Machine Learning Enabled Distributed Mobile Edge Computing Network

Junchao Ma
jcma@vt.edu
Bradley Department of Electrical and Computer Engineering, Virginia Tech.
Blacksburg, Virginia
Key Lab of Information Coding and Transmission, Southwest Jiaotong University
Chengdu, China

Hao-Hsuan Chang
haohsuan@vt.edu
Bradley Department of Electrical and Computer Engineering, Virginia Tech.
Blacksburg, Virginia

Pingzhi Fan
pzfan@swjtu.edu.cn
Key Lab of Information Coding and Transmission, Southwest Jiaotong University
Chengdu, China

Lingjia Liu
ljliu@vt.edu
Bradley Department of Electrical and Computer Engineering, Virginia Tech.
Blacksburg, Virginia

ABSTRACT

In this work, we propose to establish a mobile edge computing (MEC) network that considers computation, caching and communication jointly. Depending on the demanding categories, users in the network are partitioned into computation-driven and caching-driven users, both of which need memory resource to improve their quality of experiences (QoEs). Thus, a memory resource allocation problem is aroused to maximize the performance of the whole network. Due to the fact that the users' characterization plays an important role to the resource allocation scheme and with the help of machine learning techniques, we propose to study and predict the users' patterns by distributed learning methods which take the heterogeneity of base station type and users' mobility, etc into consideration. The proposed machine learning based distributed MEC system can maximize the efficiency of the network by optimizing the resource allocation scheme and perfectly predicting users' pattern.

CCS CONCEPTS

• **Networks** → *Network performance modeling*; *Network performance analysis.*

ACM Reference Format:
Junchao Ma, Hao-Hsuan Chang, Pingzhi Fan, and Lingjia Liu. 2019. Machine Learning Enabled Distributed Mobile Edge Computing Network. In *SEC '19: ACM/IEEE Symposium on Edge Computing, November 7–9, 2019, Arlington, VA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3318216.3363454

ACM ISBN 978-1-4503-6733-2/19/11...$15.00
https://doi.org/10.1145/3318216.3363454

1 INTRODUCTION

The synergy of Computing, Caching, and Communication (3C) has received much attention recently in 5G networks [4]. However, designing a network that jointly considers these three aspects is still an open problem [6]. The difficulty of the collaboration relies on the observation of user cases that relate to the three aspects simultaneously. In our work, we are the first to propose a machine learning based distributed mobile edge computing (MEC) network that jointly considers these three aspects. The users in the network are partitioned into two categories: caching-driven users with repeatedly requests of the same content (e.g. video on demand) [3] and computing-driven users with high computing power requirements (e.g. on-line gaming). The behaviors of user groups under the coverage of different base stations (BSs) are learned and predicted using advanced machine learning techniques [1, 5]. Based on the prediction, we propose a resource allocation scheme to allocate the limited memory resource at BSs to caching resource and computation resource. Then the caching replacement and computation task offloading policy are developed to maximize the quality of experience (QoE) of caching-driven and computation-driven users, respectively. In addition, we propose an utility function to evaluate the QoE of users under the machine learning based MEC network.

2 SYSTEM MODEL

In our work, we consider a heterogeneous network consisting one macro BS (MBS) and a group of small BSs (SBSs). As shown in Fig. 1, control information is exchanged between the MBS and SBSs or adjacent SBSs. Each edge BS (MBS or SBS) is equipped with a certain amount of memory which can be used to compute the offloaded tasks from users, or to cache contents users may retrieve. Depending on the BS type and some environment parameters (BS coverage, budgets, etc.), BSs are assumed to have heterogeneous memory capacities.

The users in the network are partitioned into two categories: caching-driven users and computation-driven users. Computation-driven users need to compute a series of tasks with stringent latency

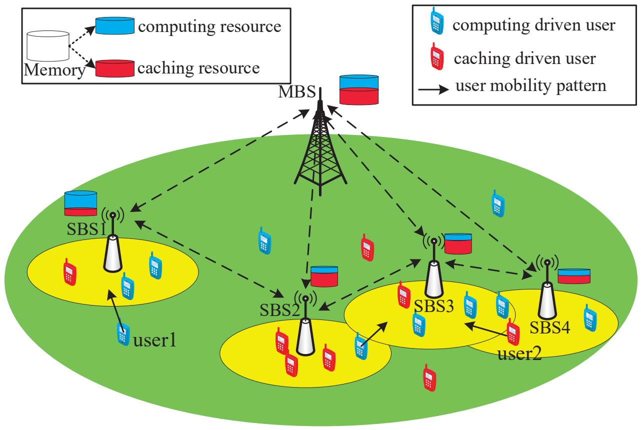

Figure 1: Applied Network Model.

requirements. To reduce the energy consumption at the users and the suffered latency which includes transmission delay and computation delay, the compute-intensive tasks are offloaded to edge BSs or other nearby users [2] and require large amounts of computation resource. On the other hand, caching-driven users need the forwarding of the edge BS when the requested streaming is from the cloud to the mobile user. To avoid redundant transmission by other requests of the same video content and to reduce the response latency, it is better that the edge BS caches some popular contents, which also requires large amount of storage resource [7]. Therefore, a problem arises: *with limited memory resource equipped at the edge BSs, how much of them should be allocated to computation resource and caching resource respectively?* The problem can be addressed by carefully designing a resource allocation policy.

The prediction of the users' behavior patterns plays an important role in the network design. In contrast to existing works that assume knowing users' behavior patterns, in our work a machine learning based method is proposed to learn and predict users' behavior patterns. Our learning method is distributively ran in different BSs and different learning agents work collaboratively with one another through exchanging control information with adjacent BSs. To be specific, the heterogeneity of different BS capacities and coverage, users' mobility pattern, as well as the difference between local statistics and global statistics are considered in our learning algorithm. Given the prediction of users' behavior by the learning agents, the allocation of computation and caching resource can be done at each BS. Then the allocated computation and caching resource are used to serve different users' requirements to maximize their QoEs. We also propose an utility function to evaluate the QoE of task offloading performance and caching replacement performance. This function can be maximized by optimizing the task offloading decision and caching replacement policy together. In addition, the impact of communication is implicit in the resource allocation and thus 3C concept can be jointly considered in our proposed machine learning based distributed MEC network [4].

ACKNOWLEDGMENTS

The work of Junchao Ma and Pingzhi Fan is supported by the NSFC project under Grant No. 61731017, NSFC-NRF project under Grant No. 61661146003, and 111 project under Grant No. 111-2-14. The work of Hao-Hsuan Chang and Lingjia Liu is supported in part by National Science Foundation under Grant No. ECCS-1802710, ECCS-1811497, and CNS-1811720.

REFERENCES

[1] H. Chang, H. Song, Y. Yi, J. Zhang, H. He, and L. Liu. 2019. Distributive Dynamic Spectrum Access Through Deep Reinforcement Learning: A Reservoir Computing-Based Approach. *IEEE Internet Things J.* 6, 2 (April 2019), 1938–1948. https://doi.org/10.1109/JIOT.2018.2872441

[2] H. Chen, L. Liu, H. S. Dhillon, and Y. Yi. 2019. QoS-Aware D2D Cellular Networks With Spatial Spectrum Sensing: A Stochastic Geometry View. *IEEE Trans. Commun.* 67, 5 (May 2019), 3651–3664. https://doi.org/10.1109/TCOMM.2018.2889246

[3] J. Ma, L. Liu, B. Shang, and P. Fan. 2019. Cache-aided Cooperative Device-to-Device (D2D) Networks: A Stochastic Geometry View. *IEEE Trans. Commun.* (2019), 1–1. https://doi.org/10.1109/TCOMM.2019.2931556

[4] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong. 2019. Joint Communication, Computation, Caching, and Control in Big Data Multi-access Edge Computing. *IEEE Transactions on Mobile Computing* (2019), 1–1. https://doi.org/10.1109/TMC.2019.2908403

[5] B. Shang and L. Liu. 2019. Machine Learning Meets Point Process: Spatial Spectrum Sensing in User-Centric Networks. *IEEE Wireless Commun. Lett.* (2019), 1–1. https://doi.org/10.1109/LWC.2019.2940442

[6] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao. 2019. Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues. *IEEE Communications Surveys Tutorials* (2019), 1–1. https://doi.org/10.1109/COMST.2019.2924243

[7] Z. Zhang, C. Lung, M. St-Hilaire, and I. Lambadaris. 2019. An SDN-based Caching Decision Policy for Video Caching in Information-centric Networking. *IEEE Transactions on Multimedia* (2019), 1–1. https://doi.org/10.1109/TMM.2019.2935683