Yu-Ju Huang

yh885@cornell.edu
https://www.cs.cornell.edu/~yjhuang
https://www.linkedin.com/in/yu-ju-huang/

Education

Cornell University, Ithaca, NY
Ph.D. in Computer Science
Advisor: Prof. Robbert van Renesse

Aug. 2019 – Present

National Chiao Tung University, Hsinchu, Taiwan
M.S. in Computer Science, GPA: 4.0
Advisor: Prof. Wei-Chung Hsu

Sep. 2013 - Jun. 2015

B.S. in Computer Science, GPA: 3.91

Sep. 2009 - Jun. 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 Projects

New shared log design (ongoing)
This project builds a **shared log** that provides cross-shard total order, achieves scalable throughput, and experiences low latency. Near-realtime stream processing is getting more important in data-intensive applications. A key building block of it is the shared, such services including Corfu, Apache Kafka, Amazon Kinesis Streams, etc. While heavily used, state-of-the-art designs don't provide all the properties that application developers usually require, which are cross-shard total order, scalable throughput, and low latency. Our solution accomplishes all these three properties.

OS for disaggregated architecture (ongoing)
We build an **OS for disaggregated architecture**. This OS views hardware resources as a dynamic pool. It can boot with a small amount of CPU, memory and can grow more powerful upon requests from the applications. With the proposal of hardware resource disaggregation, we believe that the paradigm of computer abstraction is shifting. While we used to think a computer is 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. Therefore, we argue that the OSes running on it should also be fluid.

HSA virtualization (past)
 - Built a **hypervisor** that virtualizes a new heterogeneous computing architecture called Heterogeneous System Architecture (HSA).
 - Handling **memory virtualization of IOMMU**, including address translation and I/O page fault.
 - Guest processes can share GPUs on the HSA-compliant machine (AMD's Kaveri) and achieve nearly 95% of native performance.

Professional Service

Program Committee Member and Session Chair   *Xian, China. April. 2017*
13th International Conference on Virtual Execution Environments (*VEE'17*)

Awards

* **Cornell University Fellowship**, 2019-2020
* **Best Paper Award**, 12th International Conference on Virtual Execution Environments (*VEE'16*)

Work Experience

Applied Scientist Intern - Amazon Web Service (AWS)   *Seattle, US. May. 2022-Aug. 2022*
 * *Verification systems for a high-performance transaction Key-Value Store (KVS) library*
 - Built infrastructure for verifying a ***transaction KVS library in Rust***
 - Implemented invariant checkers to validate the KVS library

System Software Engineer - MediaTek, Office of CTO   *Hsinchu, Taiwan. Dec. 2015-Jun. 2019*
* *Compiler and runtime software stack for mobile AI*
 - Developed **compiler** for in-house **deep learning accelerator** (DLA).
 - **Taskforce leader** of DLA performance optimizations.
 - Built AI framework to run AI models (**TensorFlow, Android NN**) on multiple computing devices (CPU, GPU, DLA).

* *Android runtime and compiler optimization*
 - Built a **staged compiler** using **LLVM** to optimize Android applications.
 - Improved performance by 10% on a plenty of benchmarks.

* *QoS (quality of service)-based framework for optimizing Android applications*
 - Extended the Android runtime to profile Java applications and to export **QoS** hints.
 - Enabled Android runtime to be aware of QoS hints and adjust system resources accordingly.
 - Improved user experience by removing frame drops in game scenarios.

Software Engineer Intern - Marvell Technology, Inc.   *Hsinchu, Taiwan. Jul. 2014-Aug. 2014*
* *Visualization tool for cache system*
 - Created a **cache visualizer** to show the complex CPU cache behaviors.
 - Analyzed performance issues using the cache visualizer.

Research Intern - Industrial Technology Research Institute *Hsinchu, Taiwan. Feb. 2014-Jun. 2014*
* *KVM-ARM I/O virtualization reseach*
 - Optimized ARMv8 storage I/O virtualization with VirtIO and device pass-through
 - Improved performance of live **VM cloning/migration**

Professional Skills

* Software engineer with **3+ years** of software development
* Experienced in OS, hypervisor, compiler, Android, distributed systems development
* C / C++ / Rust / Java / Go / Python / Git / Linux / KVM / Xen / Docker / Kubernetes

Teaching

Cornell University
  Grad TA for CS 6410: Advanced Systems, Fall 2022
  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