

DSPy 프레임워크를 활용한 LLM - xApp 연동 기반 ns-3 FlexRIC RAN 제어 파이프라인 구조 설계

최재형, 정태일, 홍인기

경희대학교

wogud1221@khu.ac.kr, wjdxodlf012345@khu.ac.kr, ekhong@khu.ac.kr

Design of an ns-3 FlexRIC RAN Control Pipeline Based on LLM - xApp Integration Using the DSPy Framework

Jaehyung Choi, Taeil Jung, Een Kee Hong

Kyunghee University

요약

본 논문은 O-RAN의 Near Real-Time RIC 환경에서 대규모 언어 모델(LLM)을 활용하여 RAN 제어를 간소화하고 동적으로 수행할 수 있는 파이프라인을 제안한다. 제안하는 구조에서는 KPI 모니터링 xApp이 무선 네트워크의 성능 지표를 수집하고, LLM이 이를 해석하여 제어 규칙 파일을 자동으로 갱신한다. 이후 규칙 기반 핸드오버 xApp은 해당 파일의 변화를 실시간으로 감지하여 핸드오버 동작을 수행한다. 이러한 방식은 기존 ns-3 시뮬레이터 기반 xApp 구현에서 코드 수정 시마다 재빌드가 필요한 비효율성을 해소하며, LLM을 통한 동적이고 자율적인 RAN 제어의 가능성을 제시한다.

I. 서론

5G/6G로의 진화는 급변하는 무선 환경에 대한 자율적이고 민첩한 제어를 요구한다. 이를 위해 O-RAN의 Near-Real-Time RIC은 xApp 구조를 통해 네트워크 제어 기능을 개방하였으나, 실제 실험 환경에서는 정책 변경이 곧 코드 수정과 재빌드·재배포로 이어져 실험의 반복성과 유연성이 크게 저하되는 문제가 있다. 특히 ns-3 기반 시뮬레이션에서는 정책 수정 후 매번 재컴파일과 재실행이 필요하여 연구 효율성이 제한된다[1],[2].

본 연구에서는 이러한 문제를 해결하기 위해 LLM 기반의 동적 핸드오버 제어 구조를 제안한다. 제안된 파이프라인은 모니터링과 제어 기능을 규칙파일(rule file)이라는 텍스트 인터페이스로 연결하고, LLM이 실시간으로 KPI를 해석하여 해당 파일을 갱신함으로써 xApp의 동작을 직접 제어한다. 구체적으로, KPM 모니터링 xApp(xapp_kpm_moni)이 KPI를 수집·노출하면, LLM(Gemini-2.5-Flash)이 DSPy 프레임워크를 통해 KPI를 분석하고, 그 결과를 규칙 파일에 기록한다. 이후 규칙 기반 핸드오버 xApp은 파일의 변화를 감지하여 즉시 핸드오버를 수행하며, 파일 잠금(lock), 변경 해시 감지, 시작 지연, IMSI별 쿨다운 등의 안전장치를 통해 초기 불안정 제어나 펄스 현상을 방지한다[3].

따라서 본 구조는 정책 변경을 코드 수정이나 재빌드 과정 없이 반영할 수 있어 실험 효율성을 높인다. 또한 LLM(DSPy)을 활용한 정책 결정 과정을 RIC/xApp 코드와 분리함으로써, 제어 로직의 수정 없이 다양한 시나리오 실험이 가능하다. 기존 O-RAN 스택과의 호환성을 유지한 채 ns-3 환경에서 KPI - LLM - 규칙파일 - 제어로 이어지는 경량 페루프가 안정적으로 동작함을 확인하였으며, 이를 통해 LLM 기반 RAN 제어의 실용적 구현 가능성을 검증하였다.

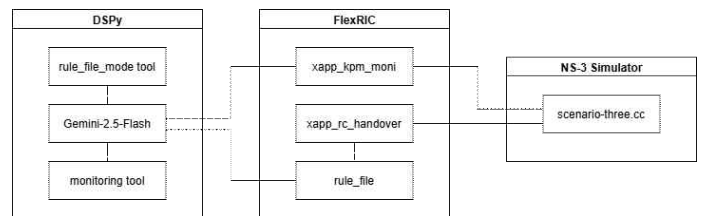
II. 본론

1. LLM 기반 RAN 제어 배경

5G/6G 네트워크는 트래픽과 무선 환경이 빠르게 변화하므로, 제어의 자율성과 적응성이 필수적이다. 이러한 요구에 대응하기 위해 O-RAN 구조는 Near-RT RIC과 xApp을 통해 제어 기능을 개방하고 있으며, 이를 기반으로 한 AI-RAN 개념이 부상하고 있다. AI-RAN은 인공지능을 활용하여 네트워크 상태를 학습하고, 자율적으로 정책을 생성·적용하는 지능형 RAN을 지향한다. 그러나 기존 xApp 기반 제어는 정책 로직이 코드

내부에 고정되어 있어, 실험 환경이나 시뮬레이터(ns-3 등)에서 정책을 자주 변경하거나 비교 실험을 수행하기에는 유연성이 부족하다. 본 연구는 이러한 한계를 완화하기 위해 LLM을 RIC 제어 파이프라인에 도입하여, KPI를 해석하고 규칙파일을 통해 RAN을 동적으로 제어하는 방안을 제시한다[4],[5],[6].

2. 구조 및 구현



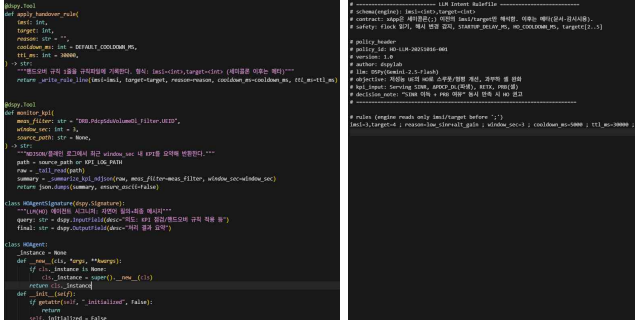
<그림1. LLM과 FlexRIC xApp의 연동 구조도>

제안된 시스템은 DSPy, FlexRIC, 그리고 NS-3 시뮬레이터의 세 모듈로 구성된다. 그림1은 전체 파이프라인의 구조를 나타낸다. DSPy 모듈은 LLM(Gemini-2.5-Flash)을 중심으로 구성되며, KPI를 기반으로 의사결정 결과를 생성하는 역할을 수행한다. 이때 DSPy 내부의 rule_file_mode tool은 LLM 출력을 규칙파일 형식으로 변환하고, monitoring tool은 KPI 모니터링 결과를 LLM 입력으로 전달한다[7].

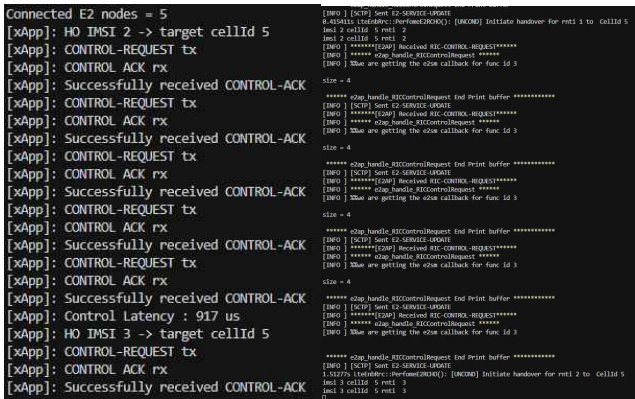
FlexRIC 모듈은 Near-RT RIC 환경을 가정한 제어 플랫폼이며, 두 개의 xApp을 사용했다. 그 중, xapp_kpm_moni는 KPM V3 포맷을 통해 셀 및 UE 단위의 KPI를 수집하며, xapp_rc_handover는 규칙파일을 주기적으로 감지하여 파일 내 명령을 파싱하고 핸드오버 메시지를 수행한다[8].

NS-3 Simulator는 RAN 환경을 모델링하며, FlexRIC에서 제공하는 예제인 scenario-three.cc를 통해 UE 이동, 셀 부하, 핸드오버 이벤트 등을 시뮬레이션한다. FlexRIC은 NS-3와 E2 인터페이스를 통해 연결되어 실시간 KPI를 전달받고, LLM의 결정에 따라 핸드오버 명령을 반영한다. 이러한 구조를 통해 KPI → LLM → 규칙파일 → xApp → NS-3로 이어지는 경량 제어 루프가 형성된다.

3. 결과



〈그림2. DSPy 프레임워크를 이용한 LLM 구조 및 규칙 파일〉



〈그림3. xapp_rc_handover의 제어 메시지와 ns3 시뮬레이션 반영 결과〉

그림 2는 DSPy 프레임워크를 기반으로 구현한 LLM 제어 구조와 규칙 파일을 보여준다. 제안된 구조는 LLM이 직접 KPI를 해석하여 제어 결정을 내리고, 이를 단순한 텍스트 형태의 규칙파일로 변환하는 경량 구조를 따른다. 규칙파일의 내용은 최소화되어 있으며, 대부분의 정책 결정 로직은 LLM 내부에서 수행된다. 이러한 단순화는 xApp이 불필요한 파싱 로직이나 복잡한 조건문 없이, 파일 변경만을 감지해 즉시 제어 명령을 수행하도록 설계하기 위함이다. DSPy 내부에는 두 가지 핵심 툴이 정의되어 있는데, monitoring tool은 KPI 수집 결과를 주기적으로 읽어 LLM 입력으로 전달하며, rule_file_mode tool은 LLM의 출력(제어 의도)을 규칙파일 형식으로 변환·기록하는 기능을 수행한다.

그림 3은 제안된 구조가 실제로 동작하는 결과를 보여준다. 로그 상에서 xApp(xapp_rc_handover)이 “CONTROL-REQUEST” 메시지를 생성해 RIC을 통해 ns-3 시뮬레이터로 전달하고, “Received RIC-CONTROL-REQUEST”를 통해 정상적으로 수신하는 과정을 확인할 수 있다. 이는 xApp에서 생성된 핸드오버 요청이 ns-3 시뮬레이션 환경에 정상적으로 반영되어 셀 전환이 수행됨을 의미한다. 또한 로그 마지막 부분에서 볼 수 있듯이, LLM이 새로운 규칙을 생성해 규칙파일을 갱신했을 때 시스템이 중단 없이 이를 감지하고 곧바로 다음 제어 루프를 실행하였다. 이 과정에서 재빌드나 재배포 없이 즉각적인 제어 정책 변경이 가능했으며, LLM 출력이 반영되는 평균 지연은 약 0.9ms 수준으로 측정되었다.

이 결과는 제안된 DSPy - FlexRIC - ns3 구조가 안정적인 실시간 동작을 수행함을 보여준다. 제어 루프는 “KPI 입력 → LLM 판단 → 규칙파일 갱신 → xApp 제어 → ns-3 반영”으로 폐루프 형태를 이룬다. 특히 LLM이 생성한 규칙은 DSPy 내부에서 검증된 형식으로만 기록되므로, 파일 파싱 오류나 잘못된 제어 명령으로 인한 충돌이 발생하지 않았다.

III. 결론

본 연구에서는 O-RAN의 Near-RT RIC 환경에서 LLM을 활용해 동적으로 RAN을 제어할 수 있는 경량 파이프라인을 제안하였다. 제안된 구조

는 DSPy 프레임워크를 활용한 LLM이 KPI를 해석해 규칙파일을 갱신하고, xApp이 이를 실시간으로 반영함으로써 코드 수정이나 재배포 없이 정책을 즉시 적용할 수 있음을 보였다. ns-3 기반 실험을 통해 LLM의 판단이 중단 없이 제어 루프에 반영되고, 핸드오버 명령이 정상적으로 수행됨을 확인하였다. 본 연구는 LLM을 RIC 제어 파이프라인에 통합하여 RAN 제어의 유연성과 실시간성을 향상시킬 수 있음을 보여주었으며, 향후에는 다중 KPI 기반 의사결정과 도메인 특화 파인튜닝된 LLM 모델을 결합해 보다 고도화된 AI-RAN 제어로 확장할 예정이다[9].

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원(IITP-2025-RS-2021-II21046, 50%)과 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2024-00397520, 5G 버티컬 서비스를 위한 가상화 기반 경량형 5G 올인원 시스템 개발, 50%)

참 고 문 헌

- [1] T. Jung, J. Choi, and E. K. Hong, “Design and implementation of an xApp integrating FlexRIC with the ns-O-RAN Framework,” Proc. KICS Summer Conference, 2025.
- [2] A. Garcia-Saavedra, M. Gramaglia, X. Costa-Pérez, and V. Sciancalepore, “ns-O-RAN: Simulating O-RAN 5G Systems in ns-3,” arXiv:2305.06906, 2023.
- [3] O. Giwa, M. Adewole, T. Awodumila, and P. Aderinto, “The LLM as a Network Operator: A Vision for Generative AI in the 6G Radio Access Network,” arXiv:2509.10478, 2025.
- [4] L. Bao, S. Yun, J. Lee, and T. Q. S. Quek, “LLM-Guided Open RAN: Empowering Hierarchical RAN Intelligent Control,” arXiv:2504.18062, 2025.
- [5] X. Wu, J. Farooq, Y. Wang, and J. Chen, “LLM-xApp: A Large Language Model Empowered Radio Resource Management xApp for 5G O-RAN,” Proc. FutureG Workshop on Security & Privacy of Next-Generation Networks, 2025.
- [6] B. Wei, R. Jiang, R. Zhang, Y. Liu, D. Niyato, Y. Sun, Y. Lu, Y. Li, S. Mao, C. Yuen, M. Di Renzo, and M. Peng, “Large Language Models for Next-Generation Wireless Network Management: A Survey and Tutorial,” arXiv:2509.05946, 2025.
- [7] O. Khatlab, A. Singhvi, P. Maheshwari, Z. Zhang, K. Santhanam, S. Vardhamanan, S. Haq, A. Sharma, T. T. Joshi, H. Moazam, H. Miller, M. Zaharia, and C. Potts, “DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines,” arXiv:2310.03714, 2023.
- [8] P. Baumgartner, L. Suresh, M. Irazabal, M. Gramaglia, A. Garcia-Saavedra, and X. Costa-Pérez, “FlexRIC: An SDK for Next-Generation SD-RANs,” in Proc. of the 2021 ACM CoNEXT Conference, pp. 449 - 462, 2021.
- [9] F. A. Bimo, M. A. C. Galdon, C.-K. Lai, R.-G. Cheng, and E. K. P. Chong, “Intent-Based Network for RAN Management with Large Language Models,” arXiv:2507.14230, 2025