

# 산업 네트워크 환경에서의 DQN 기반 지능형 라우팅 정책 연구

장여진, 장하린, 이상금\*  
\*국립한밭대학교

jyeoj251@gmail.com, hrjang713@gmail.com, \*sangkeum@hanbat.ac.kr

## A Study on DQN-Based Intelligent Routing Policy in Industrial Networks

Yeojin Jang, Harin Jang, Sangkeum Lee\*  
\*Hanbat National University

### 요 약

본 논문은 실시간성과 안정성을 동시에 요구하는 산업 네트워크 환경에서 QoS(Quality of Service) 성능을 향상시키기 위해 DQN(Deep Q-Network) 기반 지능형 라우팅 정책을 제안한다. 산업 네트워크는 복잡한 트래픽 구조로 인해 부하 변동이나 병목 발생 시 QoS 지표(지연, 패킷 손실률, 큐 점유율)가 악화되는 문제가 존재한다. 제안된 DQN 에이전트는 네트워크 상태(큐 점유율, 링크 용량 등)를 입력받아 최적의 경로를 선택하도록 학습하며 지연, 혼잡, 패킷 손실을 반영한 보상 함수를 통해 안정성과 실시간성을 모두 고려한다. 연구 결과, DQN 기반 라우팅은 기존의 고정 경로 방식 대비 지연, 손실률, 큐 점유율 측면에서 우수한 성능을 보였다. 이를 통해 강화학습 기반 접근법이 스마트 팩토리 및 같은 중규모 산업 제어망에서의 트래픽 제어에 효과적이고 향후 실시간 제어망 최적화에 기여할 수 있음을 보여준다.

### I. 서론

산업 네트워크는 스마트 팩토리, 에너지 관리 시스템 등을 구성하는 핵심 인프라로서 데이터 전송의 실시간성과 안정성을 동시에 보장해야 한다. 그러나 실제 환경에서는 트래픽 부하 변동, 링크 장애, 버스트(burst) 트래픽과 같은 요인으로 인해 지연, 손실률, 큐 길이 등의 QoS(Quality of Service) 지표를 안정적으로 유지하기 어렵다. 기존의 최단 경로 기반 라우팅이나 대역폭을 고려한 최단 경로 기법은 고정된 정책에 의존하므로 동적으로 변화하는 네트워크 상황에 효과적으로 대응하기 어렵다는 한계를 가진다[1].

본 논문에서는 동적인 산업 네트워크 환경에서 주요 QoS 지표를 안정적으로 보장하기 위해 DQN 기반 지능형 라우팅 정책을 제안한다. 이를 위해 동적 네트워크 환경을 구축하고 트래픽 부하 증가, 링크 장애 등 다양한 상황에서 성능을 분석한다. 제안하는 정책을 기존의 최단 경로(SP, Shortest Path) 및 대역폭 반영 최단 경로(SP-cap, Shortest Path with Capacity-Aware) 기법과 비교하여 주요 QoS 지표를 기준으로 성능을 평가한다. 실험을 통해 DQN(Deep Q-Network) 정책은 동적인 환경에서 고정 라우팅 기법 대비 안정성과 실시간성을 동시에 확보할 수 있음을 보여준다.

### II. 본론

#### 1. 시스템 모델 및 문제 정의

실시간성과 안정성을 동시에 요구하는 산업 네트워크에서는 센서 데이터와 같은 실시간 트래픽(RT, Real-Time traffic)과 생산 로그 전송과 같은 비실시간 트래픽(NRT, Non-Real-Time traffic)이 함께 존재한다. 산업 데이터는 일반적으로 수십~수백 바이트 크기의 패킷이 수 밀리초 주기마다 발생하며, 이러한 트래픽 혼재 환경은 트래픽 부하 증가나 특정 구간의 병목 현상을 초래해 QoS 지표가 급격히 악화될 수 있다.

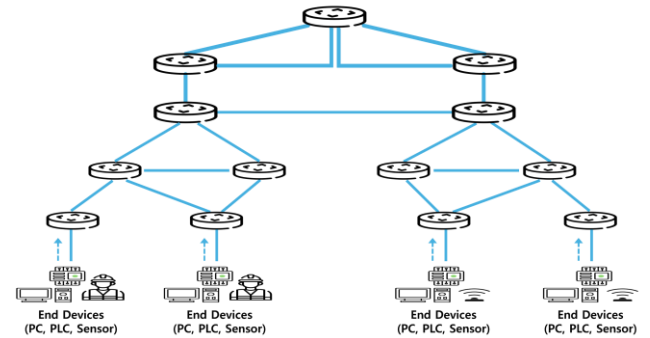


그림 1. 네트워크 토폴로지 구조

실제 산업 네트워크는 그림 1과 같이 다양한 장비가 계층적으로 연결된 복잡한 구조를 가진다. 이러한 구조는 특정 통신 경로에 예기치 않은 병목이 발생할 수 있다. 해당 환경을 재현하기 위해 시뮬레이션에서 특정 링크의 용량 축소로 지연을 증가시켜 실제 산업망의 이벤트를 모사한다.

기존의 최단 경로 및 대역폭 반영 최단 경로 라우팅 기법은 단순성과 효율성 측면에서 유리하지만, 환경 변화를 실시간으로 반영하지 못하므로 혼잡 상황에서 지연과 손실률이 증가한다. 따라서 DQN 기반 강화학습 에이전트를 적용하여 네트워크 상태에 따라 동적으로 경로를 선택하도록 하며, 이를 통해 실시간성과 안정성을 동시에 달성할 수 있는지 검증한다.

#### 2. DQN 기반 지능형 라우팅 에이전트 설계

DQN은 Q-learning의 원리를 신경망으로 확장하여 복잡한 네트워크 상태에서도 효율적으로 학습할 수 있는 강화학습 기법이다[2]. 본 연구에서는 각 노드에 탑재된 DQN 에이전트가 네트워크 상태를 기반으로 다음 홉을 동적으로 선택한다. Q-learning의 기본 업데이트 규칙은 다음과 같다.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$$

위 수식에서  $s_t$ 는 상태(state),  $a_t$ 는 행동(action),  $r_t$ 는 보상(reward),  $\alpha$ 는 학습률,  $\gamma$ 는 할인율을 의미한다. 이 규칙은 에이전트가 현재 상태에서 선택한 행동의 가치를 보상과 미래 가치 추정치에 따라 점진적으로 갱신하는 과정이다.

$$L(\theta) = \mathbb{E}[(r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) - Q(s_t, a_t; \theta))^2]$$

DQN은 Q-learning의 원리를 신경망으로 근사하기 위해 손실 함수를 정의한다. 손실 함수는 정책망이 예측한 현재 Q값과 보상 및 타깃 네트워크를 통해 계산된 목표 Q값의 차이를 최소화하는 방식으로 주어진다. 위 수식에서  $\theta$ 는 정책망의 파라미터,  $\theta^-$ 는 타깃 네트워크의 파라미터이다. 타깃 네트워크는 일정 주기마다 정책망의 가중치를 복제하여 이를 통해 학습 과정의 발산을 방지하고 안정성을 확보한다.

$$r_t = (-0.1) + (-0.5 \times (q(u, v))^2) + (20 \times 1[\text{deliver}]) + (-20 \times 1[\text{drop}])$$

또한 산업 네트워크의 특성을 반영하기 위해 지연, 혼잡, 성공적인 전달, 패킷 손실을 고려한 보상 함수를 설계한다. 위 수식에서  $q(u, v)$ 는 노드  $u$ 에서 인접 노드  $v$ 로 향하는 링크의 큐 점유율을 의미하며  $1[\cdot]$ 은 해당 조건이 발생했을 때 1을 반환하는 지시 함수이다. 이를 통해 에이전트는 단순히 지연을 줄이고 링크 혼잡을 회피하면서 패킷 전달 성공률을 높이는 방향으로 학습한다.

### 3. 실험 환경 및 성능 평가

파라미터	값
에피소드 수	150
에피소드 길이	2000 steps
최대 큐 크기	30
링크 기본 용량	15
병목 링크 용량	1
전송 지연	1 step (정상), 30 steps (병목)
트래픽 발생 분포	Poisson ( $\lambda$ )
트래픽 부하 범위	$\Lambda = \{10, 20, 30, 40, 50\}$
학습률 ( $\alpha$ )	$1e-4$
할인율 ( $\gamma$ )	0.98
배치 크기	128
타깃 네트워크 갱신 주기	매 20 steps

표 1. 시뮬레이션 파라미터 설정값

본 연구의 시뮬레이션은 30개 노드로 구성된 토폴로지에서 진행된다. 파라미터는 표 1의 값으로 설정하였으며 500 steps마다 무작위로 선택된 특정 링크의 통신 비용을 급증시켜 병목 현상을 유발한다. 이를 통해 산업망에서 발생할 수 있는 트래픽 폭주나 링크 장애 상황을 재현한다.

DQN 에이전트는 네트워크의 상태(큐 점유율, 링크 용량 등)를 입력받아 최적의 다음 홉을 결정하도록 학습된다. 학습은 총 150개 에피소드 동안 진행되며 학습 안정성을 위해 타깃 네트워크는 매 20 steps마다 갱신된다.

학습이 완료된 DQN 에이전트의 성능을 전통적인 라우팅 방식인 SP와 SP-cap을 비교군으로 설정하여 평가한다. 세 가지 정책을 모두 시뮬레이션한 뒤 평균 지연 시간, 패킷 손실률, 큐 길이라는 핵심 QoS 지표를 측정하여 정량적으로 비교 분석한다.

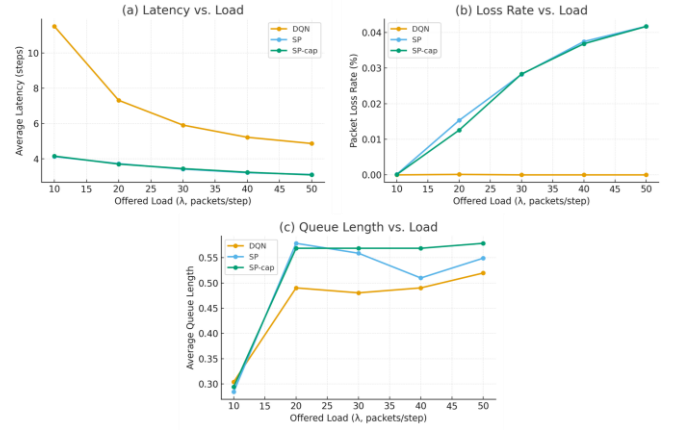


그림 2. 라우팅 정책별 QoS 성능 비교

그림 2는 부하( $\lambda$ )에 따른 SP, SP-cap, DQN 세 라우팅 방식의 QoS 지표 변화를 보여준다. 평균 지연 그래프에서 SP와 SP-cap은 혼잡을 인지하지 못해 패킷을 손실하면서 전송하므로 지연이 낮게 나타나지만, 실제 성능 향상을 의미하지 않는다. 반면에 DQN은 초기에 지연이 다소 높지만, 부하가 증가할수록 안정적으로 감소한다. SP, SP-cap과 달리 DQN은 트래픽 부하와 링크 상태 변화를 반영해서 동적으로 경로를 선택하므로 효율적인 트래픽 분산을 수행한 것이다.

패킷 손실률 그래프에서 DQN은 부하 증가 상황에서 0%에 가까운 손실률을 유지하지만, SP와 SP-cap은 급격히 증가한다. 이는 DQN이 병목 구간을 회피하는 정책을 학습했기 때문이다. 또한 DQN은 평균 큐 길이도 낮으므로 모든 측면에서 우수한 성능을 보인다.

### III. 결론

본 논문에서는 실시간성과 안정성을 동시에 요구하는 산업 네트워크 환경에서 QoS 성능을 향상시키기 위한 DQN 기반 지능형 라우팅 정책을 제안하였다. 실험을 통해 제안된 DQN 라우팅은 부하 증가나 병목 발생 등 동적인 상황에서 기존의 고정 경로 방식보다 지연, 손실률, 큐 점유율 측면에서 우수한 성능을 보였다. 이는 DQN 에이전트가 네트워크 상태를 실시간으로 인지하고 병목 구간을 회피하도록 학습한 결과이다. 본 연구의 결과는 강화학습 기반 라우팅이 산업 네트워크와 같은 실시간 제어 환경에서도 효과적으로 적용될 수 있음을 보여주며 향후에는 학습 확장을 통해 보다 현실적인 네트워크 제어로 발전시킬 수 있을 것이다.

### ACKNOWLEDGMENT

본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음. (2022-0-01068)

### 참고 문헌

- [1] Y. Song, X. Qian, N. Zhang, W. Wang, and A. Xiong, "QoS routing optimization based on deep reinforcement learning in SDN," *Computers, Materials & Continua*, vol. 79, no.2, pp. 3007–3021, 2024.
- [2] I. Kim, S. Lee, H. Jeong, S. H. Nengroo, and D. Har, "Frequency Hopping Synchronization by Reinforcement Learning for Satellite Communication System," *arXiv preprint arXiv:2503.04266*, 2025.