

블록체인 전파 최적화를 위한 LLM 기반 시맨틱 패킷 분석 및 피어 평가

박현수, 박준범, 함승현, *정지원, 박종서

한국항공대학교, *하나금융티아이

park970803@kau.kr, jbpark@kau.kr, gka9710@kau.kr, jwjeong@hanafn.com, jspark@kau.ac.kr

LLM-Based Semantic Packet Analysis and Peer Evaluation for Blockchain Propagation Optimization

Park Hyeonsu, Park Junbeom, Ham Seunghyun, Jeong Jiwon*, Park Jongsou

Korea Aerospace Univ., *Hana Institute of Technology

요약

블록체인 네트워크에서 블록과 트랜잭션의 전파 효율은 블록의 메인 체인 포함 여부와 non-canonical block 발생에 직접적인 영향을 미친다. 그러나 기존의 단순 통신 지표 기반 피어 선택 방식은 지속적인 지연, 불안정, 또는 Free-riding 패턴을 효과적으로 구분하지 못한다. 본 논문에서는 전파 효율에 부정적인 영향을 미치는 피어를 식별하기 위해, 단일 패킷 기반 LLM 분류와 이를 집계한 피어 평가 기법을 제안한다. 제안 기법은 패킷 헤더 및 전파 과정에서 수집된 메타데이터를 문장 기반 입력으로 변환하고, 각 패킷을 정상, 지연, 불안정, Free-riding 상태로 확률적으로 분류한다. 패킷 단위 분류 결과는 시간 구간 단위로 집계되어 피어의 전반적인 전파 패턴을 반영하는 평판 점수로 갱신된다. 실험 결과, 제안 기법은 단일 패킷 분류에서 99.1%의 정확도를 달성하였다. 이는 기존의 단순 지표 기반 또는 무작위적 피어 선택 방식에서 나아가, 전파 과정에 실질적으로 기여하는 피어를 식별함으로써 보다 효율적인 전파 환경을 구성할 수 있음을 보여준다.

I. 서론

블록체인 네트워크에서 전파 지연(Propagation Delay)은 블록이 메인 체인(Canonical Chain)에 포함될 확률을 좌우하는 핵심 요인이다. 탈중앙화 환경에서 노드는 보상을 위해 블록 생성·전파에 경쟁하며, 전파가 지체될수록 유효 블록이더라도 채택 가능성이 낮아진다. 전파가 지체될 경우 유효한 블록이라 하더라도 non-canonical block으로 처리되어 보상을 얻지 못하게 된다 [1]. 블록체인 네트워크는 P2P(gossip) 통신을 통해 빠른 전파를 목표로 설계되지만, 전파에 기여하지 않는 패턴을 보이는 피어가 선택될 경우 블록 전파 과정에서 병목이 발생할 수 있다 [2]. 기존 블록체인 클라이언트는 RTT(Round Trip Time)와 같은 단순한 통신 지표를 기준으로 피어를 선택하기 때문에, 시간에 따라 누적되는 전파 패턴의 차이를 효과적으로 구분하지 못하는 한계를 가진다 [3].

본 연구에서는, 전파 효율에 부정적인 영향을 미치는 피어를 정확하게 식별하기 위한 패킷 단위 학습 기반 접근법을 제안한다. 노드 간 패킷 교환 로그를 기반으로 전파 패턴을 학습·구분함으로써, 단일 지표 기반 피어 평가의 한계를 보완한다.

II. 제안 기법

본 장에서는 전파 효율 저하 피어를 식별하기 위한 제안 기법의 구조를 설명하고, 패킷 단위 분류와 피어 단위 판단 절차를 제시한다.

II-I. 제안 기법 구조

각 노드는 이더리움 블록체인의 시간 구간별 패킷 로그에서 전파 관련 특징을 추출한 뒤, Lightweight Transformer-based Model(DistilBERT [4])로 단일 패킷을 정상·지연·불안정·Free-riding [5]의 4개 상태로 분류한다. 패킷 수준의 분류 결과는 시간 구간 단위로 집계되어 피어 단위의 전파 패턴을 평가하는 데 활용된다. 패킷 단위 추론과 피어 단위 판단을 분리함으로써, 일시적 변동에 덜 민감하게 전파 기여도가 낮은 피어를 식별할 수 있다.

II-II. 패킷 전파 패턴 분류 기준

DistilBERT은 단일 패킷을 대상으로 정상, 지연, 불안정, Free-riding의 네 가지 전파 상태 중 하나로 분류한다. 각 상태는 패킷 전송 지연 수준, 전송 안정성, 재전파 여부를 포함하여, 패킷 헤더 및 전파 과정에서 나타나는 다양한 특성을 종합적으로 반영하여 정의된다.

II-III. 패킷 분류 결과의 피어 단위 평가

패킷 수준에서 도출된 분류 결과는 피어 단위의 전파 패턴을 평가하기 위한 입력으로 활용된다. 일정 시간 구간 동안 동일 피어와 교환된 패킷들의 분류 결과를 집계함으로써, 해당 피어가 전파 과정에서 보이는 전반적인 특성을 추정한다. 이때 패킷 분류 결과는 단순 빈도 계산이 아니라, 시간에 따른 누적과 갱신을 통해 해석된다. 이를 통해 일시적인 네트워크 변동으로 인해 피어가 과도하게 부정적으로 평가되는 것을 방지하며, 전파에 기여하지 않는 피어를 안정적으로 식별할 수 있다.

III. 성능 평가

III-I. 문장 기반 패킷 표현 및 단일 패킷 분류 정의

[표 1. 패킷 종류]

Feature Name	Source	Type	Description & Role in Detection Model
Time Delta (Δt)	Metadata	Float	Time elapsed since the last packet. Indicates latency and network jitter.
Message ID	Header	Cat.	Packet type (e.g., NewBlock). Identifies peer's intent.
Block Number	Payload	Int	Block height. Checks synchronization status and stale data propagation.
Total Difficulty	Payload	Int	Cumulative difficulty. Validates contribution to the canonical chain.
Gas Used	Payload	Int	Computational work. Verifies the substance of the block content.
Tx Count	Payload	Int	Number of transactions. Key indicator for Empty Block Spamming.
Avg Nonce	Payload	Float	Average transaction nonce. Detects sequence gaps and Replay attacks.
Avg Gas Price	Payload	Float	Average gas price. Identifies low-cost DoS spam transactions.

Note: Features are extracted by parsing RLP-decoded eth messages. Metadata is derived from packet arrival timing, while Payload refers to the semantic data within the RLP structure. Cat. denotes Categorical data type.

DistilBERT 패킷 분류 모델의 성능 평가를 위해, 블록체인 네트워크에서 발생하는 패킷을 문장 형태로 표현하는 방법과 단일 패킷 분류 결과의

정의 및 라벨링 방식을 설명한다.

패킷들은 사전에 정의된 템플릿에 따라 하나의 문장으로 직렬화되며, 이는 단일 패킷의 전과 상태를 판단하기 위한 최소 단위 입력으로 사용된다. 단일 패킷 x_i 가 주어질 때, DistilBERT의 출력은 다음과 같은 상태별 확률 벡터로 정의된다.

$$q_i = \text{DistilBERT}(x_i), q_i = [q_i^{(N)}, q_i^{(D)}, q_i^{(U)}, q_i^{(F)}]$$

여기서 N, D, U, F 는 각각 정상, 지연, 불안정, Free-riding 상태를 의미하며, 각 확률은 다음 조건을 만족한다.

$$\sum_{k \in N, D, U, F} q_i^{(k)} = 1$$

이와 같이 DistilBERT는 단일 패킷에 대해 상태별 확률 분포 형태의 분류 결과를 출력하며, 해당 결과는 이후 시간 구간 단위로 집계되어 피어 단위 전과 패턴 판단에 활용된다.

III-II. 비교 기법 및 평가 설정

DistilBERT를 기반으로 한 본 연구의 분류 기법을, 기존의 규칙 기반 및 학습 기반 분류 방식과 비교한다.

[표 2. DistilBERT 분류 모델 성능 평가표]

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Static Threshold	65.2	55.0	42.0	45.0
Random Forest	87.6	88.0	87.0	87.0
LSTM	92.8	95.0	91.0	92.3
DistilBERT (Proposed)	99.1	99.0	99.0	98.8

표 2는 각 기법의 단일 패킷 분류 성능을 비교한 결과를 나타낸다. 제안한 DistilBERT 기반 분류 기법은 정확도 99.1%와 F1-score 98.8%를 기록하여 모든 비교 기법 대비 가장 우수한 성능을 달성하였다.

III-III. 패킷 집계 기반 피어 단위 판단

단일 패킷의 전과 상태를 DistilBERT로 분류한 뒤, 시간 구간 단위로 결과를 집계·누적하여 피어의 전과 패턴을 판단한다.

피어 B에 대해 시간 구간 t 동안 기록된 패킷의 개수를 N_t 라 할 때, 해당 구간에서의 상태별 확률 벡터 p_t 는 다음과 같이 정의된다.

$$p_t = \frac{1}{N_t} \sum_{i=1}^{N_t} q_i$$

시간 구간 t 에서 계산된 상태별 확률 벡터 p_t 를 기반으로, 피어의 전과 기여도를 나타내는 평판 점수 S_t 는 다음과 같이 갱신된다.

$$S_t = (1 - \alpha)S_{t-1} + \alpha(w \cdot p_t)$$

여기서 S_{t-1} 은 이전 시간 구간의 피어 평판 점수이며, $\alpha \in (0,1)$ 은 최신 구간의 반영 비율을 나타낸다. w 는 상태별 전과 기여도를 반영하는 가중치 벡터로, 정상, 지연, 불안정, Free-riding 상태 순으로 감소하도록 설정된다. 갱신된 평판 점수 S_t 를 기준으로, 피어는 다음과 같이 분류된다.

$$Peer_t = \begin{cases} \text{Normal}, & S_t \geq \theta \\ \text{Unnormal}, & S_t < \theta \end{cases}$$

Free-riding 패턴은 단순한 성능 저하와 구분되어야 하므로, 본 연구에서는 별도의 규칙 기반 판정을 적용한다. 시간 구간 t 동안 Free-riding 상태로 분류된 패킷의 개수는 다음과 같이 정의된다.

$$C_t^{(F)} = \sum_{i=1}^{N_t} I(\arg \max_k q_i^{(k)} = F)$$

만약 $C_t^{(F)} \geq \tau_F$ 를 만족할 경우, 해당 피어는 평판 점수와 무관하게 Free-riding 피어로 판정되며, 이후 전과 대상에서 제외된다.

[그림 1. 패킷 집계 기반 시간별 피어 점수 변화]

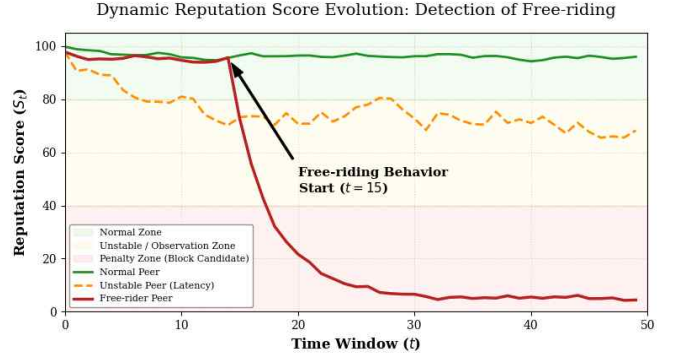


그림 1은 시간 구간에 따른 피어 점수 변화를 통해, Free-riding 패턴이 발생한 피어의 점수가 급격히 감소하며 정상 피어 및 불안정 피어와 명확히 구분되는 과정을 보여준다.

IV. 결론

본 논문은 블록체인 네트워크에서 전과 효율을 저해하는 피어를 식별하기 위해, 패킷 로그를 문장 기반 입력으로 변환해 DistilBERT로 분류하고 그 결과를 집계하는 피어 평가 기법을 제안하였다. 이를 통해 단일 통신 지표만으로는 구분하기 어려운 전과 특성의 차이를 반영한다. 정량 평가에서 제안한 DistilBERT 기반 분류는 단일 패킷 기준 정확도 99.1%와 F1-score 98.8%를 달성했으며, 규칙 기반 및 기존 학습 기반 기법보다 우수한 성능을 보였다. 또한 Free-riding 시작 시점($t=15$) 이후 해당 피어의 평판 점수는 80 이상에서 20 이하로 급락한 반면 정상 피어는 80 이상을 유지하여, 점수 기반 누적 평가가 일시적 변동과 지속적 전과 비기여 패턴을 구분할 수 있음을 확인했다. 본 접근법은 추가적인 프로토콜 변경 없이 기존 이더리움 환경에 적용 가능하며, 향후 피어 선택 정책과의 결합을 통해 전과 효율 향상으로 확장될 수 있다.

ACKNOWLEDGMENT

이 논문은 2026년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (No.RS-2021-KI002499, 2026년 산업혁신인재성장지원사업)

참고 문헌

- [1] M. Decker and R. Wattenhofer, "Information propagation in the Bitcoin network," IEEE P2P 2013.
- [2] M. Zitterbart, J. K. Möller, and G. Carle, "Performance analysis of peer selection strategies in blockchain networks," Technische Universität München, NET-2025-05-1, 2025.
- [3] J. Lee, H. Kim, and J. Park, "Blockchain-based peer evaluation for efficient block propagation," IEEE Access, vol. 9, pp. 1 - 12, 2021, doi: 10.1109/ACCESS.2021.9426434.
- [4] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.
- [5] M. Karakaya, I. Körpeoğlu, and Ö. Ulusoy, "Free riding in peer-to-peer networks," IEEE Internet Computing, vol. 13, no. 2, pp. 96 - 97, 2009, doi: 10.1109/MIC.2009.33.