

학습 가치 기반 선택적 통신을 통한 통신 오버헤드 감소형 교통 신호 제어

김건민, 장현지, 최민준, 김경백

전남대학교 인공지능융합학과

geonminkim@jnu.ac.kr, gka1225@jnu.ac.kr, larrychoi@jnu.ac.kr, kyungbaekkim@jnu.ac.kr

Learning Value-Based Selective Communication for Traffic Signal Control with Reduced Communication Overhead

Geonmin Kim, Hyeonji Jang, Minjun Choi, Kyungbaek Kim

Dept. of Artificial Intelligence Convergence, Chonnam National University

요약

본 논문은 다중 에이전트 강화학습 기반 교통 신호 제어에서 발생하는 통신 오버헤드를 줄이기 위해 학습 가치 기반 선택적 통신 프로토콜인 Adaptive-SCP를 제안한다. 기존 GNN-MARL 기반 교통 신호 제어는 인접 교차로 간 상태 정보를 항상 교환하는 always-on communication 구조를 전제로 하며, 이는 교차로 수가 증가할수록 불필요한 브로드캐스트와 높은 통신 비용을 유발한다. 또한 기존 선택적 통신 기법은 주로 큐 길이, 혼잡 임계값과 같은 고정된 휴리스틱에 의존하여 교통 수요 변화에 대한 적응성이 제한적이다. 제안하는 Adaptive-SCP는 TD-error로부터 유도된 학습 가치 신호를 기반으로 통신 우선순위를 동적으로 결정하며, emergency override, Top-K selection, cooldown, hold 전략을 통해 학습적으로 중요한 노드에만 선택적으로 통신을 수행한다. 실험 결과, Adaptive-SCP는 always-on communication 대비 평균 통신 비율을 약 27~33% 수준으로 유지하여 전체 통신량을 크게 절감하였다. 이러한 통신 절감에도 불구하고 평균 총 대기시간은 대부분의 교통 수요 조건에서 유사한 수준을 유지하였으며, 특히 교통 혼잡 조건에서는 항상 통신보다 더 낮은 총 대기시간을 보였다. 이는 학습 가치 기반 선택적 통신이 통신 비용과 교통 제어 성능 간의 균형을 효과적으로 유지할 수 있음을 보여준다.

I. 서론

최근 교통 신호 제어는 다수의 교차로가 동시에 운영되는 분산 환경에서 실시간 의사결정을 요구한다. 특히 그래프 신경망(Graph-Neural Network, GNN) 기반 다중 에이전트 강화학습 Multi-Agent Reinforcement Learning, MARL)[1-5]과 같은 학습 기반 제어는 인접 노드의 상태 정보를 교환함으로써 더 나은 추정과 정책 결정을 기대할 수 있다. 그러나 대역폭, 지연, 전송 비용 등 실제 교통 통신 인프라는 제한적이기 때문에, 모든 노드가 모든 단계에서 항상 통신하는 항상 통신 방식은 운영 측면에서 비효율적일 수 있다[6]. 이러한 문제를 해결하기 위해 최근에는 선택적 통신[7-9]이 활발히 연구되어 왔다. 대표적으로 큐 길이, 혼잡 임계값, 혹은 고정된 스케줄링 규칙에 따라 통신 여부를 결정하는 접근이 사용된다. 하지만 교통 수요가 변화하거나 혼잡이 지속되는 구간에서는 큐 통계가 상시적으로 기준을 넘기며, 휴리스틱 기반 스케줄은 불필요한 브로드캐스트를 증가시키거나 반대로 중요한 정보 전파 타이밍을 놓칠 위험이 있다. 더 근본적으로, 통신은 단순 관측을 늘리는 비용이 아니라 에이전트의 가치 추정과 정책 업데이트에 실제로 도움이 되는 정보에 선택적으로 배분되어야 한다[10-14]. 하지만 기존 기준이 혼잡 신호 자체에 고정되어 있을 경우, 학습 과정에서 변화하는 정보 가치를 충분히 반영하기 어렵다[15]. 따라서 트래픽 조건 변화에도 통신 효율성을 유지하면서 제어 성능을 안정적으로 보호하는 통신 프로토콜이 필요하다. 이에 본 논문은 교통 신호 제어를 위한 학습가치 기반 통신 절감형 선택적 통신 프로토콜인 Adaptive-SCP(Selective - Communication Protocol)를 제안한다.

Adaptive-SCP는 노드별 큐 혼잡만이 아니라, 학습 과정에서의 학습가치 신호에 기반해 통신 우선순위를 산정한다. 이를 통해 top-k 스케줄링 [16-17]으로 통신 대상을 제한하면서도, 긴급 상황에서는 하드 오버라이드[18-19]를 적용해 중요한 이벤트의 정보 전파를 보장한다. 평가는 10x10 GNN-MARL 환경과 실제 도심 환경 시나리오에서 다중 수요 패턴을 조건으로 수행하였다. 결과적으로 Adaptive-SCP는 항상 통신 대비 교통 패턴별 통신 비율을 약 27-33%로 절감하면서도, 총 대기시간은 전반적으로 항상 통신과 유사한 수준을 유지한다. 또한, 실험 횟수를 누적할수록 통신 절감 특성이 수렴하며 안정적으로 재현됨을 확인하였다.

II. 관련 연구

1. 다중 에이전트 기반 교통 신호 제어

도시 규모의 교통 신호 제어 문제는 다수의 교차로가 상호 의존적으로 작용하는 복잡한 최적화 문제로, 이를 해결하기 위해 MARL 기반 접근법이 활발히 연구되어 왔다. 각 교차로를 개별 에이전트로 모델링함으로써, 지역적인 교통 상황에 기반한 분산형 의사 결정이 가능하며, 이는 중앙 집중식 제어방식 대비 확장성과 적응성을 동시에 확보할 수 있다는 장점을 지닌다. 특히 최근 연구에서는 GNN을 결합하여 교차로 간 공간적 상관관계를 효과적으로 반영하는 방식이 제안되고 있다[1-4]. GNN 기반 MARL 구조는 인접 교차로로부터 전달되는 상태 정보를 활용하여 보다 정교한 정책 학습을 가능하게 하며, 이는 도시 전반의 교통 흐름을 개선한다. 그러나 기존 연구들은 대부분 에이전트 간의 정보 교환을 항상 수행하는 것

을 전제로 설계되어 있으며, 통신 비용이나 네트워크 자원의 제약을 명시적으로 고려하지 않는 한계를 가진다. 특히, 교차로 수가 증가함에 따라 메시지 교환량이 기하급수적으로 증가하여 통신 오버헤드와 지연이 발생할 수 있으며, 이는 실제 환경 적용 가능성의 저해 요인으로 작용한다.

2. 선택적 통신 및 통신 절감 기법

MARL 환경에서 통신을 효율적으로 제어하기 위한 다양한 선택적 통신 기법들이 제안되어 왔다[10-14]. 대표적으로, 에이전트 간 정보 교환 여부를 gating 메커니즘이나 임계값 기반 규칙을 통해 결정하는 방식이 있으며, 일부 연구에서는 중요도가 높은 상황에서만 메시지를 전송하도록 하는 selective broadcast 전략이 활용된다[12]. 또한, 고정된 임계값에 기반하여 통신 여부를 결정하는 방식이나 큐 길이, 교통 밀도 등의 단일 지표를 기준으로 통신을 제어하는 방식도 제안되었다. 이러한 접근법은 통신량을 일정 수준 감소시키는 데는 효과적이거나, 환경 변화에 대한 적응성이 부족하다는 한계를 지닌다. 특히, 교통 흐름이 급격히 변화하는 혼잡 상황에서는 단순한 임계값 기반 방식이 과도한 통신 또는 불충분한 통신을 동시에 유발할 수 있다. 결과적으로, 기존의 선택적 통신 기법들은 통신 자원의 효율성과 정책 학습 성능 간의 균형을 충분히 고려하지 못하며, 학습 과정에서 정보 가치를 반영하지 못한다는 한계를 지닌다.

3. 자율 네트워크 및 통신 자원 인식 제어

최근 인공지능 기반 자율 네트워크(AI-driven Networking) 및 AIOps 분야에서는 시스템 성능 뿐만 아니라 운영 자원의 효율적 활용을 동시에 고려하는 방향으로 연구가 진행되고 있다[20-21]. 특히, 네트워크 환경에서는 제한된 대역폭과 지연된 요구사항을 만족시키기 위해 통신 비용을 최소화하면서도 성능을 유지하는 것이 중요한 과제로 인식되고 있다. 이러한 관점에서, 단순한 성능 최적화가 아닌 자원 인식(Resource-Aware) 기반의 의사 결정이 강조되고 있으며[22-23], 이는 교통 신호 제어와 같은 대규모 분산 시스템에서도 동일하게 적용될 수 있다. 즉, 통신 자체를 하나의 비용 요소를 고려하고, 이를 최소화하려는 접근이 필요하다.

4. 본 연구의 차별성

기존의 MARL 기반 교통 신호 제어 연구는 통신을 항상 수행하는 구조에 의존하거나, 단순 규칙 기반 선택적 통신 방식에 머무르고 있다. 이러

한 접근은 실제 도시 환경과 같이 동적으로 변화하는 교통 상황에서 통신 효율성과 학습 성능을 동시에 만족시키기 어렵다. 이에 본 논문에서는 학습 과정에서의 정보 가치에 기반하여 통신 여부를 동적으로 결정하는 Adaptive-SCP를 제안한다. 제안 방식은 TD-error와 같은 학습 신호를 활용하여 각 시점에서의 통신 필요성을 정량적으로 평가하고, 이를 기반으로 통신 마스크를 생성함으로써 불필요한 메시지 교환을 효과적으로 줄인다. 이를 통해 통신 효율성과 교통 제어 성능을 동시에 향상시키는 새로운 선택적 통신 프레임워크를 제시한다.

III. 제안 방법: Adaptive-SCP

1. 시스템 모델

도시 교통 신호 제어 문제를 MARL 기반으로 모델링하며, 시뮬레이션 환경으로 SUMO(Simulation of Urban Mobility)[24]를 사용하고 TraCI[25] 인터페이스를 통해 신호 제어를 수행한다. 도로 네트워크는 다수의 교차로(Traffic Light System, TLS)로 구성되며, 각 교차로의 신호는 하나의 에이전트로 정의된다. 각 에이전트는 자신의 국소적인 교통 상태를 기반으로 신호 위상을 선택하며, 인접 교차로와의 정보 교환을 통해 협력적 제어가 가능하다.

각 에이전트의 관측(state)은 해당 교차로에 진입하는 차선별 정지 차량 수를 기반으로 한 큐 길이(queue length)로 정의된다. 이는 실제 대기 차량 수를 근사하는 지표로 사용되며, 시간에 따라 변화하는 교통 혼잡도를 반영한다. 관측은 노드별 feature 벡터로 구성되며, 그래프 구조 내에서 인접 교차로 간 관계를 반영하기 위해 인접 행렬이 사용된다.

행동(action)은 각 교차로에서 선택 가능한 신호 위상 중 하나를 선택하는 것으로 정의된다. 각 에이전트는 이산적인 행동 공간을 가지며, 선택된 신호 위상은 TraCI를 통해 시뮬레이터에 적용된다. 이를 통해 교차로 단위의 분산 제어가 이루어진다. 보상(reward)은 각 스텝에서 시스템 내 모든 차량의 대기 시간(waiting time)의 합을 음수로 정의된다. 이는 에피소드 전체에 걸쳐 누적될 경우 총 대기시간(total waiting time)과 직접적으로 연계되며, 결과적으로 에이전트는 전체 교통 흐름을 개선하기 위해 대기 시간을 최소화하는 방향으로 학습된다.

Fig 1은 제안하는 Adaptive-SCP 기반 교통 신호 제어 시스템의 전체 데이터 흐름과 학습-통신-제어 간의 페루프 구조를 나타낸다. SUMO 교통 시뮬레이션 환경에서 TraCI를 통해 각 교차로의 큐 길이와 차선별 정

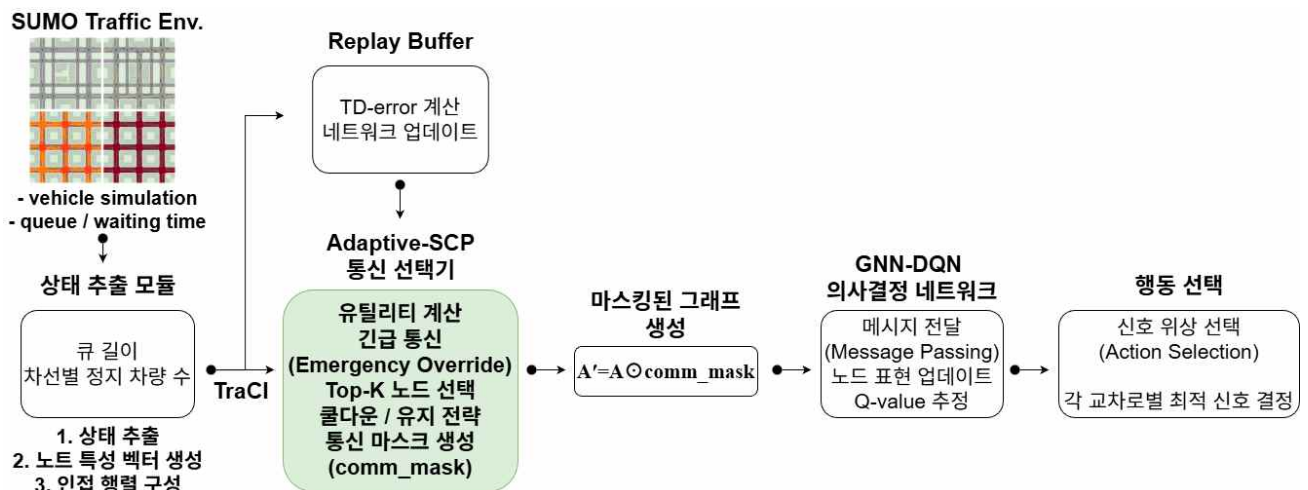


Fig 1. 선택적 통신을 위한 Adaptive-SCP 기반 다중 에이전트 교통 신호 제어 구조

지 차량 수 등의 교통 상태가 수집되며, 이를 기반으로 노드 특성 벡터와 인접 행렬이 구성된다. 동시에 상태 전이 정보는 Replay Buffer에 저장되고, 이를 이용하여 TD-error 계산과 GNN-DQN의 네트워크 업데이트가 수행된다. 제안하는 Adaptive-SCP 통신 선택기는 TD-error로부터 유도된 학습 가치 신호와 국소 혼잡 상태, 긴급 상황 여부를 함께 고려하여 통신 우선순위를 결정한다. 구체적으로, 긴급 노드를 우선 선택한 후, 학습 가치가 높은 Top-K 노드를 선택하고 cooldown 및 hold 전략을 적용하여 통신 안정성을 확보한다. 이 과정은 Algorithm 1의 통신 결정 단계와 대응되며, 최종적으로 comm_mask를 생성한다. 생성된 comm_mask는 인접 행렬에 적용되어 마스킹된 그래프 A' 를 형성하며, GNN-DQN은 해당 그래프 상에서만 메시지 패싱을 수행한다. 이를 통해 선택된 이웃 노드의 정보만 반영하여 각 교차로의 Q-Value를 추정하고, 최종적으로 최적 신호 위상을 결정한다. 이러한 구조를 통해 Adaptive-SCP는 학습적으로 중요한 노드에만 선택적으로 통신을 수행하면서도 교통 제어 성능을 안정적으로 유지할 수 있다.

2. GNN-DQN과 통신 마스킹

교차로 간의 공간적 상호작용을 효과적으로 반영하기 위해 본 연구에서는 GNN 기반의 DQN 구조[26]를 사용한다. 각 교차로는 그래프의 하나의 노드(node)로 표현되며, 노드 feature는 해당 교차로에 진입하는 차선별 정지 차량 수를 기반으로 한 큐 상태(queue length)로 구성된다. 인접행렬은 교차로 간 연결 관계를 정의하며, 이를 통해 인접 교차로의 상태 정보가 메시지 전달(message passing) 방식으로 통합된다.

Algorithm 1. Adaptive-SCP: Communication Decision

This algorithm selects a subset of nodes for communication at each step based on learning-value signals.

Input: TD-errors δ_i , queue history q_i , adjacency A

Hyperparameters: topk_ratio, cooldown, hold

- 1: for each node i do
- 2: $u_i \leftarrow \log(1 + |\delta_i|)$
- 3: end for
- 4: $S_{emergency} \leftarrow$ nodes satisfying emergency condition (queue threshold or rapid change)
- 5: $S_{candidate} \leftarrow$ nodes not in cooldown
- 6: $K \leftarrow \lfloor \text{topk_ratio} \times |V| \rfloor$
- 7: $S_{topk} \leftarrow$ top-K nodes in $S_{candidate}$ sorted by u_i
- 8: $S_{selected} \leftarrow S_{emergency} \cup S_{topk} \cup \text{hold_nodes}$
- 9: update cooldown for selected nodes
- 10: generate comm_mask based on $S_{selected}$

Output: comm_mask

GNN-DQN 구조에서 각 노드는 이웃 노드로부터 전달받은 정보를 기반으로 자신의 상태 표현(node embedding)을 갱신하고, 이를 통해 각 신호 위상(action)에 대한 Q-Value를 추정한다. 이 과정에서 인접 교차로의 정보는 정책 학습에 중요한 영향을 미치며, 개별 교차로의 국소적인 관측만으로는 반영하기 어려운 상류 및 하류 교통 흐름의 영향을 함께 고려할 수 있게 한다. 이를 통해 분산 환경에서도 협력적인 교통 신호 제어가 가능해진다.

통신은 Algorithm 1에서 정의된 comm_mask를 통해 제어된다. Adaptive-SCP는 각 시간 스텝마다 학습 가치 신호를 기반으로 통신에 참여할 노드를 선택하며, 선택된 노드들만 메시지 전달에 참여하도록 제한한다. 구체적으로, 각 노드의 TD-error를 기반으로 학습 가치 신호를 계산하고, 긴급 상황에 해당하는 노드를 우선적으로 포함한 뒤, 나머지 후보 노드 중 상위 Top-K를 선택한다. 이후 cooldown과 hold 전략을 적용하여 과도한 반복 선택을 방지하고 통신 안정성을 확보한 후 최종적으로 comm_mask를 생성한다. 통신 마스크는 인접행렬과 원소별 곱(Hadamard product)으로 결합되어 메시지 전달 경로를 제한한다. 구현에서는 기본 인접행렬(자기 루프 포함)에 노드별 통신 여부 벡터를 반영하여 마스킹된 인접행렬을 구성하며, 대각 성분은 항상 1로 유지한다. 이를 통해 각 노드는 자신의 관측 특성을 항상 node embedding에 포함할 수 있으며, 자기 상태 정보는 통신 여부와 관계없이 지속적으로 유지된다. GNN 인코더의 각 레이어에서는 마스킹된 인접행렬에 대해 행 기반의 차수 정규화를 수행한 후, 정규화된 인접행렬과 노드 특성 행렬의 곱을 통해 이웃 메시지를 집계한다. 이 과정에서 통신이 비활성화된 이웃 노드와의 연결 가중치는 0이 되므로 해당 방향의 메시지는 전달되지 않으며, 선택되지 않은 이웃 노드는 해당 스텝의 node embedding 업데이트에 기여하지 않는다. 또한, 별도의 학습 가능한 빈 이웃 임베딩을 사용하지 않고, 활성화된 이웃 노드들에 대해서만 정규화된 가중 합 형태의 aggregation이 수행된다.

따라서 동일한 GNN 구조를 유지하면서도 학습적으로 중요한 이웃 정보만 선택적으로 반영할 수 있으며, 불필요한 브로드캐스트를 줄이면서도 정책 학습의 안정성을 확보할 수 있다. 이러한 통신 마스킹 메커니즘은 통신 비용을 절감하면서도 교통 제어 성능 저하를 최소화하는 핵심 요소로 작동한다.

3. Adaptive-SCP

Adaptive-SCP는 각 시간 스텝에서 통신에 참여할 노드를 선택하여 comm_mask를 생성하는 선택적 통신 프로토콜이다. 기존의 고정 규칙 기반 통신과 달리, 본 프로토콜은 단순한 혼잡도 지표가 아닌 학습 과정에서의 학습 가치 신호를 통신 우선순위에 직접 반영한다. 각 노드 i 에 대하여, 학습 가치 신호 u_i 는 TD-error의 절댓값을 이용하여 다음과 같이 정의된다:

$$u_i = \log(1 + |\delta_i|) \quad (1)$$

여기서 δ_i 는 에이전트 업데이트 과정에서 얻은 노드별 TD-error이며, 배치 평균을 통해 추정된 값을 사용한다. 이 값은 해당 노드에서의 학습 또는 정보 갱신 필요성을 나타내며, 학습 가치 신호 u_i 값이 클수록 통신의 필요성이 높다고 판단한다. Adaptive-SCP의 노드 선택 절차는 다음과 같다. 먼저, 각 노드의 최근 큐 이력으로부터 긴급 상태를 먼저 판정한다. 이때, 큐 길이가 일정 임계값을 초과하거나 급격한 변화가 발생하는 등 긴급 노드는 통

신 예산과 무관하게 즉시 통신 집합에 포함한다. 이 후, 나머지 노드 중 cooldown 상태가 아닌 노드에 대해 학습 가치 신호 u_i 를 계산하고, topk_ratio에 따라 상위 K 개의 노드를 선택하여 통신에 참여시키며, 이를 통해 통신량의 상한을 제어한다. 또한, 이전에 선택된 노드를 유지하는 hold 전략을 통해 통신의 안정성을 확보한다. Adaptive-SCP를 통해 생성된 comm_mask는 GNN 메시지 패싱 시 이웃 정보 반영 범위를 제한하고, replay buffer에 함께 저장되어 학습 단계의 상태 전이와 일관되게 사용된다.

따라서, Adaptive-SCP는 단순 규칙 기반 게이팅이 아니라, 학습 신호-통신 결정-정책 업데이트가 연결된 페루프 구조를 갖는다. 이 구조는 학습적으로 중요한 노드와 긴급 혼합 노드는 놓치지 않도록 선택적으로 통신하는데 이점을 갖는다. 또한, Adaptive-SCP는 통신 예산 제어, 안정성 확보, 안전장치를 동시에 포함하는 실용적 프로토콜로 설계되었다.

4. 통신 비용 지표 정의

통신 비율은 에피소드 동안 실제로 통신이 발생한 횟수를 전체 가능한 통신 기회수로 정규화한 지표를 사용한다.

$$Comm\ Ratio\ (\%) = \frac{comm - steps}{steps \times N_{TLS}} \quad (2)$$

여기서, $comm - steps$ 는 통신이 활성화된 총 횟수, $steps$ 는 전체 시뮬레이션 스텝 수, N_{TLS} 는 교차로(에이전트)의 수를 의미한다. 즉, 이 지표는 이론적인 최대 통신량인 모든 노드가 매 스텝 통신하는 경우를 기준으로 실제 통신 사용 비율을 나타낸다.

IV. 실험 및 결과

1. 실험 목적 및 환경

본 장에서는 제안하는 Adaptive-SCP가 교통 신호 제어에서 통신 비용을 얼마나 절감하는지, 그리고 그에 따른 총 대기시간과 기준선과 비교해 어떻게 달라지는지를 정량적으로 검증한다. 비교 기준으로는 모든 노드가 매 스텝 통신하는 always_on(통신 비율 100%), 그리고 제안 Adaptive-SCP를 사용한다. 실험은 Fig 2와 같이 SUMO 기반 10x10 교차로 그리드 환경에서 수행되며, 각 교차로는 GNN-DQN 에이전트로 제어된다. 교통 수요는 low, med, high, rush 네 가지 시나리오로 구성되며, 무작위성 제어를 위해 서로 다른 10개의 시드를 사용하고, 각 시드마다 4개의 에피소드, 에피소드당 최대 900스텝으로 실험을 수행하였다. 이에 따라 각 수요 및 통신 모드 조합에 대해 총 40개의 에피소드 결과를 집계하였다. 추가적으로, Fig 3 과 같이 OpenStreetMap 기반 실제 도심 도로 네트워크를 대상으로 동일한 구조의 실험을 수행하였으며, 기본 수요와 rush 수요에 대해 각각 2개의 시드와 4개의 에피소드를 사용하여 총 8개의 결과를 집계하였다. 모든 비교 모드는 동일한 학습 설정과 하이퍼파라미터를 공유하여 공정한 비교를 보장한다.

2. 실험 결과

Table 1은 10x10 그리드 환경에서 각 수요 시나리오에 대한 always_on(항상 통신)과 Adaptive-SCP의 평균 총 대기시간과 Adaptive-SCP의 통신 비율을 나타낸다. Adaptive-SCP는 always_on 대비 수요별 통신 비율이 약 26.9% ~ 32.6%로, 전체 통신량의 약 67% ~ 73%를 일관되게 절감한다. 이러한 통신 절감에도 불구하고 평균 총 대기시간은 전반적으로

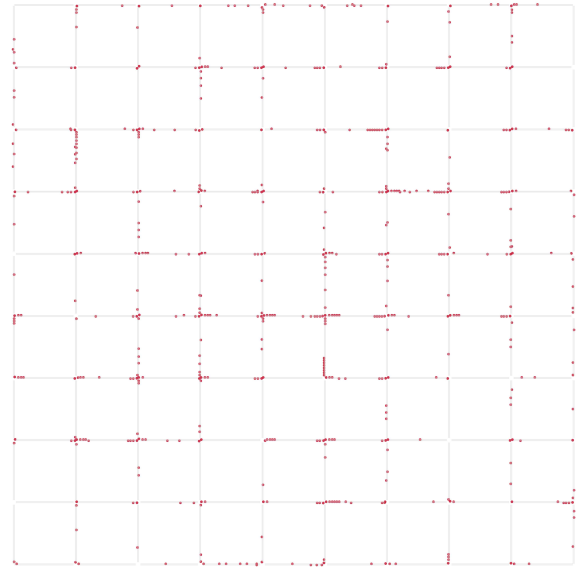


Fig 2. 10x10 그리드 도심 환경 시뮬레이션



Fig 3. OpenStreetMap 기반 실제 도심 도로 네트워크 시뮬레이션

always_on과 유사한 수준을 유지한다. 구체적으로 low 및 rush 수요에서는 두 방식 간 차이가 크지 않으며, high 수요에서는 Adaptive-SCP(112,461)가 always_on(118,104)보다 총 대기시간이 작게 나타난다. 이는 항상 통신을 수행하는 경우 불필요한 정보가 포함되어 정책 학습에 영향을 줄 수 있는 반면, Adaptive-SCP는 학습 가치가 높은 노드를 중심으로 통신을 제한하여 정보 전달의 효율을 높였기 때문으로 해석할 수 있다.

수요	always_on 총 대기시간(초)	Adaptive-SCP 총 대기시간(초)	Adaptive-SCP 통신 비율(%)
low	16,848	16,721	26.9
med	37,245	37,652	28.7
high	118,104	112,461	32.6
rush	23,301	23,258	27.5

Table 1. 수요별 평균 총 대기시간 및 통신 비율(10 x 10, n=40)

수요	always_on 총 대기시간(초)	Adaptive-SCP 총 대기시간(초)	Adaptive-SCP 통신 비율(%)
med	80,749	83,702	26.2
rush	96,438	95,443	26.5

Table 2. 실제 도심 시나리오 평균 총 대기시간 및 통신 비율(n=8)

OpenStreetMap 기반 실제 도심 도로 네트워크 시나리오 실험에서도, Adaptive-SCP는 always_on 대비 약 26%의 통신 비율을 유지한다. 또한, always_on 대비 통신량을 크게 절감함과 동시에, always_on 대비 유사한 평균 총 대기시간을 유지하며, 이러한 결과는 제안 방법이 합성 그리드 환경뿐 아니라 실제 도시 네트워크에서도 유사한 경향을 보이며, 통신 비용과 성능 간의 균형을 유지할 수 있음을 시사한다. 실험 결과를 종합하면, Adaptive-SCP는 always_on 대비 통신 비율을 약 26% 수준으로 유지하며 전체 통신량을 크게 절감한다. 이러한 절감에도 불구하고, 평균 총 대기시간 성능은 대부분의 수요 조건에서 always_on과 유사한 수준을 유지하며, 특히 높은 혼잡 조건에서는 오히려 개선되는 경향이 관측된다.

V. 결론 및 향후 연구

본 논문에서는 교통 신호 제어를 위한 다중 에이전트 강화학습 환경에서 통신 비용을 줄이기 위한 선택적 통신 프로토콜인 Adaptive-SCP를 제안하였다. 기존의 MARL 기반 접근은 에이전트 간 정보를 항상 교환하는 구조를 전제로 하여 통신 비용을 명시적으로 고려하지 않는 한계를 가지며, 단순한 규칙 기반 선택적 통신 방식 역시 환경 변화에 대한 적응성이 제한적이라는 문제가 있다. 이러한 배경에서 본 연구는 학습 과정에서의 정보 가치에 기반하여 통신 여부를 동적으로 결정하는 방법을 설계하였다. 제안 방법은 각 노드의 TD-error를 기반으로 학습가치 신호를 정의하고, 이를 이용하여 매 스텝 통신에 참여할 노드를 선택한다. 또한 emergency override, cooldown, hold와 같은 보조 메커니즘을 통해 통신 안정성과 반응성을 함께 고려하였다. 이러한 구조는 통신을 단순히 줄이는 것이 아니라, 필요한 시점과 위치에서만 선택적으로 수행하도록 하는데 목적이 있다.

실험 결과, Adaptive-SCP는 always_on 대비 통신비율을 및 약 27-33% 수준으로 유지하며, 전체 통신량을 일관되게 절감하는 것으로 나타났다. 이러한 통신 절감에도 불구하고 총 대기시간 성능은 대부분의 수요 조건에서 always_on과 유사한 수준을 유지하였으며, 일부 혼잡 조건에서는 오히려 개선되는 경향이 관측되었다. 이는 모든 노드가 항상 통신하는 구조가 항상 최적이지 않을 수 있으며, 학습가치 기반 선택적 통신이 보다 효율적인 정보 전달을 가능하게 함을 시사한다. 결론적으로, 본 연구는 교통 신호 제어에서 통신을 하나의 비용 요소로 고려하고, 이를 학습 과정과 통합하여 제어하는 새로운 접근을 제시한다. Adaptive-SCP는 통신 비용과 제어 성능 간의 균형을 유지하면서도 다양한 환경에서 안정적으로 동작할 수 있는 프레임워크로 활용될 수 있다. 향후에는 보다 다양한 교통 네트워크와 수요 조건에서의 평가를 통해 제안 방법의 일반성을 검증하고, 통신 전략과 정책 학습 간의 상호작용을 보다 정교하게 모델링하는 방향으로 확장할 수 있을 것이다.

ACKNOWLEDGMENT

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 농식품과학기술융합형연구인력양성사업의 지원을 받아 연구되었음(RS-2024-00397026).(34%) 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2026-RS-2022-00156287)(33%)본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2026-RS-2023-00256629)(33%)

참고 문헌

- [1] H. Shen, H. Zhao, Z. Zhang, X. Yang, Y. Song and X. Liu, "Network-Wide Traffic Signal Control Based on MARL With Hierarchical Nash-Stackelberg Game Model," in IEEE Access, vol. 11, pp. 145085-145100
- [2] N. Sumathi and G. Navamani, "Real-Time Traffic Optimization: A Graph-Based Reinforcement Learning Framework," 2025 International Conference on Inventive Computation Technologies (ICICT), Kirtipur, Nepal, 2025, pp. 767-771
- [3] W. Xiaopeng et al., "A Traffic Intersection Optimization Control System Based on GNN-RL Integration Framework," 2025 37th Chinese Control and Decision Conference (CCDC), Xiamen, China, 2025, pp. 737-742
- [4] X. Peng, S. Chen, H. Gao, H. Wang and H. Michael Zhang, "Combat Urban Congestion via Collaboration: Heterogeneous GNN-Based MARL for Coordinated Platooning and Traffic Signal Control," in IEEE Transactions on Intelligent Transportation Systems, vol. 26, no. 6, pp. 7667-7677, June 2025
- [5] Z. Zeng, "GraphLight: Graph-based Reinforcement Learning for Traffic Signal Control," 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, China, 2021, pp. 645-650
- [6] Das, Abhishek, et al. "Tarmac: Targeted multi-agent communication." International Conference on machine learning. PMLR, 2019.
- [7] Wei, Hua, et al. "Presslight: Learning max pressure control to coordinate traffic signals in arterial network." Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019.
- [8] Y. Zheng, W. Zhao and J. Luo, "GTALight: An IoT-Enhanced End-Edge-Cloud Collaborative Framework for Scalable Regional Traffic Signal Control," in IEEE Internet of Things Journal, vol. 13, no. 4, pp. 6863-6876, 15 Feb.15, 2026
- [9] Singh, Amanpreet, Tushar Jain, and Sainbayer Sukhbaatar. "Learning when to communicate at scale in multiagent cooperative and competitive tasks." arXiv preprint arXiv:1812.09755 (2018).
- [10] Zhu, Changxi, Mehdi Dastani, and Shihan Wang. "A survey of multi-agent deep reinforcement learning with communication." Autonomous Agents and Multi-Agent Systems 38.1 (2024): 4.
- [11] Tian, Qi, et al. "CGIBNet: Bandwidth-constrained Communication with Graph Information Bottleneck in Multi-Agent Reinforcement Learning." arXiv preprint arXiv:2112.10374 (2021).
- [12] R. Bokade, X. Jin and C. Amato, "Multi-Agent Reinforcement Learning Based on Representational Communication for Large-Scale Traffic Signal Control," in IEEE Access, vol. 11, pp. 47646-47658, 2023
- [13] Zhang, Sai Qian, Qi Zhang, and Jieyu Lin. "Efficient communication in multi-agent reinforcement learning via variance based control." Advances in neural information processing systems 32 (2019).
- [14] D. Chen, K. Zhang, Y. Wang, X. Yin, Z. Li and D. Filev,

- "Communication-Efficient Decentralized Multi-Agent Reinforcement Learning for Cooperative Adaptive Cruise Control," in *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 10, pp. 6436-6449, Oct. 2024
- [15] Zhang, K. et al. "Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents." *International Conference on Machine Learning* (2018).
- [16] Guan, Cong, et al. "Efficient communication via self-supervised information aggregation for online and offline multiagent reinforcement learning." *IEEE Transactions on Neural Networks and Learning Systems* 36.5 (2024): 9044-9056.
- [17] J. Mei, Z. Quan, W. Yang and X. Cao, "Attention based Reinforcement Learning for Efficient Communication under Constraint in Multi-Agent Systems," 2023 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 2023, pp. 3867-3873
- [18] K. Shibata, T. Jimbo and T. Matsubara, "Deep reinforcement learning of event-triggered communication and control for multi-agent cooperative transport," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 8671-8677
- [19] N. H. Ly, N. Q. Doan, A. T. Giang, H. T. Le, H. H. Nguyen and H. T. Tran, "Enhancing Urban Traffic Control with Priority-Aware Multi-Objective Reinforcement Learning," 2025 17th International Conference on Knowledge and System Engineering (KSE), Da Lat, Vietnam, 2025, pp. 1-6
- [20] Casilimas, Katherine & Caicedo, Mauricio. (2022). Deep Reinforcement Learning for Resource Management on Network Slicing: A Survey. *Sensors*. 22. 32. 10.3390/s22083031.
- [21] Y. Dang, Q. Lin and P. Huang, "AIOps: Real-World Challenges and Research Innovations," 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), Montreal, QC, Canada, 2019, pp. 4-5
- [22] M. Lu, H. Zhu, Y. Chen and P. Lin, "Resource-Aware and Cross-Layer Resource Allocation Mechanism for Power Wireless Access Networks," 2018 5th International Conference on Information Science and Control Engineering (ICISCE), Zhengzhou, China, 2018, pp. 1231-1235
- [23] K. J. Reddy, P. Udayaraju, V. P. R. Pamaiahgari and V. D. Kumar, "Implementing Resource-Aware Scheduling Algorithm for Improving Cost Optimization in Cloud Computing," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 282-287
- [24] M. Bakbergen and Y. Didar, "Predicting High Temporal Deviation in Traffic Flows Based on SUMO Simulation," 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2025, pp. 1-6
- [25] Y. Yang, S. -Y. Hao and H. -B. Cai, "Comparison and Evaluation of Routing Protocols Based on a Collaborative Simulation Using SUMO and NS3 with TraCI," 2016 International Conference on Information System and Artificial Intelligence (ISAI), Hong Kong, China, 2016, pp. 253-257
- [26] S. Kommuri, V. L. Korupu, A. Yellanki, A. Chippala and Y. Nuthalapati, "Smart Traffic Control: Leveraging DQN for Improved Urban Mobility," 2024 Control Instrumentation System Conference (CISCON), Manipal, India, 2024, pp. 1-6