# QKD 네트워크에서 세션 블록 최소화를 위한 심층강화학습 기반 랜덤키 생성 최적화

석영준[1], 김주봉[1], 허주성[1], Nurlanbek kyzy Asel[2], 한연희[1†]
[1]한국기술교육대학교 컴퓨터공학과 미래융합공학전공
[2]한국기술교육대학교 컴퓨터공학과 컴퓨터공학전공
[1]{dsb04163, rlawnqhd, chil1207, yhhan}@koreatech.ac.kr,
[2]aselbaekki@koreatech.ac.kr

# Deep Reinforcement Learning-Based Optimization of Random Key Generation to Minimize Session Blocking in QKD Networks

Yeong-Jun Seok[1], Ju-Bong Kim[1], Joo seong Heo[1], Nurlanbek kyzy Asel[2], Youn-Hee Han[1]
[1]Future Convergence Engineering, Dept. of Computer Science and Engineering, KOREATECH
[2]Dept. of Computer Science and Engineering, KOREATECH

요 약

양자 컴퓨터 및 양자 알고리즘의 발전으로 수학적 복잡성에 기반한 전통적인 암호체계는 쉽게 해독될 수 있다. 따라서 기밀성과 신뢰성을 유지하기 위한 새로운 보안 방식으로 Quantum Key Distribution (QKD)이 주목받고 있다. QKD 시스템에서 생성된 암호키는 양자 특성 때문에 사용에 제한이 존재한다. 따라서 QKD 시스템에서 생성된 암호키를 여러 인증 서비스에 효율적으로 할당하는 방법이 필요하다. 본 논문은 QKD 시스템의 랜덤키 생성 최적화를 위한 DRL방법을 제안한다. 제안한 방법은 Graph Attention Network (GAT)와 Long Short-Term Memory (LSTM)을 활용한다. DRL의 에이전트(Agent)는 QKD 시스템이 적용된 네트워크의 모든 정보를 그래프로 입력받아 원거리 서비스 지원 위한 랜덤키의 생성량 결정을 행동(Action)을 선택한다. 학습 결과를 통해 DRL방법은 서비스 블록(Blocking) 현상을 줄일 수 있고, 그리디 알고리즘(Greedy Algorithm)과 비교 실험을 통해 다양한 서비스 발생 패턴에 대응할 수 있음을 증명했다.

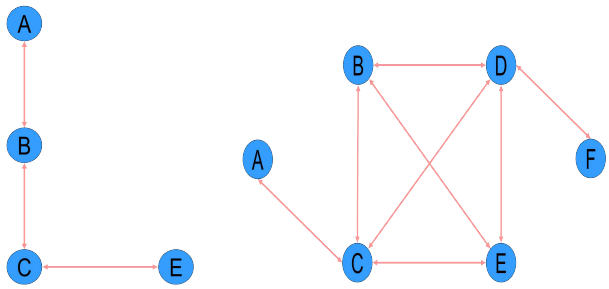
## Ⅰ. 서론

양자 컴퓨터 및 양자 알고리즘의 발전은 수학적 복잡성에 기반한 전통적인 암호화 체계에 위협이 되고 있다[1-2]. 그 결과, 금융 및 다른 산업 분야에서는 기밀성과 신뢰성을 유지하기 위한 새로운 보안 접근 방식을 모색하고 있다. 그중 하나가 Quantum Key Distribution (QKD)이다.

QKD는 양자 물리학의 원리를 활용하여 완벽한 보안을 제공하는 암호키를 배포하는 첨단 시스템이다[3]. QKD에서 생성된 암호키는 양자비트의 특성을 활용하여 대칭키를 형성한다. 그러나 QKD가 적용된 네트워크에서 생성된 암호키는 양자 특성 때문에 재사용이 불가능하며 사용 시간이 제한되어 있다. 이러한 특성은 QKD의 암호키가 보안을 보장하지만, 효율적인 사용에 제약이 존재한다는 것을 의미한다. 따라서 QKD 시스템에서 생성된 암호키를 여러 인증 서비스에 효율적으로 할당하는 방법이 필요하다.

현재, 자원 할당 문제에 대한 효과적인 해결책으로 Deep Reinforcement Learning (DRL)의 적용이 고려될 수 있다. DRL은 Deep Neural Network (DNN)를 통해 기존의 강화학습 한계를 극복하고 다양한 분야의 문제를 해결하는 데 사용된다[4]. 에이전트(Agent)는 환경(Environment)과 상호작용하여 누적 보상(Reward)을 최대화하는 행동(Action)을 선택하는 정책(Policy)을 학습한다.

본 논문은 QKD 시스템에서 Blocking 최소화를 위한 DRL기반 랜덤키 생성 최적화 방법을 제안하고, 이를 통해 QKD가 적용된 네트워크의 통신을 원활하게 하는 것을 목표로 한다. 또한 실제 양자암호통신망 구조를 환경으로 하여 그리디 알고리즘(Greedy Algorithm)과 비교실험으로 DRL 적용의 필요성을 입증한다. 또한 네트워크 구조에 활용하기 위해 Graph Attention Network (GAT)[5]와 다양한 랜덤키 요구 패턴에 반응할 수 있도록 Long Short-Term Memory (LSTM)[6]를 활용한다.



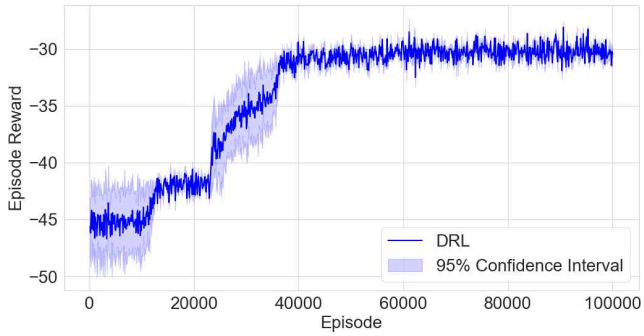

(a)베이징-상하이 (b) SECOQC

[그림 1] 본 연구에서 활용하는 실제 QKD 시스템이 적용된 네트워크 토폴로지

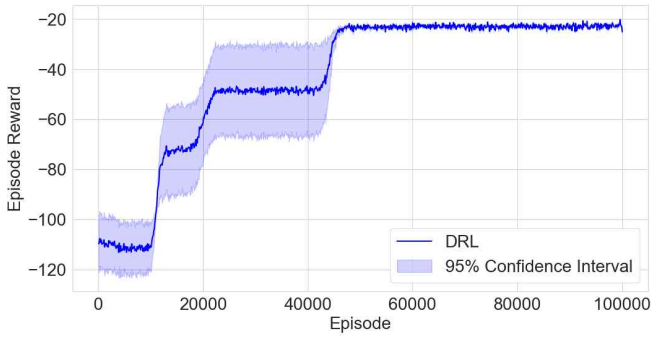
## Ⅱ. 본론

본 논문에서 QKD 시스템의 설정은 [7]을 따르며, 실제 QKD 시스템이 적용된 네트워크 토폴로지를 환경으로 사용한다. QKD 시스템이 적용된 네트워크에서 인접한 노드는 비밀키로 양자키를 사용하고, 인접하지 못한 노드는 비밀키로 랜덤키를 사용한다.

† 한연희(Youn-Hee Han, yhhan@koreatech.ac.kr): 교신저자




(a) 베이징-상하이 네트워크 토폴로지에서 훈련 중 DRL 에이전트의 에피소드 보상




(b) SECOQC 네트워크 토폴로지에서 훈련 중 DRL 에이전트의 에피소드 보상

[그림 2] Poisson Distribution 기댓값 $\lambda=1$에서 훈련 중 DRL 에이전트의 에피소드 보상

(표 1) 100회 평균 에피소드 $SB_t$

| | 베이징-상하이 | | | SECOQC | | |
|---|---|---|---|---|---|---|
| Algorithm | $\lambda$=0.8 | $\lambda$=1.0 | $\lambda$=1.2 | $\lambda$=0.8 | $\lambda$=1.0 | $\lambda$=1.2 |
| DRL | 15.53 | 28.91 | 44.59 | 6.42 | 22.02 | 46.64 |
| Greedy | 20.76 | 33.52 | 50.3 | 14.82 | 30.38 | 52.18 |

모든 비밀키는 키 풀(Key Pool, KP)에 저장되고 사용된다. 또한, 모든 비밀키는 생명 주기(Life Cycle)를 가지고 있고, 생명 주기가 만료되면 삭제된다.

서비스는 비밀키를 요구하며, 푸아송 분포(Poisson Distribution)를 따라 동적으로 생성된다. 비밀키를 할당받지 못한 서비스는 블록(Blocking)되며, 거부된다. 제안하는 DRL 방법은 원활한 서비스 지원을 위해 시간당 서비스 블록 $SB_t$을 감소시킨다. 따라서 문제의 목적함수를 다음과 같이 정의한다.

$$mininize \lim_{T\to\infty} \frac{\sum_{t=0}^{T} SB_t}{T} \quad (1)$$

에이전트는 환경으로부터 모든 노드 및 KP의 정보를 그래프 형태로 입력받고, 랜덤키 생성 수 결정을 Action으로 선택한다. 시간 $t$에서 Reward는 $r_t$로 나타내고 $SB_t$가 증가하면 누적 Reward $R$이 감소한다.

$$R = \sum_{t=0}^{T} r_t = \sum_{t=0}^{T} - SB_t \quad (2)$$

[그림 2]는 GAT와 LSTM을 결합한 모델에 Proximal Policy Optimization (PPO) 알고리즘을 적용한 학습 결과를 보여줍니다. [그림 2]는 푸아송 분포의 기댓값 $\lambda$가 1인 환경에서 5회 반복된 학습의 평균 $R$ 및 95% 신뢰 구간을 보여줍니다. 반복 학습에서 평균 $R$이 증가하며, 제안하는 DRL 방법이 에피소드 $SB_t$의 수를 줄이는 것을 알 수 있다.

[표 1]은 제안한 DRL 방법과 서비스가 랜덤키 요구량 만큼 랜덤키를 생성하는 그리디 알고리즘의 성능을 비교한 실험 결과를 보여준다. 실험은 총 100회 반복되었으며, 각 실험에서 발생한 에피소드 $SB_t$의 값을 표로 제시한다. 또한, 제안한 DRL 방법의 일반화 성능을 확인하기 위해 푸아송 분포의 기댓값 $\lambda$를 학습에서 사용하지 않은 값으로 설정해서 추가 실험했다. 실험 결과, 제안한 DRL 방법은 다양한 환경에서 평균 에피소드 $SB_t$의 값이 그리디 알고리즘보다 적은 것으로 나타났다. 따라서, DRL 방법은 그리디 알고리즘보다 효율적으로 랜덤키를 생성하여 에피소드 $SB_t$를 감소시킬 수 있으며, 다양한 서비스 발생 패턴에 대응할 수 있다.

## Ⅲ. 결론

본 논문에서는 QKD 시스템의 랜덤키 생성 최적화를 위한 DRL방법을 제안한다. 학습 결과를 통해 제안하는 기법을 활용할 때 DRL은 서비스의 발생 패턴을 학습할 수 있음을 증명한다. 또한 실험 결과를 통해 그리디 알고리즘 보다 효율적인 랜덤키 생성이 가능하고, 다양한 서비스 발생 패턴에 대응할 수 있음을 보였다.

## ACKNOWLEDGMENT

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2023R1A2C1003143 & No. 2018R1A6A1A03025526)

## 참고문헌

[1] 권오성, 김용수, 한상욱, & 문성욱. (2014). 미래통신 보안기술: 양자암호통신 연구 현황 및 전망. Telecommunications Review, 24(3), 404-418.

[2] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," Proceedings 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, 1994, pp. 124-134

[3] A. I. Nurhadi and N. R. Syambas, "Quantum Key Distribution (QKD) Protocols: A Survey," 2018 4th International Conference on Wireless and Telematics (ICWT), Nusa Dua, Bali, Indonesia, 2018, pp. 1-5

[4] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26-38.

[5] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks.

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997.

[7] ITU-T Recommendation Y.3803. Quantum key distribution networks - Key management. Technical report, 2020.