Chen Wang

IBM T. J. Watson Research Center
1101 Kitchawan Rd
Yorktown Heights, NY 10598

Office: 914-945-1773, Cell: 914-792-8658
Email: Chen.Wang1@ibm.com, wangchen615@gmail.com
LinkedIn: https://www.linkedin.com/in/chenw615/
Website: https://researcher.watson.ibm.com/researcher/view.php?person=ibm-Chen.Wang1
Google Scholar: https://scholar.google.com/citations?user=JL6iWLgAAAAJ

RESEARCH INTERESTS

- **Data driven resource management in the Cloud:** Workload profiling; Resource provisioning; resilient scheduling
- **Deep learning in the Cloud:** deep learning job runtime estimation; Deep learning platform in Container Cloud
- **Cloud measurement:** Public cloud CDN comparison; Networking measurements for Public Cloud
- **Quality of Experience (QoE) driven Management in Video Streaming System:** Adaptive server selection; QoE monitoring; Anomaly detection, localization, and diagnosis
- **Video processing and retrieval system:** Video text extraction, Video shot classification, Content-based video copy detection system

EDUCATION

Carnegie Mellon University (CMU) — Pittsburgh, PA
PhD in Electrical and Computer Engineering — August 2017
Dissertation: QoE based management and control for large-scale VoD in the Cloud
PhD Advisors: Hyong Kim (ECE, CMU), Ricardo Morla (FEUP, Porto, Portugal)

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Electrical and Computer Engineering — May 2014
GPA: 3.92/4.0

Xi'an Jiaotong University (XJTU) — Xi'an, Shaanxi, China
Master of Science in Information Engineering — July 2010
GPA: 90.2/100.0

Xi'an Jiaotong University — Xi'an, Shaanxi, China
Bachelor of Science in Information Engineering — July 2007
GPA: 87.1/100.0

WORKING EXPERIENCE

IBM Thomas J. Watson Research Center — 1101 Kitchawan Rd, Yorktown Heights, NY, 10598
Research Staff Member — 09/05/2017 – Now
Manager**:** Dr. Alaa Youssef

- **Data driven Resource Management for Container Cloud**

- Perform log analysis on large-scale Kubernetes clusters running heterogeneous workload, discovering the fact that recurring failures in Kubernetes clusters are caused by either admitting workloads with similar features or scheduling workloads to nodes with certain features. Propose and develop a reward-based admission controller to prevent admitting recurring workload prone to fail and a reward-biased scheduler extender in Kubernetes to prevent scheduling certain workload to nodes where the workload prone to fail. The scheduler and admission controller will be contributed to open source by the end of 2019. Two patents were filed about this work.

- Apply statistical analysis to profile workload in Container Cloud based on resource usage metric data. Propose an intelligent resource profile advisor to ensure efficient resource utilization in large scale Kubernetes clusters running heterogenous workload. The resource profile advisor can help reduce resource utilization for more than 50% for enterprise clusters running various machine learning services while maintaining similar service availabilities. The work has 1 poster presented in ACM SOCC 2018 and has 1 journal paper submitted in 2019.

- Collect resource usage metric data for IBM Container Cloud Platform and propose to use DoppelGANger, a synthetic data generation framework based on generative adversarial networks (GANs), to synthesize enterprise cloud resource usage data with fidelity while preserving business secrets and privacies. The synthetic dataset will be contributed to open source community and will be shared with researchers in the Cloud computing field once the approval is obtained from management within IBM. The work has 1 paper submitted to NSDI 2020.

- **Data driven Optimization for Deep Learning Service in the Cloud**

- Propose AI Gauge, a cloud service to estimate runtime and cost for training deep learning models under different configuration options on the cloud. Mainly worked on the design of AI Gauge service in container cloud platform. AI Gauge can accurately predict the remaining time of running jobs based on its runtime progress (< 10% relative error) and can accurately predict the total runtime for a job before it starts with 7-8% relative error on average. The work has 1 paper accepted in SBAC-PAD 2019.

- Performance analysis for FfDL, a deep learning platform open sourced and used at IBM. Help examine FfDL quantitatively through a detailed empirical evaluation, including the overheads introduced by the platform for various DL models, the load and performance observed in a real case study using FfDL within our organization, the frequency of various faults observed including faults that we did not anticipate. The work has 1 paper accepted in MIDDLEWARE 2019.

Carnegie Mellon University 5000 Forbes Ave, Pittsburgh, PA, 15213
Research Assistant 09/01/2016 – 08/31/2017
Advisor**:** Prof. Hyong Kim and Prof. Ricardo Morla

- **QRank: Inferring ISP Roles based on End-user QoE for Large-scale VoD service**

- Propose and develop QRank, a scalable ranking system, to rank the ISPs based on end user QoE at run time. QRank uses run-time user QoE to study and discover the roles of ISPs in the video streaming services under the assumption of abolishing network neutrality.
- QRank is evaluated in production Cloud and interesting insights are obtained from extensive experiments. 91.4% of QoE anomalies are detected on 15.32% of users. These users experience QoE anomalies persistently and recurrently. More than 95% of persistent and recurrent QoE anomalies are identified in less than 10 transit networks. QRank results indicate that the limited capacity in transit networks are the major cause of QoE anomalies.

- **Evaluation of Production Cloud CDNs for Video Streaming Services**

- Compare Production Cloud providers including Amazon Web Service (AWS), Microsoft Azure, and Google Cloud regarding their geographical coverage, scalability, stability and cost for video streaming services. Users' Quality of Experience (QoE) are used as the evaluation metric.
- Interesting findings about these providers are obtained and presented in IEEE CLOUD 2018.

INESC TEC Rua Dr. Roberto Frias, Porto, Portugal, 4200-465
Research Assistant 09/01/2013 – 05/31/2016
Advisor**:** Prof. Hyong Kim and Prof. Ricardo Morla

- **QoE based Management and Control for VoD in the Cloud**

- Model user Quality of Experience (QoE) as rewards in the Management system of VoD. Propose and develop QoE based Management System for VoD in the Cloud (QMan). QMan monitors all users' QoE at run time and applies reinforcement learning techniques to adaptively select CDNs/servers for users to optimize the overall user QoE.
- QMan is evaluated with a VoD deployed in Clouds (servers/CDNs) with 300+ DASH clients worldwide emulated in PlanetLab. Results show that the QMan (with server selection) improved 90th percentile QoE more than 40% compared to the DNS based server selection system; QMan (with CDN selection) improves 90% percentile QoE over 20% compared to VoD using single CDN.

- **QWatch: Detecting and locating QoE anomalies for VoD in the Cloud**

- **QWatch**: Design a scalable monitoring system that detects and locates anomalies based on the end-user QoE at run time. Evaluate QWatch in a controlled VoD system and in a VoD system deployed in Microsoft Azure Cloud. QWatch correctly detects and locates QoE anomalies in servers, routers and client devices.

PUBLICATIONS

1. K. R. Jayaram, V. Muthusamy, P. Dube, V. Ishakian, **C. Wang,** B. Herta, S. Boag, D. Arroyo, A. Tantawi, A. Verma, Falk Pollok, Rania Khalaf, *"D3: A Flexible Multi-Tenant Deep Learning Platform"*, *accepted to ACM/IFIP International Middleware Conference*, *Middleware 2019.*
2. P. Dube, T. Suk, **C. Wang**, "*AI Gauge: A Runtime Estimation Service for Deep Learning in the Cloud"*, *accepted to the International Symposium on Computer Architecture and High Performance Computing, SBAC-PAD 2019.*
3. **C. Wang**, H. Kim, *"Touchdown on the Cloud: Impact of the Super Bowl on Internet, Cloud and Web Services", in 2019 IEEE Fifth International Conference on Big Data Service and Applications (BigDataService)*. IEEE, 2019.

4. M. F. Aktas, **C. Wang**, A. Youssef, M. Steinder, *"Resource Profile Advisor for Containers in Cognitive Platform", in Proceedings of the ACM Symposium on Cloud Computing (SOCC).* ACM, 2018.
5. **C. Wang**, A. Yayaseelan, H. Kim, *"Comparing Content Delivery Networks for Adaptive Video Streaming", in IEEE 11th International Conference on Cloud Computing (CLOUD).* IEEE, 2018
6. **C. Wang**. *"QoE based Management and Control for Large-scale VoD System in the Cloud"*. Ph.D. Dissertation, 2017.
7. **C. Wang**, H. Kim and R. Morla. "*Identifying Persistent and Recurrent QoE anomalies for DASH streaming in the Cloud*", *in 9th IEEE International Conference on Cloud Computing Technology and Science (CLOUDCOM).* IEEE, 2017
8. **C. Wang**, H. Kim and R. Morla. "*QWatch*: Detecting and Locating QoE anomaly for VoD in the Cloud", *in 8th IEEE International Conference on Cloud Computing Technology and Science (CLOUDCOM)*. IEEE, 2016
9. **C. Wang**, H. Kim and R. Morla. *"Users Know Better: A QoE based Adaptive Control System for VoD in the Cloud"*, in *IEEE Global Communication Conference (GLOBECOM).* IEEE, 2015
10. **C. Wang**, H. Kim, and R. Morla. "QoE driven Server Selection for VoD in the Cloud", in *IEEE International Conference on Cloud Computing (CLOUD)*. IEEE, 2015.
11. **C. Wang**, H. Wang. *"Utilization of Temporal Continuity in Video Text Detection"*, in *Second IEEE International Conference on Multimedia and Information Technology (MMIT)*. IEEE, 2010.
12. **C. Wang**, H. Wang, Y. Zhang. *"Saliency map-based image segmentation using level set"*, in *International Symposium on Image Analysis and Signal Processing (IASP)*, 2010
13. N. Nan, G. Liu, **C. Wang**. *"A BoF model based CBCD system using hierarchical indexing and feature similarity constraints"*, in *Second International Conference on Internet Multimedia Computing and Service (ICIMCS).* ACM, 2010.
14. Z. Li, G. Liu, X. Qian, **C. Wang**. *"Scale and rotation invariant Gabor texture descriptor for texture classification", Visual Communications and Image Processing (VCIP)*. SPIE, 2010.
15. Z. Li, G. Liu, X. Qian, **C. Wang**, Y. Ma and Y. Yang. *"A video text detection method based on key text points", Advances in Multimedia Information Processing-PCM*. Springer, 2010.
16. N. Nan, G. Liu, X. Qian, **C. Wang**. *"An SVM-based soccer video shot classification scheme using projection histograms", Advances in Multimedia Information Processing-PCM*. Springer, 2008.

MANUSCRIPTS UNDER REVIEW

17. M. F. Aktas, **C. Wang**, A. Youssef, M. Steinder, *"kAdvisor: An Autonomic Container Sizing System for Kubernetes"*
18. Z. Lin, A. Jain, **C. Wang**, G. Fanti, V. Sekar, *"Generating High-fidelity, Synthetic Time Series Datasets with DoppelGANger", arXiv preprint arXiv:1909.13403 (2019).*

PATENTS

19. **C. Wang**, A. Kanso, S. V. Costache, A. Youssef, M. Steinder, *"Reward-based admission controller for resource requests in the Cloud", Filed*
20. **C. Wang**, S. V. Costache, A. Youssef, A. Kanso, T. Suk, A. N. Tantawi, *"Reward-biased scheduler extender for Cloud systems", Filed*

PROFESSIONAL SERVICES

Conference and Workshop Organization

- Session Chair for IEEE 11th International Conference on Cloud Computing Workshop: Cloud and Big Data Analytics

Program Committee

- The 5th International Workshop on Container Technologies & Container Clouds (WoC'19)
- The Second Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf 2019)
- International Conference on Cloud Computing (SCF CLOUD 2019)
- The IEEE Fourth International Workshop on Container Technologies and Container Clouds (WoC'18)

Reviewers

- The 5th International Workshop on Container Technologies & Container Clouds (WoC'19, 4 reviews)
- The Second Workshop on Hot Topics in Cloud Computing Performance (HotCloudPerf 2019, 1 review)
- International Conference on Cloud Computing (SCF CLOUD 2019, 1 review)
- The IEEE Fourth International Workshop on Container Technologies and Container Clouds (WoC'18, 4 reviews)
- IEEE Transaction on Cloud Computing (2017-now, 7 reviews)
- IEEE Transaction on Service Computing (2018-now, 2 reviews)
- IEEE 2018 BIGDATA Congress (1 review)

- 2017 The 5th International Workshop on Big Data (Big Data Workshop, 1 review)
- 11th International Conference on Queueing Theory and Network Applications (QTNA 2016, 1 review)
- IEEE Symposium on Industrial Electronics and Applications (ISIEA 2012, 1 review)

Conference presentations and invited talks

- Oral presentation entitled "Comparing Content Delivery Networks for Adaptive Video Streaming", in IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA. July 2018.
- Invited talk entitled "QoE based management and control for large-scale VoD in the Cloud", Research Lab of Artificial Intelligence at Accenture, San Jose, CA. February 2017.
- Invited talk entitled "QoE based management and control for large-scale VoD in the Cloud", IBM T. J. Watson Research Center, Yorktown Heights, NY. December 2016.
- Oral presentation entitled "QoE driven Server Selection for VoD in the Cloud", in IEEE International Conference on Cloud Computing (CLOUD), New York City, US. July 2015.
- Oral presentation entitled "Users Know Better: A QoE based Adaptive Control System for VoD in the Cloud", in IEEE Global Communication Conference (GLOBECOM), San Diego, US. December 2015

Award Judge:

- Reviewer for the IBM Research Accomplishment Nominations within Hybrid Cloud Pillar, 2019.

Supervisors

- September 2019 – present, work with researcher Josep Berral from Barcelona Supercomputing Center on *"Anomaly Forecasting based on Execution Graphs for Kubernetes Clusters"* and *"Online container resource resizing via resource usage similarity analysis."*
- September 2018 – present, work with PhD student, Zinan Lin, from Carnegie Mellon University on *"Generating High-fidelity, Synthetic Time Series Datasets with DoppelGANger"*.
- September 2018 – present, work with PhD student, Samarth Gupta, "Modeling and Detection of cascading failures via non-intrusive measurements in the Cloud"
- May 2018 – August 2019, work with PhD student, Mehmet Fatih Aktas, from Rutgers University on "*Resource Profile Advisor for Containers in Cognitive Platform*" and "*Identifying predictability for containers in the Cloud: statistical analysis on container resource usage*"
- January 2017 – August 2017, work with graduate student, Andal Yayaseelan, from Carnegie Mellon University on *"Comparing Content Delivery Networks for Adaptive Video Streaming"*.

MEMBERSHIPS

- IEEE Membership
- IEEE Women in Engineering Membership
- IEEE BIG DATA COMMUNITY
- IEEE CLOUD COMPUTING COMMUNITY

AWARDS and HONORS

- IBM First Patent Invention Achievement Award, 2018
- IBM Manager's Choice Award, 2017
- AWS Research Grants, 2017, AWS Cloud Credits for Research
- Microsoft Azure Award for Research, 2016, Award opportunities: Microsoft Azure for Research
- NSF Student Travel Grant Award, 2015, GLOBECOM
- Ph.D Fellowship, 2010 – 2015, The Fundação para a Ciência ea Tecnologia (Portuguese National Science & Technology Foundation)
- First class Innovation Scholarship (Top 2), 2008 – 2009, XJTU
- Excellent Graduate Award (Top 10%), 2007, XJTU
- SIYUAN Scholarship (Top 15%), 2003 – 2006, XJTU
- First class Freshman Scholarship (Top 5%), 2003, XJTU

SKILLS

Programming Language: *Proficient* in C, C++, Java, Go, Python and Matlab; *familiar with* Django, AngularJS, D3.js, Node.js, jQuery, HTML, CSS, MySQL, Perl, Shell

Cloud/CDN Platform: *Experience* in IBM Cloud, Google Cloud, Amazon AWS, Microsoft Azure, Kubernetes, Docker, Hadoop

Knowledge: *Solid foundation* in cloud computing, networking, distributed systems, information theory, signal processing, video/image processing; good understanding in large-scale distributed system, load balancing, anomaly detection, fault tolerance, DASH streaming, Video-on-Demand system, Cloud resource management, CDN system and video search system.

TEACHING EXPERIENCE

Carnegie Mellon University Pittsburgh, PA, US

18757 – Network Management and Control Spring 2013

18345 – Introduction to Telecommunication Networks Spring 2013