

Education

Ph.D. Computer Science, Cornell University, NY, USA Aug. 2019 – May. 2025 (expected)
Advisor: Prof. Robbert van Renesse

M.S. Computer Science, National Chiao Tung University, Hsinchu, Taiwan Sep. 2013 - Jul. 2015
Advisor: Prof. Wei-Chung Hsu
Thesis: A KVM-based Hypervisor for Heterogeneous System Architecture

B.S. Computer Science, National Chiao Tung University, Hsinchu, Taiwan Sep. 2009 - Jul. 2013

Publications

Disaggregated Applications Using Nanoservices

Xinwen Wang, **Yu-Ju Huang**, Tiancheng Yuan, Robbert van Renesse
Workshop On Resource Disaggregation and Serverless (WORD'21), April 2021

Building a KVM-based Hypervisor for a Heterogeneous System Architecture Compliant System

Yu-Ju Huang, Hsuan-Heng Wu, Yeh-Ching Chung, Wei-Chung Hsu.
12th International Conference on Virtual Execution Environments (VEE'16), April 2016
Best Paper Award!

Research

My general interest is to see how systems (on one or multiple machines and all layers of computer systems) work and try to make them work even better. My research thus spans operating systems, virtualization, cloud computing, distributed computing, and computer architecture.

Ziplog (ongoing)

Near-realtime stream processing is getting more important in data-intensive applications. A key building block of it is the shared log (or message broker, event stream), such services including Corfu, Apache Kafka, Amazon Kinesis Streams, etc. While heavily used, state-of-the-art shard log designs don't provide all the properties that application developers usually require, which are cross-shard total order, scalable throughput, and low latency. Ziplog project tries to achieve all three - it guarantees cross-shard total order, achieves scalable throughput, and experiences low latency!

Fluid OS (ongoing)

With the proposal of hardware resource disaggregation, we believe that the paradigm of computer abstraction is shifting. While we used to think a computer consists of a fixed set of hardware resources, the new computer abstraction is becoming fluid, where the hardware resources are continuously changing, abundant, and elastic.

We aim to build an OS that is also fluid, i.e., it views hardware resources as a dynamic pool. It can boot with a small amount of CPU, memory and can grow more powerful (including GPU and accelerators!) upon requests from the applications.

HSA virtualization (past)

- Built a hypervisor that virtualizes a new heterogeneous computing architecture called Heterogeneous System Architecture (HSA).
- Guest processes are able to use GPUs on the HSA-compliant machine (AMD's Kaveri) and achieves nearly 95% of native performance.
- GPU can be shared between guest processes on different guest OSes.

Professional Service

Program Committee Member and Session Chair

Xian, China. April. 2017

13th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE).

Employment

System Software Engineer - MediaTek, Office of CTO

Hsinchu, Taiwan. Dec. 2015 – Jun. 2019

* *Compiler and runtime software stack for edge AI*

- Built an AI various AI frameworks (TensorFlow, Android NN) and backend computing devices (CPU, GPU, DLA).
- Compiler development for in-house deep learning accelerator (DLA).
- Taskforce leader of DLA performance optimizations.

* *Android runtime (ART) and compiler optimization*

- Built a staged Java compiler using ART's optimizing compiler and LLVM to optimize Java applications.
- Improve performance by around 10% on a plenty of benchmarks.

* *QoS (quality of service)-based framework for optimizing Android runtime*

- Enhanced the Android runtime to be aware of QoS hints and adjust the system accordingly.
- Extended the Android runtime to profile Java applications and export QoS hints.
- Improve user experience by removing frame drops in game scenario.

Intern Engineer - Marvell Taiwan

Hsinchu, Taiwan. Jul. 2014 - Aug. 2014

* *Visualization tool for cache system*

- Created a cache visualizer to show the complex cache behaviors.
- Analyze performance issues using the cache visualizer.

Intern Engineer - ITRI, Cloud Computing Center

Hsinchu, Taiwan. Feb. 2014 - Jul. 2014

* *KVM-ARM I/O virtualization optimization*

- Explore different kinds of I/O virtualization techniques, including VirtIO and device pass-through, and use them to optimize ARMv8 storage I/O.
-

Awards

Cornell University Fellowship, 2019-2020

Best Paper Award, 12th International Conference on Virtual Execution Environments (VEE'16)

Teaching

Cornell University

Grad TA for CS 3410: Computer System Organization and Programming, Fall 2021

Head TA for CS 4411: Practicum in Operating Systems, Spring 2020

Head TA for CS 4410: Operating Systems, Fall 2020

Programming Languages

C, C++, Java, Go, Python

Extracurricular Activities

President of Cornell Taiwanese Student Association

Jun. 2020 – Jun 2021

Captain of Kendo Team at National Chiao Tung University

Sep. 2010 - Feb. 2012