

그래프 신경망 기반 Virtual Network Function 배치 최적화 방법

이재서, 이훈
울산과학기술원 전기전자공학과

leejss@unist.ac.kr, hoonlee@unist.ac.kr

Graph Neural Network Methods for Place Optimization of Virtual Network Functions

Jaeseo Lee, Hoon Lee

Dept. Electrical Eng., Ulsan National Institute of Science and Technology

요약

본 논문은 virtualized radio access network (vRAN) 시스템에서 virtual network function (VNF) 배치를 효율적으로 수행하기 위한 그래프 신경망 기반의 최적화 기법을 제안한다. vRAN 을 이진 그래프로 표현하고, 각 노드와 링크 간 상호작용을 효과적으로 학습하도록 모델을 설계한다. 모의실험을 통해 제안 기법의 효율성을 입증한다.

I. 서론

최근 급증하는 무선 트래픽을 효과적으로 처리하기 위해 네트워크 가상화를 수행하는 virtualized radio access network (vRAN) 시스템이 주목받고 있다 [1]. vRAN에서는 virtual network function (VNF)를 각 노드에 적절히 배치하여 비용을 최소화할 수 있다. 최적의 VNF 배치 솔루션을 전통적인 최적화 알고리즘으로 도출할 수 있지만, 계산 복잡도가 높아 대규모 네트워크나 실시간 환경에 적용하기 어렵다.

한편, 최근 무선 네트워크 자원 관리 분야에서 graph neural network (GNN)이 활발히 사용되고 있다[2]. 특히, 그래프의 노드와 간선을 통합적으로 활용하는 edge-node GNN (ENGNN)은 종래 GNN 모델의 성능을 크게 개선하였다[3].

본 논문에서는 ENGNN 을 활용하여 vRAN 시스템의 비용 최소화를 위한 VNF 배치 방식을 제안한다. 모의실험을 통해 제안 방법의 우수성을 성능과 계산 효율 측면에서 입증한다.

II. 시스템 모델

N 개의 distributed unit (DU) 및 M 개의 cloud unit (CU)로 구성된 vRAN 시스템을 고려한다. $DU\ n \in \mathcal{N} = \{1, \dots, N\}$ 은 radio unit (RU)로부터 수신한 data traffic λ_n 을 처리해야 한다. 그러나 제한된 연산 능력을 지닌 DU 만으로는 모든 네트워크 traffic 을 해결하기 어렵다. 이를 위해 DU는 할당된 network function을 가상화하여 세 개의 VNF f_1 (PHY), f_2 (MAC), f_3 (PDCP)로 분할하고 CU와 협력적으로 처리한다.

본 논문에서는 VNF의 할당 전략을 최적화하여 vRAN 시스템의 에너지 소모량과 지연시간을 동시에 최소화한다. 이진 변수 x_{1n}, x_{2n}, x_{3n} 을 각각 DU n 에 f_1, f_2, f_3 을 할당하는 상태로 정의한다. 또한, 이진 변수 $y_{1nm}, y_{2nm}, y_{3nm}$ 은 각각 CU $m \in \mathcal{M} = \{1, \dots, M\}$ 이 DU

n 의 f_1, f_2, f_3 을 처리하는 상태를 뜻한다. 그러면 DU n 이 소모하는 에너지는 다음과 같이 정의된다.

$$V_n = \kappa_n \alpha_n \sum_{f=1}^3 x_{fn} + \kappa_n \beta_n \lambda_n \sum_{f=1}^3 \rho_f x_{fn}, \quad (1)$$

여기서 κ_n [Joule/cycle]은 단위 CPU cycle 연산에 필요한 에너지, α_n 은 인스턴스 생성 비용, β_n 은 traffic에 선형적으로 비례하는 연산 비용, ρ_f [cycles/bit]은 VNF f 를 처리하는데 필요한 단위 비트 당 CPU cycle 수를 의미한다. CU m 의 에너지 소모량은 다음과 같다.

$$V_m = \kappa_m \alpha_m \sum_{n \in \mathcal{N}} \sum_{f=1}^3 y_{fnm} + \kappa_m \beta_m \sum_{n \in \mathcal{N}} \lambda_n \sum_{f=1}^3 \rho_f y_{fnm}, \quad (2)$$

한편, DU n 은 할당된 VNF를 처리한 후 남은 traffic을 CU m 과 연결된 midhaul (n, m) 을 통해 전송한다. DU n 이 전송해야 하는 traffic은 아래와 같이 표현된다[1].

$$S_n = x_{1n}(1.02\lambda_n + 1.5) - x_{2n}(0.02\lambda_n + 1.5) + 2500(1 - x_{1n})$$

Midhaul (n, m) 의 채널 용량을 c_{nm} [bps]로 정의하면, 해당 midhaul의 지연시간은 $T_{nm} = S_n / c_{nm}$ 로 계산된다.

vRAN system의 에너지 소모량과 지연시간을 동시에 최소화하는 다목적 문제를 다음과 같이 공식화한다.

$$\begin{aligned} \min_{x, y} \quad & W \left(\sum_{n \in \mathcal{N}} V_n + \sum_{m \in \mathcal{M}} V_m \right) + (1 - W) \sum_{n \in \mathcal{N}} \sum_{m \in \mathcal{M}} T_{nm}, \quad (3) \\ \text{s.t.} \quad & x_{(f+1)n} \leq x_{fn}, \quad y_{fnm} \leq y_{(f+1)nm}, \\ & x_{fn} + \sum_{m \in \mathcal{M}} y_{fnm} = 1 \end{aligned}$$

여기서 $\mathbf{x} = \{x_{fn} : \forall f, n\}$, $\mathbf{y} = \{y_{fnm} : \forall f, n, m\}$, $W \in [0, 1]$ 은 상수 weight 이다. 위 문제는 mixed integer linear programming (MILP)로 상용 소프트웨어를 통해 최적 솔루션을 도출할 수 있으나, 네트워크 크기에 비례하여 연산 복잡도가 증가하여 비효율적이다.

III. 제안하는 ENGNN 기반 VNF 배치 최적화 방법

VNF 배치 문제를 효과적으로 계산하기 위해 ENGNN 모델을 도입한다. vRAN을 DU 노드 \mathcal{N} , CU 노드 \mathcal{M} , 그리고 midhaul 간선 $\mathcal{E} = \{(n, m) : \forall n, m\}$ 으로 구성된 이진 그래프로 모형화 한다[4]. DU n 은 λ_n 을 입력 특성으로 사용하여 x_{fn} 을 추론한다. Midhaul 간선 (n, m) 은 입력 특성 c_{nm} 을 기반으로 y_{fnm} 을 결정한다.

ENGNN의 학습 과정에서 gradient vanishing 문제를 방지하기 위해 이진 변수 x_{fn}, y_{fnm} 를 0과 1 사이의 실수로 근사하는 relaxation 방법을 사용한다. 또한, 제한 조건을 만족하기 위해 적절한 activation 함수를 설계하여 ENGNN이 항상 feasible 솔루션을 도출하도록 한다. ENGNN은 문제 (3)의 목적함수를 최소화하도록 훈련한다. 학습 종료 후 다양한 크기의 vRAN에 바로 확장 적용이 가능하다.

IV. 모의실험 결과

모의실험을 통해 제안하는 ENGNN 방법의 성능을 MOSEK solver 로 계산한 최적 솔루션과 비교하여 유효성을 검증한다. vRAN 시스템을 각 DU 가 단일 CU 와 연결된 무작위 이진 그래프로 생성한다. Traffic λ_n 및 midhaul 용량 c_{nm} 은 각각 50 – 150 [Mbps], 50 – 500 [Gbps] 범위에서 균등 분포로 생성한다. ENGNN 은 $N = 10, M = 5$ 상황에서 학습한다.

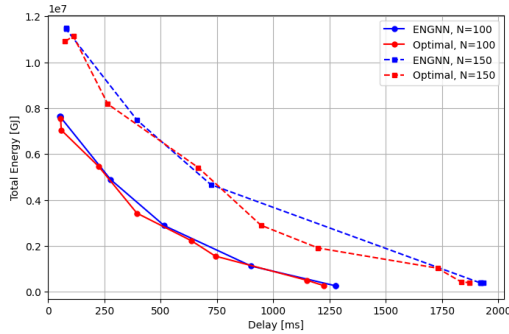


그림 1. 에너지-지연시간 tradeoff

그림 1은 vRAN의 에너지-지연시간 tradeoff를 나타낸다. ENGNN은 10개의 DU로 구성된 작은 시스템에서 학습되었으나, 100개 및 150개의 DU를 갖는 상황에서도 최적 기법과 유사한 성능을 달성한다. 이를 통해 제안하는 기법의 확장성을 검증할 수 있다.

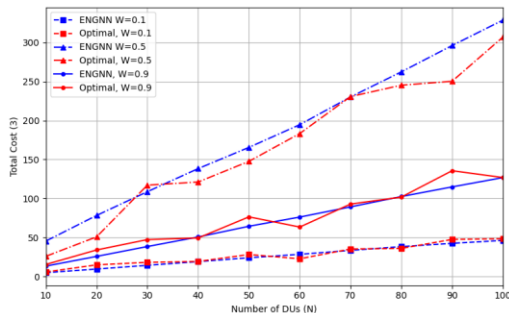


그림 2. DU 수에 따른 total cost (3) 성능

그림 2는 vRAN의 total cost를 DU 수 N 을 증가시키며 비교한 결과이다. 다양한 weight W 에 대해서 제안하는 ENGNN 기법은 모두 전역 최적 성능에 가까운 성능을 도출하여 그 유효성을 입증한다.

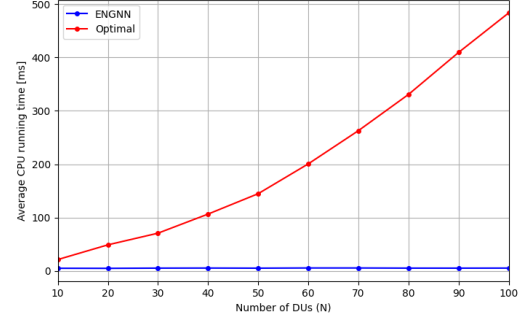


그림 3. CPU 실행 시간 비교

그림 3은 DU 수 N 에 따른 평균 CPU 연산 시간을 비교한다. ENGNN은 DU 및 CU에 대한 병렬 연산이 가능하므로 N 에 무관한 계산 복잡도를 보인다. 반면에, 기존 최적화 알고리즘은 N 이 커질수록 연산 시간이 가파르게 증가한다. 두 방법의 성능이 유사한 것을 고려하면, 제안하는 기법이 연산 효율적임을 알 수 있다.

V. 결론

본 논문은 vRAN의 최적 VNF 배치 문제를 효과적으로 해결하기 위한 ENGNN 기법을 제안한다. 모의실험 결과 제안하는 기법이 전역 최적 방식의 성능을 달성하면서도 계산 복잡도를 크게 낮추는 것을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 (2021-0-00467, 지능형 6G 무선 액세스 시스템; RS-2023-00225468, 오픈랜 지능화를 위한 무선지능화 제어기술 개발)의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] F. W. Murti, J. A. Ayala-Romero, A. Garcia-Saavedra, X. Costa-Perez, and G. Iosifidis, "An optimal deployment framework for multi-cloud virtualized radio access networks", *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2251–2265, Apr. 2021.
- [2] Y. Peng, J. Guo, and C. Yang, "Learning resource allocation policy: Vertex-GNN or edge-GNN?" *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 190–209, Jan. 2024.
- [3] Y. Wang, Y. Li, Q. Shi, and Y.-C. Wu, "ENGNN: A general edge update empowered GNN architecture for radio resource management in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 5330–5344, Jun. 2024.
- [4] J. Kim, H. Lee, S. Hong, and S. Park, "A bipartite graph neural network approach for scalable beamforming optimization", *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 333–347, Jan. 2023.