

## 큐러닝(Q-learning)을 이용한 다중 대기열 패킷 스케줄링

정현석<sup>0\*</sup>, 이태호<sup>\*</sup>, 이병준<sup>\*</sup>, 김경태<sup>\*</sup>, 윤희용<sup>\*\*</sup>

<sup>0\*</sup>성균관대학교 정보통신대학 전자전기컴퓨터공학과

<sup>\*\*</sup>성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {daun8928, leetaeho, byungjun}@skku.edu<sup>0\*</sup>, kyungtaekim76@gmail.com<sup>\*</sup>, youn7147@skku.edu<sup>\*\*</sup>

## Multiple Queue Packet Scheduling using Q-learning

Hyun-Seok Jeong<sup>0</sup>, Tae-Ho Lee<sup>\*</sup>, Byung-Jun Lee<sup>\*</sup>, Kyoung-Tae Kim<sup>\*</sup>, Hee-Yong Youn<sup>\*\*</sup>

<sup>0\*</sup>Dept. of Electrical and Computer Engineering, Sungkyunkwan University

<sup>\*\*</sup>Dept. of Software, Sungkyunkwan University

### ● 요약 ●

본 논문에서는 IoT 환경의 무선 센서 네트워크 시스템 상의 효율적인 패킷 전달을 위해 큐러닝(Q-learning)에 기반한 다중 대기열 동적 스케줄링 기법을 제안한다. 이 정책은 다중 대기열(Multiple queue)의 각 큐가 요구하는 딜레이 조건에 맞춰 최대한 패킷 처리를 미룸으로써 효율적으로 CPU자원을 분배한다. 또한 각 노드들의 상태를 큐러닝(Q-learning)을 통해 지속적으로 상태를 파악하여 기아상태(Starvation)를 방지한다. 제안하는 기법은 무선 센서 네트워크 상의 가변적이고 예측 불가능한 환경에 대한 사전지식이 없이도 요구하는 서비스의 질(Quality of service)를 만족할 수 있도록 한다. 본 논문에서는 모의실험을 통해 기존의 학습 기반 패킷 스케줄링 알고리즘과 비교하여 제안하는 스케줄링 기법이 복잡한 요구조건에 따라 유연하고 공정한 서비스를 제공함에 있어 우수함을 증명하였다.

**키워드:** 강화학습(reinforcement learning), 패킷 스케줄링(packet scheduling), 큐러닝(q-learning), 무선 센서 네트워크(wireless sensor network), 소프트맥스 알고리즘(softmax algorithm)

## I. Introduction

IoT 환경의 무선 센서 네트워크 시스템에 있어서 트래픽 컨트롤 알고리즘의 궁극적인 목적은 각 센서 노드들의 트래픽이 요구하는 QoS를 만족하면서도 Starvation이 일어나지 않고 최대한의 효율로 패킷을 전달하는 것이다. FIFO나 EDF 등의 일반적으로 널리 쓰이는 스케줄링 알고리즘의 경우 빠르고 안정적이지만 스케줄링에 있어 많은 정보를 고려하지 않기 때문에 서비스의 질과 효율 측면에서는 한계를 보인다. 이는 다중 대기열(Multiple Queue)에서 서비스의 질을 만족하는 대기열이 가용 리소스를 양보하여 다른 대기열(Queue)에 양보함으로써 해결 될 수 있다. 그렇기 때문에 이러한 알고리즘은 각 대기열의 성능(Performance)에 따라 서로 다른 서비스율(Service Rate)을 가져야 한다. 이를 위해 큐러닝(Q-learning)을 이용하였다.

본 논문은 1장 소개를 다음으로, 2장에서는 관련연구, 3장에서는 제안하는 기법, 4장에서는 결론으로 구성하였다.

## II. Preliminaries

### 1. Related works

#### 1.1 확률 기반의 스케줄링 알고리즘

먼저 확률적인 방법을 통해 다음에 서비스할 큐를 선택하는 알고리즘이 제시되어 있다[1]. 이를 위해 먼저 딜레이 요구조건(Delay Requirement)을 이용하여 현재의 상태(State)에 대해 모델링한다. 이에 따라 다음에 선택될 대기열을 확률적으로 선택한다.

#### 1.2 Q-learning 기반의 스케줄링 알고리즘

다음으로 Q-learning을 통해 가장 높은 Q값을 가지는 정책을 통해 다음에 서비스할 대기열을 선택하는 알고리즘이 제시되어 있다 [2]. 이 알고리즘 역시 각 대기열의 Delay Requirement  $R_i$  과 각 대기열의 Mean Delay  $M_i$ 에 따라 현재의 State에 대해 모델링하고, 이에 기반해 점수를 매긴 뒤 가장 높은 점수를 가진 대기열을 다음 서비스할 대기열로 선택한다.

### III. The Proposed Scheme

제안하는 방식으로는 대기열 선택에 있어 소프트맥스 함수 (Softmax function)를 이용하여 Q-learning 기반의 스케줄링 알고리즘[2] 의 대기열 선택 방식을 확률적으로 만드는 것이다. 먼저 현재 State를 평가하여 표현한다. 각 대기열  $i$ 의 딜레이 요구조건은  $R_i$ 로 정의하고, 각 대기열  $i$ 의 평균 지연시간(Mean Delay)은  $M_i$ 로 정의한다. 이에 해당하는 상태에 대해 나타내는 값을  $s_i$ 로 정의한다. 이는 다음 식에 의해 정의되며, 이를 이용한 벡터 표현이 서로 다른 Q-table의 row를 나타낸다.

$$\begin{aligned} s_i &= 1 && \text{if } M_i > R_i \\ s_i &= 0 && \text{if } M_i < R_i \end{aligned}$$

다음으로 현재 Q-value에 따라 다음 서비스할 대기열을 선택한다. 대기열의 개수를  $n$ 개로 표현하면, 각 대기열이 얻은 점수인  $\gamma_i$ 를 이용해 지수화와 정규화하여 확률 함수  $P$ 로 나타내면 다음 식과 같다.

$$P(i) = \frac{\exp(\gamma_i)}{\sum_{j=0}^{i-1} \exp(\gamma_j) + \sum_{j=i+1}^n \exp(\gamma_j)}$$

이렇게 얻어진 확률  $P$ 를 통해 확률적으로 다음에 서비스할 대기열을 선택한다.

다음으로 서비스 이후 결과에 기반하여 보상 함수  $r_i(t)$ 를 계산한다.  $C_1$ 과  $C_2$ 를 각각 보상과 페널티 계수라고 할 때, 이는 다음 식과 같이 표현되며, 적절한 Learning Rate와 함께 Q-value를 업데이트한다.

$$r_i(t) = \begin{cases} \frac{C_1 M_i}{R_i} & \text{if } M_i < R_i \\ -C_2 & \text{if } M_i > R_i \end{cases}$$

### IV. Conclusions

본 논문에서는 IoT 무선 센서 네트워크 환경에서 유연하고 안정적인 서비스를 실시간으로 제공하기 위해 강화학습의 일종인 큐리닝을 이용해 현재의 환경을 학습하고 평가하는 스케줄러를 향상시켰다. 제안하는 방식은 모의실험을 통해 기존의 큐리닝을 이용한 알고리즘보다 훨씬 공정하고 안정적으로 스케줄링하는 경향이 있음을 증명하였다. 하지만 평균 지연율과 딜레이 요구시간만을 상태 평가에 고려하고 있기 때문에 안정성이 떨어진다. 이를 보완하기 위해서 더 많은 요소를 고려할 수 있도록 상태공간(State Space)를 더 늘리는 방식에 대한 추가 연구가 이루어질 예정이다.

### ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신-방송연구 개발 사업(No. 2016-0-00133, 초연결 IoT 노드의 군집 지능화를 통한 Edge Computing 핵심 기술 연구), SW중심대학지원사업(2015-0-00914), 한국연구재단 기초연구사업(No.2016R1A6A3A11931385, 실시간 공공안전 서비스를 위한 소프트웨어 정의 무선 센서 네트워크 핵심기술 연구, 2017R1A2B2009095, 실시간 스트림 데이터 처리 및 Multi-connectivity를 지원하는 SDN 기반 WSN 핵심 기술 연구), 삼성전자, BK21PLUS 사업의 일환으로 수행되었음.

### REFERENCES

- [1] J. Hall, and P. Mars, "Satisfying QoS with a Learning Based Scheduling Algorithm," IEEE Sixth International Workshop on Quality of Service, pp. 171-173, May 1998.
- [2] H. Ferra, K. Lau, C. Leckie, and A. Tang, "Applying Reinforcement Learning to Packet Scheduling in Routers," IAAI Fifteenth Conference on Innovative Applications of Artificial Intelligence, pp. 79-84, August 2003.