Tongping Liu (CS PhD @ UMass 2014)

Contact Information

tongping.cs@gmail.com (413)-695-2437

Education Background

Ph.D., Computer Science, University of Massachusetts Amherst, May 2014.

Thesis: Safe and Efficient Multithreading. Advisor: Emery D. Berger

Professional Experience

Architect and Technical Lead, ByteDance/TikTok, August 2022 - Now.

Associate Professor, Univ. of Massachusetts Amherst, Sept. 2022 - Now.

Assistant Professor, Univ. of Massachusetts Amherst, Sept. 2019 - Aug. 2022

Technical Consulting, CoCoPIE AI, Nov. 2021 - May 2022.

Assistant Professor, Univ. of Texas at San Antonio, Aug. 2014 - Aug. 2019.

Research/Engineering Intern: NEC Labs (2014), Futurewei (2013), IBM Research (2012), Samsung (2011), VMware (2009).

Leading Projects Highlight

Machine Learning Systems (6)

- *Accelerating Machine Learning Training:* Lead a new ML training framework in sharded data parallelism with better performance and generalization. In particular, it **outperforms ColossalAI and DeepSpeed up to** $2\times$ **for using GPU only, and up to** $10\times$ **for CPU-GPU hybrid mode**.
- *New GPU Memory Management:* Developed a novel memory management policy that could further **reduce GPU memory consumption by up to 35% without affecting the training accuracy**.
- *Investigating ML Inference Performance:* Investigating the better integration of FlashAttention and PagedAttention, which could further reduce memory consumption and reduce the latency.
- *Profiling for In-Production ML Systems:* Developed a novel CPU/GPU profiler that could identify the performance and memory issues within the whole system stack, such as GPU kernels, external C/C++ libraries, Python files. This profiler does not need the change of any programs, and only introduces 10% performance overhead.
- FreeLunch (*MCHPC'21*): actively compress/decompress intermediate results to reduce the memory footprint of ML training.
- AdapMTL: a multitask model pruning method that appreciates the difference of task heads and backbone

Failure Diagnosis and System Reliability (9)

- Watcher (*OOPSLA'20*): the first automatic failure diagnosis without human's involvement. Reported by Jiqizhixin, Sina, sohu, 163.com, kknews, thepaper, matpool, 51cto, linkerresearcher,....
- iReplyer (*PLDI'18*): the in-situ and identical record-and-replay system
- UnHang (*ISSTA'22*), UnDead (*ASE'17*): efficient tools for detecting communication and resource deadlocks
- Prober (*ASE'20*), CSOD (*CGO'19*), Sampler (*Micro'18*): memory failure detectors via compiler-assisted instrumentation, page protection, and hardware performance counters
- DoubleTake (*ICSE'16*): an evidence-based failure diagnosis to identify memory issues
- Dthreads (*SOSP'11*), Grace (*OOPSLA'09*): pioneer work in deterministic multithreading

Performance Profiling and Analysis (8)

- MemPerf (*OOPSLA'23*): identifying the performance issues induced by external memory allocators
- Scaler: a holistic performance profiler via cross-flow analysis
- CachePerf (*SIGMETRICS'22*): an efficient profiler to classify different cache misses
- NumaPerf (*ICS'21*): identifying NUMA issues via compiler instrumentation
- SyncPerf (*EuroSys'17*): identifying synchronization performance issues (e.g., locks)
- Cheetah (*CGO'16*), Predator (*PPoPP'14*), Sheriff (*OOPSLA'11*): identifying cache coherence problems using hardware performance monitoring units (PMUs), compiler-based instrumentation, and page protection on top of processes-as-threads

Memory Allocators (3)

- NUMAlloc (*ISMM'23*): a NUMA allocator that outperforms SOTA allocators by 15% for end-to-end performance.
- Guarder (*USENIX Security'18*), FreeGuard (*CCS'17*): secure memory allocators with tunable and efficient features

TEACHING EXPERIENCE

Distributed and Standalone Operating Systems (graduate): S15, F15, F17, F18, S22

Operating Systems (undergraduate): F16, S17, S18, F18, S21

Kernel Programming: S17, S18

Parallel and Distributed Software Systems: S16

Systems Programming: F14, F20

PATENTS (GRANTED: 10)

1. Improved Memory Management Method for Training Transformer Models.
2. Model-Aware Method and System for Training and/or Fine-tuning a Machine Learning Model
3. Classification of Different Types of Cache Misses. Tongping Liu, Jin Zhou, Steven Tang, Hanmei Yang. US Application No. 63/281,942.
4. A System and Method for Memory Allocation and Management In Non-Uniform Memory Access Architecture Computing Environment. Tongping Liu, Xin Zhao. US Application No. 63/200,062.
5. A Precise and Fully-Automatic On-Site Failure Diagnosis Method. Tongping Liu, Hongyu Liu, Sam Silvestro. US Patent 11,599,445.
6. Low-Overhead Detection Techniques for Synchronization Problems in Parallel and Concurrent Software. Tongping Liu, Mohammad Mejbah ul Alam, Abdullah Muzahid. US Patent 11,294,652.
7. Guarder: An Efficient Heap Allocator with Strongest And Tunable Security. Tongping Liu, Sam Silvestro, Hongyu Liu. US Patent 11,593,483.
8. FreeGuard: A Faster Secure Heap Allocator. Tongping Liu, Sam Silvestro, Hongyu Liu. US Patent 10,901,828.
9. Defeating Deadlocks in Production Software. Tongping Liu, Jinpeng Zhou, Sam Silvestro, Hongyu Liu. US Patent 10,915,424.
10. System and Method for Detecting False Sharing. Tongping Liu, Chen Tian, Ziang Hu. US Patent 9,678,883.
11. System and Method for Predicting False Sharing. Chen Tian, Tongping Liu, Ziang Hu. US Patent 9,547,599.
12. Prevention of race conditions in library code through memory page-fault handling mechanisms. Daniel G. Waddington, Chen Tian, Tongping Liu. US Patent App. 13/425,312.
13. Coupled Lock Allocation and Lookup for Shared Data Synchronization in Symmetric Multithreading Environments. Daniel G. Waddington, Tongping Liu, Chen Tian. US Patent 8,868,849.
14. Mapping Guest Pages to Disk Blocks to Improve Virtual Machine Management Process. Kiran Tati, Rajesh Venkatasubramanian, Carl A. Waldspurger, Alexander Thomas Garthwaite, Tongping Liu. US Patent 10,474,369.

MAJOR AWARDS AND MEDIA REPORT

Best Paper of DSA 2023

Watcher is reported by UMass News, Jiqizhixin, Sina, sohu, 163.com, kknews, thepaper, matpool, 51cto, linkerresearcher,.....

NSF Scalable Parallelism in the Extreme (SPX) Award - 2018

CRII Award - 2016

Google Faculty Award - 2015

Research Support

1. NSF CSR Proposal: *CSR: Medium: MemDrive: Memory-Driven Full-Stack Collaboration for Autonomous Embedded Systems*. **PI**, CCF-2215193, Total $1,000,000, UMass Share: $665,765, 10/01/2023-09/30/2026.
2. NSF EDU Proposal: *An Educational Tool for Teaching and Learning Concurrent Computer Programming Techniques* **Lead PI**, CCF-2215193, Total $300, 000, UMass Share: $179,970, 07/01/2022-06/30/2025.
3. NSF CCF Proposal: *SPX: Pinpointing and Resolving Scalability Culprits Hidden in Different Components of the Whole System Stack*, **Lead PI**, CCF-1823004, Total $999,883, UTSA Share: $499,992, 10/01/2018-9/30/2022
4. Mozilla Faculty Research Grant: *Guarder: Defending Heap Vulnerabilities with Flexible Guarantee and Better Performance*, **Sole PI**, Amount $51,073, 12/2017-unlimited.
5. Google Faculty Award: *Efficient, Effective, and Intelligent False Sharing Detection*, **Sole PI**, Amount $42,355, 08/2015-unlimited.
6. NSF CCF Proposal: *CRII: Evidence-Assisted Detection and Elimination of Memory Errors*, **Sole PI**, CCF-1566154, $206,731, 03/01/2016-02/28/2019

Professional Service

Program Committee: PLDI '24, HPDC'23, BigData'23, SC'22, PPoPP'22, HPDC'22, ICDCS' 21, HPDC'21, LCTES'21, ICICS'21, HPDC'20, BigData'20, LCTES'20, COMPSAC'19, ISMM'19, TrustCom'18, SETTA '18, ICSDE'17, ICPADS'16, ICCCN'16, ICCCN'15,

External Review Committee: PLDI 2019, PPOPP 2015

Program Organizer: Publicity Chair ICDCS'21

Reviewer: IEEE Transactions on Software Engineering, TACO, Journal of Systems and Software, Journal of Computational Science, NCS17, CCGrid 2017, TDSC, TPDS, IJICT, TCC, TC, ICDCS 2013, Middleware 2013, ICDCS 2014

Publications (e.g. SOSP, PLDI, EuroSys, CCS, Micro, ICSE...)

Underlined names are students supervised by me.

1. *Exploring Performance and Cost Optimization with ASIC-Based CXL Memory*
   Yupeng Tang, Ping Zhou, Wenhui Zhang, Henry Hu, Qirui Yang, Hao Xiang, Tongping Liu, Jiaxin Shan, Ruoyun Huang, Cheng Zhao, Cheng Chen, Hui Zhang, Fei Liu, Shuai Zhang, Xiaoning Ding, Jianjun Chen.
   The 19th European Conference on Computer Systems (**EuroSys'24**). Acceptance Rate: 15.9% (39/244).
2. *Improving Resource and Energy Efficiency for Cloud 3D through Excessive Rendering Reduction*
   Tianyi Liu, Jerry Lucas, Sen He, Tongping Liu, Xiaoyin Wang, Wei Wang.
   The 19th European Conference on Computer Systems (**EuroSys'24**). Acceptance Rate: 15.9% (39/244).
3. *Profile Dynamic Memory Allocation in Autonomous Driving Software* .
   Jin Zhou, Dexin Li, Tongping Liu.
   The 10th International Conference on Dependable Systems and Their Applications (**DSA 2023**). **Best Paper Award**.

4. *NUMAlloc: A Faster NUMA Memory Allocator*.
   Hanmei Yang, Xin Zhao, Jin Zhou, Wei Wang, Sandip Kundu, Bo Wu, Hui Guan, Tongping Liu.
   The Proceedings of the 2023 ACM SIGPLAN International Symposium on Memory Management (ISMM 2023).
5. *MemPerf: Profiling Allocator-Induced Performance Slowdowns*.
   Jin Zhou, Sam Sivestro, Steven Tang, Hongyu Liu, Guangming Zeng, Bo Wu, Cong Liu, Tongping Liu.
   Proceedings of the ACM on Programming Languages (**OOPSLA'23**).
6. *UnHang: Deadlock Prediction via Generalized Dependency*.
   Jinpeng Zhou, Hanmei Yang, John Lange, Tongping Liu.
   The 2022 International Symposium on Software Testing and Analysis (**ISSTA'22**). Acceptance Rate: 24.4% (61/250).
7. *CachePerf: A Unified Cache Miss Classifier via Hybrid Hardware Sampling*.
   Jin Zhou, Steven Tang, Hanmei Yang, Tongping Liu.
   The 2022 ACM SIGMETRICS/Performance conference (**SIGMETRICS'22**). Acceptance Rate: 20.6% (21/102).
8. *FreeLunch: Compression-based GPU MemoryManagement for Deep Neural Networks*.
   Shaurya Patel, Hui Guang, Tongping Liu.
   Workshop on Memory Centric High Performance Computing.
9. *Dryadic: Flexible and Fast Graph Pattern Matching at Scale*.
   Daniel Mawhirter, Sam Reinehr, Wei Han, Noah Fields, Miles Claver, Connor Holmes, Jedidiah McClurg, Tongping Liu, Bo Wu.
   The 30th International Conference on Parallel Architectures and Compilation Techniques (**PACT'21**), pp. 289-303.
10. *GraphZero: A High-Performance Subgraph Matching System*.
    Daniel Mawhirter, Sam Reinehr, Connor Holmes, Tongping Liu, Bo Wu
    ACM SIGOPS Operating Systems Review, Volume 55, pages 21-37.
11. *NumaPerf: Predictive NUMA Profiling*.
    Xin Zhao, Jin Zhou, Hui Guan, Wei Wang, Xu Liu, Tongping Liu.
    Proceedings of The 35th ACM International Conference on Supercomputing (**ICS'21**), pp. 52-62. Acceptance Rate: 24.2% (38/157).
12. *Watcher: In-Situ Failure Diagnosis for In-Production Software*.
    Hongyu Liu, Sam Silvestro, Xiangyu Zhang, Jian Huang, Tongping Liu.
    Proceedings of the ACM on Programming Languages (**OOPSLA'20**). Acceptance Rate: 36% (109/302).
13. *Prober: Practically Defending Overflows with Page Protection*.
    Hongyu Liu*, Ruiqin Tian*, Bin Ren, Tongping Liu. [*: equally-contributed]
    Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering (**ASE'20**). Acceptance Rate: 22.5% (93/408).
14. *CSOD: Context-sensitive Overflow Detection*.
    Hongyu Liu, Sam Silvestro, Xiaoyin Wang, Lide Duan, Tongping Liu.
    Proceedings of the 2019 IEEE/ACM International Symposium on Code Generation and Optimization ((**CGO'19**). Acceptance Rate: 31% (28/69).

15. *iReplyer: In-situ and Identical Record-and-Replay for Multithreaded Applications*
Hongyu Liu, Sam Silvestro, Wei Wang, Chen Tian, Tongping Liu.
Proceedings of The 37th annual ACM SIGPLAN conference on Programming Language Design and Implementation (**PLDI'18**). Acceptance Rate: 19.8% (55/277).

16. *Sampler: PMU-based Sampling to Detect Memory Errors Latent in Production Software*
Sam Silvestro, Hongyu Liu, Tong Zhang, Changhee Jung, Dongyoon Lee, Tongping Liu.
To appear in Proceedings of The 51th International Symposium on Microarchitecture (**Micro'18**). Acceptance Rate: 21.1%.

17. *Guarder: A Tunable Secure Allocator*
Sam Silvestro, Hongyu Liu, Tianyi Liu, Zhiqiang Lin, Tongping Liu.
Proceedings of The 27th USENIX Security Symposium (**Security'18**). Acceptance Rate: 19.1% (100/524).

18. *A User Space-based Project for Practicing Core Memory Management Concepts*
Sam Silvestro, Timothy T. Yuen, Corey Crosser, Dakai Zhu, Turgay Korkmaz, Tongping Liu.
Proceedings of The 49th ACM Technical Symposium on Computer Science Education (**SIGCSE'18**).

19. *FreeGuard: A Faster Secure Heap Allocator*
Sam Silvestro, Hongyu Liu, Corey Crosser, Zhiqiang Lin, Tongping Liu.
ACM Conference on Computer and Communications Security (**CCS'17**). Acceptance Rate: 18% (151/836).

20. *UnDead: Defeating Deadlocks of Production Software*
Jinpeng Zhou, Sam Silvestro, Hongyu Liu, Yan Cai, and Tongping Liu. The 32nd IEEE/ACM International Conference on Automated Software Engineering (**ASE'17**). Acceptance Rate: 21% (65/314).

21. *SyncPerf: Categorizing, Detecting, and Diagnosing Synchronization Performance Bugs*
Mejbah ul Alam*, Tongping Liu*, Guangming Zeng, Abduallah Muzahid. [*: equally-contributed]
The 2017 European Conference on Computer Systems (**EuroSys'17**). Acceptance Rate: 20.5% (41/200).

22. *DoubleTake: Fast and Precise Error Detection via Evidence-Based Dynamic Analysis*
Tongping Liu, Charlie Curtsinger, Emery D. Berger.
The 38th International Conference on Software Engineering (**ICSE'16**). Acceptance Rate: 19% (101/530).

23. *Cheetah: Detecting False Sharing Efficiently and Effectively*.
Tongping Liu*, Xu Liu*. [*: equally-contributed]
International Symposium on Code Generation and Optimization (**CGO'16**). Acceptance Rate: 23% (25/108).

24. *Predator: Predictive False Sharing Detection (Citation: 18)*
Tongping Liu, Chen Tian, Ziang Hu, Emery D. Berger.
Proceedings of the 19th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (**PPoPP'14**). Acceptance Rate: 16% (28/179).

25. *Dthreads: Efficient Deterministic Multithreading (Citation: 233)*
Tongping Liu, Charlie Curtsinger, Emery D. Berger.

Proceedings of the 23rd ACM Symposium on Operating Systems Principles (**SOSP'11**).
Acceptance rate:18% (28/153).

26. *Sheriff: Precise Detection and Automatic Mitigation of False Sharing (Citation: 63)*
Tongping Liu, Emery D. Berger.
Proceedings of the 26th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (**OOPSLA'11**).
Acceptance rate: 37% (61/166).

27. *Grace: Safe Multithreaded Programming for C/C++ (Citation: 301)*
Emery D. Berger, Ting Yang, Tongping Liu, Gene Novark.
Proceedings of the 24th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (**OOPSLA'09**).
Acceptance rate: 17% (25/144).

28. *Redline: First Class Support for Interactivity in Commodity Operating Systems (Citation: 81)*
Ting Yang, Tongping Liu, Emery D. Berger, Scott F. Kaplan, J.Eloit B.Moss.
Proceedings of the 8th USENIX Symposium on Operating Systems Design and Implementation (**OSDI'08**).
Acceptance rate: 13% (26/193).