Accepted Manuscript

An advanced decision model enabling two-way initiative offloading in edge computing

Zhida Yin, Haopeng Chen, Fei Hu

PII: S0167-739X(17)32952-7
DOI: https://doi.org/10.1016/j.future.2018.07.031
Reference: FUTURE 4349

To appear in: *Future Generation Computer Systems*

Received date : 25 December 2017
Revised date : 25 April 2018
Accepted date : 17 July 2018

Please cite this article as: Z. Yin, H. Chen, F. Hu, An advanced decision model enabling two-way initiative offloading in edge computing, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.07.031

An Advanced Decision Model Enabling Two-way Initiative Offloading in Edge Computing

Zhida Yin[1], Haopeng Chen[1], Fei Hu[b]

[a]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
[b]China Aeronautical Radio Electronics Research Institute

Abstract

Edge computing is envisioned as a promising enabler to leverage computation capacities at the edge and address the issues faced in cloud computing. In this paper, we propose an advanced decision model to solve the computation offloading problem in edge computing. It utilizes the intrinsic hierarchical topology of the Internet and performs online scheduling in a decentralized manner to eliminate anticipated modeling. It also admits the fact that congestion can take place unexpectedly, for the possible reasons such as stale information. Our decision model is enhanced in two approaches to endow both sender and receiver (for one offloading action) with the ability of "dropping" the request, so as to avoid congestion. Random Early Detection(RED) algorithm is incorporated. The results of simulation demonstrate that our decision model can handle workloads well and these two enhancements are effective.

Keywords:
Edge Computing, Offloading Problem, Congestion Avoidance, RED Algorithm Integrated

1. Introduction

The maturity of cloud computing has tremendously influenced our lives in various aspects. Famous software as a service(SaaS) instances, such as Google applications, already have daily active users(DAU) on the scale of millions or even billions. Meanwhile, diverse infrastructures aiming at providing scalability, elasticity, and autonomy, such as Amazon EC2 and Microsoft Azure, have been developed to support these services. Our previous work [1] and its enhanced version [2] also offered some suggestions towards this topic.

In recent years, the proliferation of Internet of Things(IoT), cloudlets and mobile phones is gradually making changes. In traditional cloud computing, there are clear boundaries between data consumers and data producers; while nowadays these IoT devices can work as not only data producers but also data consumers [3]. Based on this role transition, the number of data producers surges, as well as the raw data produced. Even though the processing capacity of central servers keep developing, the growing quantity of data and the limited bandwidth of the network is becoming the bottleneck for the cloud-based computing [3].

Moreover, services, especially mobile ones, are becoming more and more sensitive to latency and require immediate response [4]. Offloading workloads to central servers, as we do in traditional cloud computing, will inevitably incur long latency due to network transmission [5] and thus have major impact on performance. Additionally, other considerations, such as privacy protection requirement, are also attracting more and more concerns.

Such trends have pushed the horizon of a new computing paradigm, edge computing. In edge computing, data should be handled in close proximity to where it is produced, namely at the edge. Since this paradigm can save network bandwidth and take full advantage of processing capacity at the edge, it is viewed as a promising enabler. However, there are still several issues to be addressed. Among them, the computation offloading problem, which concerns the trade-off between energy consumption and service latency, has drawn wide at-

Email addresses: darkness_y@sjtu.edu.cn (Zhida Yin), chen-hp@sjtu.edu.cn (Haopeng Chen), hu_fei@careri.com (Fei Hu)