

Shen Li

CONTACT INFORMATION

Facebook Inc.
770 Broadway, New York

Phone: (217) 550-2683
E-mail: cs.shenli@gmail.com

Most of my experiences lie in distributed systems. I am broadly interested in backend technologies, especially systems with some AI/ML flavor.

WORK

Facebook Inc., November 2018 - Present
Research Scientist

IBM Research, June 2016 - November 2018
Research Staff Member

University of Illinois at Urbana-Champaign, August 2010 - May 2016
Research Assistant

IBM Research, May 2015 - August 2015
Research Intern

IBM Research, May 2014 - August 2014
Research Intern

Facebook Inc., June 2013 - August 2013
Software Engineer Intern

Yahoo! Inc., May 2012 - August 2012
Software Engineer Intern

EDUCATION

University of Illinois at Urbana-Champaign, IL, USA May 2016

- Ph.D. in Computer Science
Advisor: Tarek F. Abdelzaher

Nanjing University, Jiangsu, China July 2010

- B.S. in Computer Science and Technology, Rank **1/160**

RESEARCH EXPERIENCES

IBM Research T. J. Watson Research Center June 2016-November 2018

- **IBM Streams**

IBM Streams is a high-throughput low-latency analytic platform for streaming data applications. I work on multiple research issues in IBM Streams, including transform graph translation and optimization, out-of-order event arrival processing, large window aggregation, etc. Currently, I lead a small team of four people to adopt Apache Beam model into IBM Streams, which involves filling in the model gaps between Beam and Streams, indexing/garbage-collecting Beam transform states, and managing operator parallelism. Our work produces both research papers and a product (IBM Streams Beam runner toolkit) deployed in IBM Cloud which won 2018 IBM Outstanding Technical Achievement Award.

University of Illinois at Urbana-Champaign August 2010-May 2016
Advisor: Tarek F. Abdelzaher

- **Energy Efficient Data Centers**

This project is sponsored by NSF, and in collaboration with VMWare. In this project, we built a 107-server cluster in a server room equipped with 40 UPS devices, 12 temperature sensors, a remotely controllable Computer Room Air Conditioning (CRAC) unit, and 6 smart PDUs. On this cluster, we design and evaluate energy-aware systems for data center workloads. More specifically, I am working on the following problems: (1) Thermal-aware workload placement, which assigns workloads to different servers leveraging the uneven thermal distribution in the cluster. (2) Power proportional memory cache cluster, that delivers elasticity and handles transition penalty. (3) Data center power capping using energy storage systems.

- **Real-time Scheduling for Map-Reduce Systems**

This project, in collaboration with Yahoo!, introduces the marriage between real-time scheduling and Map-Reduce systems. Map-Reduce ecosystems have become dominant big data an-

alytic solutions. Nevertheless, schedulers in the state-of-the-art Map-Reduce systems remain relatively simple, which often fails to utilize resources in large clusters efficiently. Based on the advances in the real-time community, we develop both systems and theories to improve Map-Reduce workflow scheduling by enabling the utilization-based schedulability analysis.

- **Apollo: A Data Distillation Service for Social Sensing**

This project is one of the very few “Golden Nuggets” projects of the Network Science Collaborative Technology Alliance (NS-CTA), a collaborative research alliance between the US Army Research Laboratory (ARL) and other research organizations such as UIUC, CMU, and IBM Research. In this project, we propose Apollo, a new service for social sensing applications, which aims at distilling large amounts of noisy social sensory data into smaller amounts of more trustable information. Particularly, I am working on improving the scalability of the system, and at the same time preserving the Quality of Information (QoI).

- **SmartRoad: A Crowd-Sourced Road Sensing System**

This project is sponsored by NSF, Siebel Foundation, and in collaboration with Microsoft Research. In this project, we design and develop a vehicular smartphone testbed, called SmartRoad, which can automatically collect, transport, and analyze participatory sensing data for decision making. SmartRoad is currently running on 35 external participatory vehicles, and will soon be extended to a larger-scale (100-200 vehicles) deployment with the UIUC Facilities and Services Department. SmartRoad facilitates various research applications such as (1) GreenGPS: a navigation service providing the most fuel-efficient routes for vehicles between an arbitrary pair of end-points. (2) Providing energy efficient navigation service that takes advantage of phone’s low-power MEMS sensors in place of GPS.

IBM Thomas J. Watson Research Center

May 2015-August 2015

Mentor: Raghu Ganti, Manager: Mudhakar Srivatsa

- **Stark: Optimizing In-Memory Computation For Online User Queries**

In this project, we explored the possibility of combining in-memory computation frameworks with the backend of online services to answer massive user queries in real-time. Success along this path would push online services beyond the existing store-index-retrieve design pattern, allowing the backend to afford relatively complex data analytic algorithms to serve every user query with highly optimized results. This project tackles this problem by optimizing Apache Spark for online queries.

IBM Thomas J. Watson Research Center

May 2014-August 2014

Mentor: Raghu Ganti, Manager: Mudhakar Srivatsa

- **Pyro: A Spatial-Temporal Big-Data Storage System**

In this project, we designed a spatial-temporal big-data storage system tailored for high-resolution geometry queries and dynamic workload hotspots. With the rapid growth of mobile devices and applications, geo-tagged data has become a significant workload for big data storage systems. In dealing with spatial-temporal big-data workloads, existing systems either fall short in scalability or fail to deliver high efficiency. This project attacks this problem by optimizing the HBase/HDFS stack for spatial-temporal data.

PUBLICATIONS

VLDB18

Shen Li, Paul Gerver, John MacMillan, Daniel Debrunner, William Marshall, Kun-Lung Wu, “Challenges and Experiences in Building an Efficient Apache Beam Runner For IBM Streams”.

UBICOMP18

Shuochao Yao, Yiran Zhao, Huajie Shao, Aston Zhang, Chao Zhang, **Shen Li**, Tarek Abdelzaher, “RDeepSense: Reliable Deep Mobile Computing Models with Uncertainty Estimations”.

ICDCS17

Shen Li, Md Tanvir Amin, Raghu Ganti, Mudhakar Srivatsa, Shaohan Hu, Tarek Abdelzaher, “Stark: Optimizing In-Memory Computation For Online User Queries”.

UBICOMP17

Yiran Zhao, Shuochao Yao, **Shen Li**, Shaohan Hu, Huajie Shao, Tarek F Abdelzaher, “VibeBin: A Vibration-Based Waste Bin Level Detection System”.

- ICDCS17 Huajie Shao, Shiguang Wang, **Shen Li**, Shuochao Yao, Yiran Zhao, Md Tanvir Al Amin, Tarek Abdelzaher and Lance Kaplan, “Optimizing Source Selection in Social Sensing in the Presence of Influence Graphs”.
- ICCP17 Yiran Zhao, **Shen Li**, Shaohan Hu, Lu Su, Shuochao Yao, Huajie Shao and Tarek Abdelzaher, “GreenDrive: A Smartphone-based Intelligent Speed Adaptation System With Real-time Traffic Signal Prediction”.
- EWSN17 Yiran Zhao, Shuochao Yao, **Shen Li**, Shaohan Hu, Huajie Shao, Tarek F. Abdelzaher, “Unsupervised Fill-level Estimation for Smart Trash Removal Systems.”.
- VLDB16 Yiran Zhao, **Shen Li**, Shaohan Hu, Hongwei Wang, Shuochao Yao, Huajie Shao, Tarek Abdelzaher, “An Experimental Evaluation of Datacenter Workloads On Low-Power Embedded Micro Servers”.
- ICDCS16 Shuochao Yao, Shaohan Hu, **Shen Li**, Yiran Zhao, Lu Su, Lance Kaplan, Aylin Yener, Tarek F. Abdelzaher, “On Source Dependency Models for Reliable Social Sensing: Algorithms and Fundamental Error Bound”.
- IPSN16 Shuochao Yao, Md Tanvir Amin, Lu Su, Shaohan Hu, **Shen Li**, Shiguang Wang, Yiran Zhao, Tarek Abdelzaher, Lance Kaplan, Charu Aggarwal, Aylin Yener, “Recursive Ground Truth Estimator for Social Data Streams”.
- TMC16 Fatemeh Saremi, Omid Fatemieh, Hossein Ahmadi, Hongyan Wang, Tarek Abdelzaher, Raghu Ganti, Hengchang Liu, Shaohan Hu, **Shen Li**, and Lu Su, “Experiences with GreenGPS – Fuel-Efficient Navigation using Participatory Sensing”.
- ATC15 **Shen Li**, Shaohan Hu, Raghu Ganti, Mudhakar Srivatsa, Tarek Abdelzaher, “Pyro: A Spatial-Temporal Big-Data Storage System”.
- RTAS15 **Shen Li**, Shaohan Hu, Tarek Abdelzaher, “The Packing Server for Real-time Scheduling of MapReduce Workflows”.
- UBICOMP15 Shaohan Hu, Lu Su, **Shen Li**, Shiguang Wang, Chenji Pan, Siyu Gu, Md Tanvir Amin, Hengchang Liu, Suman Nath, Romit Roy Choudhury, Tarek F. Abdelzaher, “Experiences with eNav: A Low-power Vehicular Navigation System”.
- IPSN15 Shiguang Wang, Lu Su, **Shen Li**, Shaohan Hu, Tanvir Amin, Hongwei Wang, Shuochao Yao, Lance Kaplan, Tarek Abdelzaher, “Scalable Social Sensing of Interdependent Phenomena”.
- RTSS15 Shaohan Hu, Shuochao Yao, Haiming Jin, Yiran Zhao, Yitao Hu, Xiaochen Liu, Nooreddin Naghibolhosseini, **Shen Li**, Akash Kapoor, William Dron, Lu Su, Amotz Bar-Noy, Pedro Szekely, Ramesh Govindan, Reginald Hobbs, Tarek F. Abdelzaher, “Data Acquisition for Real-time Decision-making under Freshness Constraints”.
- DCOSS15 Shaohan Hu, **Shen Li**, Shuochao Yao, Lu Su, Ramesh Govindan, Reginald Hobbs, Tarek F. Abdelzaher, “On Exploiting Logical Dependencies for Minimizing Additive Cost Metrics in Resource-Limited Crowdsensing”.
- ICAC15 Md Tanvir Amin, **Shen Li**, Muntasir Raihan Rahman, Panindra Seetharamu, Shiguang Wang, Tarek Abdelzaher, Indranil Gupta, Mudhakar Srivatsa, Raghu Ganti, Reaz Ahmed, Hieu Le, “SocialTrove: A Self-summarizing Storage Service for Social Sensing”.
- CALDAM15 Mindi Yuan, Wei Shen, Jun Li, Yannis Pavlidis, **Shen Li**, “Auction/Belief propagation algorithms for constrained assignment problem”.
- ICDCS14 **Shen Li**, Shaohan Hu, Shiguang Wang, Lu Su, Tarek Abdelzaher, Indranil Gupta, Richard Pace, “WOHA: Deadline-Aware Map-Reduce Workflow Scheduling Framework over Hadoop Cluster”.
- ICAC14 **Shen Li**, Shaohan Hu, Shiguang Wang, Siyu Gu, Chenji Pan, Tarek Abdelzaher, “WattValet: Heterogenous Energy Storage Management in Data Centers for Improved Power Capping”.
- ICCCN14 **Shen Li**, Lu Su, Yerzhan Suleimenov, Hengchang Liu, Tarek Abdelzaher and Guihai Chen, “Centaur: Dynamic Message Dissemination over Online Social Networks”.

- IPSN14 Dong Wang, Tanvir Amin, **Shen Li**, Tarek Abdelzaher, Lance Kaplan, Siyu Gu, Chenji Pan, Hengchang Liu, Charu Aggarwal, Raghu Ganti, XinLei Wang, Prasant Mohapatra, Boleslaw Szymanski, Hieu Le, “Humans as Sensors: An Estimation Theoretic Perspective”.
- DCOSS14 Shiguang Wang, Tarek Abdelzaher, Santhosh Gajendran, Ajith Herga, Sachin Kulkarni, **Shen Li**, Hengchang Liu, Chethan Suresh, Abhishek Sreenath, Hongwei Wang, William Dron, Alice Leung, Ramesh Govindan, John Hancock, “The Information Funnel: Exploiting Named Data for Information-maximizing Data Collection”.
- DCOSS14 Siyu Gu, Chenji Pan, Hengchang Liu, **Shen Li**, Shaohan Hu, Lu Su, Shiguang Wang, Dong Wang, Tanvir Amin, Ramesh Govindan, Charu Aggarwal, Raghu Ganti, Mudhakar Srivatsa, Amotz Barnoy, Peter Terlecky, Tarek Abdelzaher, “Data Extrapolation in Social Sensing for Disaster Response”.
- IPSN14 Shaohan Hu, Lu Su, **Shen Li**, Shiguang Wang, Chenji Pan, Siyu Gu, Md Tanvir Al Amin, Hengchang Liu, Suman Nath, Romit Roy Choudhury, Tarek F. Abdelzaher, “eNav – a Smartphone-based Energy Efficient Vehicular Navigation System”, poster.
- ICDCS13 **Shen Li**, Shiguang Wang, Fan Yang, Shaohan Hu, Fatemeh Saremi, Tarek Abdelzaher, “Proteus: Power Proportional Memory Cache Cluster in Data Centers”.
- ICCCN13 Shiguang Wang, Shaohan Hu, **Shen Li**, Hengchang Liu, Md Yusuf Sarwar, and Tarek Abdelzaher, “MINERVA: Information-Centric Programming for Social Sensing”.
- SENSYS13 Hengchang Liu, Siyu Gu, Chenji Pan, Wei Zheng, **Shen Li**, Shaohan Hu, Shiguang Wang, Dong Wang, Tanvir Amin, Lu Su, Zhiheng Xie, Ramesh Govindan, Charu Aggarwal, Amotz Barnoy, Tarek Abdelzaher, “Extrapolation from Participatory Sensing Data”, demo.
- ICDCS13 **Shen Li**, Hieu Le, Nam Pham, Jin Heo, and Tarek Abdelzaher, “Joint Optimization of Computing and Cooling Energy: Analytic Model and A Machine Room Case Study”.
- JSC12 **Shen Li**, Shiguang Wang, Tarek Abdelzaher, Maria Kihl, and Anders Robertsson, “Temperature Aware Power Allocation: An Optimization Framework and Case Studies”.
- RTSS12 Lu Su, Shaohan Hu, **Shen Li**, Feng Liang, Jing Gao, Tarek Abdelzaher, and Jiawei Han, “Quality of Information based Data Selection and Transmission in Wireless Sensor Networks”.
- IGCC11 **Shen Li**, Tarek Abdelzaher, and Mindi Yuan, “TAPA: Temperature Aware Power Allocation in Data Center with MapReduce”, IEEE **International Green Computing Conference**, 2011.
- ICDCS11 Mohammad Khan, Jin Heo, **Shen Li**, and Tarek Abdelzaher, “Understanding Vicious Cycles in Server Clusters”.
- ICCCN11 Xiaobing Wu, **Shen Li**, and Guihai Chen, “Aggregation-Friendly Data Collection Protocol of Wireless Sensor Networks for PoI Monitoring”.
- JISE11 **Shen Li**, Andong Zhan, Xiaobing Wu, Panlong Yang, and Guihai Chen, “Efficient Emergence Rescue Navigation with Wireless Sensor Networks”, Journal of Information Science and Engineering (JISE), Vol. 27, No. 1, 2011.
- CLUSTER11 Phil Miller, **Shen Li**, Chao Mei, “Asynchronous Collective Output with Non-dedicated Cores”.
- ICPADS09 **Shen Li**, Andong Zhan, Xiaobing Wu, and Guihai Chen, “ERN: Emergence Rescue Navigation with Wireless Sensor Networks”.
- MOBICOM09 Andong Zhan, **Shen Li**, Wenchang Zhou, Tao Li, Panlong Yang, Xiaobing Wu and Guihai Chen, “iBrush: Writing Chinese Characters with a Flashlight and a Wireless Sensor Grid”, demo.
- MASS09 Andong Zhan, **Shen Li**, Tao Li, Lubin Guan, Panlong Yang, Xiaobing Wu and Guihai Chen, “iTracking: accurate light-based location-tracking in wireless sensor networks”.

HONORS AND
AWARDS

IBM Outstanding Technical Achievement Award, 2018
ACM/IEEE ICCPS best paper award, 2017
IBM Manager's Choice Award, 2016, 2017
Feng Chen Memorial Award, 2016
IBM Ph.D. Scholarship, 2015
USENIX ICAC best paper award, 2015
ACM IPSN best poster award, 2014
Outstanding Bachelor's Degree Thesis in **Jiangsu Province**, 2011
Outstanding Bachelor's Degree Thesis in Nanjing University (9/3600 @ NJU), 2010
ACM MobiCom Demo Session Honored Mention, 2009
National Scholarship by Ministry of Education and Ministry of Finance, 2007, 2008, 2009
ACM International Collegiate Programming Contest **World Final** Stockholm, 49th Place, 2009
ACM International Collegiate Programming Contest Chengdu Site **Gold Medal**, Contest Hefei Site **Gold Medal**, 2008
"Changliangbei" **National** Collegiate Programming Contest **1st Prize**, 2008
Collegiate Programming Contest of **Jiangsu Province**, **1st Prize**, 2008
Mathematical Contest in Modeling **Honorable Mention**, 2008
ACM International Collegiate Programming Contest Nanjing Site **Gold Medal**, Changchun Site **Silver Medal**, 2007
Exceptional Student in Nanjing University (1/1000 @ NJU), 2007
Outstanding Student in Nanjing University, 2007
Outstanding Student of CS Dept, 2006, 2007

PROFESSIONAL
ACTIVITIES

Conference Committee Member
CONNET'15'16, IEEE ICCCN'17'18'19, ICDCS'18, IC2E'18; ACM EdegSys'18'19
External Reviewer
IEEE Transactions on Computers
ACM Transactions on Storage
ACM Transactions on Cyber-Physical Systems
IEEE Transactions on Mobile Computing
ACM Transactions on Autonomous and Adaptive Systems
IEEE Transactions on Industrial Informatics
IEEE Transactions on Network and Service Management
IEEE Transactions on Dependable and Secure Computing
IEEE Access
Elsevier Knowledge-Based Systems
Journal of Distributed Sensor Networks
Journal of Parallel and Distributed Computing
Journal of Information Processing Systems
Journal of Applied Science and Engineering
Frontiers of Information Technology and Electronic Engineering
Conferences: ICDCS, INFOCOM, MASS, SenSys, ICCCN, NVMSA, EdgeSys