PONGSAKORN U-CHUPALA, Ph.D.

ONLINE RESUME: HTTPS://PUCHUPALA.COM **EMAIL:** PUCHUPALA@GMAIL.COM **TEL:** (+81)-80-4243-9556

EDUCATION

Nara Institute of Science and Technology,
Nara, Japan (2013-2018)

- Doctor of Engineering, Computer Science, Graduate School of Information Science (GPA: 4.00)
- Master of Engineering, Computer Science, Graduate School of Information Science (GPA: 4.00)

Kasetsart University,
Bangkok, Thailand (2008-2013)

- Bachelor of Engineering, Computer Engineering, *Cum Laude* (GPA: 3.54)

LANGUAGE QUALIFICATIONS

- 990 Points, TOEIC IP, June 2015
- 101 Points, TOEFL iBT, October 2012
- Level N2, Japanese Language Proficiency Test, December 2021

SCHOLARSHIPS

- MEXT Scholarship *(2015-2018)*
- JASSO Scholarship *(2011-2012)*

AFFILIATIONS

- PRAGMA Student Steering Committee *(2015-2018)*
- Google Developer Group Thailand *(2012-2018)*
- Google Student Ambassador SEA, Google Inc. (2012)

LINKS

- Personal Github: https://github.com/puchupala
- Work Github: https://github.com/te-pongsakornuchupala
- LinkedIn: https://www.linkedin.com/in/puchupala/

SKILLS

- Distributed Machine Learning (PyTorch, TensorFlow, NNabla)
- GPGPU Application Optimization (CUDA, ROCm)
- Distributed Communication Algorithm
- High-Performance Computing for Machine Learning Workload
- Cloud Computing, Virtualization, and Linux Containers
- Web Technology and Web Development
- Software-Defined Networking and OpenFlow
- Linux Administration
- Programming Languages: Python (expert), C/C++ (proficient), JavaScript (prior experience), PHP (prior experience)
- Languages: Thai (Native), English (Fluent), Japanese (Business Proficiency)

EXPERIENCES

2018-present <u>Research Engineer (Distributed Deep Learning)</u>, R&D Center, Sony Group Corporation, Japan	- To enable engineers and data scientist to utilize standard AI/ML toolchain on non-standard proprietary GPGPU cluster, I developed custom distributed deep learning stack including custom collective communication solution for non-standard proprietary communication fabric. - I help design GPGPU cluster using non-standard proprietary hardware for distributed deep learning workload. - Using low-rank learning method we developed in collaboration with UW-Madison, I reduced memory footprint of several of our neural network learning tasks by up to 50%. - I coordinate our team research collaboration effort with UW-Madison, which resulted in the publication: **PUFFERFISH: Communication-efficient Models at No Extra Cost**. - I coordinate our team research collaboration effort with Georgia Tech, which resulted in the publication: **Nested Dithered Quantization for Communication Reduction in Distributed Training**. - I developed a distributed deep learning simulator, which help our team **broke world record of ImageNet/ResNet-50 training speed**. During ABCI Grand Challenge 2018, we gained access to the entire ABCI cluster only for a limited time. The simulator allows us to do dry hyper-parameter tuning, thus significantly reducing the number of experiments required on the cluster. **I work on NNabla,** Sony's high-performance deep learning framework. I am responsible for distributed learning performance optimization as well as designing next-generation distributed learning API.
2013-2018 <u>PHD Student</u>, Software Design and Analysis Laboratory, Nara Institute of Science and Technology, Japan	*Doctoral Dissertation* **Increasing Data Center Efficiency with Improved Task Scheduling and Communication** I propose several optimizations for cloud infrastructure. *Master's Thesis* **Overseer: Application-Aware Routing** OpenFlow controller for bandwidth and latency aware routing implemented with POX. **PRAGMA-ENT** Breakable international SDN testbed for PRAGMA community. I help established and maintained this network, which connect multiple institutions including NAIST, Osaka University, University of California San Diego, and University of Florida. **Applying Deep Learning to Network Traffic Identification and Categorization** I developed network traffic classification model using stacked denoising autoencoder in TensorFlow. This model is learned on the CAIDA Internet traffic dataset. The model is a part of my proposal to create automatic SDN-based data center network traffic optimizer. **Container Rebalancing** I proposed a novel scheduling mechanism with a rebalancing processing working alongside a scheduling process. A Hadoop/Hive-powered data processing technique and a Python-based simulation using Google's cluster data is performed to validate this method.
2017 <u>Internship</u>, Information Technology Research Institute, AIST, Japan	I was responsible for deploying and benchmarking an experimental multi-site GPFS cluster connecting Japan, Australia, and U.S.A. The work involves the administration and debugging of Linux environment, as well as collaborating with researchers from multiple institutions.
2014 <u>Visiting Scholar</u>, CalIT2, University of California San Diego, United States	**PRAGMA Boot** A program to instantiate VM in PRAGMA's cloud. I was responsible for OpenNebula plugin written in Ruby.
2013 <u>Part-Time Developer</u>, Innovative Extremist Co., Ltd.	**ByteArk** S3-compatible SEA-based CDN. I was a part of the team responsible for the internal API. **Nyanlive** A complete solution for creating and maintaining video streaming platform. I was responsible for streaming authentication/authorization system and the internal API implemented with Django. **Knowbita** Online lecture archive of dept. of computer eng., Kasetsart University. I was responsible for the internal API implemented with Django.
2008-2013 <u>Student</u>, High Performance Computing and Networking Center, Kasetsart University	*Thesis* **An implementation of a multi-site virtual cluster cloud** Virtual cluster over multiple OpenNebula sites.
2012 <u>Part-Time Developer</u>, Onebit Matter Co., Ltd. (now Wisesight Co., Ltd.)	**OBVOC** Social media monitoring platform. I was responsible for social media data collection using Python.
2009-2010 <u>Part-Time Developer</u>, Thoth Media Co., Ltd. (now Wisesight Co., Ltd.)	**Kpiology** Social media analytics platform. I was responsible for the early version of Twitter™ data collection and analytics using Python.

SIDE PROJECTS

- **Homebridge Nature Remo Multi Toggle Light** *(2021)*: Homebridge plugin for controlling toggle light through Nature Remo device.
- **GaineViz** *(2017)*: Web-based visualization tool for Gainesville city's open-data. Best hack award, CENTRA2 Student Hackathon.
- **eCOStamp** *(2013-2014)*, Electronics collectible stamp platform combining web service, smartphone application and 3D-printed Arduino-based hardware. Part of Creative and International Competitiveness Project (CICP2013) supported by NAIST.

PUBLICATIONS

- (Affiliate[1]) H. Wang, S. Agarwal, and D. Papailiopoulos, **"PUFFERFISH: Communication-efficient Models at No Extra Cost,"** in The Fourth Conference on Machine Learning and Systems (MLSys), 2021
- H. Mikami, H. Suganuma, P. U-chupala, Y. Tanaka, and Y. Kageyama, **"ImageNet/ResNet-50 Training in 224 Seconds",** arXiv:1811.05233 [cs.LG], 2018.
- P. U-chupala, **"Increasing Data Center Efficiency with Improved Task Scheduling and Communication",** Nara Institute of Science and Technology, 2018.
- P. U-Chupala, Y. Watashiba, K. Ichikawa, and H. Iida, **"Towards Self-Optimizing Network: Applying Deep Learning to Network Traffic Categorization and Identification in the Context of Application-Aware Network"** , IPSJ SIG Internet and Operation Technology (IOT), 2018.
- K. Ichikawa et al., **"Dynamic International SDN and Inter-Cloud Infrastructure,"** in The 2nd RICC-RIEC workshop, 2017.
- P. U-chupala, Y. Watashiba, K. Ichikawa, S. Date, and H. Iida, **"Application-aware network: network route management using SDN based on application characteristics,"** in CSI Transactions on ICT, pp. 1–11, 2017.
- P. U-chupala, Y. Watashiba, K. Ichikawa, S. Date, and H. Iida, **"Container Rebalancing: Towards Proactive Linux Containers Placement Optimization in a Data Center,"** in The 41th IEEE Computer Society International Conference on Computers, Software & Applications (COMPSAC), 2017.
- K. Ichikawa et al., **"PRAGMA-ENT: An International SDN testbed for cyberinfrastructure in the Pacific Rim,"** Concurrency and Computation: Practice and Experience, February, 2017.
- S. Date et al., **"SDN-accelerated HPC infrastructure for scientific research,"** in International Journal of Information Technology (IJIT), 2016
- S. Date et al., **"An Empirical Study of SDN-accelerated HPC Infrastructure for Scientific Research,"** in 2015 International Conference on Cloud Computing Research and Innovation (ICCCRI), 2015, pp. 89–96.
- K. Ichikawa et al., **"PRAGMA-ENT: Exposing SDN Concepts to Domain Scientists in the Pacific Rim,"** in PRAGMA Workshop on International Clouds for Data Science (PRAGMA-ICDS) 2015, 2015.
- P. U-chupala, **"Overseer: SDN-Assisted Bandwidth and Latency Aware Route Optimization based on Application Requirement,"** Nara Institute of Science and Technology, 2015.
- P. U-chupala, K. Ichikawa, H. Iida, N. Kessaraphong, P. Uthayopas, S. Date, H. Abe, H. Yamanaka, and E. Kawai, **"Application-Oriented Bandwidth and Latency Aware Routing with OpenFlow Network,"** in The 6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2014.
- P. U-chupala, K. Ichikawa, P. Uthayopas, S. Date, and H. Abe, **"Designing of SDN-Assisted Bandwidth and Latency Aware Route Allocation,"** in Summer United Workshops on Parallel, Distributed and Cooperative Processing (SWoPP), 2014.
- P. U-chupala, P. Uthayopas, K. Ichikawa, S. Date, and H. Abe, **"An implementation of a multi-site virtual cluster cloud,"** in The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2013, pp. 155–159
- P. U-chupala, K. Ichikawa, H. Abe, S. Date, and S. Shimojo, **"A Virtual Cluster Manager using a Hierarchical Management Model for Cloud Infrastructure,"** in The 6th International Conference on Ubiquitous Information Technologies and Applications (CUTE), 2011.