

# 심층 강화학습 기반 에지 캐싱 및 컴퓨팅 기술의 최신 동향

최 민 석\*

## A Survey of Deep Reinforcement Learning-Based Edge Caching and Computing Techniques

Minseok Choi\*

요 약

매달 수십 엑사바이트 이상 발생하는 데이터 트래픽과 급증하는 모바일 기기를 지원하기 위해서 분산형 에지 네트워크가 조명받고 있다. 특히 에지 캐싱과 에지 컴퓨팅 기술은 전체 트래픽의 60-70%를 차지하면서 소수의 콘텐츠가 반복적으로 요청되는 멀티미디어 서비스를 효율적으로 지원하기에 알맞아 활발히 연구되고 있다. 시간에 따라 변화하는 채널, 콘텐츠 인기도, 현재 캐시 상태에 따라 콘텐츠의 캐싱과 전송 정책을 최적화하는 연구는 에지 서비스가 다양하게 분화되고 사용자 수가 급증하면서 굉장히 높은 복잡도와 처리 시간을 요구하게 되었다. 이에 심층 강화학습을 활용한 콘텐츠 캐싱, 전송, 데이터 오프로딩 기술이 연구되었고, 본 논문에서는 다양한 관점에서 심층 강화학습 기반 에지 캐싱 및 컴퓨팅 기술의 최신 기술 동향을 소개한다. 먼저 다양한 에지 네트워크 시나리오에서 살펴보고, 각종 심층 강화학습 알고리즘이 활용되는 방식에 대해 알아본다. 또한, 다른 최신 에지 기술들과 심층 강화학습의 융합으로 발전된 에지 캐싱과 컴퓨팅 정책을 제안하는 연구를 소개한 뒤, 마지막으로 향후 연구 방향을 제시한다.

**키워드 :** 에지 네트워크, 에지 캐싱, 에지 컴퓨팅, 강화학습

**Key Words :** Edge network, Edge caching, Edge computing, Reinforcement Learning

### ABSTRACT

The distributed edge network has been getting spotlight nowadays for handling tens of exabytes of global data traffic on a daily basis and supporting increasingly exploding mobile devices. Especially, edge caching and edge computing technologies have been actively researched because they are appropriate for efficiently providing multimedia services in which a relatively small part of popular content is repeatedly requested and accounts for about 60-70% of the global wireless data traffic. As wireless edge services have become diversified and the number of mobile users sharply increases, edge caching and computing techniques optimized depending on time-varying channel, content popularity, and the current cache state require high complexity and execution latency. Accordingly, the deep reinforcement learning-based content caching and delivery, and data offloading techniques have been researched, and this paper surveys the state-of-the-art techniques. We first study the related work in a variety of edge network scenarios, and various kinds of reinforcement learning algorithms applied to edge networks are investigated. Also, the advanced edge caching and computing techniques jointly employed the deep reinforcement learning and other techniques are introduced and future research directions are lastly presented.

\* 이 논문은 2021학년도 제주대학교 교육·연구 및 학생지도비 지원에 의해서 연구되었음.

• First Author : Jeju National University Department of Telecommunication Engineering, ejaqmf@jeju.ac.kr, 정회원  
논문번호 : 202109-245-B-RN, Received September 17, 2021; Revised October 16, 2021; Accepted October 23, 2021

## I. 서 론

최근 메모리 및 비메모리 반도체 기술의 급격한 발전으로 인해 차세대 모바일 통신 기기와 소형 기지국은 높은 전력을 비축하면서 데이터를 저장하고 처리할 수 있는 능력이 향상되었다. 이는 네트워크와 데이터 처리 및 학습 프로세스를 점점 분산화시킬 수 있게 해주었고, 모바일 통신 기기의 수가 급격히 증가하면서 중앙집중형 서버나 기지국이 모든 기기를 감독하거나 데이터를 처리하기에도 어려워졌다. 이처럼 데이터의 수집, 처리, 저장, 활용 등이 무선 네트워크의 에지에서 이루어지면서, 방대한 데이터 트래픽을 좀 더 효율적으로 처리할 수 있게 되었다.

무선 에지 네트워크에서 활용되는 대표적인 기술로는 에지 캐싱과 에지 컴퓨팅이 있다. 매달 수십 엑사바이트 이상 발생하는 데이터 트래픽의 60-70%는 멀티미디어 서비스에서 발생하는데<sup>[1,2]</sup>, 멀티미디어 서비스에서는 소수의 인기 있는 콘텐츠 요청 및 전달이 대부분의 트래픽을 차지하는 특징이 있다. 여러 사용자로부터 반복적으로 발생하는 인기 있는 콘텐츠 요청에 효율적으로 응답하기 위하여, 무선 캐싱 기술이 최근 조명받고 발전되어 왔다. 무선 캐싱 네트워크에서는 개별적인 저장장치를 갖는 캐싱 헬퍼라고 불리는 독립체들을 사용자 주변에 설치하고, 인기 있는 콘텐츠를 사용자가 요청하기 전에 미리 캐싱 헬퍼에 캐싱하여 사용자가 요청하면 멀리 위치한 서버까지 알릴 필요 없이 바로 캐싱 헬퍼가 요청받은 콘텐츠를 전달할 수 있다<sup>[3,4]</sup>. 이러한 기술을 통해 반복적이고 중복된 요청이 발생할 때마다 코어 네트워크에 트래픽을 발생시키지 않고, 콘텐츠 전송 속도를 높일 수 있게 되었다.

콘텐츠 및 데이터 저장뿐 아니라 데이터 처리도 에지 네트워크에서 할 수 있게 되었는데, 이는 에지 컴퓨팅 기술을 등장시켰다. 기존에는 모바일 기기가 오버헤드가 높은 데이터 처리를 직접 하기 어려워 클라우드 서버에 오프로드(offload) 하였다. 그러나 점점 에지에서의 데이터 처리 능력이 향상되면서, 포그 컴퓨팅(Fog computing)을 거쳐 모바일 기기가 바로 자신과 가까운 에지 서버에 데이터 오프로드 할 수 있는 에지 컴퓨팅 기술까지 내려왔다. 더 나아가 최근에는 이웃한 모바일 기기에게까지 데이터 오프로딩 할 수도 있다. 이러한 에지 컴퓨팅 기술은 무선 캐싱 네트워크에서 전송할 콘텐츠의 처리에 활용될 수 있다. 특히, 온라인 비디오 서비스에서는 비디오 콘텐츠들의 비트율(bitrate)을 무선 채널 환경과 이용 가능한 자원에

따라 조절하는 트랜스코딩(transcoding)이 필요할 때 에지 컴퓨팅 기술이 활용될 수 있다.

통신 기기들의 성능이 좋아지고 무선 네트워크가 점점 분산화되어 가면서, 캐싱 헬퍼뿐 아니라 스몰 셀 네트워크의 소형 기지국, 차량 네트워크의 노변 장치(Roadside Unit, RSUs), 모바일 기기에도 콘텐츠를 캐싱하는 기술이 등장하였고<sup>[5]</sup>, 트랜스코딩 또는 데이터 처리를 직접 할 수도 있게 되었다. 다양한 장치 내 효율적인 콘텐츠 캐싱 기법들은 사용자들이 어떤 콘텐츠를 요청할 것인지, 즉 콘텐츠의 인기도와 사용자의 선호도에 큰 영향을 받는다. 또한, 콘텐츠의 전송 및 컴퓨팅까지 고려할 경우 사용자의 위치, 채널 환경, 이용 가능한 자원도 중요한 요소가 된다. 기존에는 이와 같은 네트워크 특성들을 반영한 콘텐츠 캐싱과 전송에 대한 최적화 문제를 수학적으로 풀어낸 기법이 많았으나<sup>[6-8]</sup>, 실제 환경에서는 콘텐츠의 인기도나 사용자의 선호도는 일반적으로 알려지지 않고 시간에 따라 변화하며, 채널 환경 역시 추정하기 어려운 경우가 많다. 또한, 고려해야 할 네트워크 특성이 많아지면서 최적의 캐싱 방법을 찾아내는 수학적 알고리즘의 복잡도도 매우 높아졌다.

따라서, 콘텐츠의 인기도와 사용자 선호도 등 알려지지 않은 사용자 및 네트워크 특성을 예측하고 콘텐츠 캐싱과 전송 기법을 학습하는 방법이 최근 활발히 연구되고 있다. 특히 사용자의 콘텐츠 요청 패턴이나 시간에 따른 콘텐츠 인기도의 변화, 사용자의 위치 등은 모두 마르코프 특성을 띠고 있어서, 많은 연구자가 주로 강화학습을 활용한 기술에 초점을 맞추고 있다. 강화학습을 활용할 시 현재의 콘텐츠 인기도나 사용자 선호도, 채널 환경, 사용자 위치 등을 관찰하고 에이전트에게 가장 높은 이득을 줄 수 있는 콘텐츠 캐싱, 전송, 자원 할당 등에 대한 일련의 결정을 내려줄 수 있다.

강화학습 기반의 에지 캐싱<sup>[9,10]</sup>, 기계학습 기반의 에지 컴퓨팅<sup>[11,12]</sup>에 대한 리뷰 논문들은 여럿 있으나, 대부분 에지 캐싱이나 에지 컴퓨팅 등 한 가지 기술에 초점을 맞춘다. 본 논문에서는 두 기술을 통합하여 활용하는 일반적인 에지 네트워크로 확장하여 강화학습의 쓰임새를 살펴보고자 한다. 따라서 에지 캐싱 및 에지 컴퓨팅 시스템에서의 심층 강화학습의 효용성에 대해 먼저 살펴보고, 다양한 심층 강화학습 알고리즘들이 어떤 이유로 어떻게 활용되는지 연구 동향을 조사한다. 또한, 다른 최신 기술들과 에지 네트워크가 융합된 시나리오에서의 강화학습 연구도 살펴본 후, 점점 분산화되고 복잡해지는 콘텐츠 전송을 위한 강

화학습 기반의 에지 기술의 미래 연구 방향성에 대해서 논의한다.

본 논문의 세부 구성은 다음과 같다. II 장에서는 본 논문에서 고려하는 에지 네트워크에서의 캐싱과 컴퓨팅 시스템을 소개한다. III 장에서는 다양한 에지 네트워크 상황에 따라 심층 강화학습을 활용하는 기술을 소개한다. IV 장에서는 심층 강화학습 알고리즘 별로 에지 캐싱과 컴퓨팅에 적용되는 방식을 살펴보고, V 장에서는 다른 최신 기술들과 심층 강화학습을 융합하여 에지 네트워크에 활용한 기술을 살펴본다. VI 장에서는 미래 연구 방향성에 대해 논의하고, VII 장에서는 논문의 결론을 지으며 마무리한다.

## II. 에지 네트워크에서의 캐싱과 컴퓨팅

대부분의 무선 캐싱 기술에 관한 연구는 그림 1과 같은 계층적인 콘텐츠 전송 네트워크(content delivery network, CDN)를 고려한다. 본 논문에서 소개할 심층 강화학습 기반의 에지 캐싱 및 컴퓨팅 기술 최신 연구들도 그림 1과 비슷한 형태의 네트워크를 고려하기 때문에, 그림 1과 같은 에지 캐싱을 활용한 콘텐츠 전송 네트워크에 대해 본 장에서 살펴본다.

그림 1에서 모바일 사용자, 에지 서버, 클라우드 서버가 계층적인 CDN을 구성한다. 일반적으로, 에지 서버와 사용자는 무선 채널 링크로, 클라우드 서버와 에지 서버 간에는 유선 백홀 링크로 연결되어 있다. 사용자는  $N$ 명, 에지 서버는  $K$ 개, 콘텐츠는 총  $F$ 개가 존재한다고 가정하자. 에지 서버는 제한된 크기  $M$ 의 캐시를 갖고 있으며, 에지 컴퓨팅 서버도 구축하여 전송할 콘텐츠의 비트율을 조절할 수도 있다. 각 콘텐츠는 총  $L$ 개의 비트율을 가질 수 있고, 비트율에 따른

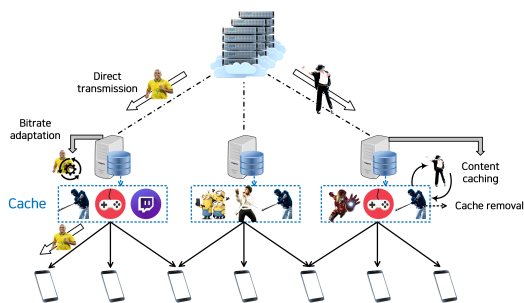


그림 1. 에지 캐싱을 활용한 콘텐츠 전송 네트워크  
Fig. 1. Edge caching-assisted content delivery network

콘텐츠의 파일 크기는 모든  $i \in \tilde{F} \equiv \{1, \dots, F\}$ 와  $l \in \tilde{L} \equiv \{1, \dots, L\}$ 에 대하여  $S_{i,l}$ 이라고 가정하자. 에지 서버  $k$ 의 비트율  $l$ 의 콘텐츠  $i$  캐싱 여부 지표는  $x_{k,i,l} \in \{0, 1\}$ 라 하면, 모든 에지 서버에의 캐싱 기법은  $\mathbf{x} = [x_{k,i,l} : k \in \tilde{K}, i \in \tilde{F}, l \in \tilde{L}]$ 를 구하는 것으로 표현할 수 있다. 이때, 에지 서버는 제한된 캐시 크기를 갖고 있으므로, 아래 조건을 만족해야 한다.

$$\sum_{i=1}^F \sum_{l=1}^L x_{k,i,l} S_{i,l} \leq M \quad \forall k \quad (1)$$

가장 낮은 계층에서는 모바일 사용자들이 콘텐츠를 요청한다. 각 사용자가 요청할 콘텐츠의 종류는 확률적으로 모델링 될 수 있으며, 콘텐츠의 인기도 및 사용자의 선호도에 영향을 받는다. 모바일 기기의 캐싱이 가능할 경우, 사용자는 본인 기기에 해당 콘텐츠가 있는지 확인하고, 없다면 주변 사용자에게 콘텐츠를 요청한다. 주변 모바일 기기에도 콘텐츠가 없을 때는, 본인이 소속된 영역의 인접 에지 서버에게 요청한다. 이때, 에지 서버  $k$ 의 영역 내 사용자  $n$ 의 소속 여부 지표를  $y_{kn} \in \{0, 1\}$ 라 하자. 주변 모바일 기기나 인접 에지 서버의 캐시에 해당 콘텐츠가 캐싱되어 있다면 ( $y_{kn} = 1, x_{k,i,l} = 1$ ), 클라우드 서버에게 요청할 필요 없이 바로 타겟 사용자에게 콘텐츠를 전달해 줄 수 있다. 에지 서버는 에지 컴퓨팅 서버를 함께 구축하여 전송하려는 콘텐츠를 트랜스코딩하거나 데이터 처리를 수행할 수 있다. 캐싱과 마찬가지로 비디오 트랜스코딩은 서버, 기지국, 모바일 기기 모두 가능할 수 있으며, 사용자는 본인 기기의 컴퓨팅 자원이 부족한 경우 그림 2에서처럼 데이터 처리를 서버나 주변 기기에 오프로딩을 하게 된다.

그러나, 에지 서버가 요청한 콘텐츠를 캐싱하고 있지 않을 경우 ( $y_{kn} = 1, x_{k,i,l} = 0$ ), 모든 콘텐츠를 저장하고 있는 라이브러리에 접근 가능한 클라우드 서버로부터 직접 전송받아야 한다. 클라우드 서버는 언제나 어떤 콘텐츠 요청이건 응답해 줄 수 있지만, 값비싼 백홀 링크 사용이 필요하고 비교적 전송 시간이 오

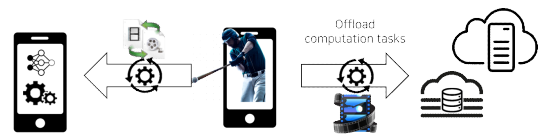


그림 2. 에지 캐싱을 활용한 콘텐츠 전송 네트워크  
Fig. 2. Edge caching-assisted content delivery network

래 걸린다는 단점이 있다. 따라서, 최대한 네트워크 설계자는 에지 서버와 모바일 기기의 제한된 캐시 크기에 주변 사용자가 많이 요청할 거라 예상되는 콘텐츠 위주로 캐싱하는 효율적인 방법을 찾고자 하며, 이를 캐싱 문제라고 한다.

캐싱 기술을 평가하는 성능 지표로는 대표적으로 캐시 히트율<sup>[6]</sup>과 콘텐츠의 평균 비트율<sup>[7]</sup>, 전송 지연 시간<sup>[8]</sup>, 네트워크 자원 소모 비용 등이 있다. 캐시 히트는 어떤 사용자가 임의의 콘텐츠를 요청했을 때 인접 에지 서버 혹은 캐싱 헬퍼가 해당 콘텐츠를 캐싱하고 있어 바로 응답 가능한 사건으로 정의된다. 시간  $t$ 에서 사용자  $n$ 이 요청한 콘텐츠의 인덱스를  $i_n(t)$ , 콘텐츠  $i$ 의 인기도를  $f_i$ 라 하면, 시간  $t=1, \dots, T$  동안 사용자들의 평균 캐시 히트율은 다음과 같이 표현될 수 있다.

$$\frac{1}{NT} \sum_{t=1}^T \sum_{n=1}^N f_{i_n(t)} \cdot \left( \sum_{k=1}^K \sum_{l=1}^L x_{k,i_n(t),l} \cdot y_{k,n} \right) \quad (2)$$

캐시 히트율을 성능 지표로써 콘텐츠 캐싱을 최적화한 연구<sup>[9]</sup>에서 찾아볼 수 있다.

콘텐츠의 전송 지연 시간을 성능 지표로 삼은 연구에서<sup>[8]</sup>, 에지 서버와 사용자 간의 무선 링크 전송률과 클라우드 서버와 에지 서버 간의 백홀 링크의 전송 지연 시간을 고려해야 한다. 콘텐츠 전송은 에지 서버가 캐싱한 콘텐츠를 그대로 전송하는 방법, 에지 서버가 캐싱한 콘텐츠의 비트율을 조절하여 전송하는 방법, 에지 서버가 캐싱하지 않아 클라우드 서버가 전달해주는 방법으로 나뉜다. 비트율  $l$ 의 콘텐츠  $i$ 를 에지 서버에서 사용자에게 무선 링크로 전송하는 지연 시간을  $\tau_{i,l}^{comm}$ , 클라우드 서버에서 백홀 링크로 에지 서버에 전송하는 지연 시간을  $\tau_{i,l}^{back}$ , 비트율  $l$ 의 콘텐츠  $i$ 를 비트율  $l$ 로 트랜스코딩 하는데 걸리는 시간을  $\tau_{i,l,l}^{comp}$ 라 하자. 특정 사용자가 서버  $k$ 에게 비트율  $l$ 의 콘텐츠를 요청할 때 겪는 평균 지연 시간은 다음과 같이 적을 수 있으며,

$$\sum_{i=1}^F f_i (\tau_1 + \tau_2 + \tau_3) \quad (3)$$

위 식에서  $\tau_1$ 은 요청한 콘텐츠  $i$ 의 비트율  $l$ 을 에지 서버가 캐싱할 때의 전송 지연 시간,  $\tau_2$ 는 에지 서버가 콘텐츠  $i$ 의 다른 비트율  $l'$ 을 캐싱할 때의 트랜스

코딩과 전송 지연 시간,  $\tau_3$ 은 클라우드 서버로부터 전송받는 지연 시간을 의미하며 각각 아래와 같이 표현된다.

$$\tau_1 = x_{k,i,l} \tau_{i,l}^{comm} \quad (4)$$

$$\tau_2 = (1 - x_{k,i,l}) \cdot \min_{\substack{l' \neq l \\ x_{k,i,l'} \neq 0}} \{ \tau_{i,l',l}^{comp} + \tau_{i,l}^{comm} \} \quad (5)$$

$$\tau_3 = \prod_{l'=1}^L (1 - x_{k,i,l'}) (\tau_{i,l}^{back} + \tau_{i,l}^{comm}) \quad (6)$$

전송 지연 시간뿐만 아니라, 에너지 소모 혹은 네트워크 비용 발생도 콘텐츠 전송과 컴퓨팅 과정에서 발생하므로, (3)-(6)과 같이 표현될 수 있으며 해당 값을 최소화하는 캐싱과 컴퓨팅 정책을 얻고자 한다.

### III. 다양한 에지 네트워크 시나리오에서의 심층 강화학습 활용

본 장에서는 다양한 에지 네트워크 시나리오에서 심층 강화학습이 어떻게 사용되고 그 효용성에 대해 연구한 논문을 살펴본다. 에지 컴퓨팅, 에지 캐싱, 차량 네트워크, 비디오 스트리밍 시스템에서 각 시나리오의 특성을 반영한 심층 강화학습 기반의 연구를 소개한다.

#### 3.1 에지 캐싱과 에지 컴퓨팅

논문<sup>[13]</sup>에서는 사물 인터넷(Internet of Things, IoT) 에지 컴퓨팅 시스템에서 IoT 기기가 수집한 데이터 처리를 위한 태스크 스케줄링 기술을 처리 지연 시간을 줄이면서 컴퓨팅 자원을 효율적으로 이용하도록 설계한다. 수집된 센서 데이터가 에지 서버에 업로드될 때마다 네트워크의 job queue에 데이터 처리 태스크가 쌓이고, 태스크 스케줄링은 job queue에 대기 중인 태스크 처리를 위해 할당할 자원의 크기와 할당할 시간 간격의 결정을 의미한다. 에지 서버는 job queue에 쌓인 태스크의 종류와 크기, 현재 이용 가능한 컴퓨팅 자원의 개수, 이용 가능한 자원 개수의 변화를 관찰하고, 태스크 스케줄링 액션을 결정하거나 지나친 자원 소모 시 이용 가능한 자원의 개수를 변화시킬 수도 있다. 처리 지연 시간과 컴퓨팅 자원 소모의 가중합으로 Markov decision process(MDP)의 보상을 정의하였으며, 잘 알려진 Deep Q-Network(DQN) 알고리즘으로 최적의 액션을 얻어냈다. 저자들은 현재

이용 가능한 자원의 수가 다를 때 얻은 경험 데이터들은 서로 연관성이 적음을 실험적으로 확인하고, 리플레이 메모리(replay memory)를 이용 가능한 자원의 개수마다 따로 만들어 DQN을 작동시켜 수렴 속도를 빠르게 하였다.

일반적인 에지 캐싱 네트워크에서의 콘텐츠 캐싱 방법<sup>[14]</sup>에서 큐러닝(Q-learning)으로 풀어진 바 있다. 기지국에 매 시각 새로운 콘텐츠를 캐싱하는 액션을 서버에서 내려주는데, [14]에서는 전체 네트워크에 대한 콘텐츠의 글로벌 인기도와 각 기지국 영역 내 로컬 인기도를 모두 고려하였다. 글로벌과 로컬 인기도가 마르코프 특성을 갖는다고 가정된 뒤 인기도와 기지국들의 캐시 상태를 관찰하여 시간에 따라 변화한 인기도에 따라 새롭게 캐싱할 콘텐츠를 결정한다. 이때 MDP의 보상은 캐시를 업데이트하기 위한 비용, 글로벌과 로컬 인기도와 캐시 상태의 유사도의 가중합으로 정의된다.

에지 캐싱과 컴퓨팅을 모두 고려한 시스템에서 캐싱과 콘텐츠 전송을 큐러닝과 DQN으로 결정하는 기술은 [15]에 제안되었다. [15]에서는 백홀 비용과 무선 자원 비용을 동시에 줄이는 콘텐츠 캐싱과 전송 최적화 문제를 두 하위 문제로 나눈 뒤 계층적 강화학습 구조를 통해 문제를 풀어나갔다. 사용자 디바이스도 캐싱이 가능하여 본인이 직접 콘텐츠 제공을 할 수도 있고, 인근 기지국에 요청한 콘텐츠가 캐싱 되었을 경우 바로 전송받거나(푸싱, pushing), 그렇지 않으면 서버로부터 기지국을 통해 즉각 전송받는다(reactive transmission). 네트워크 비용은 이 세 종류의 콘텐츠 전송에 요구되는 비용의 가중합으로 정의된다. 서버는 네트워크 비용을 최소화하기 위해 사용자의 콘텐츠 요청과 기지국의 캐시 상태를 관찰하여 어떤 콘텐츠를 미리 푸싱하고, 캐싱된 콘텐츠 중 어떤 것을 제거할지 결정한다. 저자들은 기지국 캐시 정보를 고정된 채로 먼저 선형 근사를 이용한 큐러닝으로 사용자 디바이스의 캐싱 결정을 내리고, 이후에 사용자 디바이스 캐싱을 고정된 채로 기지국 캐싱을 DQN으로 결정하는 계층적 강화학습 구조를 설계하였다.

위 연구<sup>[13-15]</sup>는 기초적인 강화학습 알고리즘인 큐러닝 또는 DQN으로도 기존의 최적화 기반 캐싱 및 컴퓨팅 방법보다 더 낮은 복잡도로 비슷하거나 더 좋은 성능을 낼 수 있음을 보였다. 따라서, 연구자들이 좀 더 다양한 에지 네트워크 시나리오에 최신 강화학습 알고리즘을 적용하는 데 관심을 가지도록 만들었다.

### 3.2 차량 네트워크

차량 네트워크에서는 RSU가 캐싱 헬퍼의 역할을 할 수 있으며, 이동성 때문에 사용자의 위치 파악이 중요하다. 차량 네트워크에서 에지 캐싱과 컴퓨팅이 활용되는 시나리오에서 차량 데이터 스케줄링을 결정을 심층 강화학습으로 내리는 연구는 [16]에서 된 바 있다. 각 차량에서 생성된 데이터는 스스로 처리하거나, 주변의 다른 차량 또는 소속된 RSU에게 오프로딩하거나, 어느 곳에서도 현재 데이터 처리를 할 수 없다면 캐싱 큐에 저장한다. 데이터 스케줄링은 위 옵션 중 하나의 데이터 처리 옵션을 선택하는 것을 의미하고, 차량과 RSU의 위치와 큐 상태에 따라서 처리 지연 시간과 에너지 소모를 최소화하는 스케줄링 결정을 내린다. 해당 문제는 DQN 알고리즘을 적용하여 데이터 처리와 캐싱 간의 균형을 잘 맞추어낸다.

연구 [17]에서는 차량에도 캐싱할 수 있는 경우를 고려했고, 서로 다른 두 시간대에 내리는 액션을 분류하여 MDP 문제를 만들었다. 콘텐츠 전송은 매우 짧은 지연 시간 조건을 갖고 단시간에 자주 발생하는 반면, 캐싱이나 캐시 상태 업데이트는 네트워크 혼잡이 적거나 콘텐츠 인기도에 변화가 있을 때, 비교적 긴 주기로 이루어진다. 따라서, 본 논문에서는 큰 시간대로 캐싱을, 작은 시간대로 콘텐츠 전송 결정을 내린다. 이때, 차량의 셀의 이동을 확률적으로 모델링하여 차량의 위치 이동을 전이 확률로 표현하게 된다. 저자들은 콘텐츠를 캐싱하고 접근하는 데 드는 비용과 지연 시간이 길면 발생하는 페널티를 합친 네트워크 비용을 최소화하도록 다른 시간대에 콘텐츠 캐싱과 전송을 결정하는 문제를 deep deterministic policy gradient(DDPG)로 풀어낸다.

### 3.3 비디오 스트리밍 시스템

비디오 스트리밍 시스템에서는 콘텐츠 전송 네트워크와 달리 비디오 스트림 파일의 특성을 반영하여 비디오 캐싱 및 전송 기술을 설계해야 한다. 비디오 스트림의 가장 큰 특징은 동일 콘텐츠가 여러 퀄리티 레벨을 가질 수 있다는 점이다. 하나의 스트림 파일은 수많은 비디오 조각(chunk)들로 나뉘고 각 조각은 순서대로 동영상 재생의 일부분을 담당한다. 특히 동적 스트리밍 서비스의 경우 각 비디오 조각들을 서로 다른 퀄리티 레벨을 가질 수도 있어 송신기가 매 시각 전송하는 비디오 조각의 비트율(bitrate)을 조절할 수 있다. 또한, 수신자는 모든 비디오 조각들은 다 받기 전에 먼저 스트리밍을 할 수 있다. 논문 [18]의 저자들은 채널과 사용자의 수신 큐 상태에 따라서 비디오를

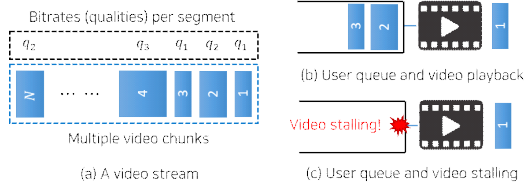


그림 3. 비디오 스트림 파일과 비디오 스트리밍  
Fig. 3. A video stream file and video streaming

전송할 헬퍼를 선택하고, 매 시각 전송할 비디오 조각의 수와 비트율을 조절하는 문제를 MDP로 정립했다. 사용자가 끊기지 않는 재생을 즐기기 위해서는 사용자 큐에 항상 다음에 재생할 비디오 조각이 있어야 한다. 따라서, 논문 [18]에서는 사용자 큐가 비어있지 않으면서 평균 전송 비트율을 최대화하는 전송 기술을 구하고자 했으나, 해당 문제를 동적 계획법(dynamic programming)으로 풀어 수많은 채널의 상태를 다루기 힘들었고, 확장성이 낮았다.

논문 [19]에서는 위와 유사한 비디오 전송 문제를 차량 캐싱 네트워크에 단말 간 직접 통신(Device-to-Device, D2D)이 공존하는 D2D underlaid vehicle caching network에서 강화학습으로 풀어냈다. 결정론적인 정책(deterministic policy)을 업데이트하는 DDPG를 사용하여 복잡도를 낮추면서도 연속적인 액션 공간(action space)을 다룰 수 있게 하여 비디오를 전송해주는 주변 차량의 전력까지 제어할 수 있도록 했다.

앞서 언급한 [18, 19]에서는 비디오 전송만을 보았는데, [20]에서는 비트율별 비디오 캐싱 방법을 제시하였다. 저자들은 모바일 에지 클라우드에 모든 콘텐츠가 가장 높은 비트율로 저장되었는데, 채널과 사용자의 상태에 따라서 비트율을 낮춰야 하는 상황을 고려하였다. 따라서, [20]에서의 캐시 히트율은 원하는 콘텐츠가 모바일 에지 클라우드에 있는지가 아닌 원하는 비트율의 콘텐츠가 있는지로 정의된다. 트랜스코딩에 소모되는 자원과 시간을 절약하기 위해 다른 비트율의 동일 콘텐츠를 추가로 캐싱할지를 결정한다. 저자들은 사용자 요청에 따른 비트율 조절, 대역폭과 전력 제어를 위한 문제와 사용자에게 전송하고자 트랜스코딩한 파일을 캐싱할지 결정하는 문제로 나누어 순차적인 강화학습 구조를 만들었고, asynchronous advantage actor-critic(A3C) 알고리즘을 이용하여 풀 어냈다.

## IV. 심층 강화학습 알고리즘별 에지 네트워크 기술

강화학습이 에지 캐싱과 에지 컴퓨팅을 포함한 다양한 에지 네트워크 시나리오에서 좋은 성능과 효율적인 네트워크 비용을 제공할 수 있는 도구로서의 가능성을 확인하였다. 이어서 본 장에서는 다양한 심층 강화학습 알고리즘이 어떤 이유로 어떻게 사용되는지 살펴본다. 가치 기반 알고리즘, 정책 그래디언트 알고리즘(policy gradient algorithm), 다중 에이전트 강화학습으로 나누어 관련 최신 캐싱 및 컴퓨팅 연구 동향을 소개한다.

### 4.1 가치 기반 알고리즘

에지 네트워크가 점점 더 다양한 서비스를 제공하고 많은 사용자를 서비스하면서, 에지 캐싱과 컴퓨팅 결정을 내리기 위한 MDP의 상태와 액션 공간의 크기가 매우 커져 차원의 저주를 겪게 되었다. 이를 해결하기 위해 간단한 형태의 강화학습 알고리즘인 큐러닝과 DQN 등의 가치 기반 알고리즘을 활용한 에지 기술들이 등장했다.

앞서 소개한 [14]에서는 큐러닝을 사용하였는데, 큐함수를 예측하기에 높은 복잡도가 요구되어 큐함수의 선형 근사를 통해 학습 속도를 높이고 확장성을 높였다. 논문 [15]에서도 사용자 디바이스와 기지국의 콘텐츠 캐싱 정책을 두 개의 하위 문제로 나누어 풀 때, 사용자 디바이스 캐싱 문제는 큐러닝으로 풀 바 있다. 저자들은 디바이스 캐싱 정책에 대해서는 DQN을 사용하였을 시 수렴 성능이 좋지 않아 큐러닝을 사용하였다고 언급하였으며, 대신 복잡도를 낮추기 위하여 가치 함수를 근사하는 과정을 수차례 거쳤다.

이처럼, 상황에 따라서 큐러닝과 DQN 중에 수렴 성능이 더 좋은 것이 다를 수 있지만, 일반적으로는 DQN이 더 좋은 수렴성과 학습 속도를 보여 최근 연구들<sup>[13,15,16,22]</sup>에서는 DQN을 활용하고 있다. 그러나 DQN에서도 여전히 상태 공간과 액션 공간이 방대할 경우 학습 속도가 느릴 수 있어 리플레이 메모리를 상태에 따라 분리하거나<sup>[13]</sup>, 에지 캐싱 또는 컴퓨팅의 제한 조건을 이용해 액션 공간을 축소하는 과정을 거친다<sup>[15]</sup>.

### 4.2 정책 그래디언트 알고리즘

환경의 상태나 액션 공간이 연속적일 경우, 가치 이터레이션을 사용하게 되면 무한의 가치 함수를 예측해야 하므로 딥러닝을 활용하더라도 태생적인 한계가 있다. 반면, 정책 자체를 경사 하강법으로 최적화할

수 있다면, 연속적인 상태와 액션 공간을 다루기 쉬워진다.

앞서 보았던 [18]에서는 비디오 스트리밍 시스템에서 트랜스코딩을 통해 비트율을 조절하는 액션 공간이 연속적인 문제를 다루기 위해서 A3C 알고리즘을 사용하였다. 저자들은 연속적인 액션 공간에 대해 정책을 다변량 정규분포(multivariate normal distribution)로 근사하고, 해당 확률 밀도 함수의 평균 값과 분산을 예측하는 것으로 연속적인 액션 공간에서의 정책을 학습했다.

차량 네트워크에서 서로 다른 시간대에 콘텐츠 캐싱과 전송을 최적화한 [17]에서는 확률적인 정책이 아닌 결정론적인 정책을 기반으로 학습하는 DDPG를 사용하였다. 이 논문에서는 차량 간의 협력을 고려하고 하나의 콘텐츠 전송이 크기에 따라 여러 시간 슬롯에 걸쳐 전송된다고 가정하여, 주변 차량의 상태와 현재 콘텐츠 전송 상태가 추가되어 상태 공간이 매우 광범위해진다. 결정론적 정책의 학습은 확률적인 정책 학습과 달리 전체 상태와 액션 공간에 대해서 정책을 완성할 필요가 없어 넓은 상태와 액션 공간을 다룰 수 있다. 그러나 결정론적 정책을 학습할 경우 에이전트가 ‘탐험(exploration)’하지 못하므로 인위적으로 잡음을 주입하는 방식을 사용해야 한다.

#### 4.3 다중 에이전트 강화학습

에지 네트워크가 다양한 서비스를 지원하기 위해 분화되고, 규모가 증가하면서 강화학습 에이전트가 콘텐츠 캐싱 및 관련 네트워크 결정을 내려주기 위해 매 시각 수집해야 할 네트워크 내 모든 상태 정보가 매우 방대해졌다. 이는 매 시각 매우 큰 데이터 트래픽을 유발하고, 긴 지연 시간을 발생시키게 되었다. 이를 해결하기 위해, [14-16]에서는 각 기지국이 본인 영역 내의 네트워크 정보만을 수집하여 독립적으로 강화학습 알고리즘으로 네트워크 결정을 내렸다. 그러나, 여러 기지국에 적용 가능한 에지 캐싱 및 컴퓨팅 알고리즘을 설계하거나 인접한 기지국 사이에 관계성이 있을 때는 각 기지국이 독립적으로 학습하는 것으로는 효율적인 캐싱 및 컴퓨팅 정책을 얻기 어렵다.

논문 [21]에서는 인접한 기지국이 광섬유로 연결되어 서로 협력할 수 있는 CDN을 고려하였고, 다중 에이전트 강화학습 알고리즘으로 인접 기지국과의 협력을 반영한 비디오 캐싱 기술을 제안하였다. 이 논문에서는 인접 기지국의 영역에서 동일 파일 요청이 얼마나 자주 발생하는지를 content similarity ratio(CSR)로 수치화하고, 캐싱 상태에 따라 콘텐츠를 소속 기지

국, 협력 기지국, CDN 서버로부터 전송받는 평균 지연 시간과 네트워크 비용을 최소화하는 캐싱 정책을 고안했다. 각 기지국 에이전트는 협력할 수 있는 인접 기지국의 캐싱 정책을 상태 정보로 받아들여야 하는데, [21]에서는 상태 공간이 지나치게 방대해지지 않도록 낮은 차원을 갖는 인접 기지국의 캐싱 정책의 확률 심플렉스(probability simplex)만을 상태 정보로 이용하는 fingerprint 기반 방식을 사용한다. 각 기지국은 연속적인 액션 공간을 갖는 확률론적 캐싱 정책을 학습하기 위해 advantage actor-critic(A2C) 알고리즘을 사용한다.

### V. 심층 강화학습과 다른 기술들과의 융합

본 장에서는 다양한 최신 통신 및 학습 기술들이 어떻게 심층 강화학습 알고리즘과 함께 에지 캐싱 및 컴퓨팅 기술에 활용되는지 살펴본다. 단일 에이전트 강화학습을 분산된 형태로 가능하게 하는 연합 학습, 다중 에이전트 강화학습을 도울 수 있는 생성적 적대 신경망, 콘텐츠 전송 네트워크의 높은 데이터 트래픽을 처리하기 위한 비동기 다중 접속 기술이 심층 강화학습과 함께 사용된 연구 동향을 소개한다.

#### 5.1 연합 학습

연합 학습은 분산 학습의 일종으로, 분산된 사용자들이 본인의 데이터셋을 공유하지 않고 각자 학습을 진행한 뒤 학습된 모델만을 서버에 업로드하여 사용자의 프라이버시를 보호하는 기술이다. 이때, 서버에서는 수신한 사용자들의 로컬 학습 모델을 응집하여 글로벌 모델을 생성하고, 이 모델은 특정 사용자의 편향된 데이터셋이 아니라 일반적으로 모든 사용자에게 활용될 수 있도록 한다. 에지 캐싱과 에지 컴퓨팅은 각각 분산 스토리지, 분산 컴퓨팅의 일종으로 연합 학습과 더불어 시너지 효과를 낼 것으로 기대되며, 차세대 분산 시스템을 구축하는 데 중요한 역할을 할 것으로 보인다.

논문 [22]에서는 DQN 알고리즘으로 캐싱과 데이터 오프로딩, 자원 할당을 결정하는데, 이 중 캐싱 결정을 연합 학습 형태로 내릴 수 있게 알고리즘을 설계했다. 여러 종류의 서비스를 기지국이 에지 노드들을 거쳐서 모바일 사용자에게 제공하는데, 각 서비스에 필요한 데이터를 캐싱할 수도 있고, 필요한 태스크들을 처리하기도 해야 한다. 태스크 처리는 실시간으로 해야 하지만, 서비스 캐싱은 트래픽이 적은 시간대에 가끔 진행하므로 서로 다른 두 시간대의 액션을 내리

도록 MDP를 모델링한다. 저자들은 두 개의 시간대별로 각각 DQN 알고리즘을 작동하는 two-timescale DRL(2Ts-DRL) 방식을 제안한다. 하나의 DQN은 빠른 시간대에 해당하며, 태스크 오프로딩과 오프로딩에 필요한 통신 자원을 할당한다. 여러 번의 빠른 시간대의 결정을 내린 후에, 그동안 소모했던 자원과 오프로딩 결과를 바탕으로 다른 DQN에서 서비스 캐싱 결정을 내리는 계층적 강화학습 프레임워크를 구성한다. 이때, 모든 에지 서버 영역에 적용 가능한 서비스 캐싱 정책을 학습하기 위해 연합 학습 세팅을 활용한다. 각 에지 서버는 본인 영역 내 사용자의 데이터를 활용하여 서비스 캐싱을 위한 DQN 모델을 학습하고, 학습한 모델은 클라우드 서버에 업로드된다. 클라우드 서버에서는 여러 에지 서버들이 보낸 모델들을 응집하여 글로벌 모델을 생성하고, 다시 모든 에지 서버들에게 글로벌 모델을 내려주게 된다. 이러한 과정을 반복하여 모든 에지 서버들에 적용되는 서비스 캐싱 정책 학습 모델을 생성한다. 해당 논문 [18]에서는 평균 처리 시간과 평균 에너지 소모의 균형을 잘 맞추면서 두 지표 모두 절약하는 결과를 보여주었다.

## 5.2 생성적 적대 신경망

생성적 적대 신경망(Generative Adversarial Network, GAN)은 데이터셋의 분포를 알아내고 모사하려는 생성자(Generator)와 입력으로 들어온 값이 실제 데이터셋으로부터 나온 것인지 아니면 생성자가 만들어낸 것인지 평가하는 판별자(Discriminator)가 서로 경쟁하면서 서로의 성능을 점차 개선해나가는 인공 신경망 알고리즘이다. 이러한 GAN의 개념은 큐함수의 값을 직접 예측하지 않고 큐함수의 분포를 예측하는 distributional 강화학습에 활용될 수 있다.

논문 [23]에서는 캐싱 헬퍼 역할을 하는 스몰 셀 기지국들이 서로 광섬유로 연결되어 협력할 수 있는 네트워크에서의 캐싱 정책을 GAN을 활용한 강화학습으로 제안하였다. 사용자는 본인이 소속된 기지국으로부터 요청한 콘텐츠를 받거나, 소속 기지국이 콘텐츠를 캐싱하고 있지 않다면 인접 기지국의 도움을 받거나, 중앙 서버로부터 받을 수 있다. 이때, 인접 기지국과 중앙 서버로부터 콘텐츠를 받을 때 발생하는 통신 비용의 가중합을 최소화하는 캐싱 방법을 찾고자 한다. 저자들은 시간에 따라 변화하는 콘텐츠의 인기도와 캐시 상태를 관찰하여 캐싱 결정을 내리도록 했는데, 콘텐츠의 인기도가 연속적인 상태 공간(state space)을 갖게 되어 큐함수의 값을 직접 근사하지 않고 분포를 추정하여 캐싱 정책을 학습하고자 했다. 이

때 큐함수의 분포를 예측하기 위해서 GAN이 사용되었고, GAN 내부에서 생성자가 흉내낸 분포를 판별자가 평가하기 위해서 DQN에서 리플레이 메모리에 저장한 경험 값을 사용하도록 하였다. 미리 저장된 경험을 토대로 얻은 큐함수의 분포를 기준으로 생성자가 현재 상태만 입력으로 받고 흉내낸 분포를 검사한다. 이처럼 GAN을 활용한 distributional DQN 알고리즘인 QGAN으로 캐싱 정책 모델을 학습하였다.

## 5.3 비직교 다중 접속

비직교 다중 접속(Nonorthogonal Multiple Access, NOMA) 기술은 동일 시간과 주파수 자원으로 여러 수신자에게 신호를 보내는 기술로, 고속 통신과 많은 연결성을 지원하는 차세대 네트워크의 핵심 기술이다. 대표적인 NOMA 기술은 같은 자원으로 송신하려는 신호들에 서로 다른 전력을 할당한 뒤 합한 중첩된 신호를 송신하고, 일부 수신자는 할당된 전력 비율을 이용하여 순차적 간섭 제거(successive interference cancellation, SIC)로 본인이 원하는 신호를 복원해낸다. 이때, 사용자 디바이스가 직접 캐싱을 할 수 있는 경우, NOMA에서의 SIC를 할 때 간섭 신호를 예측할 필요 없이 본인이 캐싱한 데이터를 활용하여 바로 제거할 수 있게 된다. 이러한 특징을 바탕으로 캐싱 기술이 NOMA의 성능을 향상하기도 하고, NOMA를 활용하여 캐싱 네트워크에서의 콘텐츠 전송 시간을 단축하기도 하여 서로 시너지를 낼 수 있음이 연구되었다<sup>24)</sup>.

이처럼 디바이스 캐싱된 파일을 이용하여 NOMA의 SIC를 완벽하게 할 수 있는 경우 NOMA의 처리량을 최대화할 수 있는 전력 할당 기법을 [25]에서는 mixed-integer programming과 심층 강화학습 알고리즘 기반으로 제안하였다. 심층 강화학습을 이용하여 인공 신경망 모델을 학습시켜 놓으면 최적에 가까운 전력 할당 기법을 빠르게 내려주어 매번 시간 소모가 큰 mixed-integer programming 기반 방식보다 장점을 보였다. 논문 [26]에서는 더 나아가 캐싱이 지원하는 NOMA 에지 컴퓨팅 시스템을 고려하였는데, 사용자들이 주변 AP에 태스크 오프로딩을 하고, 에지 컴퓨팅 서버에 접근 가능한 AP는 태스크 컴퓨팅을 한 후 컴퓨팅 결과의 캐싱 여부를 결정한다. 사용자들의 AP로의 태스크 오프로딩 과정에서 NOMA가 사용되는데, 분산 강화학습을 사용하여 각 사용자가 에이전트로서 태스크 오프로딩 결정을 각기 큐러닝 기반으로 내리도록 하여 상태 공간과 액션 공간을 크게 줄였다. 저자들은 제안한 다중 에이전트 큐러닝 알고리즘을



활용하여 에너지 소비를 최소화하는 태스크 오프로딩, 컴퓨팅 및 캐싱 자원 할당 정책을 고안하였다.

## VI. 심층 강화학습 기반 에지 네트워크 기술의 방향성

지금까지 살펴본 바로는, 에지 캐싱 및 에지 컴퓨팅 기술에의 심층 강화학습의 효용성을 보이는 연구가 주를 이루고, 현재는 이를 좀 더 넓은 시나리오로 확장하고 있는 단계라고 판단된다. 기본적으로 강화학습 에이전트는 액션을 내려야 하는 모든 범위의 네트워크 정보를 수집해야 하는 중앙집중 형태를 띠는데, 점점 분산화되는 에지 네트워크에서는 좀 더 분산된 형태로 강화학습을 수행하는 연구가 필요하다. 앞 장에서 언급한 연합 학습 프레임워크 내에서 강화학습을 진행하거나<sup>[22]</sup>, 다중 에이전트 강화학습 내에서 GAN을 활용하여 에이전트 간 의존성을 보완해주는 연구<sup>[23]</sup>가 그 예시가 될 것이다.

또한, 단순 에지 캐싱 및 에지 컴퓨팅 시나리오를 넘어 통합된 분산 시스템에의 강화학습 적용이 필요할 것이다. 예를 들어, 시간에 따라 변화하는 콘텐츠 인기도, 사용자 선호도 등을 학습하는 분산 학습 기술이나, 캐싱, 컴퓨팅, 통신 간의 오버헤드 및 비용 트레이드-오프<sup>[27]</sup>를 조절하는 알고리즘 개발이 요구된다.

마지막으로, 실제 데이터와 에지 네트워크 환경에 대한 심층 강화학습 알고리즘의 적용 및 검증이 필요하다. 일부 강화학습 알고리즘은 수렴하기 위한 조건이 까다로운데, 제안하는 강화학습 알고리즘이 다른 환경이나 시간에 따라 변화하는 네트워크에서도 보편성과 확장성을 보여줄 수 있는지 실제 데이터를 바탕으로 구현을 동반한 엄밀한 검증이 필요할 것이다.

## VII. 결 론

본 논문에서는 차세대 통신 네트워크에서 높은 데이터 트래픽을 유발하고 많은 사용자 연결성을 요구하는 멀티미디어 서비스를 효율적으로 지원할 수 있는 심층 강화학습 기반의 에지 캐싱 및 컴퓨팅 기술에 대해 살펴보았다. 특히 다양한 에지 네트워크 시나리오에 대해 어떤 심층 강화학습 알고리즘이 어떻게 적용되는지 살펴보았다. 또한, 최신 기계 학습 및 통신 기술들과 심층 강화학습을 공동으로 활용하여 에지 네트워크에서 시너지 효과를 내는 방법에 대해서도 살펴보아 추후 지능형 에지 네트워크가 나아가야 할 방향에 대해 알아볼 수 있었다.

## References

- [1] *Visual Networking Index: Global Mobile Data Traffic Forecast Update*, 2018-2023, Cisco, San Jose, CA, USA, Mar. 2020.
- [2] X. Cheng, J. Liu, and C. Dale, "Understanding the characteristics of Internet short video sharing: A YouTube-based measurement study," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1184-1194, Aug. 2013.
- [3] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402-8413, Dec. 2013.
- [4] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131-139, Feb. 2014.
- [5] M. Ji, G. Caire, and A. F. Molisch, "Wireless Device-to-Device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176-189, Jan. 2016.
- [6] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *Proc. IEEE ICC*, pp. 3358-3363, London, U.K., Jun. 2015.
- [7] M. Choi, J. Kim, and J. Moon, "Wireless video caching and dynamic streaming under differentiated quality requirements," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1245-1257, Jun. 2018.
- [8] M. Choi, A. F. Molisch, D. -J. Han, D. Kim, J. Kim, and J. Moon, "Probabilistic caching and dynamic delivery policies for categorized contents and consecutive user demands," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2685-2699, Apr. 2021.
- [9] X. Xiong, K. Zheng, L. Lei, and L. Hou, "Resource allocation based on deep reinforcement learning in iot edge computing," *IEEE J. Sel. Areas in Commun.*, vol. 38, no. 6, pp. 1133-1146, Jun. 2020.

- [10] N. Nomikos, S. Zoupanos, T. Charalambous, I. Krikidis, and A. Petropulu, "A survey on reinforcement learning-aided caching in mobile edge networks," *arXiv preprint arXiv:2105.05564*, 2021.
- [11] H. Zhu, Y. Cao, W. Wang, T. Jiang, and S. Jin, "Deep reinforcement learning for mobile edge caching: Review, new features, and open issues," *IEEE Netw.*, vol. 32, no. 6, pp. 50-57, Nov.-Dec. 2018.
- [12] B. Cao, L. Zhang, Y. Li, D. Feng, and W. Cao, "Intelligent offloading in multi-access edge computing: A state-of-the-art review and framework," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 56-62, Mar. 2019.
- [13] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: a comprehensive survey," *IEEE Commun. Surv. & Tuts.*, vol. 22, no. 2, pp. 869-904, Second quarter 2020.
- [14] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and scalable caching for 5g using reinforcement learning of space-time popularities," *IEEE J. Sel. Topics in Sign. Process.*, vol. 12, no. 1, pp. 180-190, Feb. 2018.
- [15] Y. Qian, R. Wang, J. Wu, B. Tan, and H. Ren, "Reinforcement learning-based optimal computing and caching in mobile edge network," *IEEE J. Sel. Areas in Commun.*, vol. 38, no. 10, pp. 2343-2355, Oct. 2020.
- [16] Q. Luo, C. Li, T. H. Luan, and W. Shi, "Collaborative data scheduling for vehicular edge computing via deep reinforcement learning," *IEEE Internet of Things J.*, vol. 7, no. 10, pp. 9637-9650, Oct. 2020.
- [17] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and N. Ansari, "Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks," *IEEE Internet of Things J.*, vol. 7, no. 1, pp. 247-257, Jan. 2020.
- [18] M. Choi, A. No, M. Ji, and J. Kim, "Markov decision policies for dynamic video delivery in wireless caching networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5705-5718, Dec. 2019.
- [19] M. Choi, M. Shin, and J. Kim, "Dynamic video delivery using deep reinforcement learning for device-to-device underlaid internet-of-vehicle caching networks," *IEEE/KICS J. Commun. and Netw.*, vol. 23, no. 2, pp. 117-128, Apr. 2021.
- [20] J. Luo, F. R. Yu, Q. Chen, and L. Tang, "Adaptive video streaming with edge caching and video transcoding over software-defined mobile networks: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1577-1592, Mar. 2020.
- [21] F. Wang, F. Wang, J. Liu, R. Shea, and L. Sun, "Intelligent video caching at network edge: A multi-agent deep reinforcement learning approach," *IEEE INFOCOM 2020 - IEEE Conf. Comput. Commun.*, pp. 2499-2508, 2020.
- [22] S. Yu, X. Chen, Z. Zhou, X. Gong, and D. Wu, "When deep reinforcement learning meets federated learning: Intelligent multitimescale resource management for multiaccess edge computing in 5G ultradense network," *IEEE Internet of Things J.*, vol. 8, no. 4, pp. 2238-2251, Feb. 2021.
- [23] H. Weng, L. Li, Q. Cheng, W. Chen, and Z. Han, "Content caching policy based on GAN and distributional reinforcement learning," *ICC 2020*, pp. 1-5, 2020.
- [24] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, "NOMA assisted wireless caching: strategies and performance analysis," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4854-4876, Oct. 2018.
- [25] K. N. Doan, M. Vaezi, W. Shin, H. V. Poor, H. Shin, and T. Q. S. Quek, "Power allocation in cache-aided NOMA systems: Optimization and deep reinforcement learning approaches," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 630-644, Jan. 2020.
- [26] Z. Yang, Y. Liu, Y. Chen, and N. Al-Dhahir,

“Cache-Aided NOMA mobile edge computing: A reinforcement learning approach,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6899-6915, Oct. 2020.

- [27] E. K. Markakis, K. Karras, A. Sideris, G. Alexiou, and E. Pallis, “Computing, caching, and communication at the edge: The cornerstone for building a versatile 5G ecosystem,” *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 152-157, Nov. 2017.

최 민 석 (Minseok Choi)



2011년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학과 졸업

2013년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학과 석사

2018년 8월 : 한국과학기술원 (KAIST) 전기 및 전자공학과 박사

2018년 9월~2019년 2월 : 중앙대학교 박사후연구원

2019년 3월~2020년 2월 : University of Southern California (USC), CA, USA, 방문 연구원

2020년 3월~현재 : 제주대학교 통신공학과 조교수

<관심분야> 차세대 이동통신 시스템, 기계학습 기반 네트워크 최적화

[ORCID:0000-0001-7027-1920]