Future Generation Computer Systems 90 (2019) 149–157

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

# Autonomic computation offloading in mobile edge for IoT applications

Md Golam Rabiul Alam [a], Mohammad Mehedi Hassan [b,*], Md. ZIa Uddin [c], Ahmad Almogren [b], Giancarlo Fortino [d]

[a] Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[b] College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[c] Department of Informatics, University of Oslo, Norway
[d] Department of Informatics, Modeling, Electronics, and Systems, University of Calabria, 87036 Rende, Italy

## HIGHLIGHTS

- An autonomic computation offloading model for mobile edge/fog is proposed.
- A deep reinforcement Q-learning model is used for computation offloading.
- Our method significantly improves the performance of the computation offloading.

## ARTICLE INFO

*Article history:*
Received 23 February 2018
Received in revised form 13 July 2018
Accepted 25 July 2018
Available online xxxx

*Keywords:*
Computation offloading
Autonomic computing
Mobile edge/fog computing
Deep Q- learning

## ABSTRACT

Computation offloading is a protuberant elucidation for the resource-constrained mobile devices to accomplish the process demands high computation capability. The mobile cloud is the well-known existing offloading platform, which usually far-end network solution, to leverage computation of the resource-constrained mobile devices. Because of the far-end network solution, the user devices experience higher latency or network delay, which negatively affects the real-time mobile Internet of things (IoT) applications. Therefore, this paper proposed near-end network solution of computation offloading in mobile edge/fog. The mobility, heterogeneity and geographical distribution mobile devices through several challenges in computation offloading in mobile edge/fog. However, for handling the computation resource demand from the massive mobile devices, a deep Q-learning based autonomic management framework is proposed. The distributed edge/fog network controller (FNC) scavenging the available edge/fog resources i.e. processing, memory, network to enable edge/fog computation service. The randomness in the availability of resources and numerous options for allocating those resources for offloading computation fits the problem appropriate for modeling through Markov decision process (MDP) and solution through reinforcement learning. The proposed model is simulated through MATLAB considering oscillated resource demands and mobility of end user devices. The proposed autonomic deep Q-learning based method significantly improves the performance of the computation offloading through minimizing the latency of service computing. The total power consumption due to different offloading decisions is also studied for comparative study purpose which shows the proposed approach as energy efficient with respect to the state-of-the-art computation offloading solutions.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

The massive growth of mobile devices (e.g. smart phones, laptops, tablet pc's, mobile IoT's and automobiles) and their computation demands imposed a huge scarcity in communication network and computation resources. Some of the application services e.g. image processing and real-time translation services require extensive computation, the resource-constrained mobile devices are not the feasible domiciles to process those applications. Therefore, to meet the computation demands of such type of mobile devices and applications the outsourcing of computation is the demand in need.

Computation offloading is a relocation mechanism of processes or modules of software applications or systems from resource-constrained devices to the resource-rich platforms. Mobile cloud is the well-known platform for computation offloading of mobile devices. Mobile cloud computing is becoming a popular method

* Corresponding author.
*E-mail address:* mmhassan@ksu.edu.sa (M.M. Hassan).

https://doi.org/10.1016/j.future.2018.07.050
0167-739X/© 2018 Elsevier B.V. All rights reserved.

for mobile services e.g. mobile video games, video streaming, education, social networking, messenger and mobile healthcare services [1].

However, the key barriers to offloading computation in mobile cloud are the network bandwidth and latency. Data travels a longer hazardous path from mobile device to the mobile cloud during offloading and thus consumes huge network bandwidth [2]. The bandwidth scarcity, and internet bottlenecks and traffic congestions are the catalysts for the higher latency of offloading computation. Real-time applications are highly latency sensitive and thus it requires to compute data in a close proximity of mobile devices or users. So, mobile fog can be the effective and suitable platform for offloading mobile computation.

Fog computing [3] is introduced by Cisco Systems Inc. to extend the cloud computing paradigm to the edge of network especially for Internet of Things (IoT) services. Mobile Fog is the complementary model of fog computing especially prototyped for seamless and latency-aware mobile services [4]. However, the key research questions for offloading computation in mobile fog are (1) How to offload computation in the mobile fog? (2) Which module or process of mobile application should offload? (3) Where to offload the module or process for minimizing the latency of service computing? Moreover, the mobility, heterogeneity and geographical distribution mobile devices impose additional challenges of computation offloading in mobile fog. This research contributes to finding the answer to the above questions. The key contributions of this research are as follows.

- A code offloading framework is proposed for computation offloading in mobile fog environment. The code analyzer unit of the framework determines which basic blocks of the code are computation hungry and subject to offload.
- A deep Q-learning [5] based computation offloading method is proposed for the autonomic management of massive offloading request. The trained code offloader unit of the proposed framework takes the offloading decision considering resource demand, resource availability and network status to minimize the latency of service computing.
- The performance of the proposed model studied through simulation. The performance gain in terms of latency and energy efficiency justifies the dominance of the proposed autonomic offloading model.

Rest of the paper is organized as follows. In Section 2, we discussed the related works. The system model of mobile fog is presented in Section 3. The deep Q-learning based autonomic code offloading method in mobile fog is illustrated in Section 4. We presented the simulation and performance study results in Section 5. Finally, we concluded the paper in Section 6 with some future directions.

## 2. State-of-the-arts computation offloading methods

Mobile fog interplays with tradition cloud to access its huge computational resources. Thus, this section discussed state-of-the-arts resource provisioning methods of legacy cloud computing paradigm. Afterwards, the pioneer works on computation offloading in the mobile cloud are discussed in this section.

The elasticity and scalability of cloud computing are achieved through virtualization of cloud resources. The resources of cloud data centers are managed through VM configuration and placement methods. The optimized placement of virtual machines in cloud brokering architecture is proposed in [6]. The paper presented very detail architecture of cloud service broker. The optimized selection of virtual resources of cloud brokers through cloud scheduler was one of the primary objectives of the paper. The holistic approach, OPTIMIS, is proposed in [7] to optimize the service lifecycle of cloud service provisioning. The paper introduced a toolkit for reliable, sustainable and trustful service provisioning.

The cost-effective deployment of computing clusters in multi-cloud infrastructure is presented in [8]. They provided analysis on the viewpoint of performance and cost. The proposal is only for loosely coupled many-task computing (MTC) applications. The proposal overlooks tightly coupled MTC applications, where facts are highly interdependent and synchronization among the computational units is necessary.

The optimal allocation of computing and networking resources in cloud computing networks is proposed in [9]. The authors of this paper used mixed integer programming to formulate optimal networked cloud mapping problem. In the proposal, the authors modeled cloud request as undirected graph of virtual nodes and virtual network links and then allocate QoS-aware virtual resources according to networked cloud request.

Energy-aware resource allocation and provisioning methods are discussed in [10]. They proposed a green cloud architecture with power model, VMs placement and migration algorithm. The proposal is fully devoted to power-aware policy development by minimizing the migration of VMs among multi-cloud infrastructure.

A joint or coordinated VM resource provisioning and maintenance scheduling method is proposed by the authors of [11]. They formulated the problem as an Integer Linear Programming problem and then transformed it into an equivalent problem to obtain linear programming relaxation solution, then they apply LIST rounding algorithm towards a final approximate solution. CoTuner [12] is the model-free reinforcement learning based VM configuration framework. It can configure VM's on the fly with changing workloads.

In mobile cloud computing, most of the pioneer works proposed the VM migration mechanism in a surrogate cloud server. The cloudlets [13] are the trusted, resource-rich and nearby computing box to offload mobile data for extensive processing. Cloudlet is well-connected to the central cloud through internet and it is considered within one hop communication range of mobile devices. The cloudlets are also called the little clouds, which act as the surrogates of centralized mobile cloud to process latency sensitive application services. The adjacent cloudlets are connected with each other through mesh connectivity [14] and can communicate and migrate virtual machines (VMs) with one other to support mobility. Each cloudlet is connected with centralized mobile cloud to fetch, store, and process necessary data through VM placement. The physical servers of host mobile cloud and cloudlets are placed in the fixed geographical locations but because of the arbitrary user requests and resource requirements the states of VMs change dynamically. The cloudlets are placed usually in coffee shops, subway stations and other public places. The dense deployment of cloudlet requires a huge investment.

CloneCloud [15] proposed a solution of offloading computation in cloud servers by introducing an automatic application partitioner, which portioned the mobile application at runtime and deploys it onto device clones in the computational cloud. The communication latency and VM formation cause jitter in latency-sensitive applications. The mobility of mobile users is ubiquitous and thus we need more efficient solutions which ensure seamless mobile services.

MobiCloud [16] proposed the Mobile Ad Hoc Networks (MANATs) as the mobile cloud computing units, where each of the mobile nodes acts like the service node. Every service node is mirrored in virtualized cloud servers to provide secure service architecture. Scavenger [17] is the mobile cyber-foraging system to offload resource intensive jobs. The framework provides an opportunity to offload computation to nearby surrogate devices.

Download English Version:

https://daneshyari.com/en/article/6872774

Download Persian Version:

https://daneshyari.com/article/6872774

Daneshyari.com