Cloud-Scale Data Analytics

SEMINAR ANNOUNCEMENT

Speaker: Haiying (Helen) Shen
Date: Monday, May 16
Time: 10:45
Location: Rice Hall 242
Host: Samira Khan

Title: Optimizing Virtual Resource Management in Cloud Systems

Abstract

Cloud systems employ virtualization technology to provide resources in physical machines (PMs) in the form of virtual machines (VMs). Because overloaded PMs result in low application performance and underloaded PMs result in resource underutilization, cloud providers must devise methods to fully utilize cloud resources while still ensuring high application performance. In this presentation, I will introduce three approaches to reduce the presence of overloaded and underloaded PMs in order to achieve load balance in a cloud system. The first approach achieves load balance in the initial VM allocation. It consolidates complementary VMs with spatial/temporal-awareness to reduce the number of PMs needed and the number of VM migrations. The second approach achieves load balance by VM migration. It dynamically assigns different weights to different resources according to their usage intensities in the overloaded PM, thus reducing the time and cost needed for achieving load balance. The third approach achieves long-term load balance by VM migration. It uses the Markov Decision Process to allow a PM to proactively find an optimal action to transit to a long-term load balanced state. This research contributes to achieving higher resource utilization, with higher profit for cloud providers and higher quality-of-service for cloud users.

About the speaker: Dr. Haiying (Helen) Shen is currently an Associate Professor in the Department of Electrical and Computer Engineering at Clemson University. She received the 2015 IEEE Technical Committee on Scalable Computing (TCSC) Mid-career Award, the 2010 Microsoft Faculty Fellowship Award, the 2013 NSF CAREER Award, the 2013 Sigma Xi Clemson Chapter Young Investigator Award, and the 2010 Faculty Excellence Award of Clemson University. Her research interests include Cloud computing and datacenters, Big data, Information retrieval, Distributed systems, High performance computing, Cyber-physical systems, Internet of things, Mobile computing, and Social networks. Dr. Shen has made substantial contributions to her field with over 200 publications in prestigious conferences and journals such as IEEE/ACM Transactions on Networking (TON), IEEE Transactions on Parallel and Distributed System (TPDS), IEEE Transactions on Computers (TC), IEEE Transactions on Mobile Computing (TMC), ACM Symposium on Cloud Computing (SOCC), ACM Multimedia, IEEE International Conference on Computer Communications (INFOCOM), and IEEE International Conference on Distributed Computing Systems (ICDCS). She currently advises eleven Ph.D. students, one MS student and one undergraduate student. Dr. Shen is a subject area editor for the Scalable Computing Journal, and an associate editor for the International Journal of Parallel, Emergent and Distributes Systems (IJPEDS) and for the Journal of Information Science and Engineering. She is also a program committee member of many leading conferences, and the former program co-chair for a number of international conferences.