Microprocessors and Microsystems 000 (2016) 1–11

Contents lists available at ScienceDirect

Microprocessors and Microsystems

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

Energy-Aware on-chip virtual machine placement for cloud-supported cyber-physical systems

Xuanzhang Liu[a,c], Huaxi Gu[a,*], Haibo Zhang[b], Feiyang Liu[b], Yawen Chen[b], Xiaoshan Yu[a]

[a] State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China
[b] Department of Computer Science, University of Otago, Dunedin, New Zealand
[c] Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, Shijiazhuang 050081, China

A R T I C L E I N F O

Article history:
Received 6 February 2016
Revised 15 June 2016
Accepted 15 July 2016
Available online xxx

Keywords:
Cyber-physical systems
Virtual machine placement
Network-on-Chip
Ant colony optimization

A B S T R A C T

Recent trends in the design of cyber-physical systems (CPS) are moving towards heterogeneous multi-core architectures with cloud support. In this paper, we propose an energy-aware scheme for virtual machine placement in cloud-supported CPS with Network-on-Chip (NoC) architecture. We formulate the energy-aware on-chip virtual machine placement problem as an optimization problem, and design a heuristic scheme based on ant-colony optimization. We address problems of slow convergence speed and easily falling into stagnation in ant-colony algorithm by employing pheromone diffusion model that makes the proposed scheme more efficient. Simulation results show that our scheme achieves much higher energy efficiency compared with previous schemes with different network sizes and traffic models.

1. Introduction

Cyber-Physical Systems (CPS) are composed of services and applications deployed across a range of communication topologies, computing platforms, and sensing and actuation devices. Since numerous applications can be deployed in CPS platforms, cloud computing has become a promising computing paradigm which can provide a good support for CPS platforms in terms of cost-efficiency, scalability, and safety [1]. In cloud-supported CPS, the sensors and actuators communicate with the cloud through the cybernet, whereas information processing, control decision making, and system virtualization are all done at the cloud using the high-performance multi-core servers. This ensures quick responses to physical dynamics, and effective and stable control of the physical environment. In addition, applications in the CPS are inherently heterogeneous, real time, reactive and networked with hard deadlines [2]. Every virtual machine in an application is interdependent and has its own execution, arrival and time periods [3]. A key challenge for cloud-based CPS is to optimally execute different applications in the cloud. Server virtualization is an effective method in cloud computing that allows various applications being deployed more flexibly and feasibly in different locations around the world. Through virtualization technology, the physical machine (server or computer) can be logically divided into multiple virtual execution environments, with each acting as a full-function system. This isolated execution environment is so called virtual machine (VM). Therefore, one single server can simultaneously run multiple applications on separated virtual machines that leverage physical resources of the server [4,5].

With the increase number of cores integrated in a single server, the interconnection among these cores can significantly affect the system performance. The ever-advancing integration technology with billions of transistors integrated on a single chip enables on-chip multiple microprocessors based commercial servers, e.g., 80 cores in Intel Teraflops, 188 cores in Cisco/IBM SPP. Moreover, recent developments in Network-on-Chip (NoC), a technology of on-chip interconnection network, provide some efficient communication schemes for these multi-core servers. Compared with traditional bus architecture, NoC leverages the principle of interconnection network and packet switching to achieve low latency, high performance, and low power consumption [6]. For a high-performance server based on the multi-core processor, each virtual machine can be implemented on one single processor core, and the NoC architecture provides efficient communications for these virtual machines, e.g., data flows, cache coherence protocols. At present, the inter-core communication power consumption has already taken an important part of the total power budget owing to the long distance global on-chip communication and ultra-high bandwidth requirement (tens to hundreds of terabits per second) [7]. The cost of managing the power consumption and the associated cooling devices drives the need to design some application-level energy-efficient schemes and algorithms. Careless

* Corresponding author.
E-mail address: hxgu@xidian.edu.cn (H. Gu).

http://dx.doi.org/10.1016/j.micpro.2016.07.013
0141-9331/

on chip virtual machine placement, which leads to unreasonable traffic distribution and hotspots, may cause high communication energy consumption and deteriorate the communication performance. For example, placing two virtual machines with large volume communication requirement too far will lead to high power consumption and long communication delay. In this paper, we propose an energy-efficient on chip virtual machine placement algorithm in one single server to solve this challenging problem.

The objective of this work is to design an energy-aware on chip virtual machine placement scheme that is capable of implementing multiple virtual machines on a multi-core server with high power efficiency and desirable performance. The main contributions of this paper include:

- We propose an energy-aware on chip virtual machine placement scheme for cloud computing which sustains the strength of ant colony algorithm in accuracy and efficiency but offsets its weakness in slow convergence speed. By placing virtual machines running the same application to closer cores on the multi-core system according to their traffic rates, the energy consumption and communication delay can be reduced significantly.
- We formulate the on chip virtual machine placement problem as a binary integer programming (BIP) problem that aims to minimize the power consumption for inter-virtual-machine communications. To address this problem, we propose an improved ant colony algorithm that employs the pheromone diffusion model to solve these intrinsic problems of slow convergence and search stagnation. We customize the ant colony algorithm according to the application property and the character of Network on Chip communication architecture to achieve global energy efficiency.
- We carry out extensive simulations to evaluate the energy consumption of our proposed algorithm with both synthetic traffics and realistic data traces. Simulation results show that our algorithm can achieve better energy efficiency compared with some existing algorithms.

The remainder of the paper is organized as follows. Section 2 reviews some related works. In Section 3, we give a motivating example for the on chip virtual machine placement problem. Then the statement of the problem is described in Section 4. We present the energy efficient virtual machine placement algorithm in Section 5. Simulation results are shown in Sections 6, and Section 7 concludes this paper.

2. Related works

2.1. Virtual machine placement

Virtual machine placement has significant influence on cloud computing systems. For example, if two virtual machines with application dependency are non-optimally placed on two servers that locate on different racks or even different cities, the communication delay and energy consumption may be unacceptable. Generally, previous researches have focused on dynamic inter-server virtual machine migration and static server-level virtual machine placement [4,8]. Using these schemes, all the active virtual machines are migrated or placed onto a small number of servers with respect to the performance requirements and resource constraints, whereas the unused servers which have no active virtual machine can be shut down to save power. However, with the rapid development of high performance multi-core servers, tens of hundreds of virtual machines can be implemented on a single server. The on chip virtual machine placement which deals with the intra-server communications between the virtual machines has become an open issue that needs to be well addressed. Grot et al. [9] proposed the Kilo-NoC architecture which guarantees the service requirements of data flows by placing the virtual machines on a shared CMP (chip-level multiprocessor). Wang et al. [13] proposed a virtual machine placement algorithm for the heterogeneous multi-core system that exploits the different properties of each core to optimize the overall system performance and energy efficiency. Hu et al. [10] presented a virtual machine scheduling model to solve the I/O performance bottleneck based on the multicore dynamic partitioning. Especially, for multicore platforms using CPS, Kanduri et al. explore the impact of application mapping on network contention and predictability [11]. All the current researches of on chip virtual machine placement for multi-core systems mainly target on some specific architectures and applications. It still lacks some general mathematical formulations and optimal solutions for the on-chip virtual machine placement problem, which is extremely important for designing a high performance and scalable multi-core system for cloud-based CPS.

2.2. Placement algorithms

Virtual machine placement can be formulated as optimization problems with objectives to minimize communication delay or maximize throughput or energy efficiency. Some current existing inter-server virtual machine placement algorithms in data centers use Linear Programming model [12,14], Bin Packing algorithm [15] and Artificial Intelligent algorithm [4,16,17]. The basic ideas of these algorithms are described in the followings.

The Linear Programming (LP) schemes assume that the performance goal is linearly related to the placement of virtual machines. For example, Chaisiri et al. [14] proposed an algorithm which places virtual machines on different physical servers with the assumption that the minimal number of servers required and the resources in each server subject to a linear function. In [12], the authors also designed some extension constraints for the linear programming model, such as restricting the number of virtual machines in a single physical server, limiting the number of virtual machine migrations, etc. The main advantage of the Linear Programming based schemes is its simplicity.

The Bin Packing schemes assume that the virtual machine placement can be formulated as a variant of the vector bin-packing problem and various heuristic solutions have been used to solve this problem. Zhang et al. [15] designed several heterogeneity-aware heuristic algorithms for virtual machines placement, which explores the heterogeneity of the requirements of virtual machines for different resources and utilizes the dominant resources (e.g. CPU, memory) as constraints to assist inter-server virtual machine placement. However, this kind of schemes has higher complexity because of the heuristic algorithms.

Artificial Intelligent algorithm is derived from some natural activities and can be used to achieve an optimal virtual machine placement. Xu and Fortes [16] proposed a two-level control system that adopts a modified genetic algorithm with fuzzy multi-objective evaluation to deal with the problem of allocating workloads to virtual machines and virtual machines to physical servers. Ant colony algorithm has been applied to the field of multi-objective optimization for virtual machine placement. Gao et al. [4] formulated the problem of virtual machine placement as a multi-objective combinatorial optimization problem aiming to simultaneously optimize total resource wastage and power consumption. A modified version of the ant colony system algorithm is proposed to effectively deal with the potential large solution space for large-scale data centers. Similarly, Liu et al. [17] proposed an approach based on the ant colony optimization to solve the virtual machine placement problem so as to effectively use the physical resources and to reduce the number of running physical servers.

Please cite this article as: X. Liu et al., Energy-Aware on-chip virtual machine placement for cloud-supported cyber-physical systems, Microprocessors and Microsystems (2016), http://dx.doi.org/10.1016/j.micpro.2016.07.013