is previous work insufficient?
- Solution or Approach
  - What is their approach?
  - How does it solve the problem?
  - How is the solution unique and/or innovative?
  - What are the details?
- Evaluation is unfair/biased
  - How do they evaluate their solution?
  - What questions do they anser?
  - What are the strength/weakness of the system and evaluation itself?

S. Hand'10   24

Elements in Review Comments

- What do YOU think?
  - Where you finally get to explain your opinion!
  - You should aim to give *a judgement* on the work
  - Your judgement should be backed by your argument
- Questions for the authors

S. Hand'10   25

How to Review a Paper Aid…

- S. Keshav: How to Read a Paper, ACM SIGCOMM Computer Communication Review 83 Volume 37, Number 3, July 2007.

- T. Roscoe: Writing Reviews for Systems Conferences, 2007.

- Simon Peyton-Jones: How to write a great paper and give a great talk about it, Microsoft Research Cambridge.

- David A. Patterson: How to Have a Bad Career in Research/Academia, 2001.

  See course web page for the paper links.

26

Structure of Presentation

- Cover 3 things in your presentation
1. Background/context
   - What motivated the authors?
   - What else was going on in the research community?
   - How have things changed since?
2. What is problem to be tackled?
   - What is the problem they tried to solve?
   - What are the key ideas?
   - What did the authors actually do?
   - What were the results?
3. Your opinion of the paper
   - What you agree and what you disagree?
   - What is the strength and weakness of their approach?
   - What are the key takeaway?
   - What was the impact (possible impact)?

S. Hand'10   27

Preparing...

- Not too much basics: remember, everyone will have read the paper
   - Brief overview
   - Do not make exact repeat of the paper

- Aim: generate discussion – spit your straight opinion about the paper to stir the discussion
   - Explore the arguments they make and the conclusions they draw. What is your opinion on it?
   - When you argue, state clearly the point of argument

S. Hand'10   28

Presenting...

- Practice beforehand to ensure length of your presentation

- Getting nervous is normal!
  - We are in the same boat and we help each other to understand the paper
  - Presentation is a tool to provide a discussion forum

- Try not to get defensive or angry at questions
  - It is not your paper !

S. Hand'10   29

Listening Presentation...

- You need to get involved

- Ask questions from your review – bring your *review_log* copy

- Always be respectful of the speaker

S. Hand'10   30

How to write Survey paper

- Demonstrate a summary of recent research results in a novel way that integrates and adds understanding to work in the research area

- Must expose relevant details associated, but it is important to keep a consistent level of details and to avoid simply listing the different works

- For example:
  - Define the scope of your survey
  - Classify and organize the trend
  - Critical evaluation of approaches (pros/cons)
  - Add your analysis or explanation (e.g. table, figure)
  - Add reference and pointer to further in-depth information

31

Summary

- R202 course web page:

  http://www.cl.cam.ac.uk/~ey204/teaching/ACS/R202_2012_2013

- Slides of presentation, forms, other information will be on the web

- Enjoy the course!

32