Title: Autonomic Cloud Management for Power and Performance Constraints
Speaker: Cihan Tunc, University of Arizona

Abstract

Large data centers and cloud systems have an increasing demand from the end-users. However, management of such systems with both power and performance constraints have been a major challenge requiring complex management. It has been reported that the power consumption of data centers and cloud systems have increased almost three times between 2007 and 2012. Typically, pessimistic over-provisioning techniques are applied to meet the peak workloads of the Virtual Machines (VMs). However, without autonomic dynamic power management methods, the high power consumption is inevitable. Therefore, in our work, we present an autonomic power and performance management method for cloud systems in order to dynamically match the application requirements with "just-enough" VM resources at runtime that will reduce power reduction while meeting the quality of service requirements of the cloud applications. Our solution uses the following functions: 1) Real-time cloud resource monitoring and workload behavior running on VMs; 2) Characterization of workload behaviors with multiple attributes; 3) Determination of the current workload operating point on VMs running these workloads; 4) Dynamic management of the VM resources (scaling up and down the number of cores, CPU frequency, and memory amount) at run time to reduce power consumption without sacrificing the QoS requirements of cloud workloads. We validate the effectiveness of our approach using RUBiS benchmark (an auction model emulating eBay transactions that generates a wide range of workloads such as browsing and bidding with different number of clients). Our experimental results show that our approach can lead to reduction in power consumption up to 87% when compared to the static resource allocation strategy, 72% compared to adaptive frequency scaling strategy and 66% compared to a similar multi-resource management strategy.