J. Parallel Distrib. Comput. 111 (2018) 126–135

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

journal homepage: www.elsevier.com/locate/jpdc

Energy-aware task assignment for mobile cyber-enabled applications in heterogeneous cloud computing

Keke Gai [a], Meikang Qiu [b,c,*,1], Hui Zhao [a,d]

[a] Department of Computer Science, Pace University, New York City, NY 10038, USA
[b] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China
[c] Department of Electrical Engineering, Columbia University, New York City, NY, 10027, USA
[d] Software School, Henan University, Kaifeng, Henan, 475000, China

HIGHLIGHTS

- We propose a novel approach for reducing the computation energy costs for heterogeneous MES in cloud systems. Our algorithm can intelligently assign the tasks to on-premise cores or remote cloud servers within an adaptive time period.
- We present a method of the adjustment that is designed to transfer sub-optimal solutions to optimal solutions at a high success rate.
- We propose a feasible solution to the proposed task assignment problem for heterogeneous MES that is a NP-hard problem. The proposed approach can be used in other application scenarios.

ARTICLE INFO

Article history:
Received 31 December 2016
Received in revised form 26 May 2017
Accepted 7 August 2017
Available online 18 August 2017

Keywords:
Energy-aware
Task assignment
Cloud computing
Mobile embedded systems
Cyber-enabled applications
NP-hard

ABSTRACT

Recent remarkable growth of mobile computing has led to an exceptional hardware upgrade, including the adoption of the multiple core processors. Along with this trend, energy consumptions are becoming greater when the computation capacity or workload grows. As one of the solutions, using cloud computing can mitigate energy costs due to the centralized computation. However, simply offloading the workloads to the remote side cannot efficiently reduce the energy consumptions when the energy costs caused by wireless communications are greater than that of on mobile devices. In this paper, we focus on the energy-saving problem and consider the energy wastes when tasks are assigned to remote cloud servers or heterogeneous core processors. Our solution aims to reduce the total energy cost of the mobile heterogeneous embedded systems by a novel task assignment to heterogeneous cores and mobile clouds. The proposed model is called *Energy-Aware Heterogeneous Cloud Management* (EA-HCM) model and the main algorithm is *Heterogeneous Task Assignment Algorithm* (HTA2). Our experimental evaluations have proved that our approach is effective to save energy when deploying heterogeneous embedded systems in mobile cloud systems.

1. Introduction

Contemporary mobile technologies have been enabling cyber-enabled systems to become a ubiquitous existence changing people's lives in multiple dimensions, from smart phones to mobile vehicular systems [16,25,30]. Different multimedia systems have brought many new or strengthened accesses to the computing resources. As an emerging technology, *Mobile Embedded Systems* (MES) with *Cyber-Enabled Applications* (CEA) have become a mainstream of mobile computing that balance the high-performance and the cost. This change has been empowered by the deployments of cloud computing in recent years [8,12]. Many cloud-based solutions support the offloads of the heavy workloads to the remote cloud servers, by which the energy consumptions are reduced [14,19]. Nevertheless, most contemporary approaches are confronting the contradictions deriving from the wireless communications and energy consumptions, since the local energy costs are generally less than that of the wireless data transmissions [6]. A fixed working mode that offloads jobs to the cloud-based servers cannot satisfy the requirement of the energy-saving. Multiple factors need to be addressed when creating the strategy of offloading tasks to clouds, such as heterogeneous computing capacities and energy consumptions in different phases.

* Corresponding author.
E-mail addresses: kg71231w@pace.edu (K. Gai), qiumeikang@gmail.com, mq2203@columbia.edu (M. Qiu), hz58694w@pace.edu (H. Zhao).
[1] M. Qiu is with Shenzhen University, China and Columbia University, USA.

http://dx.doi.org/10.1016/j.jpdc.2017.08.001
0743-7315/

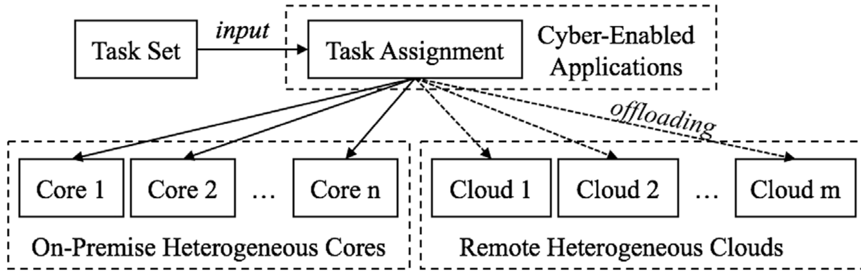

Fig. 1. The architecture of the *Energy-Aware Heterogeneous Cloud Management* (EA-HCM) model.

Moreover, contemporary energy consumptions on MES are facing a variety of challenges, from app executions to wireless communications [9,23]. For instance, smart phones can not only deliver the wireless calls but also provide a mobile app platform. Heterogeneous computing systems enable mobile devices to achieve a high performance of computations, such as synchronously assigning a chain of tasks to multiple cores [21]. Nevertheless, the energy costs go up when the large sized tasks are loaded, which may require less energy while some tasks are offloaded to the clouds [11]. Therefore, an optimized deterministic mechanism of task assignments can aid the system to reduce the total energy costs when the energy requirements are varied between local and remote executions.

Focusing on this urgent demand, we propose a novel model, named *Energy-Aware Heterogeneous Cloud Management* (EA-HCM) model, in order to achieve the high performance computation capacity by using the reduced energy. Fig. 1 represents the architecture of EA-HCM model, by which the input tasks are dynamically assigned to heterogeneous cores and remote processors. This model is an extension of Gai et al.'s [10] work. Distinguishing prior work, this model further emphasizes the role of cyber-enabled applications that empower the task assignment for mobile embedded systems in heterogeneous cloud computing. The core algorithm creating near-optimal solutions is proposed in this model. In addition, as shown in the figure, the assignment uses the estimated energy costs to reduce the total energy consumption, which is controlled by the cyber-enabled applications. The expected goal is creating the plan of the task assignment that can reduce the total energy consumption while considering the both computation and communication energy costs.

Furthermore, we emphasize the implementation of the cyber-enabled application in our proposed model. The reason for using cyber-enabled applications is that it enables feasible configurations on various mobile devices for meeting different demands of data collections and task distributions. In line with the cyber-enabled applications, we propose our algorithm, *Heterogeneous Task Assignment Algorithm* (HTA2), in order to achieve our designed goal. The algorithm aims to dynamically assign the tasks to various cores and remote clouds. The tasks assignment is based on the cost mapping and our proposed optimal alternatives and adjustments. The algorithm consists of two steps: (1) generate a sub-optimal solution, and (2) improve the obtained sub-optimal solution and try to transfer it to the optimal solution via a set of adjustments. Our approach is an attempt of achieving high performance heterogeneous MES with lower-level energy requirements.

The main contributions of this paper include the following:

1. We propose a novel approach for reducing the computation energy costs for heterogeneous MES in cloud systems. Our algorithm can intelligently assign the tasks to on-premises cores or remote cloud servers within an adaptive time period. Distinguishing from prior work, the approach proposed by this work considers all available computing resources the individual options for task assignment, such that each cloud source is an objective for task scheduling.
2. We present a method of the adjustment for generating near-optimal solutions, which is designed to transfer sub-optimal solutions to optimal solutions at a high success rate.
3. We propose an adaptive solution to the task assignment problem for heterogeneous MES, which has been proved as an NP-hard problem. The proposed approach can be used in other application scenarios.

The remainder of this paper is organized by the following sections. We accomplish a survey of cloud resource management in Section 2. Next, a motivational example is represented in Section 3. Moreover, we explain the mechanism and system definitions in Section 4. Furthermore, in Section 5, we illustrate the main algorithm proposed for our model. In addition, we display and analyze partial experimental results in Section 6. Finally, we give the conclusions in Section 7.

2. Related work

Previous research have explored a variety of dimensions in the field of cloud resource management and task scheduling provisions.

First, formulating a proper resource provision strategy was a research direction in demand. Rodriguez and Buyya [20] investigated the issue of resource provisioning and scheduling strategy for meeting the demands of resource management in the cloud environment. This approach used meta-heuristic optimization technique to minimize the total operation costs when the timing constraints are configured. However, this work is restricted by the designed operating scenario which lacks parallel computations. There were some other completed studies using different approaches, such as applying *k*-mean methods to divide workloads into clustered tasks [29], considering response time to increase the level of customer satisfactions [27], and using stochastic programming to optimize the multiple provision stages [4].

Moreover, some studies addressed green computing issues and balanced a variety of factors, such as energy consumptions, system performances, and execution time. Addis et al. [1] designed distributed hierarchical framework for the mixed-integer nonlinear optimizations using multiple timing constraints. Addressed the same issue in the cloud context, another approach [28] introduced a new layer named *Skewness* to examine the status of the multidimensional server utilizations, by which different workload types could be formulated. This approach used a heuristic algorithm to attempt the minimization of the energy consumption. Other similar work confronting green computing applied other methods, such as Greedy algorithms [7], genetic algorithms [18] and dynamic programming [11]. However, the weakness of these approaches was that the configuration of the resource controller could rarely prove its relationship with the system performance,