Abstract

Cloud data centres improve CPU utilization of their servers (physical machines or PMs) through Virtualization (virtual machines or VMs). Over virtualised and under virtualized PMs suffer performance degradation and power dissipation respectively. This work presents a stochastic modular scheme for allocating VM requests to a PM by avoiding overloading of PM and keeping the global load characteristics under specified QoS goal. The proposed approach categorizes PMs into three groups (Under Load, Normal Load, Over Load) in a way that minimizes number of PMs in Under Load and Over Load groups and maximizes number of PMs in Normal Load group. We compute VM request rejection probability, response time, service time and number of PMs that are overload or under loaded for evaluating the performance of our model. The results show that these parameters do not degrade with increasing arrival rate. Thus the proposed model is simple yet efficient approach for VM placement problem.
Overload Avoidance Model using Optimal Placement of Virtual Machines in Cloud Data Centers

- Anton Beloglazov, Jemal Abawajy, Rajkumar Buyya, "Energy-aware Resource
Overload Avoidance Model using Optimal Placement of Virtual Machines in Cloud Data Centers


Overload Avoidance Model using Optimal Placement of Virtual Machines in Cloud Data Centers


**Index Terms**

Computer Science  
Cloud Computing

**Keywords**

Markov Chain  
Virtual Machine  
Physical Machine  
Virtual Chunks  
VM Consolidation