강화학습 기반 다중 양자보안 서비스 스케줄링 기법 연구

임현교, 이찬균, 김주봉, 이원혁
한국과학기술정보연구원, 양자통신연구센터
{hk.lim, chankyunlee, jjbong, livezone}@kisti.re.kr

A Study on Reinforcement Learning-Based Multiple Quantum Security Service Request Scheduling

Hyun-Kyo Lim, Ju-Bong Kim, Wonhyuk Lee
Korea Institute of Science and Technology Information (KISTI)

요 약

본 논문은 양자암호통신 네트워크 환경에서 다중 양자보안 서비스 요청을 효율적으로 처리하기 위한 강화학습 기반 스케줄링 기법을 제안한다. 현재 양자암호통신 네트워크는 물리적 한계로 인해 제한된 양자키 자원을 활용해 다수의 서비스 요청을 동시에 처리하는 데 한계가 있다. 이를 해결하기 위해 전반적인 네트워크 상태와 큐에 대기 중인 양자보안 서비스 요청 특성을 고려한 Markov decision process 정의를 기반으로, PPO 알고리즘과 Set transformer 구조의 정책망을 결합한 강화학습 모델을 설계한다. 제안하는 강화학습 기반 스케줄링 기법의 성능 평가를 위해 FIFO, Random, Min-path 기반 스케줄링 기법 대비 평균 할당량과 양자키 자원 활용 측면에서 NSFNET 환경을 활용해 실험 평가를 수행한다.

Ⅰ. 서 론

양자암호통신은 양자 키 분배(Quantum Key Distribution) 기술을 통해 외부의 도청으로부터 안전한 네트워크 시스템을 구축[1]함으로써, 생성된 양자키 자원을 활용해 기존의 암호 시스템보다 높은 안전성을 제공한다. 그러나 현재 양자암호통신은 물리적 한계로 인해 양자키 자원의 생성률이 높지 않아 자원 부족으로 인한 양자보안 서비스 제공의 한계가 존재한다. 부족한 양자암호통신의 자원을 보다 효율적으로 활용함으로써, 보다 많은 양자보안 서비스 요청을 처리하기 위한 연구[2, 3]는 아직 상대적으로 부족하다.

양자암호통신 네트워크에서 서비스 계층의 양자보안 서비스 요청은 일정 시간 내에 양자키 자원을 활용해 암호화 데이터 전송을 요구한다. 양자암호통신 네트워크의 양자 키 분배 기술은 높은 보안성을 제공하지만, 실제 활용에서는 거리적 한계와 양자 채널의 손실로 인해 중요 자원인 양자키 생성률이 낮으며, 현재 기술로는 약 100km 이상의 거리를 양자 채널로 직접 제공하기 어렵다. 이에 따라 양자키 릴레이 기법을 통해 중간에 신뢰할 수 있는 양자 중계기를 거쳐 원거리 노드 간 양자키 교환을 수행하여 양자보안 서비스 요청을 위한 양자키 자원을 할당한다. 하나의 양자보안 서비스 요청을 처리하기 위해서는 최소 1개 이상의 양자키 자원을 소모하며, 원거리 요청일수록 더 많은 양자키 자원을 필요로 한다. 또한 매 시간마다 다중 양자보안 서비스 요청이 동시에 생성될 수 있으며, 다중 요청들을 효과적으로 처리하기 위해서는 스케줄링 기법이 필수적이다.

본 논문에서는 서비스 계층에서 발생하는 양자보안 서비스 요청을 동시에 처리하기 위한 강화학습 기반 스케줄링 기법을 제안한다. 제안하는 강화학습 기반 스케줄링 기법은 양자암호통신 네트워크의 남은 양자키 자원과 현재 대기중인 요청들을 상태 정보로 활용해 가장 적절한 대기 중인 다중 요청들의 할당 순서를 결정한다. 강화학습 기반 스케줄링을 위해 본 논문에서는 실제 양자암호통신 양자키 자원 개수, 키의 수명(lifetime)과 양자보안 서비스 요청의 출발지 및 도착지 노드, 요청 대기 시간(waiting time)을 동시에 고려할 수 있는 강화학습 모델을 설계한다.

환경은 매 에피소드마다 초기화되며, 일정량의 양자키가 링크별로 공급되고 시간이 지남에 따라 키의 수명은 감소한다. 매 스텝에서 에이전트는 현재 대기 중인 양자보안 서비스 요청 중 하나를 선택하고, 선택된 요청이 네트워크 경로 상에서 할당 가능한 양자키 자원을 보유하고 있을 경우 성공적으로 서비스를 제공한다. 이때 경로에 포함된 링크의 양자키 수량은 1씩 감소하고, 서비스가 성공적으로 제공되면 에이전트는 보상을 획득한다. 반대로 요청이 실패하거나 자원이 부족할 경우 차단(blocking)으로 처리한다.

Ⅱ. 본론

제안하는 강화학습 기반 스케줄링 기법은 양자암호통신 환경과 상호작용을 통해 누적 보상을 최대화하는 정책을 학습한다. 강화학습 기반 스케줄링 문제를 해결하기 위해 MDP(Markov Decision Process) 정의를 다음과 같이 한다.

- **상태(State)**: 양자암호통신 네트워크의 상태 정보와 현재 큐에 남은 양자보안 서비스 요청들의 정보로 나누어진다. 먼저 양자암호통신 네트워크의 상태 정보는 남은 키 풀 상태와 토폴로지 구성을 인지하기 위해 각 링크의 남은 양자키 개수로 이루어진 $N \times N$ 인접 행렬(Adjacency Matrix)으로 구성 구성된다. 또한, 또한 현재 대기중인 양자보안 서비스 요청들의 스케줄링을 위한 순서 결정을 위해 각 요청별 특성(출발 노드, 목적 노드, 경로 길이(hop 수, 현재 키 풀 기반), 최소 잔여 키, 대기 시간)을 포함한다.
- **행동(Action)**: 현재 큐에 대기 중인 양자보안 서비스 요청들의 순서를 결정한다.
- **보상(Reward)**: 강화학습 에이전트의 행동으로 결정된 순서로 양자키가 할당되어 양자보안 서비스가 제공되었으면 +1, 양자키 자원 부족으로 인해 차단될 경우 -1을 얻는다.

본 논문에서는 PPO(Proximal policy optimization) 알고리즘을 기반으로 강화학습 모델을 학습시켰으며, 정책망은 대기중인 다중 양자보안 서비스 요청들의 대기 순서에 상관 없이 특성을 반영하기 위해 Set Transformer 구조[4]를 기반으로 하여, 요청 집합을 인코딩하고 양자암호통신 네트워크 상태와 결합해 의사결정을 수행하도록 구현한다.

실험은 NSFNET 토폴로지를 기반으로 진행하였으며, 다양한 요청 발생 패턴과 키 생성률 조건에서 강화학습 기반 스케줄링 기법의 성능을 평가하였다. 표 1은 양자암호통신 실험 환경 파라미터 설정이며, 표 2는 강화학습의 하이퍼 파라미터이다.

표 1. NSFNET 실험 파라미터 설정

파라미터	값
노드	14
링크	21
초기 키 수명	3
키 생성량	2
키 생성 노이즈	1.0
요청 생성량	10

표 2. 강화학습 하이퍼 파라미터 설정

파라미터	값
학습률(Learning rate)	0.0001
할인율(Discount factor)	0.98
GAE λ	0.95
PPO clipping 범위	0.1
Entropy	0.05
미니배치 크기	64
총 에피소드	10,000

또한, 강화학습 기반 스케줄링 기법과의 비교를 위해 1) FIFO, 2) 무작위(Random), 3) 최소 경로(Min-hop)와 같은 기준 알고리즘들과의 비교를 통해 제안 기법의 우수성을 검증하였다.

표 3과 그림 1의 NSFNET 토폴로지를 기반으로 진행한 비교 실험에서, 제안된 강화학습 기반 스케줄링 기법은 평균 할당량 측면에서 Min-path, FIFO, Random과 같은 기존 기준 알고리즘 대비 우수한 성능을 보였다. 특히 평균 할당량은 4.0으로 가장 높게 나타났으며, 이는 제한된 양자키 자원을 보다 효율적으로 활용하여 더 많은 요청을 성공적으로 처리했음을 의미한다. 또한, 기준 알고리즘들과 할당량의 표준편차가 상대적으로 작았으며, 이는 강화학습 기반이 요청 패턴과 자원 상태 변화에 따라 변동성이 적다는 점을 보여준다. 또한 만료된 키 개수 측면에서 Random 기법은 무작위로 요청을 처리하기 때문에 특정 경로에 집중되지 않고, 자원이 빠르게 소모되어 키가 사용되기 전에 만료되는 키의 개수가 적다. 그러나 제안하는 RL 기법은 할당량을 증가시키기 위해 경로에 따라 적절한 요청의 순서를 정함으로써 만료된 키의 개수가 상대적으로 낮은 것을 볼 수 있다.

표 3. 성능 비교 평가 결과

알고리즘	할당량	할당량 표준편차	만료된 키 수
RL	**4.00**	**6.18**	695
Min-path	3.30	7.73	926
FIFO	3.68	7.65	842
Random	3.04	7.60	**574**

Ⅲ. 결론

본 논문에서는 양자암호통신 네트워크 환경에서 다중 양자보안 서비스 요청을 동시에 처리하기 위한 강화학습 기반 스케줄링 기법을 제안하였다. 제안된 기법은 PPO 알고리즘과 Set Transformer 기반 정책망 모델을 활용하여, 네트워크 상태(토폴로지, 잔여 키 자원)와 요청 특성(출발/목적 노드, 경로 hop 수, 최소 잔여 키, 대기 시간)을 동시에 고려하는 정책을 학습하였다. 성능 평가를 위해 NSFNET 토폴로지 환경에서 실험을 진행했으며, Min-path, FIFO, Random 알고리즘들 대비 강화학습 기반 스케줄링 기법이 양자보안 서비스 할당량에서 보다 나은 성능을 보이는 것을 증명했다. 이러한 결과는 제안된 기법이 단순히 평균 성능 향상뿐 아니라 자원 활용과 서비스 제공 효율성 측면과 양자키 자원의 로드 밸런싱 측면에서도 효율적임을 보인다.

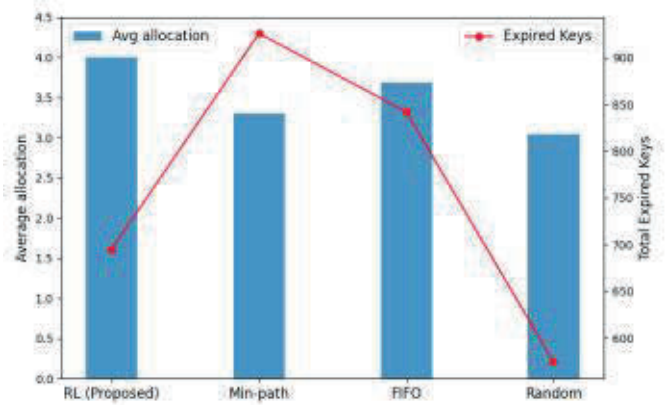

그림 1. 각 알고리즘 별 평균 할당량 및 만료된 키 개수 비교

향후 연구에서는 (1) 더 다양한 토폴로지와 대규모 네트워크 환경에서의 확장성 검증, (2) 키 자원 예측 기반 proactive scheduling 기법과의 결합, (3) 실제 QKD 장비 성능 특성을 반영한 시뮬레이션을 통한 검증을 수행하고자 한다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. RS-2025-02263666)과 2025년도 한국과학기술정보연구원(KISTI)의 기본사업의 지원(과제번호: (KISTI)K25L5M2C2)을 받아 수행된 연구임

참 고 문 헌

[1] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dusek, N. Lutkenhaus, and M. Peev, “The Security of Practical Quantum Key Distribution,” Reviews of Modern Physics, vol. 81, no. 3, pp. 1301 - 1350, Sep. 2009.

[2] Y. Fu, Y. Hong, T. Q. S. Quek, H. Wang and Z. Shi, “Scheduling Policies for Quantum Key Distribution Enabled Communication Networks,” in IEEE Wireless Communications Letters, vol. 9, no. 12, pp. 2126-2129, Dec. 2020.

[3] Y. Cao, Y. Zhao, Y. Wu, X. Yu and J. Zhang, “Time-Scheduled Quantum Key Distribution (QKD) Over WDM Networks,” in Journal of Lightwave Technology, vol. 36, no. 16, pp. 3382-3395, 15 Aug.15, 2018.

[4] J. Lee, Y. Lee, J. Kim, A. Kosiorek, S. Choi, Y.W. Teh, “Set transformer: A framework for attention-based permutation-invariant neural networks,” International conference on machine learning. PMLR, 2019.