

To Cloud or not to Cloud: Optimizing Cloudbursting Costs

Technical Report

ABSTRACT

The emerging hybrid cloud architectures allow organizations (users) to augment the private infrastructure with practically unlimited public cloud resources in order to cost effectively meet their intermittent peak demands. In such scenarios, users first utilize their already paid private computation infrastructure and offload selected tasks to the public cloud when the private resources become overloaded. Consequently, there is a need to devise efficient on-line task offload algorithms that optimize the overall user cost while maintaining adequate quality of service. Such algorithms should take into account the difference in communication and computation requirements between the different task types. For example, it is clear that between two tasks with the same computational requirements, the task with the lower migration cost is a better candidate to be offloaded to the cloud. In this work, we devise optimal on-line decision algorithms by modeling and solving several associated multi-dimensional Markov Decision Process problems. We address the case in which arriving tasks have multiple communication costs and prove the structural properties of the optimal threshold policy. In addition, we also apply the MDP framework for the complement problem facing cloud providers. If certain cloud resources are not pre-allocated to users, it makes sense for the cloud provider to offer them for opportunistic on-demand usage. We model and provide optimal policies for the buildup of a task backlog by accepting or rejecting tasks that carry user offered price or by dynamically changing advertised prices for tasks. The analytical results are supported by numerical evaluations. We demonstrate the practical advantage of threshold type policies and provide an insight of their dependence on system parameters.

Keywords

Cloud Computing, Markov Decision Processes, Offloading algorithms.

1. INTRODUCTION

The emergence of cloud computing is changing the ways in which organizations address their information technology (IT) needs. The concept of cloud computing brings out a new way of increasing computational and storage capacity or adding capabilities on the fly without investing in a new infrastructure, training new personnel, or licensing new software. In theory, the cloud agility and elasticity make the cloud the best IT solution model. However, many practical issues such as limited network speeds, lack of strict SLA guarantees, lack of cloud standards, information regulatory compliance and the wish to preserve the full control over their core IT resources and know-how limit the full adoption of the cloud model. Consequently, there is a topical trend to leverage the best of both worlds by keeping the minimal essential legacy IT infrastructure while adopting the public cloud where it is more cost effective. One of the terms which is frequently used to describe this paradigm is a "hybrid cloud". Essentially, the hybrid cloud refers to a business that keeps some of its server operations on-premise, while also utilizing the services of a cloud provider to augment or supplement the internal infrastructure. Another topical term associated with the hybrid cloud is "cloudbursting". According to this concept, in case the internal data center runs out of computing resources, the organization "bursts" or "offloads" the additional workload to an external cloud on an on-demand basis. The internal computing resource is the "Private Cloud", while the external cloud is typically a "Public Cloud", for which the organization gets charged on a pay-per-use basis. Effective cloudbursting offers the organization a solution to a critical and very basic performance-related and economical dilemma. The pre-cloud old-fashioned solution utilizes over-provisioning, e.g., stacking a computational gear inventory that meets the peak demands. However, if the influx of tasks is characterized by high variability with rare peaks, such a solution is extremely resource wasteful. At the same time, the organization revenues and its business conduct may be considerably harmed if the peak demands are not met. ([1]). Therefore, the hybrid cloud architecture combined with effective "cloudbursting" offers an effective solution to offload the task load peaks from the resource-limited and cost reduced private cloud to a seemingly infinite cloud.

The hybrid cloud environment poses new research challenges associated with effective task offloading from the private data center to the cloud. If all the computational tasks were identical, a simple cloudbursting mechanism would track the backlog of tasks waiting for local execution; if the backlog