



**INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS**

Name of the Student : ANIRBAN LEKHARU

Roll Number : 186101001

Programme of Study : Ph.D

Thesis Title:
Deep Learning Approach for Efficient Mobile Edge Computing

Name of Thesis Supervisor(s) : Prof. Arijit Sur
Dr. Moumita Patra

Thesis Submitted to the Department/ Center : Computer Science and Engineering

Date of completion of Thesis Viva-Voce Exam : 07th August 2023

Key words for description of Thesis Work : Mobile Edge Computing, Deep Learning, Reinforcement Learning, Quality of Experience, Video Popularity

SHORT ABSTRACT

Mobile data traffic has increased enormously in recent years with an increase in mobile and smart devices. Global mobile data traffic is set to increase manifold in the coming years. With the exponential rise in mobile video traffic and dynamic request patterns, maintaining a decent Quality of Experience (QoE) for end-users is challenging for content service providers. MEC provides an opportunity for caching the most requested content closer to the end-users, thereby reducing the overall traffic cost and access delay. Therefore, the primary objective of such a caching strategy is to increase the cache hit rate at the edge server, aiming to improve the end-users' overall QoE. In recent scientific literature, it has been observed that various heuristic and Machine Learning-based caching strategies at the MEC server, have been presented. However, most of the existing caching techniques are not adaptive enough to handle diverse and complicated requests across temporal and geographical dimensions. Intuitively, the above problem can be formulated as a multi-objective optimization problem. To solve such a hard problem, learning-based solutions have recently gained popularity, especially using Deep Learning (DL) techniques. Keeping the massive success of DL techniques in mind, in this dissertation, Deep Reinforcement Learning (DRL) is used to design an efficient and robust caching mechanism at the MEC server. The first contribution of this thesis considers the users viewing profile, which frequently changes in different time slots of the day. For this, a DL-based content-aware caching model (called DCache) has been proposed, which is deployed at the MEC server. In the second contributory chapter, a Reinforcement Learning (RL)-based ABR caching mechanism at the MEC server within the purview of a single BS has been proposed. This work introduces a novel joint optimization framework using RL called ABRCache, which improves the overall QoE for a video streaming session and reduces the traffic load on the backhaul links and the overall access delay simultaneously. To further improve the caching performance, the third contributory chapter presents a collaborative caching called ColabCache, where MECs within a given cluster collaborate to serve the requested content. In ColabCache, a novel Deep Reinforcement Learning (DRL) using Asynchronous Advantage Actor Critic Network (A3C) network for caching at the edge server has been proposed. The first part of the final chapter focuses on caching the video content in a federated manner. In the proposed FedCache, user data is not centrally collected; instead, the model is trained on the individual data of each user, and the central server aggregates the updates from each user. The second part of this chapter focuses on users' viewing experience for a video streaming session. In this work, a Deep Neural Network (DNN)-based model is proposed that selects the

appropriate video bitrates to maximize the overall QoE of a user for a video streaming session. Finally, the thesis is concluded by summarizing the significant contributions and proposing some relevant future research directions.

