

# 차세대 5G 네트워크에서 LLM(BERT)을 활용한 Network Slicing

서동준, 김기천\*

건국대학교

ggaebi99@konkuk.ac.kr, \*kckim@konkuk.ac.kr

## Network Sliding Using LLM(BERT) in Next Generation 5G Networks

Seo Dong Jun, Kim Kee Cheon\*

Konkuk Univ.

### 요약

본 논문은 차세대 5G, 6G 환경에서 급속도로 증가하는 네트워크 트래픽을 처리하는 것에 현재 주목도가 높은 분야인 생성형 AI, 중에서도 응용도가 높은 LLM을 사용하여 해결하기 위하여 보안, 인텔리 기반에서의 LLM 사용 최신 연구 동향을 조사 및 분석하였다. LLM의 높은 이해력은 네트워크 트래픽 패킷을 이해할 수 있고, 그에 따라 네트워크 환경을 설계하는 방법은 신뢰성 및 비용적 측면에서 효과적이다. 네트워크 환경에서의 특수성을 고려하여 모델과 방법을 선택했을 때, GPT를 사용한 방법에는 비용적, 시간적 소모가 커 적합하지 않고, BERT 모델을 사용하여 트래픽 패킷의 임베딩을 활용하는 것은 효과적인 것을 확인했다. 두 가지 사실을 조합하여 BERT 아키텍처를 활용한 네트워크 트래픽 패킷에 대한 Network Slicing 자동 할당을 제안한다.

### I. 서론

2022년 그림 인공지능(DALL-E 2)의 등장으로 성능 및 활용도가 높아져 생성형 인공지능의 주목도가 높아졌고, 최근에는 Sora같은 Text-To-Video 생성형 인공지능이 나오는 등 생성형 인공지능이 유망한 연구 주제로 대두되고 있다. 그중에서도 LLM은 OpenAI의 GPT[1]와 구글의 BERT [2] 메타의 LLaMA[3] 등이 대표적으로 많은 분야에서 활용되고 있다. GPT는 버전 4가 현재는 최신이나 올해 7월에 출시될 것이라는 소식이 있고, LLaMA는 올해 4월 18일 버전 3을 출시하였다.

이렇게 LLM의 발전이 진행됨에 따라, 비용적 문제와 더 효과적인 사용을 위해 사용하는 방법에 따라 프롬프트 엔지니어링, 퓨샷러닝, 생성증강 검색(RAG) 등 많은 기법으로 활용되고 있고, 이에 따른 연구 또한 이미 많은 성취를 내고 있다. 또한 IoT 기술 및 5G의 탄력적인 성장이 입증과 현대 사회의 디지털 서비스 의존도의 증가 등의 이유로 네트워크 트래픽이 급속도로 증가하고 있다. 클라우드 기업인 클라우드 플레이어는 2023년 전 세계 트래픽이 전년도에 비해 25% 증가했다는 보고서를 제출하였고, 에릭슨 모빌리티라는 회사에서 2029년까지 네트워크 성장을 예측하고 성장이 계속될 것이라는 보고서를 제출하였다.

따라서 이 논문에서는 증가하는 네트워크 트래픽을 효과적으로 처리하기 위하여 현재 다양한 분야에서 효과를 입증한 LLM을 사용하여 해결하고자 한다. 그러나 국내에는 5G 및 6G 등 차세대 네트워크 환경에서의 LLM 사용에 대한 연구가 드물어, 국외의 논문 중 차세대 네트워크에서의 사용법에 대해 조사 및 분석한다. 분석한 결과 GPT는 비용적 한계가 있어 오픈소스 모델인 BERT 아키텍처의 작업에 대한 이해 능력 및 속도와 비용적 이점, Network Slicing의 네트워크 관리 능력을 사용하여 네트워크 관리를 자동화할 것을 제안한다.

### II. 관련연구

[4]에서는 IoT 네트워크에서 사이버 보안 공격의 빈도와 다양성이 증가

함에 따라 네트워크 기반 공격을 자율적으로 식별할 수 있는 효율적인 기술로 LLM을 사용한 방법을 제안한다. 기존의 ML, DL 방식을 뛰어넘는 정확도를 보였으며, 실사용을 위한 속도 또한 확인하였다.

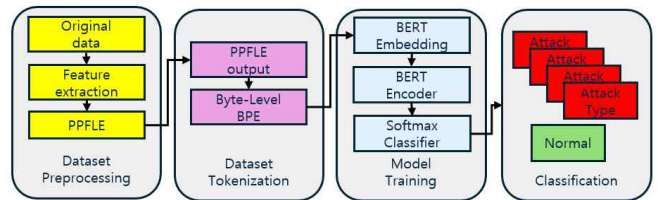


그림 1 SecurityBERT흐름도

BERT 아키텍처를 활용하여 사이버 위협을 효과적으로 탐지하는 SecurityBERT를 제안한다. SecurityBERT는 네트워크 패킷이 주어지면 PPFL E(Privacy-Preserving Fixed-Length Encoding)를 이용하여 전처리하고 BBPE(Byte level Byte Pair Encoding) 방법으로 토큰화하여 OOV(Out of Vocabulary)문제에 효율적으로 대처한다. 앞서 설명한 방법으로 얻은 결과를 토대로 BERT 모델에 Pre-training을 진행하고 정교화하기 위하여 Fine-Tuning 또한 진행한다. 그림[1]은 앞서 설명한 SecurityBERT의 흐름도이다.

AI type	AI Model	Accuracy
Traditional ML	KNN	93.78%
Deep learning models	CNN-LSTM	97.14%
Large Language model	SecurityBERT	98.20%

표 1. SecurityBERT와 기존 방법들 비교

[표 1]과 같이 SecurityBERT의 정확도는 98.2%로 정통적인 방법 중 가장 높은 정확도를 보이는 KNN의 93.78%와 Deep learning 방법 중 가장 높은 정확도를 보이는 CNN-LSTM의 97.14%보다 높은 정확도로 PPFL E와 BERT 아키텍처를 사용하는 것이 유의미함을 보여준다.

[5,6]은 5G 네트워크 기술 중 물리적인 하나의 네트워크를 효율적으로

사용하기 위해 논리적인 여러 개의 네트워크로 분할하여 사용자, 네트워크 서비스 제공자의 의도에 따라 네트워크 리소스를 할당하는 방법인 네트워크 슬라이싱 기술과 LLM의 의도 분류 기술을 사용하여 네트워크 자동화를 제안한다. 네트워크 슬라이싱과 LLM의 의도 분류 기술의 결합이 비용적 절감을 끌어내고, 제로 네트워크를 달성할 수 있는 방법이라는 것을 입증하였다. 그림[2]는 LLM을 사용한 인텐트 기반 자동화 네트워크의 흐름도를 보여준다.

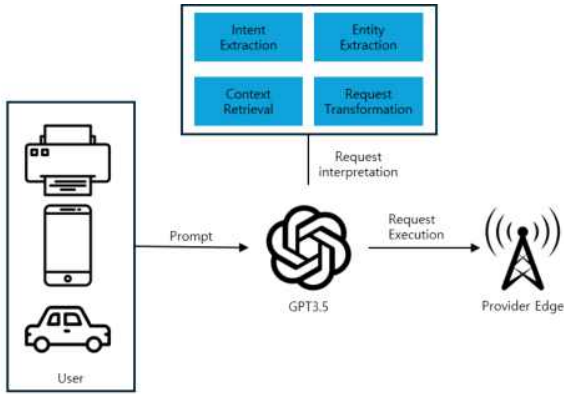


그림 2. 인텐트 기반 자동화 네트워크 흐름도

GPT를 사용하여 진행한 결과 속도에 있어서는 강점을 드러내지는 못하며, 이를 해결하는 것을 네트워크 환경에서 LLM을 사용하는 데에 중요한 과제로 제시한다.

### III. BERT를 활용한 Network Slicing 자동화

본 논문에서는 BERT 아키텍처를 사용한 Network slicing의 자동화를 제안한다. [그림 3]은 BERT를 사용하는 Network slicing 자동 할당에 대한 구성 및 설계를 보여준다.

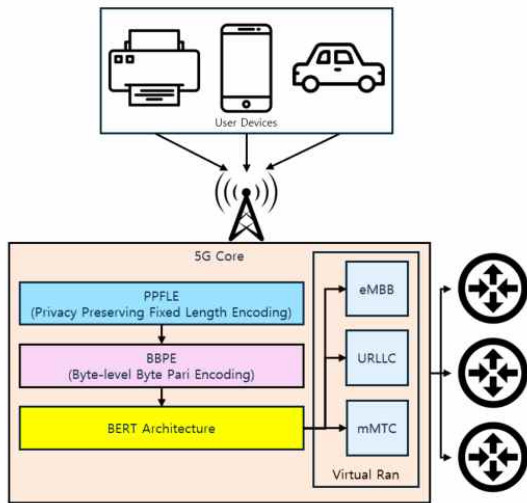


그림 3. BERT를 통한 Network Slicing

User Devices에서 패킷을 보내게 되면 5G Radio를 통해서 들어오게 되고 5G Core는 Network Slice Controller 역할을 수행한다. 들어온 패킷은 PPFLE, BBPE 전처리를 통하여 토큰화되고 BERT 아키텍처에서 서비스 제공자가 고려해야하는 정책과 패킷의 특성에 따라 각각의 필요한 가상 망에 할당하여준다.

#### 3-1. BERT 활용 근거

BERT는 트랜스포머의 인코더 부분을 활용하여 만든 아키텍처로 디코더

부분을 활용하여 만든 GPT가 주어진 정보를 바탕으로 다음에 올 단어를 예측하여 생성하는 작업에 특화된 것과는 달리, 양방향의 정보를 활용하고 문맥을 파악하는 자연어를 이해하는 작업(감정 분석, 개체 관계 추출 등)에 특화되어 만들어진 아키텍처로써 생성보다 이해의 작업이 더 중요한 작업에는 더 활용도가 높다. 또한 폐쇄형 모델인 GPT와는 다르게 BERT는 오픈소스로 공개되어 있다. 그러한 두 가지 이유로 본 논문은 BERT를 사용할 것을 제안한다.

#### 3-2. Network Slicing 활용 근거

Network Slicing 기술을 사용하여 하나의 물리적인 망상에 여러 개의 논리적인 망을 만들어 각 패킷에 맞는 Virtual Network를 할당하여 사용한다. 서비스 제공자가 클라이언트의 요청을 받아 여러 가지 정책을 고려하여 할당할 때 LLM이 고려하여 할당하게 되면, 사람의 도움이 필요없이 제로 터치 네트워크를 완성할 수 있다. 이러한 이유로 본 논문은 Network Slicing 기술을 사용할 것을 제안한다.

### IV. 결론

본 논문에서는 차세대 5G, 6G 네트워크 환경에서 급속도로 증가하는 네트워크 트래픽을 처리함에 있어 LLM, 그 중에서도 BERT 아키텍처를 Network slicing 기술에 사용하여 처리할 것을 제안한다. BERT는 PPFL 기술과 함께 사용하면 네트워크 패킷을 이해할 수 있고, 그에 근거하여 네트워크 슬라이싱 기술에 적용한다면 시간적 비용적 자원을 사람이 직접 하는 것보다 효율적으로 사용할 수 있다. 본 논문에서는 실험은 하지 않고 제안만 하였다. 앞으로의 연구는 구현을 통하여 실험하고, 결과값을 분석하여 실제 환경에서 사용 가능성을 확인해서 결과를 토대로 실제 환경에서 구현할 것을 제안한다.

### ACKNOWLEDGMENT

본 과제는 과학기술정보통신부 및 정보통신기획평가원(IITP)의 지원을 받아 수행중에 있습니다. (IITP-2024-2020-0-01834)

### 참고 문헌

- [1] Brown, Tom, et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
- [2] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [3] Touvron, Hugo, et al. "Llama: Open and efficient foundation language models." *arXiv preprint arXiv:2302.13971* (2023).
- [4] Ferrag, Mohamed Amine, et al. "Revolutionizing Cyber Threat Detection with Large Language Models: A privacy-preserving BERT-based Lightweight Model for IoT/IIoT Devices." *IEEE Access* (2024).
- [5] Manias, Dimitrios Michael, Ali Chouman, and Abdallah Shami. "Towards Intent-Based Network Management: Large Language Models for Intent Extraction in 5G Core Networks." *arXiv preprint arXiv:2403.02238* (2024).
- [6] Orlandi, Barbara, et al. "Intent-based network management with user-friendly interfaces and natural language processing." *2024 27th Conference on Innovation in Clouds, Internet and Networks (ICIN)*. IEEE, 2024.