Advanced Research Readings in Distributed Systems and Networks

Dr. Colin Perkins
<csp@dcs.gla.ac.uk>

6 January 2006

1 Introduction

The aims of this module are to immerse students in some seminal topics in distributed systems and networks via papers from the literature, reinforce critical reading and reviewing skills, and enable students to develop their critical thinking and group discussion capabilities.

The intended learning outcomes for the module are that students should: 1) exhibit understanding of a broad range of seminal research in the area of distributed systems and networking by summarising relevant papers, and by answering examination questions that require an overall understanding of distributed systems and networks as a whole; and 2) demonstrate the ability to present a critical analysis of relevant literature by presenting verbal summaries of selected research papers to the class, and leading class discussions.

2 Module Description

The module comprises 10 tutorial sessions held from 1-2pm on Thursdays in room F092. The first session will be on 12 January 2006, continuing weekly until 16 March. The module is structured around discussion of one research paper each week (with a review of several papers discussed in the networks and distributed systems part of the Research Readings module in the first week). Students *must* read the papers in advance of each session, and come prepared to discuss the subject. All students are expected to participate in the discussion.

The module will focus on providing students with a deep understanding of the behaviour of the Internet and how this effects the design of network protocols and distributed systems. Accordingly, you are expected to read additional material relating to the topics under discussion, rather than limiting yourself to a narrow reading of the papers we will consider in the tutorials.

Week 1: Introduction

The module begins with a study of the design of the Internet protocol suite. Three papers will be discussed this week: one ("The Design Philosophy of the DARPA Internet Protocols") is new to you, we have previously discussed the other two papers in the Research Readings in Computing Science module last semester. These papers set the scene for the remainder of the module, describing the structure of the Internet protocol and the design philosophy underpinning that structure. You should read these papers to understand why the designers made particular choices, how those choices impact the structure of the network, and what could be done differently.

- D. D. Clark "The Design Philosophy of the DARPA Internet Protocols", Proceedings of ACM SIGCOMM 1988, Stanford, CA, USA, August 1988.

- J. H. Saltzer, D. P. Reed and D. D. Clark, “End-to-End Arguments in System Design”, ACM Transactions on Computer Systems, volume 2, number 4, November 1984.
- V. G. Cerf and R. E. Kahn, “A Protocol for Packet Network Intercommunication”, IEEE Transactions on Communications, volume 22, number 5, May 1974.

Week 2: Measuring Network Performance – The Core Network

We begin our survey of network performance by studying some recent measurements of the performance of the core network. You should read this paper to understand how the measurement infrastructure has been developed, how the network behaves, and how the performance of the network measured in this paper might differ from other networks.

- C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, C. Diot, “Packet-level Traffic Measurement from the Sprint IP Backbone”, IEEE Network Magazine. November 2003.

Week 3: Measuring Network Performance – Measurement Infrastructure

This week we consider the difficulties in building and operating a large-scale distributed computing infrastructure to collect network performance measurements. You should read this paper with a view to how the infrastructure could be improved, considering which issues are inherent to the problem domain and which result from the setup used in these particular experiments. Consider how other large-scale distributed applications (e.g. Grid computing, peer-to-peer systems) might be impacted by the same issues. How does the system described here differ from that discussed last week? Are these differences important?

- V. Paxson, A. K. Adams and M. Mathis, “Experiences with NIMI”, Proceedings of the Passive and Active Measurement Workshop, Hamilton, New Zealand, April 2000.

Week 4: Measuring Network Performance – Simulation and Modelling

The first paper to be discussed this week (“Why We Don’t Know How To Simulate The Internet”) was previously read as part of the Research Readings in Computing Science module. It outlines challenges encountered when developing realistic simulator models of the Internet. The discussion will cover the difficulties of capturing scale, topology and heterogeneity of protocols and traffic, along with strategies to cope with these difficulties.

The second paper (“Performance Evaluation with Heavy Tailed Distributions”) discusses the evidence for heavy tailed distributions as a realistic model for network traffic as well as computation, and discusses implications on the design of future networks and servers. The following points will be covered during the discussion: characteristics of heavy tailed distributions, characteristics of self similar traffic, evidence of this traffic model, and performance implications network and server design.

- S. Floyd and V. Paxson, “Why We Don’t Know How to Simulate the Internet”, Proceedings of the 1997 Winter Simulation Conference, December 1997.
- M. E. Crovella, “Performance Evaluation with Heavy Tailed Distributions (Extended Abstract)”, Lecture Notes in Computer Science 1786, March 2000.

Week 5: TCP Congestion Control

This week we will study how network performance affects the dynamics of transport protocols. The paper describes how the ARPANET suffered congestion collapse in the late 1980s, and how TCP was modified to prevent that collapse. In addition to understanding the problem of congestion control for TCP in the ARPANET, you should consider how the design choices made affect today's networks. With the vast increases in network performance, and the growing prevalence of wireless networks and non-TCP traffic, are the design choices expressed in this paper still relevant today? How does the presence of network features such as RED, ECN, and QoS affect the congestion response of the network and the design of transport protocols?

- V. Jacobson and M. J. Karels, "Congestion Avoidance and Control", Proceedings of ACM SIGCOMM '88, Palo Alto, CA, USA, August 1988.

Week 6: TCP Friendly Congestion Control

Following on from our discussion of TCP congestion control last week, we now consider alternative congestion control algorithms for applications which cannot use TCP. The focus is on TCP Friendly congestion control: algorithms that are fair to TCP traffic on average, but have different dynamics of variation. You should consider why it is important that alternative congestion control algorithms are TCP friendly, what classes of application these alternative algorithms suit, and when non-TCP friendly algorithms might be used.

- S. Floyd, M. Handley, J. Padhye and J. Widmer, "Equation-Based Congestion Control for Unicast Applications", Proceedings of ACM SIGCOMM 2000, Stockholm, Sweden, August 2000.

Week 7: Multicast Congestion Control

We begin our discussion of application scalability by considering how to make congestion control algorithms that scale for one-to-many applications, for example large scale streaming video. You should consider how such applications differ from the traditional client-server model, and how these differences impact congestion control. What does it mean for a multicast application to be TCP friendly?

- S. McCanne, V. Jacobson and M. Vetterli, "Receiver-driven Layered Multicast", Proceedings of ACM SIGCOMM 1996, Stanford, CA, USA, August 1996.

Week 8: Scalable Multicast Applications

Another issue which is difficult for large scale many-to-many applications is reliability. We discuss one novel model for building reliable multicast applications. You should consider how this protocol differs from other reliable data transfer protocols you have seen, how those differences make it scaleable, and what is the trade-off between reliability, consistency, and scalability. Is that trade-off appropriate for other reliable multicast applications? How might the design of other reliable multicast protocols differ?

- M. Handley, J. Crowcroft, "Network Text Editor (NTE): A scalable shared text editor for the MBone", Proceedings of ACM SIGCOMM 1997, Cannes, France, 1997

Week 9: Peer-to-Peer Applications

Other important problems in the design of large scale multicast and peer to peer applications are naming and addressing. We discuss one approach to building a large scale distributed naming lookup service for peer to peer systems. What are the challenges in distributed name lookup? How might one design other solutions? How might protocols such as Chord be used?

- I. Stoica, R. Morris, D. Karger, M. F. Kaashoek and H. Balakrishnan, “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications”, Proceedings of ACM SIGCOMM 2001, San Diego, CA, USA, August 2001.

Week 10: Future Directions

We conclude the module with a discussion of the future of the network, looking at how the design of networked systems may evolve, and how we can attempt to direct that evolution through careful system design.

- D. D. Clark, K. R. Sollins, J. Wroclawski and R. Braden, “Tussle in Cyberspace: Defining Tomorrow’s Internet”, Proceedings ACM SIGCOMM 2002, Pittsburgh, PA, USA, August 2002.

3 Assessment

Assessment is by a presentation given to the Embedded, Networked, and Distributed Systems research group (30%) and by an open book examination (70%). This is a level M module, worth 5 credits.

Presentations should be given during weeks 7-12 of Semester 2. Students must contact Dr. Olufemi Komolafe <femi@dcs.gla.ac.uk> to arrange the date of their presentation. Each student is responsible for choosing the subject for his own presentation, related to the topics under discussion in the module. Subjects must be approved by the module co-ordinator in advance; students may not choose a topic already chosen by another student.