

심층강화학습 기반 서비스 그룹별 큐 관리 메커니즘

정설령* · 이성근**

A Queue Management Mechanism for Service groups based on Deep Reinforcement Learning

Seol-Ryung Jung* · Sung-Keun Lee**

요 약

인터넷을 기반으로 다양한 종류의 응용 서비스들을 제공하기 위해서 각 흐름 별로 서비스 품질을 보장하는 것은 이상적이지만, 이를 실현하는 것은 매우 어려운 일이다. 서비스 품질 요구조건이 같거나 비슷한 여러 흐름들을 동일한 그룹으로 지정하고, 그룹별로 서비스 품질을 제공하는 방안이 효율적이다. 라우터에서 적용되는 큐 관리 메커니즘은 데이터의 효율적으로 전송하고, 서비스 별로 차별화된 서비스 품질을 지원하기 위하여 매우 중요한 역할을 수행한다. 다양한 멀티미디어 서비스를 효율적으로 지원하기 위해서 지능적이고 적응적인 큐 관리 메커니즘 기능이 필요하다. 본 논문은 일정 기간 유입되는 각 흐름 그룹의 트래픽 정보와 현재의 네트워크 상태 정보를 기반으로 그룹별 패킷의 전달 여부를 결정하는 심층강화학습 기반의 지능형 큐관리 메커니즘을 제안한다.

ABSTRACT

In order to provide various types of application services based on the Internet, it is ideal to guarantee the quality of service(QoS) for each flow. However, realizing these ideas is not an easy task.. It is effective to classify multiple flows having the same or similar service quality requirements into same group, and to provide service quality for each group. The queue management mechanism in the router plays a very important role in order to efficiently transmit data and to support differentiated quality of service for each service. In order to efficiently support various multimedia services, an intelligent and adaptive queue management mechanism is required. This paper proposes an intelligent queue management mechanism based on deep reinforcement learning that decides whether to deliver packets for each group based on the traffic information of each flow group flowing in for a certain period of time and the current network state information.

키워드

Deep Reinforcement Learning, Adaptive Queue Management Mechanism, Flow Groups, Quality of Service
심층강화학습, 적응적 큐 관리 메커니즘, 흐름 그룹, 서비스 품질

* 순천대학교 멀티미디어공학과(tjffud97@naver.com) · Received : Oct. 26, 2020, Revised : Nov. 21, 2020, Accepted : Dec. 15, 2020
** 교신저자 : 순천대학교 멀티미디어공학과 · Corresponding Author : Sung-Keun Lee
· 접수일 : 2020. 10. 26 · Dept. Multimedia Eng., Suncheon National University,
· 수정완료일 : 2020. 11. 21 · Email : sklee@scnu.ac.kr
· 게재확정일 : 2020. 12. 15

I. 서 론

최근 5G 서비스의 제공과 함께 다양한 스마트 기기의 보급으로 전례없는 수준의 모바일 멀티미디어 데이터 트래픽이 급증하고 있다. 시스코의 예측에 따르면 2017년부터 2022년까지 7배 이상의 모바일 데이터 트래픽 성장이 예상됨에 따라 인터넷 통신 프로토콜에 큰 영향을 미치게 되었다. 5G 네트워크는 기존망에 비해 최고 20배 정도 빠른 초고속 전송 속도를 제공할 뿐 아니라 저지연 및 대용량 특성을 지원한다 [1]. 4G LTE 서비스는 최대 1Gbps의 빠른 통신 속도를 제공하므로 초고속 인터넷이나 고화질 동영상 서비스를 제공하는 것이 가능하였다. 그러나, 자율주행차나 스마트 공장에 사용되는 지능형 로봇과 같이 반응속도가 빨라야 하는 서비스를 지원하기 위해서는 매우 낮은(latency)을 보장하여야 한다[2]. 또한, 스마트 시티와 같은 사물인터넷 서비스를 제공하기 위해서 하나의 이동통신 기지국에서 수만 개에서 수십만 개 이상의 디바이스를 수용할 수 있어야 한다. 또한 5G 네트워크의 기본 설계 사상은 하나의 단일한 통신망을 통해 다양한 특성을 갖는 여러 종류의 서비스를 서비스 품질(Quality of Service, QoS)을 보장하면서 동시에 제공하는 것이다. 이를 위해 ITU-T에서는 eMBB, mMTC, URLLC 등의 서비스를 정의하였다. eMBB는 모바일 환경에서 고속 데이터 통신을 제공하며, mMTC는 다수의 기기 접속을 허용하는 대용량 접속을 허용하는 특성을 갖는다. URLLC는 초저지연 및 고신뢰 통신을 제공한다. 따라서 유선 및 무선 통신 프로토콜은 다른 서비스 품질을 요구하는 응용 서비스에 대해서 공정하게 서비스 품질을 보장하는 기능을 반드시 제공하여야 한다. 하나의 통합된 네트워크를 통해 다양한 특성을 갖는 모든 서비스들을 제공하여야 한다. 다양한 서비스들을 수용하면서 개별 서비스들의 서비스 품질을 보장하기 위해서는 서비스별로 서로 논리적으로 구분되는 네트워크를 이용해야만 한다. 네트워크 슬라이싱(Network Slicing)은 이러한 목적을 만족시키기 위해서 제안된 개념이다.

인터넷을 기반으로 다양한 종류의 응용 서비스들을 제공하기 위해서 각 흐름(flow) 별로 서비스 품질을 보장하는 것은 이상적이지만, 이를 실현하는 것은 매우 어려운 일이다. 따라서 서비스 품질 요구조건이 같

거나 비슷한 여러 흐름들을 동일한 그룹으로 지정하고, 각 그룹 별로 서비스 품질을 제공하는 방안이 효율적이면서 상대적으로 쉽게 구현이 가능하다. 이러한 개념은 차별서비스 모델을 기반으로 네트워크 슬라이싱 기능으로 확장되어 왔다. 인터넷에서 서비스 품질을 보장하기 위해서는 IP(Internet Protocol)의 큐 관리 메커니즘과 TCP(Transmission Control Protocol)의 혼잡 제어 알고리즘이 중요한 역할을 수행한다. 본 논문에서는 IP의 패킷 전달 정책의 적응적인 제어에 대해 논한다. IP 프로토콜은 각 패킷 헤더에 표시된 서비스 품질 정보를 기반으로 패킷을 목적지 노드 전달하는 기능을 수행한다. 서비스 품질은 동일한 트래픽 특성을 갖는 흐름 그룹으로 분류하여 각 그룹별 차별화된 QoS 레벨은 보장한다.

일정 기간 유입되는 각 흐름 그룹의 트래픽 정보와 현재의 네트워크 상태에 기반하여 그룹별 패킷의 폐기 확률을 계산하고, 그에 따라 스케줄링을 수행하는 적응적 큐 관리 알고리즘을 적용한다. 심층강화학습(Deep Reinforcement Learning : DRL)은 최종적으로 가장 큰 보상을 획득하기 위하여 환경과 상호 작용을 통해 얻게 되는 경험으로부터 스스로 학습할 수 있다. 이러한 DRL을 적용한 네트워킹 분야 최적화에 대한 연구가 시도되었고, 많은 분야에서 우수한 성능 향상 결과를 나타내었다[3]. 본 논문은 일정 기간 유입되는 각 흐름 그룹의 트래픽 정보와 현재의 네트워크 상태 정보를 기반으로 그룹별 패킷의 전달 여부를 결정하는 DRL 기반의 지능형 큐 관리 메커니즘을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 분석하고, 3장에서는 적응형 큐 관리 메커니즘 설계에 대해 상세히 다루고, 4장에서는 본 논문의 결론을 맺는다.

II. 관련 연구

2.1 큐 관리 메커니즘

라우터에서 가장 기본적인 패킷 포워딩 방법은 FIFO(First-In First-Out) 큐를 구성하여 패킷을 전달하는 것이다. FIFO 큐의 처리 방법은 트래픽의 버스트 특성을 완화시킬수 있도록 일정 크기의 버퍼를 두고, 패킷이 들어오는 순서대로 처리를 한다. 만일

패킷의 도착률이 출력 링크의 처리율을 초과하게 되면 버퍼의 점유율이 증가되고, 버퍼가 가득 차게 되면 그 이후에 수신되는 패킷은 버려지게 되는 droptail 현상이 나타난다. 이러한 FIFO 기반의 큐 관리 메커니즘은 단순하며 구현이 쉬운 반면에 몇 가지 문제점을 나타낸다. 첫째, 여러 흐름 중에서 전송률이 높은 흐름이 큐를 많이 점유하게 되므로 라우터의 대역폭은 전송률이 높은 흐름이 많이 차지하게 되어 서비스 제공의 불균형이 일어난다. 둘째, 큐의 길이를 상대적으로 크게 설정하면, 큐에서는 손실이 발생할 확률은 낮아지게 되지만 패킷의 지연시간은 상대적으로 증가하게 된다. 셋째, 혼잡 상황이 발생하여 큐가 가득 찼을 때 이후에 유입되는 모든 패킷들은 연속해서 버려지게 된다. TCP 혼잡제어 알고리즘에 의해 각 흐름들은 일시적으로 패킷 전송률은 크게 낮추게 되고, 혼잡 상황은 완화가 되지만 링크의 이용 효율을 낮아지게 된다. 혼잡 상황이 해소되었다고 판단한 호스트들은 동시에 전송률을 증가시키므로 일정한 기간 후에 또다시 혼잡상태가 발생하여 패킷 처리율이 크게 낮아지는 상황이 주기적으로 반복되는 전역 동기화현상이 발생한다.

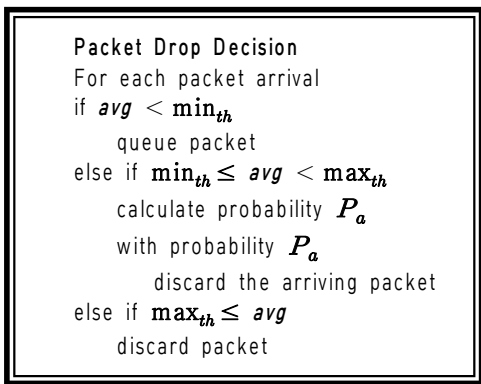


그림 1. RED 큐 관리 알고리즘
 Fig. 1 RED queue management algorithm

이러한 문제를 개선하기 위하여 다양한 큐 관리 메커니즘이 제안되었다. 특히, RED(Random Early Detection), RIO(RED with IN/OUT) 등과 같이 버퍼에 대기 중인 패킷의 개수와 데이터 도착률에 따라 확률적으로 패킷을 폐기하는 다양한 적응적 큐 관리 메커니즘이 제안되었다[4]. RED의 기본동작을 그림 1에 나타내었다. RED 방식은 큐의 길이가 어느 수준을 넘어서면 혼잡이 발생할 가능성이 있다고 판단하여 측정된 평균 큐 길이를 기반으로 폐기 확률을 계산하여 들어오는 패킷을 공정하고, 안정적으로 폐기한다. RED는 혼잡 상황이 발생할 때 대응하는 방법이 아니라 혼잡이 발생하지 않도록 예방함으로써 TCP 전역 동기화 문제를 완화시킬 수 있다. RIO는 RED를 개선한 방법으로서 두 개의 패킷 그룹(IN, OUT)에 대해 각각 RED 알고리즘을 적용하며, 각각에 대해 서로 다른 파라미터 값을 설정한다. 두 개의 패킷 그룹에 대해 다른 패킷 폐기 확률을 적용함으로써 차별화된 서비스 품질을 지원할 수 있다[5][6]. 최근에 CoDel 등 다양한 능동적 큐 관리 메커니즘이 제안되었다[7].

2.2 심층강화학습

강화학습은 그림 2에 나타난 바와 같이 에이전트와 환경이라는 두 개의 개체로 구성되며, 이들 간의 상호 작용은 지속적으로 환경에 영향을 미치고, 에이전트는 환경과의 상호 작용을 통해 얻게 되는 보상값을 통해 학습한다[8]. 강화학습은 순차적 행동 결정 문제를 풀기 위해 최적의 정책을 구하는 과정이다.

각 에피소드 동안에, 에이전트는 상태 정보 s_t 를 관찰하고, 해당 상태에서 정의된 정책 $\pi(s_t)$ 에 따라 행동 a_t 를 결정한다.

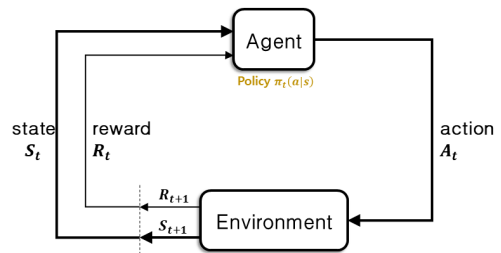


그림 2. 심층강화학습 구성요소
 Fig. 2 Components of deep reinforcement learning

에이전트는 행동을 수행하고, 환경으로부터 스칼라 형태의 보상값 $r(s_t, a_t)$ 을 받고, 환경의 변화된 다음 상태를 관찰한다. 에이전트 전체 에피소드 동안 환경으로부터 얻어지는 누적 보상값을 최대화하는 정책을 찾도록 학습한다. 정책 $\pi(a_t | s_t)$ 또는 $\pi(s_t)$ 는 상태공간을 행동공간 ($S \rightarrow A$)에 확률적 분포나 또는 결정적인 값의 형태로 대응하며, 특정 상태에서 취할 행동을 결정한다. 예를 들어, 우리는 출력 행동을 이산 행동 공간 $P(a)$ 에 따른 확률 분포로 모델링 할 수 있다. 또한 수행할 행동은 확률 분포에 따라 확률적으로 샘플링하거나 가장 큰 확률값을 가진 것을 선택할 수 있다. [9]에서 처음으로 소개된 DRL 알고리즘들은 비슷한 구조를 기반으로 하지만 가치함수 또는 정책함수를 학습하기 위해 심층신경망을 사용한다. 정제되지 않고 관찰된 데이터(raw sensory data) 그대로 상태 s_t 로 인식하여 심층신경망에 입력된다. 이러한 심층신경망은 심층 Q-네트워크(Deep Q-Network : DQN) 형태이며, Q-함수 $Q(s_t, a_t)$ 를 근사화 하는데 활용되고, 이는 주어진 모든 상태 s_t 에 대해 가능한 모든 행동에 대한 큐함수 값을 계산한다. 주어진 정책 π 에 대한 Q 가치함수 $Q\pi(s_t, a_t)$ 는 식 (1) 과 같이 정의된다[10].

$$Q\pi(st, at) = E[Rt|st, at] \tag{1}$$

ω 는 DQN의 가중치를 나타내고, $Q(st, at, \omega) \approx Q\pi(st, at)$ 라고 가정하면, 손실함수를 식 (2) 와 같이 정의할 수 있다.

$$J(\omega) = E[(r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))^2] \tag{2}$$

III. 적응형 큐 관리 메커니즘 설계

3.1 시스템 구성

본 논문에서 설계한 DRL 기반 큐 관리 메커니즘의 구성도를 그림 3 에 나타내었다. 제안한 메커니즘은 네트워크 에지에서 동일한 트래픽 특성을 갖는 흐름들을 동일한 그룹으로 분류하고, 라우터에서 각 그룹 별로 차별화된 QoS 레벨은 보장한다. 동일하거나 유사한 서비스 품질을 요구하는 모바일 어플리케이션

을 동일한 흐름 그룹으로 배정하고, 패킷의 헤더 부분에 흐름 그룹 식별자 정보를 마킹한다[11]. 각 라우터는 일정 기간 유입되는 각 흐름 그룹의 트래픽 정보와 현재의 네트워크 상태에 기반하여 그룹별 패킷의 폐기 처리 행동을 예측하고, 그에 따라 스케줄링을 수행하는 적응적 큐 관리 알고리즘을 적용한다.

그림 3에 나타난 바와 같이 DQN을 기반으로 QoS 제어 학습을 모델링하였으며, 시스템 모델은 POMDP (:Partially observable Markov decision process)을 이용하였다. 본 논문에서 정의한 각 컴포넌트의 기능은 다음과 같다.

- 환경(Environment) : 각 라우터가 상호 연결된 인터넷을 의미하며, 각 라우터는 큐 관리 메커니즘으로 추상화되며, N 개의 흐름 그룹별 차별화된 패킷 폐기 확률에 따라 패킷을 처리한다. 에이전트에게 네트워크 특성을 나타내는 상태 정보를 제공하고, 에이전트 행동에 따라 각 흐름 그룹별 패킷 전달 정책을 수행한다.

- 에이전트(Agent) : 현재의 상태 정보와 보상값을 기반으로 강화학습을 통해 얻어진 행동을 결정하여 큐 관리 메커니즘의 패킷 관리 정책에 반영한다. 학습 알고리즘은 DQN 방법을 통해 수행되며, T_s 주기마다 보상값과 상태 정보를 기반으로 각 흐름 그룹별 패킷 폐기 확률에 대한 행동을 결정하여 이 값을 환경에 적용한다. 행동 결정을 위한 정책은 e-greedy 방법에 따르므로, 학습 초기에는 다양하고, 충분한 탐험(exploration)을 수행하며, 학습이 충분히 이루어지면 이용(exploitation) 과정을 균형적으로 실현한다.

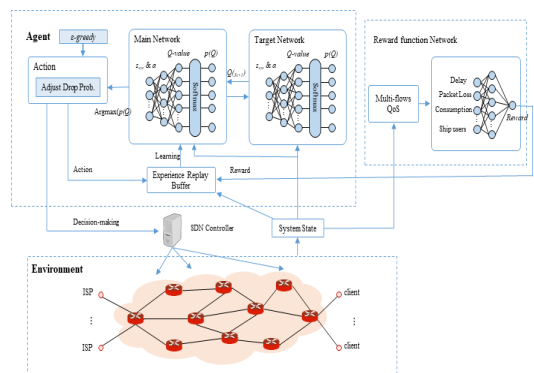


그림 3. 큐 관리를 위한 DRL 프레임워크
Fig. 3 DRL framework of queue management

- 상태(State) : 환경의 상태를 추상화한 정보로서, 각 흐름 그룹의 트래픽 특성, 각 흐름 그룹별 큐 상태 정보 등으로 구성된다. 트래픽 변화를 판단할 수 있도록 Ts 주기 동안 M 개의 히스토리 정보를 구성하고, 상태 정보는 상태 정보 수집기를 통해 QoS 제어 학습 알고리즘과 보상값 평가 알고리즘 입력으로 전달한다.
- 행동(Action) : 에이전트는 상태 공간 정보에 따라 행동 공간에 정의된 적절한 행동을 대응시키는 역할을 수행하며, 흐름 그룹별로 패킷 전달 정책을 나타낸다. 패킷 전송 또는 패킷 폐기 등 두 가지 행동으로 지정하였다. 향후 성능분석 결과에 따라 이산화된 확률 분포 형태로 조정될 수 있다.
- 보상 함수(Reward Function) : 에이전트에서 수행한 행동의 적절성의 평가로서, 각 흐름 별 QoS 메트릭스에 대한 만족도를 나타내며, QoS 메트릭스 만족도에 따라 (양수), (0), (음수) 값을 반환한다.

3.2 큐 관리 메커니즘 동작

그림 4에 DRL 기반 큐 관리 알고리즘을 나타내었으며, 세부 동작은 다음과 같다.

- 1단계 : 에이전트는 일정주기(Ts) 마다 환경으로부터 상태 정보를 수집한다.
- 2단계 : 에이전트는 DQN 메인 신경망을 사용하여 현재 상태에서의 각 행동에 대한 가치 확률값을 구한 후, e-greedy 방법에 의해 행동을 결정한다.
- 3단계 : 에이전트는 선택된 행동에 따라 큐 관리 파라미터를 업데이트 하도록 환경에 요청한 후, 행동 재생 메모리에 <상태, 행동, 타임스탬프> 정보를 저장한다.
- 4단계 : 다음 주기에 에이전트는 환경으로부터 다음상태 정보를 수집하고, 보상값 평가 네트워크를 통해 보상값을 받고, 보상값 재생 메모리에 <타임스탬프, 다음상태, 보상값> 정보를 저장한다. 타임스탬프 정보는 행동과 보상값 정보를 연결하는 링크 역할을 수행한다.
- 5단계 : 재생 메모리에 저장된 샘플 데이터가 일정 크기 이상 축적이 되면, 행동 및 보상값 재생 메모리로부터 배치 사이즈 크기 만큼 무작위로 추출하여, 학습을 수행한다. 샘플 데이터의 크기가 지정된 파라미터 값보다 적으면 학습을 수행하지 않고 1단계

과정을 반복한다.

- 6단계 : 티켓 모델 업데이트 주기가 되면 메인 신경망의 파라미터 값을 티켓 신경망의 파라미터 값으로 복사한다. 그렇지 않을 경우에는 티켓 신경망 파라미터는 변화시키지 않고, 메인 신경망 파라미터 값을 업데이트 한다.

```

DQN Based Queue Management Algorithm
Initialize an evaluation NN and a target NN
Initialize a replay memory buffer.
for episode ep = 0 to N do
    Receive the incoming packets.
    Obtain the observation oi(si) of system state si.
    Choose an action by e-greedy policy
    Execute the selected action, and observe the reward Ri(si; ai).
    Store the transition information in the buffer.
    Samples random mini batch of transitions from the memory buffer as the input of NN.
    Get the label value from the target NN
    Calculate the loss function with MSE
    Update the weights of evaluation NN by minimizing the loss function
    if update cycle then
        Update the target NN by resetting weights.
    end if
end for
    
```

그림 4. 큐 관리를 위한 DQN 알고리즘
Fig. 4 DQN Algorithm for queue management

IV. 결 론

라우팅, 서비스 품질 제어, 혼잡 제어 등 다양한 분야에서 딥러닝과 강화학습을 통한 성능 향상 및 지능화 연구가 매우 활발히 진행되고 있다. 본 논문은 최근 다양한 분야에서 많은 장점을 나타내고 있는 심층 강화 학습 기술을 적용하여 새로운 인터넷 환경에서 적용적으로 동작할 수 있는 DQN 기반 큐 관리 메커니즘에 설계에 대해 논했다. 현재 각 컴포넌트에 대해 구현 중에 있으며, 제안한 알고리즘에 대해서 다양한 환경에서의 시뮬레이션을 통한 성능 검증을 통하여

제한한 알고리즘의 성능이 흐름의 대역 보장 및 공정성을 달성할 수 있는지에 대한 성능 분석을 수행할 계획이다.

감사의 글

“이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임. (No. 2019R1I1A3A0106291)“

References

[1] K. Zhu and E. Hossain, “Virtualization of 5G cellular networks as a hierarchical combinatorial auction,” *IEEE Transaction on Mobile Computing*, vol. 15, no. 10, Oct. 2016, pp. 2640 - 2654.

[2] S. Lee, “Design and Application of LoRa-based Network Protocol in IoT Networks,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 14, no. 6, pp. 1089 - 1095, 2020.

[3] Z. Xu et al. ,“Experience-Driven Networking: A Deep Reinforcement Learning Based Approach,” *Proc. IEEE INFOCOM*, Honolulu, HI, USA, Apr. 2018.

[4] S. Floyd and V. Jacobson, “Random early detection gateways for congestion avoidance,” *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, Aug. 1993, pp. 397-413.

[5] J. Kim, S. Lee, Koh, and J. Park, “Traffic Control Algorithm for Periodic Traffics in WSN,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 5, no. 1, pp.44-50, 2010.

[6] J. Kim, S. Lee, Koh, and C. Jung, “A Marking Algorithm for QoS Provisioning in WMSN,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 5, no. 2, pp.193-204, 2010.

[7] K. Nichols and V. Jacobson, “Controlling queue delay,” *ACM Queue*, vol. 10, no. 5, May 2012, pp.1-14.

[8] K. Arulkumaran, M. Deisenroth, M. Brundage and A. Bharath, “Deep reinforcement learning : a brief survey,” *IEEE Signal Process. Magazine*, vol. 34, no. 6, Nov. 2017, pp.26 - 38.

[9] X. Huang, T. Yuan, G. Qiao and Y. Ren, “Deep reinforcement learning for multimedia traffic control in software defined networking,” *IEEE Network*, vol. 32, no. 6, Dec. 2018, pp. 35 - 41.

[10] B. Guo, X. Zhang, Y. Wang, and H. Yang, “Deep Q-network based multimedia multi-service QoS optimization for mobile edge computing systems,” *IEEE Access*, vol. 7, Nov. 2019, pp. 160961 - 160972.

[11] K. Xiao, S. Mao, and J. K. Tugnait, “TCP-Drinc: Smart congestion control based on deep reinforcement learning,” *IEEE Access*, vol. 7, Jan. 2019, pp. 11892 - 11904.

저자 소개



정설령(Seol-Ryung Jung)

2020년 : 순천대학교 컴퓨터공학과 (공학사)

2020년 3월 ~ 현재 : 순천대학교 대학원 멀티미디어공학과(석사과정)

※ 관심분야 : 인공지능 응용, 심층강화학습, 스마트 농업



이성근(Sung-Keun Lee)

1985년 고려대학교 전자공학과 졸업(공학사)

1987년 고려대학교 대학원 전자공학과 졸업(공학석사)

1995년 고려대학교 대학원 전자공학과 졸업(공학박사)

1987년 ~ 1992년 : 삼성전자 정보통신연구소

1996년 ~ 1997년 : 삼성전자 네트워크 연구팀

2017년 ~ 2018년 : Georgia Institute of Technology ECE 방문교수

1997년 ~ 현재 순천대학교 멀티미디어공학과 교수

※ 관심분야 : 강화학습 기반 QoS 보장 기술, WSN, 스마트농업, 멀티미디어통신