

Nathan Kwan-Ho Ng

Ph.D. Candidate, Computer Science
University of Massachusetts Amherst

✉ kwanhong@cs.umass.edu

☎ (347) 615-8340

🌐 [linkedin.com/in/nathan-kwanho-ng](https://www.linkedin.com/in/nathan-kwanho-ng)
🌐 khngnathan.github.io/personal-website

Education

Ph.D. in Computer Science

09/2019 – 12/2026 (Expected)

University of Massachusetts Amherst

Advisor: Prof. Prashant Shenoy

B.Sc. in Computer Science

09/2014 – 06/2018

Hong Kong Baptist University

Advisor: Prof. Byron Choi

Research Interests

My research spans distributed systems and systems for machine learning, with a current focus on AI systems across the device–edge–cloud continuum. I design, build, and optimize systems and resource-management techniques that deliver predictable low-latency performance, covering both average-case and tail-latency behavior, for emerging AI applications.

Publications

HYDRA: Multi-Level Ensemble Learning for Failure-Resilient Edge Inference

Krishna Praneet Gudipaty, Walid Hanafy, **Nathan Ng**, Kaan Ozkara, Qianlin Liang, Jesse Milzman, David Irwin, Suhas Diggavi, Prashant Shenoy

ACM/IEEE Symposium on Edge Computing (SEC'26). To appear.

Collaborative Processing for Multi-Tenant Inference on Memory-Constrained Edge TPUs

Nathan Ng, Walid Hanafy, Prashanthi Kadambi, Balachandra Sunil, Ayush Gupta, David Irwin, Yogesh Simmhan, Prashant Shenoy

IEEE International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT'26) **Best Paper Award**

To Offload or Not To Offload: Model-driven Comparison of Edge-native and On-device Processing in the Era of Accelerators

Nathan Ng, David Irwin, Ananthram Swami, Don Towsley, Prashant Shenoy

ACM/SPEC International Conference on Performance Engineering (ICPE'26)

TailClipper: Reducing Tail Response Time of Distributed Services Through System-Wide Scheduling

Nathan Ng, Abel Souza, David Irwin, Don Towsley, Prashant Shenoy

ACM Symposium on Cloud Computing (SoCC'24)

Collaborative Inference in Resource-Constrained Edge Networks: Challenges and Opportunities

Nathan Ng, Abel Souza, Suhas Diggavi, Tarek Abdelzaher, Don Towsley, Prashant Shenoy

IEEE Military Communications Conference (MILCOM'24), IoT for Adversarial Environments Workshop

Unlocking Efficiency: Understanding End-to-End Performance in Distributed Analytics Pipelines

Abel Souza, **Nathan Ng**, Tarek Abdelzaher, Don Towsley, Prashant Shenoy

IEEE Military Communications Conference (MILCOM'23), IoT for Adversarial Environments Workshop

FGreat: Focused Graph Query Autocompletion

Nathan Ng, Peipei Yi, Zhiwei Zhang, Byron Choi, Sourav S. Bhowmick, Jianliang Xu

IEEE International Conference on Data Engineering (ICDE'19), Demonstration Track

Answering the Why-Not Questions of Graph Query Autocompletion

Guozhong Li, **Nathan Ng**, Peipei Yi, Zhiwei Zhang, Byron Choi

International Conference on Database Systems for Advanced Applications (DASFAA'18)

Preprints & Under Submission

Computation Placement in the Device-Edge Continuum: Model-Driven Comparison of On-Device, Edge-Native, and Split Processing with Accelerators

Nathan Ng, David Irwin, Ananthram Swami, Don Towsley, Prashant Shenoy

Under review at *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*

Tuning the Tail Latency of Distributed Queries Using Replication

Nathan Ng, Hung Le, Prashant Shenoy, Marco Serafini

Under review at the *Distributed and Parallel Databases Journal (DAPD)*. arXiv:2212.10387

Patents

Dynamic user profile management

Nathan Ng, Tung Mai, Thomas Greger, Kelly Quinn Nicholes, Antonio Cuevas, Saayan Mitra, Somdeb Sarkhel, Anup Bandigadi Rao, Ryan A. Rossi, Viswanathan Swaminathan, Shivakumar Vaithyanathan

US Patent US12,339,916B2. Published June 2025.

Dynamic user profile projection

Nathan Ng, Tung Mai, Thomas Greger, Kelly Quinn Nicholes, Antonio Cuevas, Saayan Mitra, Somdeb Sarkhel, Anup Bandigadi Rao, Ryan A. Rossi, Viswanathan Swaminathan, Shivakumar Vaithyanathan

US Patent US12,130,876B2. Published Oct 2024.

Awards and Honors

Best Paper Award	2026
-------------------------	------

DCOSS-IoT'26

Donald F. Towsley Graduate Scholarship in Computer Science	2026
---	------

University of Massachusetts Amherst

James Kurose Scholarship in Computer Science	2020
---	------

University of Massachusetts Amherst

Undergraduate Overseas Summer Research Scholarship	2018
---	------

Hong Kong Baptist University

Outstanding Student Scholarship	2017
--	------

Hong Kong Baptist University

Research Experience

Research Assistant, *University of Massachusetts Amherst* 09/2022 – Present
Advisor: Prof. Prashant Shenoy

- Developed SwapLess, a system enabling multi-tenant collaborative inference on memory-constrained Edge TPUs.
- Developed an analytic-model-driven resource manager for adaptive inference offloading in edge environments.
- Developed TailClipper, a queueing-theory-inspired scheduler that reduces tail latency of distributed applications.

Research Scientist Intern, *Adobe* 06/2026 – 09/2026
Mentors: Dr. Haoliang Wang, Dr. Kunjal Panchal, Dr. Jack Wang

- Developing privacy-preserving LLM inference offloading strategies within Adobe Acrobat.

Research Scientist Intern, *Adobe* 05/2022 – 08/2022
Mentor: Dr. Tung Mai

- Developed an Observability Data Management System deployed in production for Adobe engineering teams.
- Developed a URL knowledge extraction system using Random Forest to improve metrics reporting.

Data Science & Machine Learning Intern, *Adobe* 05/2021 – 12/2021
Mentor: Dr. Tung Mai

- Proposed a cache admission algorithm using a predictive statistical model to forecast customer revisits.
- Proposed dynamic Cloud–Edge user profile replication algorithms to maximize cache hit rates.

Student Research Collaborator, *Meta* 01/2021 – 05/2021
Mentor: Dr. Ritwik Yadav

- Benchmarked performance of RocksDB and LiveGraph using LinkBench on AWS EC2 instances.
- Collaborated on a potential deployment of LiveGraph as a Database-as-a-Service model at Meta.

Research Assistant, *University of Massachusetts Amherst* 09/2019 – 05/2022
Advisor: Prof. Marco Serafini

- Proposed data replication algorithms to satisfy user-defined tail latency bounds for distributed queries.
- Developed a distributed graph query processing middleware that coordinated query execution across nodes.

Research Assistant, *Hong Kong Baptist University* 09/2018 – 08/2019
Advisor: Prof. Byron Choi

- Proposed algorithms for efficient indexed maximum common subgraph mining in graph databases.

Undergraduate Student Researcher, *Hong Kong Baptist University* 09/2017 – 08/2018
Advisor: Prof. Byron Choi

- Proposed ranking and search algorithms to address the ‘why-not’ questions in graph query autocompletion.

Undergraduate Research Intern, *Nanyang Technological University* 05/2017 – 08/2017
Advisor: Prof. Sourav S. Bhowmick

- Developed FGGreat, an interactive graph query autocompletion system that assists query formulation.

Teaching Experience

CS 377: Operating Systems, Undergraduate Teaching Assistant Spring 2025
University of Massachusetts Amherst

- Led lab sessions for 50 undergraduates covering OS concepts and hands-on assignments.

CS 677: Distributed & Operating Systems, Graduate Teaching Assistant Spring 2023, 2024
University of Massachusetts Amherst

- Held office hours, graded coursework, and provided Piazza support for students.

CS 532: Systems for Data Science, Graduate Teaching Assistant Fall 2019, 2022, 2023
University of Massachusetts Amherst

- Designed and led lab sessions on Hadoop and Spark for 120 graduate students.

CS 2007: Object-Oriented Programming, Undergraduate Teaching Assistant Spring 2018
Hong Kong Baptist University

- Assisted students with hands-on debugging and Java programming exercises.

CS 1005: Essence of Computing, Undergraduate Teaching Assistant Fall 2017
Hong Kong Baptist University

- Assisted freshman students with programming exercises and debugging.

Mentoring

Sribatscha Maharana, B.Sc. and M.S. in Computer Science (UMass'27)	Spring 2026
Naga Surya Prathik Vunnava, M.S. in Computer Science (UMass'26)	Spring 2026
Balachandra Devarangadi Sunil, M.S. in Computer Science (UMass'26)	Fall 2025
Ayush Gupta, M.S. in Computer Science (UMass'26)	Fall 2025
Khai Nguyen, M.S. in Computer Science (UMass'25)	Spring 2025
Tanay Joshi, M.S. in Computer Science (UMass'25)	Spring 2025
Rishi Patel, B.Sc. in Computer Science (UMass'25)	Spring 2025

Service & Outreach

Artifact Evaluation Committee Member Fall 2026
ACM/IEEE International Symposium on Code Generation and Optimization (CGO'27)

Reviewer 2026
PeerJ Computer Science

External Reviewer Fall 2025
ACM/IEEE Symposium on Edge Computing (SEC'25)

Shadow PC Member Fall 2025
The European Conference on Computer Systems (EuroSys'26)

Instructor Summer 2024, 2025
UMass Summer Turing Program

References

Prashant Shenoy Distinguished Professor <i>University of Massachusetts Amherst</i> shenoy@cs.umass.edu	Don Towsley Distinguished Professor <i>University of Massachusetts Amherst</i> towsley@cs.umass.edu	David Irwin Professor <i>University of Massachusetts Amherst</i> irwin@ecs.umass.edu
Hui Guan Assistant Professor <i>University of Massachusetts Amherst</i> huiguan@cs.umass.edu	Tung Mai Senior Research Scientist <i>Adobe</i> tumai@adobe.com	