Video on Demand Streaming Using RL-based Edge Caching in 5G Networks

Rasoul Nikbakht, Sarang Kahvazadeh, Josep Mangues-Bafalluy
Centre Tecnològic Telecomunicacions Catalunya (CTTC), Castelldefels, Spain
{rnikbakht, skahvazadeh, jmangues}@cttc.es

Abstract—Edge caching can significantly improve the 5G networks' performance both in terms of delay and backhaul traffic. We use a reinforcement learning-based (RL-based) caching technique that can adapt to time-location-dependent popularity patterns for on-demand video contents. In a private 5G, we implement the proposed caching scheme as two virtual network functions (VNFs), edge and remote servers, and measure the cache hit ratio as a KPI. Combined with the HLS protocol, the proposed video-on-demand (VoD) streaming is a reliable and scalable service that can adapt to content popularity.

Index Terms—Edge caching, Video-on-demand streaming, Reinforcement Learning (RL), 5G, OSM

I. INTRODUCTION

It is expected that by 2022, media content will account for more than 80% of Internet traffic, and popular media will make up the majority of the media traffic. This naturally opens the door for caching and storage in the cloud and edge, as media files are usually large and frequently accessed. In this work, we implemented edge-caching concept as virtual cache servers (VCaches) in the 5G infrastructure. Each VCache is a virtual network function (VNF) that can be controlled by an open source management and orchestration (OSM) software stack.

The main contributions of this work are as following:

• An RL-based framework is used for edge caching. We take the number of requests for the cached content alongside their IDs as the state for the RL agent. Then, we use state-of-the-art soft actor-critic networks to obtain an efficient caching strategy.
• We deploy an end-to-end 5G network with a core residing in the cloud and the cache server in the edge.
• Our solution is scalable and cloud-native.

In [1] centralized and decentralized caching using deep deterministic policy gradient is proposed. We adopt a similar workflow but with two major differences. First, we use the state-of-the-art soft actor-critic network for RL-based caching. Second, the state dimension in our proposed approach is smaller than the one in [1], hence actor and critic networks are significantly simpler. A containerized implementation of proposed video-on-demand streaming using RL-based caching is publicly available in [2].

II. SYSTEM ARCHITECTURE

In this demonstration, we show how the RL-based edge-caching can be deployed, trained, and tested in a 5G network.

This work is founded by the H2020 5GSolutions (Grant Agreement no.856691).

We use Open5GS on the top of a Kubernetes cluster as core network function; OSM as orchestrator; Open-Stack-Ansible (OSA) and Linux container (LXC) as virtual infrastructure manager (VIM); and Amarisoft as radio access network and simbox for emulating user equipment [3].

The edge and remote servers are VNFs and launched using OSM alongside a network slice (NS). The OSM deployment is based on VNF and NS descriptors where the day zero operation for VNFs is conducted using the cloud.init script, which installs the required software, downloads the source code, and finally launches the remote and edge servers. In this implementation, we assumed an edge server located in the network edge and a remote server in the cloud. Since we don’t have a dedicated server outside our organization, we implement both VNFs in the same infrastructure and introduce a reasonable random delay in the link between edge and remote servers.

III. VIDEO-ON-DEMAND SERVICE

Th HLS protocol consists of a playlist and video segments. The user client asks for video playlist (video.m3u8) through 5G link. If content is cached (the playlist and video segments are available in the cache storage), the Nginx in edge server receives the request and serves the user. In addition, the Nginx mirrors the user request to the ML algorithm. If the content is not cached in the edge-server, the edge server re-routes the user request to the remote server. Meanwhile, the application server that includes the RL-based caching algorithm updates the cached content on the edge-server.

IV. SUMMARY

We implemented a video-on-demand service using RL-based caching for 5G networks. Firstly, the RL-based solution outperforms the traditional approach like LFU based caching. Second, the HLS-based video-on-demand service is implemented and managed by OSM, making it a good fit for the new design paradigm of the 5G networks.

REFERENCES

[1] C. Zhong, M. C. Gursoy, and S. Velipasalar, “Deep reinforcement learning-based edge caching in wireless networks,” IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 1, pp. 48–61, 2020.
[2] R. Nikbakht, “Docker-based implementation for video on demand streaming using rl-based edge caching,” https://github.com/RasoulNik/VoD, accessed: 2022-07-27.
[3] S. Kahvazadeh, H. Khalili, R. Nikbakht, and J. Mangues-Bafalluy, “Vertical oriented 5g platform-as-a-service: user-generated content case study,” in FNWF22, 2022.