香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. YE Shujin
Date	29 August 2023 (Tuesday)
Time:	4:00 pm – 6:00 pm (35 mins presentation and 15 mins Q & A)
Venue:	1) DLB637, 6/F, David C Lam Building, Shaw Campus 2) ZOOM (Meeting ID: 931 5603 4931) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 6:00 pm, 28 August 2023)

Resource and Energy Management for Clouds and Data Centers

Abstract

In large-scale cloud computing, resource and energy management are important for better cost-effectiveness and environmental-friendliness. This thesis tackles the following three resource and energy management problems for clouds and data centers.

1) Energy management in data centers: A data center consumes a large amount of energy and hence its energy management is important. Typically, the resource demand of a virtual machine (VM) is time-varying. This feature could be exploited such that multiple VMs could statistically share the resources of a physical machine (PM) for better energy efficiency. To realize this goal, we address two issues. First, we propose a new prediction method to model and predict the resource demand of a VM based on its recent resource utilization. This method applies Markov chain, non-parametric clustering and Gaussian mixture to model and predict the resource demand, where all the parameter values are automatically determined. Second, we formulate a new problem of assigning VMs with time-varying resource demands to heterogeneous PMs in a data center. The objective is to minimize the energy consumption rate while supporting multiple service level agreements (SLAs), where each SLA specifies that the resource demand of a VM is satisfied with a probability higher than a pre-specified threshold. We design an efficient approximation algorithm for this VM-to-PM assignment problem, and theoretically prove that this algorithm achieves an approximation ratio of $(\lambda+1)/2$ for homogeneous PMs, where λ is the maximum number of VMs that can be assigned to a PM, $\lambda \leq 1/(1-P)$ and P is the smallest probability threshold defined in the SLAs. We conduct simulation experiments using the real-world resource demands from the Google Cluster traces. The results show that: i) the proposed prediction method is significantly more accurate than the existing ones, and ii) the proposed approximation algorithm could effectively reduce the energy consumption rate.

2) Renewable energy in data centers: The increasing demand for cloud computing has incurred a significant rise in energy consumption and carbon emissions of data centers. To address this issue, researchers have explored various approaches to power data centers with renewable energy. However, these approaches have limitations in terms of the amount of renewable energy provided and the distance between renewable energy sources and data centers. To overcome these challenges, we propose a crowdsourcing approach where users with solar panels or wind turbines provide the nearby data center with renewable energy through energy trading. In this approach, users connect their renewable energy sources to the power grid and set their own selling prices of renewable energy, which can be purchased by the cloud service provider (CSP). However, purchasing renewable energy from users is challenging due to the uncertainty of renewable energy generation, as well as the correlation among geographically-close renewable sources. To model uncertainty and correlation, we introduce the random forests based on generalized least squares (RF-GLS) for renewable energy prediction. In particular, this method leverages Gaussian processes to obtain the joint probability distribution of renewable source combinations and considers the correlation among renewable sources. We formulate the problem as a biobjective optimization problem, aiming to maximize renewable energy supply while minimizing energy cost and considering the risk associated with uncertainty. To solve the optimization problem, we propose a heuristic algorithm based on the concept of maximum clique. Simulation results based on real-world datasets demonstrate the effectiveness of our algorithm.

3) Cloud federation for collaboration services: Cloud federation paradigm can improve cloud service providers' (CSPs) profits by renting their idle resource to other federation members. However, these CSPs have the risk that they cannot fulfill their scalability commitment when some of their customers have large short-term resource demand. To reduce this risk, we design a reinsurance-emulated collaboration mechanism in a broker-based cloud federation. Reinsurance is an insurance policy which transfers all or part of insurance business in order to scatter the risk to other insurers. Similar to insurance companies, in our proposed model, each CSP determines its resource retention for its future demand. We design an exact method to determine each CSP's retention, with the aim of maximizing its expected profit. Once the CSP's retention cannot meet its future demand, it reduces the risk by outsourcing part of the requests to others. After every CSP determines its retention, the broker will make an assignment to maximize the resource utilization so as to reduce the risk of the CSPs. We design an algorithm to maximize the resource utilization, with a guarantee of not worse than half of the optimal solution. Simulation results show that our proposed algorithm is efficient.

***** ALL INTERESTED ARE WELCOME *****