

# Wei Xu

[weixu@tsinghua.edu.cn](mailto:weixu@tsinghua.edu.cn)  
Tsinghua University

## WORKING EXPERIENCE

### **Tsinghua University**

Assistant Professor at Institute for Interdisciplinary Information Sciences IIIS  
Research Area: Distributed Systems + Machine Learning

Administrative work:

- Assistant Dean of IIIS
- Director of International Partnership, Tsinghua Online Education Office
- Supervisory Committee Member, Tsinghua Data Science Institute
- Director, Open Compute Lab at Tsinghua
- Associate Director, Institute for Financial Technology and Director of its Blockchain Research Center, Tsinghua University

Beijing, China  
Feb, 2013 – Now

Jun, 2014 - Now  
Jun, 2013 – Dec, 2016  
Sept, 2015 – Now  
Jan, 2015 – Dec, 2017  
Nov, 2017 - Now

### **Google**

Software Engineer

-Design and implementation of Google's next generation debug logging and cluster monitoring infrastructure.

Mountain View, CA  
Aug, 2010 to Feb, 2013

### **Netease Beijing R&D Center**

Software Research Engineer

- Designer and lead engineer for a large scale, distributed storage system
- Designer and lead engineer for image search (<http://image.yodao.com>)

Beijing, China  
Jun – Dec, 2006

### **HP Labs**

Research Intern

- Research on predictive control in virtual machine resource allocations

Palo Alto, CA  
Jun - Aug, 2005

## EDUCATION

### **University of California at Berkeley**

**Ph.D. in Computer Science**

Advisors: Prof. David Patterson and Prof. Armando Fox

Dissertation: Detecting Large Scale System Problems by Mining Console Logs

Certificate in Management of Technology (MOT)

Berkeley, CA  
Dec. 2005 – July, 2010

### **University of California at Berkeley**

**M.S. in Computer Science, GPA: 3.95 / 4.00**

Research Advisor: Prof. David Patterson

Berkeley, CA  
Aug, 2003- Dec. 2005

### **University of Pennsylvania**

**B.S. Eng. in Computer Science, Summa Cum Laude, GPA 4.00 / 4.00**

with Electrical Engineering Minor

Philadelphia, PA  
Sept. 2001- May. 2003

### **Tsinghua University**

Major in Computer Science

Beijing, China  
Sept. 1999 - Aug. 2001

## PUBLICATIONS

Charley Chen, Guosai Wang, Jiao Sun and Wei Xu. **Data-Driven Data Center Temperature Modeling and Prediction**. In Proc. of *ACM SIGOPS Asia-Pacific Workshop on Systems (APSys) 2018*, Jeju Island, South Korea.

Zhengyang Song, Yongzheng Jia and Wei Xu. **Exploring Business Models and Dynamic Pricing Frameworks for SPOC**

**services.** In Proceedings of the *2nd International Workshop on Data Management and Mining on MOOCs (DMMOOC)* 2018

Guosai Wang, Shiyang Xiang, Yitao Duan, Ling Huang and Wei Xu. **Do Not Pull My Data for Resale: Protecting Data Providers using Data Retrieval Pattern Analysis.** In Proceedings of *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)*. (Short paper)

Jiao Sun, Qixin Zhu, Zhifei Liu, Xin Liu, Yueming Wang, Jihae Lee, Lei Shi, Ling Huang and Wei Xu. **FraudVis: Understanding Unsupervised Fraud Detection Algorithms.** In Proc. of the *11th IEEE Pacific Visualization Symposium (PacificVis 2018)*

Yiran Li, Da Wei, Xiaoqi Chen, Ziheng Song, Ruihan Wu, Yuxing Li, Xin Jin and Wei Xu. **DumbNet: A Smart Data Center Network Fabric with Dumb Switches.** In Proc. of the *European Conference on Computer Systems (EuroSys)* 2018

Yongzheng Jia, Xue Liu and Wei Xu. **When Online Dating Meets Nash Social Welfare: Achieving Efficiency and Fairness.** In Proc. of the *Web Conference (WWW)* 2018

Shun Zheng, Jialei Wang, Fen Xia, Wei Xu and Tong Zhang. **A General Distributed Dual Coordinate Optimization Framework for Regularized Loss Minimization.** In *Journal of Machine Learning Research (JMLR)* 2017

Yi Li, Yitao Duan, and Wei Xu. **PEM: Practical Differentially Private System for Large-Scale Cross-Institutional Data Mining.** In Proceedings of *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD)* 2017

Shuhao Wang, Cancheng Liu, Xiang Gao, Hongtao Qu and Wei Xu. **Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks.** In Proceedings of *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD)* 2017

Yuhan Su, Zhongming Jin, Ying Chen, Xinghai Sun, Yaming Yang, Fangzheng Qiao, Fen Xia and Wei Xu. **Improving Click-Through Rate Prediction Accuracy in Online Advertising by Transfer Learning.** In Proceedings of the *IEEE/WIC/ACM International Conference on Web Intelligence (WI)* 2017

Yong Cheng, Qian Yang, Yang Liu, Maosong Sun and Wei Xu. **Joint Training for Pivot-based Neural Machine Translation.** In Proceedings of the *International Joint Conference on Artificial Intelligence (IJCAI)* 2017

Yuxi Dong, Yuchao Pan, Jun Zhang and Wei Xu. **Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN.** In Proceedings of the *2nd International Workshop on Big Data Analytics for Smart and Connected Health (BIGDATA4HEALTH)* 2017

Yuxi Dong, Yuchao Pan, Xihai Zhao, Rui Li, Chun Yuan and Wei Xu. **Identifying Carotid Plaque Composition in MRI with Convolutional Neural Networks.** In Proceedings of the *3rd IEEE International Conference on Smart Computing (SMARTCOMP)* 2017

Guosai Wang and Wei Xu. **What Can We Learn from Four Years of Data Center Hardware Failures?** In Proceedings of the *47th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* 2017 **[BEST PAPER AWARD]**

Yongzheng Jia, Wei Xu and Xue Liu. **An Optimization Framework For Online Ride-sharing Markets.** In Proceedings of the *37th IEEE International Conference on Distributed Computing (ICDCS)* 2017

Yang Zhang, Tingjian Zhang, Yongzheng Jia, Jiao Sun, Fangzhou Xu and Wei Xu. **DataLab: Introducing Software Engineering Thinking into Data Science Education at Scale.** In Proceedings of the *39th International Conference on Software Engineering - Software Engineering Education and Training Track (ICSE-SEET)* 2017

Yongzheng Jia, Zhengyang Song, Xiaolan Bai and Wei Xu. **Towards Economic Models for MOOC Pricing Strategy Design** In Proceedings of the *1st International Workshop on Data Management and Mining on MOOCs (DMMOOC)* 2017

Yong Cheng, Yang Liu and Wei Xu. **Maximum Reconstruction Estimation for Generative Latent Variable Models.** In Proc. of *AAAI*, 2017

Qingtang Xia, Tianjia Chen and Wei Xu. **CIDS: Adapting Legacy Intrusion Detection Systems to the Cloud with Hybrid Sampling.** In Proceedings of *the 6th IEEE International Symposium on Cloud and Service Computing (SC2)* 2016

Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun and Yang Liu. **Semi-supervised Learning for Neural Machine Translation.** In Proc. of *ACL'2016*.

Yong Xiang, Hu Li, Sen Wang, Charley Peter Chen and Wei Xu. **Debugging OpenStack Problems Using a State Graph Approach.** In Proc. of *ACM SIGOPS Asia-Pacific Workshop on Systems (APSys) 2016* Hong Kong, China, 2016. [BEST PAPER AWARD]

Peilun Li and Wei Xu. **Optimizing Hash-based Distributed Storage Using Client Choices.** In Proc. of *ACM SIGOPS Asia-Pacific Workshop on Systems (APSys) 2016* Hong Kong, China, 2016.

Xin Jin, Yiran Li, Da Wei, Siming Li, Jie Gao, Lei Xu, Guangzhi Li, Wei Xu, Jennifer Rexford. **Optimizing Bulk Transfers with Software-Defined Optical WAN.** In Proc. of *ACM Sigcomm 2016* Brazil, August, 2016

Yi Li, Hong Liu, Wenjun Yang, Dianming Hu, Xiaojing Wang and Wei Xu. **Predicting Inter-Data-Center Network Traffic Using Elephant Flow and Sublink Information.** In *IEEE Transactions on Network and Service Management*, 13, no. 4 (2016): 782-792

Da Wei, Lei Xu, Xin Jin, Yiran Li and Wei Xu. **A 12-Rack, 180-Server Datacenter Network (DCN) Using Multiwavelength Optical Switching and Full Stack Optimization.** In proceedings of *Optical Fiber Communication Conference (OFC), (Postdeadline Paper - PDP)* Anaheim, USA, March, 2016

Yang Zhang, Fangzhou Xu, Erwin Frise, Siqi Wu, Bin Yu and Wei Xu. **DataLab: A Version Data Management and Analytics System,** in proceedings of *ICSE first International Workshop on BIG Data Software Engineering*, Austin, USA, May, 2016

Hu Li and Wei Xu. **Improving Spark Performance with Zero-copy Buffer Management and RDMA.** In proceedings of *IEEE INFOCOM First International Workshop on Big Data Sciences, Technologies and Applications (BDSTA 2016)*, San Francisco, USA, Apr, 2016

Yuchao Pan, Yuxi Dong, Jingtian Zhou, Mark Hallen, Bruce R. Donald, Jianyang Zeng and Wei Xu. **cOSPREY: A Cloud-Based Distributed Algorithm for Large-Scale Computational Protein Design.** In *Journal of computational biology (JCB)*, 2016

Guosai Wang, Shuhao Wang, Bing Luo, Xin Jin, Yinghang Zhu, Wenjun Yang, Longbo Huang, Weisong Shi, Dianming Hu and Wei Xu. **Increasing Large-Scale Data Center Capacity by Statistical Power Control.** In proceedings of *the European Conference on Computer Systems (EuroSys 2016)*, London, UK, Apr, 2016

Xiaofeng Lin, Yu Chen, Xiaodong Li, Junjie Mao, Wei Xu, Jiaquan He, and Yuanchun Shi, **Scalable Kernel TCP Design and Implementation for Short-Lived Connections.** In proceedings of *the 21th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2016)*, Atlanta, USA, Apr, 2016

Yi Li, Hong Liu, Wenjun Yang, Dianming Hu and Wei Xu. **Inter-Data-Center Network Traffic Prediction with Elephant Flows.** In proceedings of *IEEE/IFIP Network Operations and Management Symposium (NOMS 2016)*, Istanbul, Turkey, Apr, 2016

Quanquan Zhi, Wei Xu. **MED: The Monitor-Emulator-Debugger for Software-Defined Networks.** In proceedings of *IEEE International Conference on Computer Communications (INFOCOM 2016)*, San Francisco, USA, Apr, 2016

Yichao Zhou, Wei Xu, Bruce R. Donald, Jianyang Zeng. **An Efficient Parallel Algorithm for Accelerating Computational Protein Design.** In proceedings of *ISMB 2014*. Boston, Massachusetts, USA, July 2014. *Bioinformatics*.

Adam Oliner, Archana Ganapathi, Wei Xu. **Advances and Challenges in Log Analysis.** In *Communications of ACM (CACM)*, Invited article, Feb, 2012

Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan. **Experience on Mining Google's Production Console Logs.** In proceedings of the *Workshop on Managing Systems via Log Analysis and Machine Learning Techniques (SLAML '10)*, Vancouver, BC Oct, 2010

Ariel Rabkin, Avani Wildani, Randy Katz, Wei Xu, Armando Fox. **A graphical representation for identifier structure in logs.** In proc. of the *Workshop on Managing Systems via Log Analysis and Machine Learning Techniques (SLAML '10)*, Vancouver, BC

Oct, 2010

Wei Xu. **Detecting Large Scale System Problems by Mining Console Logs**. PhD dissertation, UC Berkeley, July, 2010

Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan. **Using Machine Learning Techniques in Console Log Analysis**. Invited paper in *the 27th International Conference on Machine Learning (ICML'10)*, Haifa, Israel, June 2010

Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan. **Online System Problem Detection by Mining Patterns of Console Logs**. In Proc. of *the IEEE International Conference on Data Mining (ICDM'09)*, Miami, FL, December 2009

Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan. **Large-scale system problem detection by mining console logs**. In Proc. of *the 22nd ACM Symposium on Operating Systems Principles (SOSP'09)*, Big Sky, MT, October 2009

Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan. **Mining console logs for large-scale system problem detection**. In Proc. of *the 3<sup>rd</sup> workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML'08)*, San Diego, CA, December 2008

Wei Xu, Zhangxi Tan, Armando Fox and David Patterson. **Regulating workload in J2EE application servers**. In Proc. of *the 1st International Workshop on Feedback Control Implementation and Design in Computing Systems and Networks (FeBID'06)*, Vancouver, Canada, April 2006

Wei Xu, Xiaoyun Zhu, Sharad Singhal, and Zhikui Wang. **Predictive control for dynamic resource allocation in enterprise data centers**. In Proc. of *the 10th IEEE/IFIP Network Operations & Management Symposium (NOMS'06)*, Vancouver, BC, Apr. 2006

Wei Xu. **Feedback control theory and processing system log streams**. Master thesis, EECS Department, UC Berkeley, December, 2005

Wei Xu, Joseph L. Hellerstein, Bill Kramer and David Patterson. **Control considerations for scaling event processing**. In Proc. of *the 16th IFIP/IEEE Distributed Systems: Operations and Management (DSOM'05)*, Barcelona, Spain, October 2005

Wei Xu, Peter Bodik and David Patterson. **A flexible framework for statistical learning and data mining from system log streams**. In Proc. of *Workshop on Temporal Data Mining: Algorithms, Theory and Applications at The Fourth IEEE International Conference on Data Mining (ICDM'04)*, Brighton, UK, Nov, 2004

Bjorn Knutsson, Honghui Lu, Wei Xu and Bryan Hopkins. **Peer-to-Peer support for massively multiplayer games**. In Proc. of *the 23rd Conference of the IEEE Communications Society (INFOCOM'04)*, Hong Kong, March 2004

## NON-TECHNICAL PUBLICATIONS ON EDUCATION

徐葳. **建立与研究生的信任——写在我的第一个博士生毕业之际**. In *科技导报* 2017, 35 (18) (In Chinese)

Wei Xu, Yongzheng Jia, Armando Fox and David Patterson. **From MOOC to SPOC: Lessons from MOOC at Tsinghua and UC Berkeley** (In Chinese). In *Modern Distance Education Research*, 2014 (In Chinese)

## PROFESSIONAL ACTIVITIES

### Program Committee Chair / Organizer

|  |           |
|--|-----------|
| <i>Workshop "Hot Topics in Big Data and Networking (HotData-1), collocated with ICCCN</i>    | 2014      |
| <i>IEEE Big Data Science Conference</i>  | 2014      |
| <i>Workshop "Dependability and Security of System Operation" (DSSO), collocated with DSN</i> | 2014-2016 |

### Guest Editor

|  |           |
|--|-----------|
| <i>Chinese Science Bulletin, Special issue on Data Science</i> | 2014-2015 |
|--|-----------|

### Program Committee Member

|   |            |
|---|------------|
| <i>Strata + Hadoop World</i>  | 2016, 2017 |
| <i>IEEE International Conference on Parallel and Distributed Systems, ICPADS'14</i>         | 2014       |
| <i>International Conference on Computer Communications and Networks, ICCCN'14</i>           | 2014       |
| <i>Asia-Pacific Workshop on Systems</i>   | 2014       |
| <i>19th IEEE International Conference on Parallel and Distributed Systems (ICPADS 2013)</i> | 2013       |
| <i>Workshop on Managing Systems via Log Analysis and Machine Learning Techniques</i>        | 2010-2012  |

## **HONORS AND AWARDS**

|   |           |
|---|-----------|
| <b>Beijing Pioneer in Teacher's Ethics</b>                                      | 2018      |
| <b>Tsinghua University Lab Design Award (3<sup>rd</sup> class)</b>              | 2018      |
| <b>Tsinghua University Excellent Performance Employee Award</b>                 | 2017      |
| <b>DSN Best Paper Award</b>   | 2017      |
| <b>Tsinghua University Award for Graduate Student Advising (Special grade)</b>  | 2016      |
| <b>Tsinghua University Undergrad Class Advisor Award (2<sup>nd</sup> class)</b> | 2016      |
| <b>APSys Best Paper Award</b>   | 2016      |
| <b>IBM Faculty Award, IBM</b>   | 2015      |
| <b>Microsoft Research Asia Collaborative Research Awards, Microsoft</b>         | 2014      |
| <b>Google Faculty Research Award, Google</b>                                    | 2014      |
| <b>"Youth Talent 1000" Scholar Award, Chinese central government program</b>    | 2013      |
| <b>Department Fellowship, EECS Department, UC Berkeley,</b>                     | 2004      |
| <b>Faculty Appreciation Award, Univ. of Penn</b>                                | 2003      |
| <b>Dean's List, School of Engineering and Applied Science, Univ. of Penn</b>    | 2002-2003 |
| <b>Tau Beta Pai Honor Society and Golden Key Honor Society</b>                  | 2002      |