

Q-xApp: O-RAN 준실시간 최적화를 위한 양자 기반 xApp

이욱진, 송영진, 장용훈, 이상현*
고려대학교

{mekdugi, thd4090, disclose, sanghyunlee}@korea.ac.kr

Q-xApp: A Quantum-Based xApp for Near-Real-Time Optimization in O-RAN

Lee Wookjin, Song Youngjin, Jang Yong Hun, Lee Sang Hyun*
Korea Univ.

요약

본 논문은 양자 컴퓨팅 기반 준실시간 네트워크 최적화 가속을 위한 양자 알고리즘을 설계하고, 이를 O-RAN(Open Radio Access Network) 구조하에서 적용하기 위해 가상화 소프트웨어 형태의 Q-xApp(Quantum xApp)으로 구성한다. 또한 ns-3 기반 O-RAN 시뮬레이션 환경에 Q-xApp을 적용하여 제안 기법의 실현 가능성을 검증한다.

I. 서론

차세대 통신은 초저지연, 고신뢰성, 대규모 기기 접속 등 복합적인 요구를 수용할 수 있는 유연하고 확장 가능한 무선접속망 구조를 필요로 한다. O-RAN(Open Radio Access Network)은 일체화된 전통적 RAN을 기능별로 분리 및 가상화하고 표준화된 인터페이스를 적용하여, 사업자가 다양한 벤더(vendor)의 모듈을 조합해 최적 RAN 구성을 실현하도록 설계되었다. 또한, 다양한 RAN 구성요소를 일관되게 제어하기 위해 별도로 RIC(RAN Intelligent Controller)가 도입되었다 [1]. RIC 구현을 위한 핵심적인 과제는 RAN의 near-RT(Real-Time) 모니터링 및 제어로 이를 위해 xApp이라는 내부 구성 요소를 포함한다. 특히, 미래 통신 환경에서는 통신 지연에 민감한 서비스가 다수를 차지하여 더욱 RAN 제어의 실시간성이 강조되며, xApp은 이를 지원하기 위한 고성능의 연산 체계를 필요로 한다.

양자 컴퓨팅은 양자 연산 단위인 큐비트(qubit)의 중첩 특성을 기반으로 기존 컴퓨팅 대비 지수적인 연산 가속을 구현할 수 있다 [2]. 중첩을 통한 양자 연산은 전용 양자 하드웨어 내에서 특수한 논리 구조를 설계해야 하기에 기존의 일체화된 RAN 구조에서는 활용이 불가능했다. 그러나 기능 분리 및 가상화를 기반으로 하는 O-RAN 구조는, 집약적 연산이 수행되는 xApp에 가상화된 모듈로서 양자 컴퓨팅을 통합할 수 있는 가능성을 제시했다. 한편, 현재 존재하는 양자 알고리즘은 연속적인 파라미터를 다루고, 엄격한 제한 조건을 만족해야 하는 네트워크 최적화 문제의 특성을 고려하지 못한다. 기존의 대표적인 양자 알고리즘인 양자 어닐링, QAOA(Quantum Approximate Optimization Algorithm)는 연속 파라미터를 이산화한 양자 표현으로 변환하고, 제약 조건을 목적함수에 페널티(penalty) 항으로 포함하기 때문에 엄격한 제한 조건 만족을 보장할 수 없다. 따라서 양자 컴퓨팅 기반 xApp 구현을 위해서는 네트워크 최적화의 요구조건을 만족하는 양자 알고리즘의 설계가 필수적이다.

본 논문에서는 네트워크 최적화 문제의 특성을 만족하는 양자 알고리즘을 설계하고, 이를 가상화 소프트웨어 형식의 xApp으로 구현한 Q-xApp(Quantum xApp)을 제안한다. 또한 ns-3 기반의 O-RAN 시뮬레이션 환경에서 Q-xApp을 개발하여 기법의 실현 가능성을 검증한다.

II. 본론

O-RAN은 CU(Central Unit), DU(Distributed Unit), RU(Radio Unit)로 RAN 기능이 분리된 3GPP NR(New Radio) 기반의 NG(Next Generation)-RAN 구조를 따르며, RAN 제어 기능을 분리하기 위한 RIC를 추가하여 가상화 네트워크 관리를 가능하게 한다. RIC는 제어 시간 스케일에 따라 두 계층으로 구성되며, 장기적인 정책 수립을 담당하는 non-RT RIC(1s 초과)와 준실시간 제어를 수행하는 near-RT RIC(10ms~1s)가 이에 해당한다. Near-RT RIC는 여러 개의 독립된 제어 기능 단위인 xApp을 통해 다양한 제어 의사결정을 수집 및 종합한 후, 이를 E2 인터페이스를 통해 CU, DU, RU로 구성된 E2 노드에 전달함으로써 near-RT 제어 루프(loop)를 형성한다. 병렬적으로 운영되는 다수의 최적화 xApp 중 하나라도 제어 주기 내에 연산을 완료하지 못한다면, 이후 루프에서는 이를 반영하지 못한 채로 다른 xApp의 제어가 진행되므로 단일 오류가 전체 네트워크 동작에 영향을 미친다. 따라서, 제어 주기를 준수하는 것은 xApp의 핵심적인 설계 요구사항이다.

O-RAN Work Group 3의 Use cases and Requirements 표준 문서 [3]에서는 near-RT RIC 내부 xApp의 여섯 가지 유즈케이스를 제시하고 있다. 이들 유즈케이스에서의 near-RT 최적화는 요구되는 계산 구조에 따라 다음과 같이 네 가지 유형으로 분류할 수 있다: (1) 제어 파라미터의 범위 내에서 설정값을 조정하는 **정적 파라미터 조정**, (2) 선택지 수가 제한되어 우선순위 또는 정책 규칙 기반으로 결정이 가능한 **제한적 선택 기반 최적화**, (3) 복수의 제어 대상 간 매핑(mapping) 구조를 고려해야 하는 **일반 다대다 매핑 최적화**, (4) AI/ML 기법을 활용하거나 수리적 모델로 정형화하기 어려운 **예측 또는 학습 기반 최적화**이다. 이중 정적 파라미터 조정과 제한적 선택 기반 최적화는 복잡한 연산이 필요하지 않으며, 예측 또는 학습 기반 최적화는 O-RAN에서 AI/ML 기법을 활용하기 위한 계층적 구조를 따로 제시하고 있기 때문에, Q-xApp의 적용 대상으로 일반 다대다 매핑 최적화를 고려한다.

무선 네트워크의 주요 객체 그룹 간 관계 설정은 전체 성능에 큰 영향을 미친다. 다대다 매핑 최적화의 예시로는 BS(Base Station)-UE(User Equipment), 슬라이스(slice)-UE, DRB(Data Radio Bearer)-QFI(QoS Flow Identifier), GoB(Grid-of-Beam)-UE가 있다.

이러한 여러 매핑 최적화 문제의 공통점은 연속적인 목적 함수(objective)와 여러 제한조건(constraints) 하에서 객체 그룹간 할당 변수를 결정하는 조합 최적화 문제로 수식화할 수 있다는 것이다. 대표적으로, 트래픽 스티어링 (traffic steering) 유즈케이스에서의 BS-UE 할당 문제 **P1**은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \mathbf{P1}: \max_{x_{bu}} & \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} R_{bu} x_{bu} \\ \text{subject to} & \sum_b x_{bu} = 1, \forall u \in \mathcal{U} \\ & \sum_u x_{bu} \leq C_b, \forall b \in \mathcal{B} \\ & x_{bu} \in \{0,1\}, \forall (b,u) \in (\mathcal{B}, \mathcal{U}) \end{aligned}$$

P1은 BS 집합 \mathcal{B} 의 원소 b 와 UE 집합 \mathcal{U} 의 원소 u 간의 할당 이진변수 x_{bu} 를 결정하는 문제이다. b 와 u 가 할당될 때의 데이터 전송량 R_{bu} 의 전체 합을 최대화하는 연속 목적함수와 한 UE는 한 BS만 선택해야 하고, 각 BS b 는 최대 C_b 개의 UE를 수용할 수 있는 트래픽 제한조건을 가진다.

최적화 문제 **P1**을 해결하기 위한 양자 알고리즘을 설계한다. Q-xApp은 해당 양자 알고리즘이 소프트웨어화된 구조로, 지역(regional) 양자 하드웨어에서 제어 루프 내에 작동한다. C_b 는 non-RT RIC에서 결정되고 R_{bu} 는 E2 노드에서 측정되어 near-RT RIC 내부 Q-xApp으로 전달된다. Q-xApp은 near-RT 제어 루프 내에 최적화 문제를 풀어 x_{bu} 를 결정하여 BS-UE 할당 결과를 E2 노드로 전달한다.

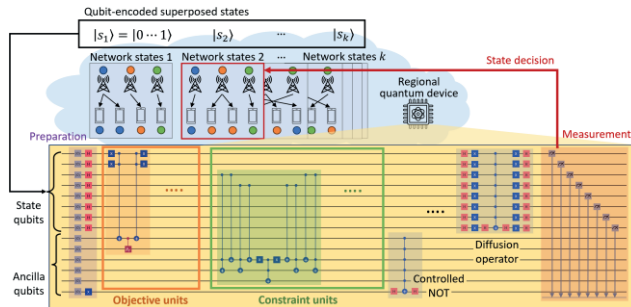


그림 1. Q-xApp 내 양자 알고리즘의 구조

그림 1은 Q-xApp의 양자 알고리즘의 구조를 도시한다. 양자 알고리즘은 모든 BS-UE 간 할당 상태를 상태 큐비트(state qubits)의 배열로 표현하며, 설계된 양자 게이트를 모두 통과하면 문제 **P1**을 만족하는 할당 상태의 양자 검출 확률이 가장 높도록 한다. 준비 단계(preparation)에서 Hadamard 게이트를 통해 모든 가능한 할당 상태가 동일한 확률을 갖도록 한다. 그 후, 작은 단위의 여러 목적함수(objective units) 및 제한조건(constraint units) 양자 게이트가 존재하여 조건 만족 여부는 보조 큐비트(ancilla qubits)에 차례로 만족할 시 1, 아닐 시 0으로 기록된다. 모든 보조 큐비트가 1임을 검사하는 controlled NOT 게이트를 통해 모든 조건이 만족된 것이 확인된 할당 상태는, 확산 연산자(diffusion operator)에서 검출 확률이 반복적으로 증폭된다. 이는 N 개의 가능한 할당 상태가 존재할 때 $O(\sqrt{N})$ 의 연산복잡도가 요구되어 고전 연산 기반 $O(N)$ 보다 상당히 작음을 확인할 수 있다. 최종 측정(measurement) 단계에서 할당 결과를 결정한다.

목적함수 및 제한조건 양자 게이트의 자세한 작동원리는 다음과 같다. 목적함수 양자 게이트는 각

BS-UE 쌍에 대해 정규화된 R_{bu} 값이 보조 큐비트가 1이 될 확률의 형식으로 반영된다. BS-UE 할당 결과에 해당하는 모든 BS-UE 쌍의 R_{bu} 값이 확률의 형식으로 더해지게 되어 최종적으로 총 데이터 전송량이 높은 할당 결과의 양자 검출 확률이 제일 높아지게 된다. 해당 구조를 통해 기존 양자 알고리즘이 다룰 수 없었던 연속 값을 검출 확률로 반영할 수 있다. 또한, 제한조건 양자 게이트는 BS-UE 쌍 사이의 상태 큐비트 일치를 검사하는 원리로 서로 다른 BS가 같은 UE를 선택하거나 트래픽 제한조건을 위반한 BS-UE 할당 결과를 아예 확률적으로 검출할 수 없도록 제외시킨다.

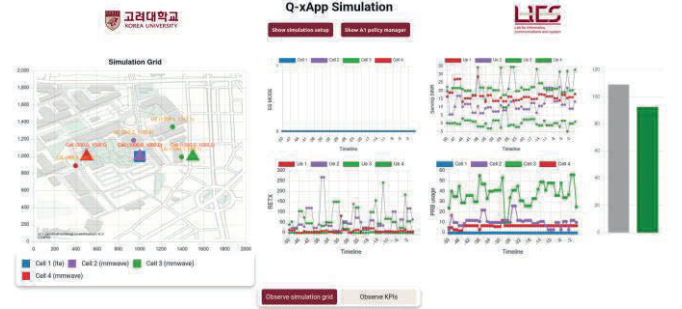


그림 2. ns-3를 통한 Q-xApp 시뮬레이션 플랫폼

그림 2는 네트워크 시뮬레이터 ns-3 상에 O-RAN 환경을 구성하여, 제안한 Q-xApp의 동작을 검증하기 위한 시뮬레이션 플랫폼이다. BS 세 개, UE 네 개가 존재하는 네트워크를 랜덤 생성하여 1,000번 반복 최적화한 결과 제안하는 알고리즘이 최적 할당의 데이터 전송률 총합 대비 1% 미만의 성능 격차를 낼 확률은 91%로 이산 목적함수 양자 알고리즘 [2]의 57.3%보다 높은 결과를 보였다.

III. 결론

본 논문에서는 양자 컴퓨팅을 통해 O-RAN 구조 내의 near-RT 네트워크 최적화를 수행하기 위한 Q-xApp을 제안하였다. 또한, ns-3 O-RAN 시뮬레이션 환경에서 Q-xApp을 구현하여 기존 양자 알고리즘 대비 네트워크 최적화 성능 우위를 실증하였다.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2021-II210467, Intelligent 6G Wireless Access System, and No.RS-2024-00397216, Development of the Upper-mid Band Extreme massive MIMO (E-MIMO)).

참고 문헌

- [1] M. Polese et al., "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," IEEE Commun. Surv. Tut., vol. 25, no. 2, pp. 1376–1411, May 2023.
- [2] Y. Seo, Y. Kang, and J. Heo, "Quantum search algorithm for weighted solutions," IEEE Access, vol. 10, pp. 16209–16224, Feb. 2022.
- [3] O-RAN Working Group 3, "Use Cases and Requirements," O-RAN Alliance, Alfter, Germany, document O-RAN.WG3.TS.UCR-R004-v08.00, Feb. 2025.