MPhil in Advanced Computer Science
Data Centric Networking

Leader: Eiko Yoneki
Timing: Lent
Prerequisites: ACS Network Architectures (suggested)
Structure: 8 × 2-hour seminar sessions (combination of lectures and reading club)

AIMS

This module provides an introduction to data centric networking, where data is a communication token in networking. Data centric networking in distributed systems relies on content addressing instead of host addressing, thus providing network independence for applications. Integration of complex data processing with networking is a key vision for future computing. This course provides various aspects in data centric networking ranging from content-based routing to data-flow programming, providing a solid basis to work on the next generation of communication paradigm.

SYLLABUS

The module consists of 8 sessions, of which 6 sessions focus on a specific aspect of the topic in data centric networking research. Each session discusses 2-3 papers, led by the assigned students. Each student will present about 2 paper reviews during the course. The first session advises how to read/review a paper and a brief introduction of different perspectives in data centric networking. The last session is dedicated to the presentation of the open source project studies present by the students. Three guest lectures are planned, covering inspiring current research in the data centric networking domain.

1. Introduction to Data Centric Networking
    - Data as communication token
    - Data centric networking from different perspectives
2. Content-Based Networking and Content Distribution Networks
    - Content-based routing
    - Content distribution overlay
    - Channel vs. Content filtering
3. Content-Centric Networking and Named Data Networking
    - Naming
    - Publish/Subscribe
    - Caching – Network as a storage
    - Multicast and Anycast
4. Programming/controlling distributed computation
    - Network meets data flow programming
    - Parallel data processing (e.g. Map/Reduce, Hadoop, Dryad/LINQ))
    - Declarative networking (e.g. P2, Declarative Sensor Network)
5. Stream Data Processing and Data/Query Model

- Stream data processing and continuous query processing
- Advanced data processing in networks (e.g. data model)

6. Network holds Data in Delay Tolerant Networks
    - Delay tolerant data
    - Networked storage
    - Opportunistic networking
7. Network Structure/Characteristics and Contexts
    - Network graph for data flow
    - Optimising network topology
8. Presentation of Open Source Project Study

OBJECTIVES

On completion of this module, the students should:

- Understand key concepts of data centric approaches in future networking
- Obtain a clear understanding of building distributed systems using data centric communication.

COURSEWORK

Reading Club:

The reading club will require the students to read between 1 and 3 papers every week. At each session, around 3 papers are selected under the session topic, and the students present their review work. The paper includes following two types and focuses on the specified aspects upon reviewing the paper are:

1. Full length papers
    - What is the significant contribution?
    - What is the difference from the existing works?
2. Short length papers
    - What is the novel idea?
    - What is required to complete the work?

Reports:

The following three reports are required, which could be extended from the reading assignment of the reading club or a different one within the scope of data centric networking.

1. Review report on a full length of paper (2500 words)
    - Describe the contribution of paper in depth with criticism
    - Crystallise the significant novelty in contrast to the other related work
    - Suggestion for future work

2. Survey report on sub-topic in data centric networking (3500 words)
    - Pick up to 5 papers as core papers in the survey scope
    - Read the above and expand reading through related work
    - Comprehend the view and finish as an own survey paper

3. Project study and exploration of a prototype (3500 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

The reports 1 and 2 should be handed in by the end of 5th week and 7th week of the course (not in any particular order). The report 3 should be by the end of the Lent term.

ASSESSMENT

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

1. 25%: for a reading club (participation)
2. 75%: for the three reports
    - 20%: Intensive review report
    - 25%: Survey report
    - 30%: Project study

RECOMMENDED READING

[1] D. J. Abadi, D. Carney, U. Cetintemel, M. Cherniack, et al. A New Model and Architecture for Data Stream Management. VLDB Journal, 12(2), 2003.
[2] A. Arasu, B. Babcock, S. Babu, et al. Stream: The Stanford Stream Data Manager. ACM SIGMOD, 2003.
[3] A. Carzaniga, M. Rutherford, and A. Wolf. A Routing Scheme for Content-based Networking. IEEE INFOCOM, 2004.
[4] P. Dagand, D. Kostic, and V. Kuncak. Opis: Reliable Distributed Systems in Ocaml. TLDI, 2009.
[5] J. Dean and S. Ghemawat. Mapreduce: Simplified Data Processing on Large Clusters. OSDI, 2004.
[6] EU FP7 Haggle Project. http://www.haggleproject.org/
[7] Van Jacobson Diana K. Smetters James D. Thornton Michael F. Plass Nicholas H. Briggs Rebecca L. Braynard, Networking Named Content, ACM CoNext, 2009.
[8] B. Karp and H. Kung. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. Computing and Networking, pages 243–254, 2000.
[9] X. Li, Y. J. Kim, R. Govindan, and W. Hong. Multi-dimensional Range Queries in Sensor Networks. Sensys, 2003.
[10] LINQ Project. http://msdn.microsoft.com/en-gb/library/bb308959.aspx.
[11] S. Madden, M. J. Franklin, J. Hellerstein, et al. TAG: a tiny aggregation tree for ad-hoc sensor networks. OSDI, 2002.
[12] E. Nordström, P. Gunningberg, and C. Rohner. A Search-based Network Architecture for Mobile Devices. Tech Report Uppsala University, 2009.
[13] P2. http://p2.berkeley.intel-research.net/.
[14] M. Welsh and G. Mainland. Programming Sensor Networks using Abstract Regions. NSDI, 2004.
[15] Y. Yu, M. Isard, D. Fetterly, M. Budiu, U. Erlingsson, P. K. Gunda, and J. Currey. DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing Using a High-Level Language. OSDI, 2008.