Editorial

ICSES Interdisciplinary Transactions on Cloud Computing, IoT, and Big Data

Journal Homepage: www.i-cses.com/iitcib/

An Efficient SDN-based Edge Computing Framework for Computation Offloading and Service Augmentation

Ting Wang [a,*] and Yangzhou Wang [b]

[a] Huawei Technologies Co., Ltd., Shanghai, China
[b] College of Computer Science, Chongqing University, Chongqing, China

* Corresponding Author E-mail: twangah@connect.ust.hk

INTRODUCTION

INTERNET of Things (IoT) as a new Internet evolution greatly impels the revolutionary development of network. It is conservatively estimated that there will be 50 billion connected devices by 2020 [1]. More and more new applications designed for terminal devices such as interactive gaming, GPS navigation, 4K videos, natural language processing, face recognition, virtual reality and augmented reality are emerging, where these kinds of applications are usually resource-hungry, demanding intensive computation and high energy consumption. With the explosion of "things", current computing models and network architecture face many challenges and limitations in meeting the higher demands of ultra-low latency, huge data processing/computation, and high bandwidth.

Mobile cloud computing is considered as a computing paradigm to provide elastic resources to augment computation capabilities of mobile devices for resource hungry applications by offloading the computation to remote resource rich cloud data centers [2], as shown in Fig. 1. Although this approach partially addresses such a challenge to a certain extent with some potential benefits, it also brings numerous additional severe issues, which can be listed as follows. Firstly, huge amount of data is generated every second by end devices (e.g. more than 400 zettabytes data a year is expected by 2019 [3]). If all the data are sent to remote cloud to be processed, it will significantly increase the burden of network core. Secondly, huge amount of traffic volume across the Internet induces high transport cost. Thirdly, processing data at remote sites leads to high network latency, especially in case of inefficient hierarchical routing where packets travel for a significant distance in the mobile core network. This is a critical problem for latency sensitive services and applications. Fourthly, as known, cloud computing has limited mobility support. Fifthly, remote cloud solutions cannot provide location aware services for mobile applications. Last but not least, cloud data centers are geographically highly centralized, where the robustness and reliability are big concerns. All these issues have all along been suffered by mobile applications and IoT services, where a more efficient and practical solution is required to be designed.

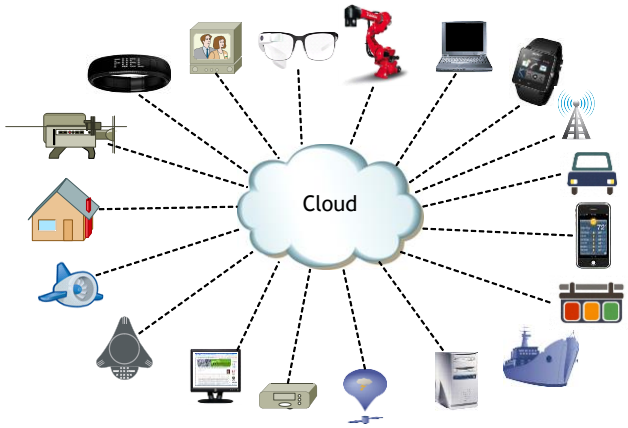

Figure 1. Mobile cloud computing: computation is offloaded to the remote cloud

To deal with these challenges, the concept of edge computing is proposed, which attempts to push the computing intelligence and capabilities to the network edge closer to the end users [4]. The recently emerging edge computing models include Mobile Edge Computing (MEC, standardized by ETSI), Fog Computing (proposed by Cisco), Cloudlet (developed by CMU), Micro data center (developed by Microsoft Research), and so on. The standardized 5G network architecture design also follows the CUPS (Control and User Plane Separation) principle, where the control plane (like SMF/AMF/NSSF/NRF, etc.) is centralized deployed in the central data center and user plane (i.e., UPF) is distributed deployed in the edge data

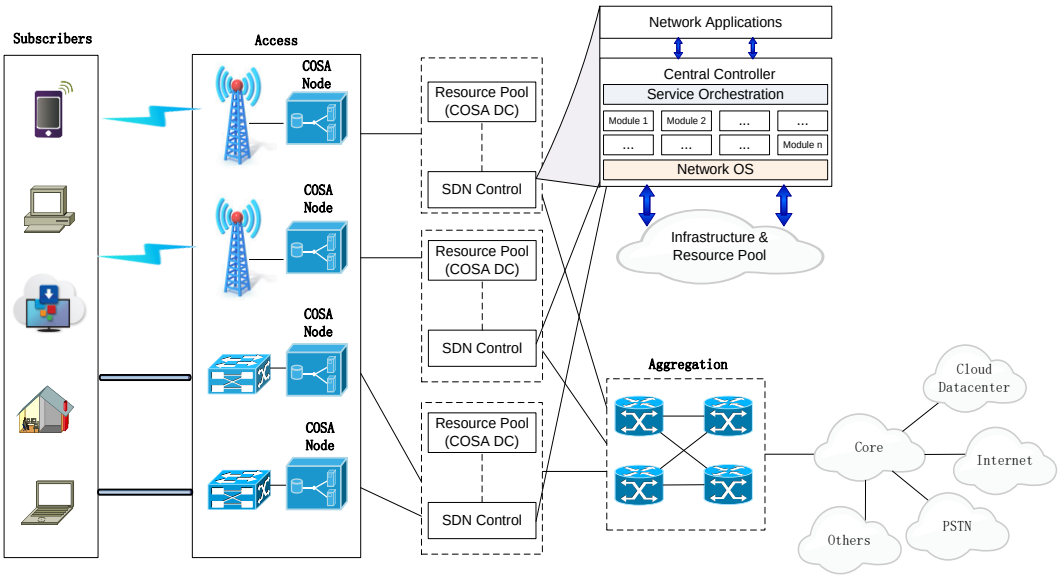

Figure 2. SDEC Architecture

center being close to the users. These approaches, which all aim to move computation, data and control from remote cloud DC to the network edge, to some extent, conceptually partially overlaps and are also complementary. The edge computing enables IoT services to meet the requirements of low latency, high scalability and energy efficiency, as well as mitigate the traffic burdens of the transport network. However, there are still many challenges and open issues needed to be solved in edge computing, such as mobility-aware service migration problems, security and privacy concerns, the deployment optimization of edge servers, the cooperation between Fog/Cloudlet/Edge computing and cloud computing, the resilience of edge computing system, the network automation, the computation offloading mechanism, and cost-effective edge computing architectures, and so on.

In line with the CUPS principle, this paper proposes an effective and practical architectural solution, which is a SDN (software defined network) based architecture for computation offloading and service augmentation at the network edge.

SDN-BASED MEC ARCHITECTURE DESIGN

In this section, an effective and practical architectural solution for edge computing, named SDEC, is proposed. SDEC is a SDN-based (software defined network) architecture for Computation Offloading and Service Augmentation at the network edge.

The key principle of the SDEC is to geographically distribute computation resources at the network edge, which is closer to end subscribers, who can offload their computation intensive tasks to the nearest "optimal" SDEC nodes. The edge nodes can monitor and cache the Internet/cloud services to provide end users faster services with lower end-to-end delay. Fig. 2 illustrates an example of multiple-tired SDEC architecture, where a number of edge nodes with adequate computation/storage capabilities are distributed at the network edge to provide elastic resources for computation offloading and service caching. There are several different roles of these SDEC nodes, which are common node (termed as CNode), regional agency node (termed as RANode), super node (termed as SNode), and certificate authority node (termed as CA).

Fig. 3 gives an example of control plane of SDN-based SDEC. In the SDEC framework, the service monitoring module of SDN controller can monitor the network traffic, and cache the certain services in proper SNodes within the data center. The service providers can also proactively deploy their services in SDEC data centers. Besides, according to actual needs, the SNodes can further push some cached services down to some selected CNodes, which are even more closer to end users and better accessible. With the benefit of service offloading, not only the users can experience much faster services, but also the traffic can be offloaded from the network core, accordingly, which can greatly help reduce the traffic burden of the network core

CONCLUSION

As the explosion of mobile devices and Internet of Things (IoT) applications, the traditional mobile cloud solution is already not efficient enough. What's more, the 5G is coming, and the era of zettabytes is upon the world in the near future, it will be impractical to send all the huge amount of data to remote cloud data center across the whole transport network to be processed, which will expose a big challenge to the traffic burden of network core. Consequently, SDEC is proposed to solve these issues, and push the resources closer to end users and limit the traffic only within the edge network without traversing to the network core. This offers numerous benefits for latency sensitive services or real time interaction applications.

Manuscript ID = 253 — Copyright © 2019 International Computer Science and Engineering Society (ICSES) — ISSN: 2467-295X
Website: www.i-cses.com — E-mail: info@i-cses.com

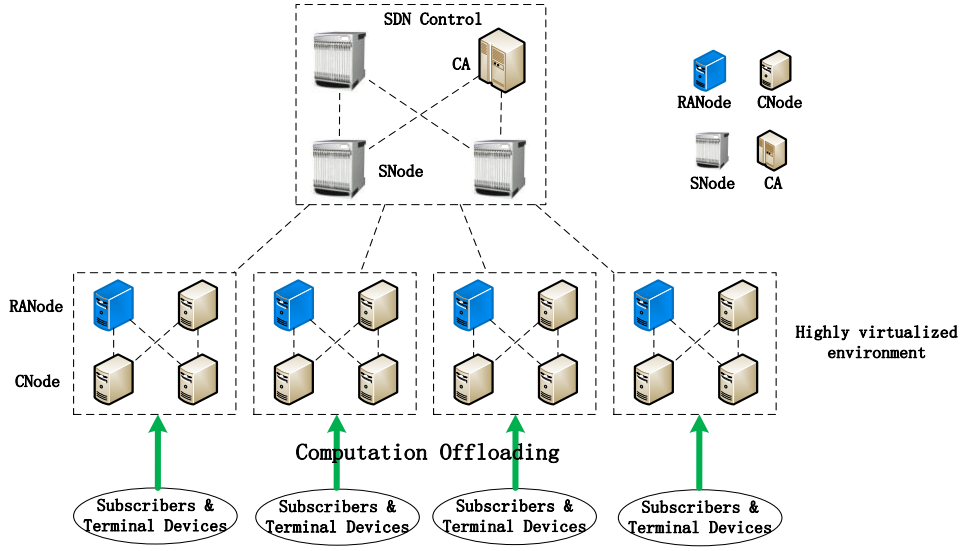

Figure 3. SDEC's Distributed Nodes

Moreover, SDEC can better support mobility and location-aware services.

REFERENCES

[1] Evans, D., "The internet of things: How the next evolution of the internet is changing everything," *CISCO* white paper1.2011, pp. 1-11, 2011.

[2] Wang, T., Su, Z., Xia, Y., and Hamdi, M., "Rethinking the data center networking: Architecture, network protocols, and resource sharing," *IEEE access*, vol. 2, pp. 1481-1496, 2014.

[3] Worth, D, "Internet of Things to generate 400 zettabytes of data by 2018," *Retrieved September*, vol. 7, pp. 2015, 2014.

[4] Pan, J. and McElhannon, J., "Future edge cloud and edge computing for internet of things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439-449, 2018.