Dingming Wu

+1-281-928-8685
dingming.wu@rice.edu
Dingming Wu
dmwu
https://dmwu.github.io

Education

Rice University 2015 –
PhD Computer Science *Houston, TX*

Nanjing University 2012 – 2015
MS Computer Science *Nanjing, China*

Wuhan University 2008 – 2012
BS Computer Science *Wuhan, China*

Experience

Alibaba Group Aug. 2018 – Jan. 2019
Research Intern *Sunnyvale, CA*

- Designed and implemented network service function chaining on programmable switches for Alibaba's cloud infrastructure.

Microsoft Azure May. 2018 – Aug. 2018
Intern *Redmond, WA*

- Built a high-frequency IO performance tracing tool for multi-threaded cache driver of Azure host OS.

Microsoft Azure May. 2017 – Aug. 2017
Intern *Redmond, WA*

- Designed and Implemented a data-driven model to detect and predict memory leak of Azure system software.

Microsoft Research Asia May 2014 – August 2014
Research Intern *Beijing, China*

- Designed a semantic-oriented programming framework for data management in Internet-of-Things.

Selected Projects and Publications

Accelerated Service Function Chaining on Programmable Switches

- Network functions and service function chaining are prevalent in cloud and ISP networks. In traditional software-based solutions, scaling up the capacity of these functions requires a large number of server cores. However, edge clouds are severely resource constrained in terms of space, power, and budget, so traditional methods incur a high cost. We present a system that can offload a service chain to a programmable switch to achieve high performance and resource efficiency. Our system can compose multiple network functions into a single program that preserves the original chaining requirements, and exploit features of the switch ASIC to efficiently deploy the composed program on a single switch.
- **Dingming Wu**, Ang Chen, T. S. Eugene Ng, Guohui Wang, Haiyong Wang, Accelerated Service Function Chaining on Programmable Switches, **Under Review.**

Towards a Rackless Network Architecture for Data Centers

- We propose a rackless architecture that removes the *rack boundary* in data center networks 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. Our work optimizes the network topology for the changing workloads, and achieves lower flow completion times and improved link utilization over realistic workloads.
- **Dingming Wu**, Weitao Wang, Ang Chen, T. S. Eugene Ng, Say No to Rack Boundaries: Towards a Reconfigurable Pod-Centric DCN Architecture **ACM SOSR 2019**

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.
- **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**

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.
- **Dingming Wu**, Xiaoye Sun, Yiting Xia, Xin Huang, T. S. Eugene Ng, HyperOptics: A High Throughput and Low Latency Multicast System for Datacenters, **Usenix HotCloud 2016**.

Skills

Programming Languages C, C++, Python, Java, P4

Tools Git, SVN, Latex, Matlab, Makefile, Linux

Big Data Analytics Spark, Hadoop, Tez

Honors and Rewards

Graduate Fellowship at Rice University 2015

Outstanding Graduate Student at NJU Rank first of the CS department, 2015

National Scholarship at NJU top 3%, 2014