

Presents the Spring 2012 EECS Seminar Series

Shaolei Ren

Electrical Engineering Department, UCLA

“Exploiting Heterogeneity for Greening Data Centers”

Thursday, January 19, 2012 • 2:30 p.m. • HEC 101

ABSTRACT

With the emergence of cloud computing, there has been a growing trend toward large-scale data centers, each housing tens of thousands of servers and having an enormous appetite for energy. Fueled by the climbing energy cost as well as the call to action to reduce carbon emissions, greening data centers has, therefore, been of paramount importance to IT companies such as Google and Microsoft. Traditional approaches to data center deployment rely on homogeneity and uniformity in hardware, thereby simplifying system maintenance. Nevertheless, hardware heterogeneity, ranging from heterogeneous cores to heterogeneous data centers, is becoming increasingly pervasive and hence, poses new challenges to greening data centers, as the existing resource management in data centers is not tailored to the prevailing hardware heterogeneity.

In this talk, I will show that exploiting heterogeneity enables new advances in energy saving and service optimization. First, I focus on interactive services (e.g., web search, online gaming) and introduce efficient yet practical resource management techniques to exploit core heterogeneity. Compared to batch workloads, interactive services are more challenging to manage because of their responsiveness and quality requirements. In my research, I develop an optimal resource management algorithm that minimizes the average energy consumption of a single interactive request without knowing the service demand while satisfying the quality requirement. As a dual problem of energy minimization, I also investigate response quality maximization under an energy budget and propose an efficient algorithm for managing multiple interactive requests on a heterogeneous multi-core processor. The effectiveness of the proposed algorithms is demonstrated by a simulation study modeling a large commercial interactive service. Second, I turn to batch services and exploit server heterogeneity as well as data center heterogeneity by presenting a provably-efficient online scheduling algorithm minimizing the energy cost subject to latency constraints. Finally, I will conclude this talk by discussing my research in other areas such as social media with heterogeneous content.

BIOGRAPHY

Shaolei Ren is a final-year Ph.D. student in Electrical Engineering Department, University of California, Los Angeles (UCLA), working with Prof. Mihaela van der Schaar. He received his B.E. degree and M.Phil. degree from Tsinghua University in 2006 and from Hong Kong University of Science and Technology in 2008, respectively, both in electrical engineering. His research interests include cloud computing, network economics and smart grid. He received the Best Paper Award at IEEE International Conference on Communications in 2009, and was selected by IBM T. J. Watson Research as one of the 10 worldwide Emerging Leaders in Multimedia and Signal Processing in 2010. In recognition of his research, he also received the prestigious Dissertation Year Fellowship 2011-2012 from UCLA.