Dingming Wu

Phone: +1(281)9288685 Email: dw40@rice.edu
Department of Computer Science, Rice University
6100 Main Street, Duncan Hall 3006, Houston, Texas 77005-1827

RESEARCH INTERESTS

My research interest generally lies in the modeling, designing and implementation of networking systems. In particular, I work on problems that revolve around the performance, fault-tolerance, and reliability issues that arise within modern datacenters. I'm interested in exploiting emerging underlying technologies such as optical devices and programmable switches to solve those problems.

EDUCATION

Rice University *August 2015 - present*
PhD in Computer Science
Advisor: T. S. Eugene Ng

Nanjing University *September 2012 - June 2015*
M.S. in Computer Science
Thesis Title: Research on Simultaneous Queries in Wireless Sensor Networks

Wuhan University *September 2008 - June 2012*
B.S. in Computer Science
Thesis Title: The Design and Implementation of a Cross-language Information Retrieval System

PROJECTS & PUBLICATIONS

Building a Rackless Network Architecture for Data Centers

We propose a rackless datacenter (RDC) architecture that removes the"rack boundary" on server communications and allows servers to talk to each other with uniform high bandwidth, regardless of their topological locations. We achieve this goal by inserting circuit switches at the network edge, and dynamically reconfiguring the circuits to allow servers from different racks to form locality groups. RDC optimizes the network topology for the changing workloads, and achieves lower flow completion times and improved link utilization over realistic workloads

- **Publication**:
- Dingming Wu, Ang Chen, T. S. Eugene Ng, The Case for a Rackless Data Center Network Architecture **ACM SIGCOMM Poster Program 2018**.

Ultra-Fast and Full-Capacity Failure Recovery in Data Center Networks

- We promote the concept of shareable backup in datacenters. Shareable backup is an economical and effective way to mask failures from application performance. A small number of backup switches are shared network-wide for repairing failures on demand so that the network quickly recovers to its full capacity without applications noticing the failures. This approach avoids complications and ineffectiveness of rerouting. We propose ShareBackup as a prototype architecture to realize this concept and present the detailed design. We implement ShareBackup on a hardware testbed. Its failure recovery takes merely 0.73ms, not disrupt routing; and it accelerates Spark and Tez jobs by up to 4.1 under failures. Large-scale simulations with real datacenter traffic and failure model show that ShareBackup reduces the percentage of job flows prolonged by failures from 47.2% to as little as 0.78%. In all our experiments, the results for ShareBackup have little difference from the no-failure case.
- **Publication**:

- **Dingming Wu**, Yiting Xia, Xiaoye Steven Sun, Simbarashe Dzinamarira, Xin Sunny Huang, T. S. Eugene Ng, Masking Failures from Application Performance in Data Center Networks with Shareable Backup, **ACM SIGCOMM 2018**

Convertible Data Center Network Architectures

- We propose a convertible datacenter network architectures, which can dynamically change the network topology to combine the benefits of multiple architectures. We use the fat-tree prototype architecture as the first step to realize this concept. Flat-tree can be implemented as a Clos network and later be converted to approximate random graphs of different sizes, thus achieving both Clos-like implementation simplicity and random-graph-like transmission performance. We present the detailed design for the network architecture and the control system. Testbed evaluation shows the network core bandwidth is increased by 27.6% just by converting the topology from Clos to approximate random graph. This improvement can be translated into an acceleration of applications as we observe reduced communication time in Spark and Hadoop jobs.
- **Publication**:
- Yiting Xia, Xiaoye Steven Sun, Simbarashe Dzinamarira, **Dingming Wu**, Xin Sunny Huang, T. S. Eugene Ng, A Tale of Two Topologies: Exploring Convertible Data Center Network Architectures with Flat-tree, **ACM SIGCOMM 2017**

Machine Level Task Scheduling in Large Shared Clusters

- Large scale data analytics frameworks are shifting towards handling the consolidation of diverse workloads such as batch data analytics jobs, machine learning jobs and interactive queries. How to maintain high cluster utilization while providing low and predictable job response latency with vastly heterogeneous job types is crucial for cluster management. We design a novel multi-level feedback queue scheme that mimics the shortest job first without starving longer tasks. The key observation behind this design is that today's machines usually have many CPU cores such that each queue can occupy one or multiple cores. The multi-queue scheduler allows the longer tasks to migrate from one queue to another and guarantees that shorter tasks always have the resource to run without prior knowledge of its duration. We exploit the rich system APIs of the Linux platform and implement it as a lightweight program on top of the Linux Kernel.
- **Dingming Wu** Runway: Worker-level Task Scheduler for Large Shared Clusters, **Master Thesis, Rice University**).

Building a High Throughput and Low latency Multicast System for Datacenters

- Multicast has long been a performance bottleneck for datacenters. In this project, we propose an unconventional optical architecture design that directly interconnects top of rack switches by low-cost optical splitters, thereby eliminating the need for optical switches. The ToRs are organized to form the connectivity of a regular graph. A key property of this architecture is its link failure tolerance and low hop-count for any source-destination pair. Preliminary results from our analysis and simulation show that this architecture is scalable and highly efficient multicasts.
- **Publication**:
  Dingming Wu, Xiaoye Sun, Yiting Xia, Xin Huang, T.S. Eugene Ng, HyperOptics: A High Throughput and Low Latency Multicast System for Datacenters, **HotCloud 2016**.

Energy Efficient Counting and Identification in Wireless Sensor Networks

- In this project, we show that wireless signals can be superimposed constructively in sensor networks and the energy of the superimposed signal will increase with the increasing number of component signals. Motivated by this observation, we design two energy efficient protocols, Poid and Poc, to estimate the neighbourhood cardinality and the identification of neighbours. The testbed contains 1 USRP and 50 TelosB nodes. Results show that the counting accuracy is above 97.9% and the identification accuracy is above 96.5%, both with constant delay.
- **Publication**:
  Dingming Wu, Guihai Chen, Chao Dong, Shaojie Tang, Haipeng Dai, Simultaneous Query for WSNs: A Power Based Solution, **IEEE Transactions on Mobile Computing, 2015**.

Dingming Wu, Chao Dong, Shaojie Tang, Haipeng Dai and Guihai Chen, Fast and Fine-grained Counting and Identification via Constructive Interference in WSNs, ACM/IEEE **IPSN 2014**.

PROFESSIONAL EXPERIENCES

Research Intern *August 2018 - December 2018*
Alibaba Group, Sunnyvale, CA
Working on programmable networks in the Alibaba Infrasturacture Service team.

PhD Intern *May 2018 - August 2018*
Microsoft Azure, Redmond, WA
Built a high-frequency tracing tool for multi-threaded cache driver of Azure host OS.

PhD Intern *May 2017 - July 2017*
Microsoft Azure, Redmond, WA
Developed a data-driven model to detect and predict memory leak of Azure system software.

Research Intern *May 2014 - August 2014*
Microsoft Research Asia, Beijing, China
Designed a semantic-oriented programming framework for data management in Internet-of-Things.

TECHNICAL STRENGTHS

Computer Languages	C, C++, Java, Python
Parallel Programming Models	Pthread, Cilk++, OpenMP, OpenMPI
Big Data Processing	Spark, Hadoop Yarn
Data Analysis	Matlab, R
Tools	SVN, Git, Makefile, Eclipse, IntelliJ, Latex

SELECTED COURSES

Advanced Topics in Distributed Systems (Spring 18)
Introduction to Computer Security (Spring 18)
Production Programming (Fall 2017)
Convex Optimization (Fall 16)
Introduction to Parallel Programming (Spring 16)
Introduction to Computer Networks (Fall 15)
Artificial Intelligence (Fall 15)
Design and Analysis of Algorithm (Fall 15)
Randomized Algorithm (FALL 14)
Combinatorics (FALL 14)
Distributed Systems (Spring 13)

HONORS & REWARDS

Graduate Fellowship at Rice University, 2015
Outstanding Graduate Student at NJU, 2015 (**Rank First** of the CS department).
National Scholarship at NJU (**top 3%**), 2014