



Load balancing for dynamically scalable web services

Rhodes Hall 310: November 14, 2011 @ 12:00PM



ISN Seminar Speaker:

Yi Lu

University of Illinois at Urbana-Champaign

◇

Abstract

Web services have traditionally benefited from efficient load balancing. With multi-tenancy and the need for dynamic scalability in the cloud environment, existing hardware load balancers are expensive, difficult to scale and difficult to divide among users. Software load balancers are cost-efficient, scalable and divisible, but lack efficient load balancing algorithms. We explore two load balancing algorithms with distributed software load balancers: 1. We propose a new framework of analysis for the randomized load balancing algorithm, also known as “Power-of-Two”, with general service time distributions. We found that “two” is not always the magic number and characterized the threshold at which the behavior of the algorithm experiences a qualitative change. Exact characterization of the threshold has been an open problem for 15 years. 2. We propose a novel algorithm called Join-Idle-Queue (JIQ), which significantly outperforms the “Power-of-Two” algorithm. The performance characterization of the JIQ algorithm involves an interesting analysis of load balancing “in the reverse direction.”

◇

Biography

Yi Lu is an assistant professor in the ECE department at UIUC. She received her doctorate from the EE department at Stanford University. She is a recipient of the Stanford Graduate Fellowship, the Sigmetrics Best paper award in 2008, and the Performance Best paper award in 2011. Her work focuses on developing scalable architectures and algorithms for large and complex networks such as modern web services with dynamic content, Cloud computing, and social networks. Her work spans fundamental analysis and algorithm implementation, and emphasizes design of low-complexity distributed algorithms.