Accepted Manuscript

Holistic energy and failure aware workload scheduling in Cloud datacenters

Xiang Li, Xiaohong Jiang, Peter Garraghan, Zhaohui Wu

PII: S0167-739X(17)31565-0
DOI: http://dx.doi.org/10.1016/j.future.2017.07.044
Reference: FUTURE 3574

To appear in: *Future Generation Computer Systems*

Received date : 18 July 2016
Revised date : 29 June 2017
Accepted date : 16 July 2017

Holistic Energy and Failure Aware Workload Scheduling in Cloud Datacenters

Xiang Li[a] (corresponding author), Xiaohong Jiang[a], Peter Garraghan[b], Zhaohui Wu[a]
[a] College of Computer Science, Zhejiang University, Hangzhou 310027, China. {lixiang2011, jiangxh, wzh}@zju.edu.cn
[b] School of Computing and Communications, Lancaster University, Lancaster LA1 4WA, UK. p.garraghan@lancaster.ac.uk

ABSTRACT

The global uptake of Cloud computing has attracted increased interest within both academia and industry resulting in the formation of large-scale and complex distributed systems. This has led to increased failure occurrence within computing systems that induce substantial negative impact upon system performance and task reliability perceived by users. Such systems also consume vast quantities of power, resulting in significant operational costs perceived by providers. Virtualization – a commonly deployed technology within Cloud datacenters – can enable flexible scheduling of virtual machines to maximize system reliability and energy-efficiency. However, existing work address these two objectives separately, providing limited understanding towards studying the explicit trade-offs towards dependable and energy-efficient compute infrastructure. In this paper, we propose two failure-aware energy-efficient scheduling algorithms that exploit the holistic operational characteristics of the Cloud datacenter comprising the cooling unit, computing infrastructure and server failures. By comprehensively modeling the power and failure profiles of a Cloud datacenter, we propose workload scheduling algorithms *Ella-W* and *Ella-B*, capable of reducing cooling and compute energy while minimizing the impact of system failures. A novel and overall metric is proposed that combines energy efficiency and reliability to specify the performance of various algorithms. We evaluate our algorithms against *Random*, *MaxUtil*, *TASA*, *MTTE* and *OBFIT* under various system conditions of failure prediction accuracy and workload intensity. Evaluation results demonstrate that *Ella-W* can reduce energy usage by 29.5% and improve task completion rate by 3.6%, while *Ella-B* reduces energy usage by 32.7% with no degradation to task completion rate.

KEYWORDS

Energy Efficiency; Thermal Management; Reliability; Failures; Workload Scheduling; Cloud Computing

1. INTRODUCTION

Cloud datacenters are core infrastructure required to provision digital services globally, forming compute facilities composed by thousands of interconnected servers. Such systems consume vast amounts of energy to operate - the carbon emission of datacenters globally is comparable to that of Argentina, and continues to grow by 11% annually [1]. Datacenter energy usage can be categorized as predominantly stemming from computing (processors, storage, etc.) and cooling (air conditioner, fans, etc.) [2]. In order to reduce computing energy, Virtual Machines (VMs) are commonly scheduled to consolidate workload onto the fewest servers as possible, and then shutting down idle servers [3]. Such scheduling is a direct threat towards minimizing cooling energy, which is highly affected by skewed spatial temperature distribution of the facility (i.e. the hottest server). This requires careful workload balancing [4] in order to distribute workload evenly amongst servers to minimize the highest temperature, avoid hot spots and reduce cooling energy [5]. However, the aims of these two scheduling approaches results in a contradiction that attempts to simultaneously consolidate workload onto as fewest active servers, yet attempts to reduce the formation of system hotspots directly created by such consolidation. This does not accord with our intuition, as it is typically assumed that computing energy and cooling energy are positively correlated to each other: more computing energy results in additional cooling requirement for computing infrastructure heat rejection (and vice versa). However, this does not occur when considering the reduced cooling efficiency when lowering the Computer Room Air Conditioner (CRAC) supply air temperature to address hotspot formation.

It is imperative for Cloud providers to schedule workload to minimize total system energy of both computing and cooling to reduce operational costs. This must be achieved whilst adhering to implicit and explicit user requirements for high levels of application performance and reliability. The complexity and scale of modern day Cloud datacenters has resulted in failures becoming the normal rather than the exception [6], manifesting within both system software and hardware. Such failures cause Service Level Agreement (SLA) violation, resulting in energy [7] and monetary loss [8]. Work in [7] demonstrates the prevalence of software and hardware failures within production Cloud datacenters, with server Mean Time Between Failure (MTBF) as short as 6.5 hours [9]. [6] declares that the success of large-scale (e.g. petascale) computing will depend on its ability to provide dependable service at scale.

A challenge arises within workload scheduling as the objective to minimize total system energy does not result in optimal task reliability. Assume that S_1, S_2, S_3 correspond to scheduling solutions that produce the minimal computing energy, cooling energy, and highest reliability as shown in Equation (1).

$$
\begin{aligned}
E_{computing}(S_1) &= \min\left[E_{computing}(S)\right],\\
E_{cooling}(S_2) &= \min\left[E_{cooling}(S)\right],\\
E_{reliability}(S_3) &= \max\left[E_{reliability}(S)\right].
\end{aligned}
\tag{1}
$$