# 지역적 복원가능 부호를 활용한 데이터 가용성 샘플링 구조의 확장 가능성 연구

조승현, 김영식
대구경북과학기술원
seunghyuncho@dgist.ac.kr, ysk@dgist.ac.kr

# Exploring the Applicability of Locally Recoverable Codes in Data Availability Sampling Schemes

Seung Hyun Cho, Young-Sik Kim
DGIST

요 약

본 논문은 데이터 가용성 샘플링(DAS) 구조에서 기존 Reed-Solomon 부호의 한계를 지적하고, 이에 대한 대안으로 Locally Recoverable Codes(LRC)의 적용 가능성을 개념적으로 고찰하였다. LRC는 로컬 그룹 단위의 복원을 지원함으로써, 전체가 아닌 일부 데이터 복원이 필요한 상황에서 더 유연한 부호화 전략을 제공할 수 있다. 이를 통해 DAS의 구조적 다양성과 설계 유연성을 확장할 가능성을 제시한다. 시뮬레이션 결과 기존 방식 대비 복원 최소 수신량, 샘플 수, 총 통신량 측면에서 각각 11.5배, 19.8배, 19.8배 감소하는 것을 확인하였다.

## Ⅰ. 서 론

블록체인은 데이터의 무결성을 보장하는 구조로 되어 있지만, 탈중앙성, 보안성, 확장성 간의 균형을 달성하기 어렵다는 블록체인 트릴레마 문제를 안고 있다. 이 중 확장성과 밀접하게 관련된 요소로 데이터 가용성(data availability)이 주목받고 있다. 이는 데이터를 모두 내려받지 않고도 유효성을 검증할 수 있어야 한다는 요구에서 비롯된다.

이러한 배경에서 데이터 가용성 샘플링(Data Availability Sampling, 이하 DAS) 스킴이 제안되었으며, [1]에서는 그 개념을 형식적으로 정의하였다. DAS는 KZG [2]와 같은 다항식 커밋먼트 스킴을 일반화하여, 오류 정정 부호의 코드워드 전체에 대해 커밋할 수 있는 소거 코드 커밋먼트를 활용한다. 또한, 일부 데이터만으로 전체를 복구할 수 있는 MDS 코드의 특성을 기반으로 하여, 전체 데이터를 전송하지 않고도 신뢰성을 유지하면서 네트워크 부하를 줄일 수 있도록 설계된다.

[3]에서는 DAS 구현에 $(n,k)$ Reed-Solomon 부호(이하 RS 부호)를 적용하여 블록체인 데이터를 이차원으로 확장한 구조를 제안한다. 이는 각 노드가 행 또는 열 단위의 데이터를 검증할 수 있음을 시사한다. 다만, 이 과정에서 복원에 필요한 최소 심볼 수는 기존과 같다는 한계를 가진다.

본 논문은 기존 DAS 구조와의 정합성을 유지한 채, 더욱 효율적인 부호화 방식을 모색하며 기존에 주로 활용된 부호 외의 구조적 대안을 함께 고려한다. 이를 통해 확장성 중심의 블록체인 설계에서 데이터 가용성을 정교하게 확보할 수 있는 새로운 방향을 제시한다. 시뮬레이션 결과 제안하는 방법은 복원 최소 수신량, 샘플 수, 총 통신량 측면에서 기존 방법 대비 각각 11.5배, 19.8배, 19.8배 더 적은 크기로 감소하는 것을 확인하였다.

## Ⅱ. 본론

DAS는 [1]에서 제시한 것과 같이 그림 1의 세 단계로 구분할 수 있다. 첫째, 원본 데이터에 오류 정정 부호를 적용하여 부호화된 코드워드를 생성한다. 둘째, 생성된 코드워드에 소거 코드 커밋먼트 스킴을 적용하여 각 심볼에 대한 오프닝 정보를 생성한다. 셋째, 클라이언트는 사전에 정의된 샘플링 전략에 따라 임의의 부호 심볼과 해당 심볼에 대한 오프닝 값을 획득하고, 커밋 검증 절차를 통해 이들의 일치 여부를 확인함으로써 전체 데이터의 가용성이 통계적으로 보장되는지를 판단한다.

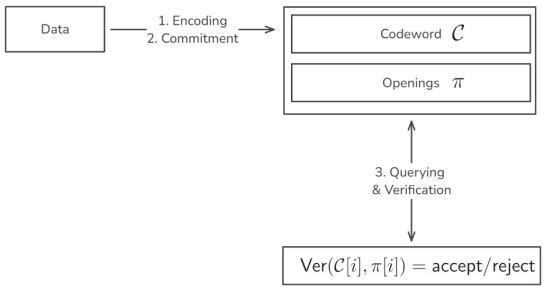


그림 1. Data Availability Sampling의 구성

현재까지 제안된 DAS 구현은 주로 RS 부호를 기반으로 한다. RS 부호는 이론적으로 MDS 특성을 만족하기 때문에, 일정 개수 이상의 심볼이 확보되면 전체 메시지를 정확히 복원할 수 있다. 그러나 이러한 특성은 전체 메시지 복원을 전제로 설계되어 있으므로, 특정 부분 데이터만 선택적으로 복원하려는 경우에는 비효율적일 수 있다. 전체가 아닌 일부 데이터 복원이 주된 목적이 되는 환경에서는, RS 부호 기반 구조가 요구하는 심볼 수가 과도하게 클 수 있으며, 이는 실용적인 제약으로 작용할 수 있다.

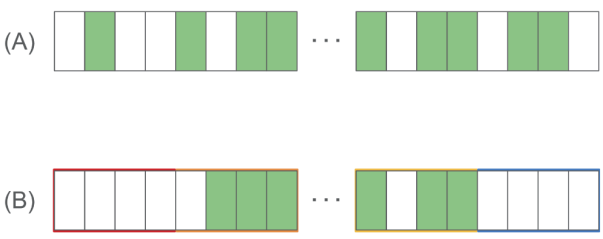


그림 2. Locally Recoverable Code

이러한 배경에서, DAS에 Locally Recoverable Codes(LRC)의 적용 가

능성을 고려할 수 있다. LRC는 전체 코드워드 수준이 아닌, 각 부호 심볼이 속한 로컬 그룹 내의 다른 심볼들을 통해 해당 심볼을 복원할 수 있도록 설계된 부호 구조이다. 그림 2의 (A)는 전통적인 오류 정정 부호의 구조로, 전체 메시지를 복원하기 위해 일정 수 이상의 심볼이 필요하다. 반면 (B)는 코드워드가 일정한 단위로 나뉘어 있어, 각 구간 내에서는 더 적은 수의 심볼만으로 해당 구간의 데이터를 복원할 수 있는 지역적 복원(local recovery) 구조를 보여준다. 예를 들어, (A)는 데이터 복구를 위해 총 9개의 심볼을 필요로 하지만, (B)는 6개만을 가지고도 특정한 범위의 데이터를 복구할 수 있다. 이와 같은 지역적 복원의 개념은, DAS에서 전체 데이터를 복원하지 않고도 일정 수준의 가용성을 확인하려는 목적과 일정 부분 정합성을 가진다. 특히, [4]에서 제안된 Local Secret Sharing 기법은 이러한 로컬 복원 구조를 효율적인 분산 환경에 적용한 사례로, LRC 기반 접근법의 활용 가능성을 시사한다.

| 입력 데이터 크기 (bit) | 코드워드 길이 | 로컬 그룹 개수 | 로컬 그룹 크기 (bit) |
|---|---|---|---|
| 80,000,000 | 833,336 | 46 | 18,116 |

| 기반 부호 | 커밋먼트 크기 (KB) | 인코딩 크기 (MB) | 통신 당 쿼리 크기 (KB) | 복원 최소 수신량 | 샘플 수 | 총 통신량 (MB) |
|---|---|---|---|---|---|---|
| RS | 0.05 | 80 | 0.1 | 208,334 | 356,833 | 35.31 |
| LRC | 0.05 | 80 | 0.1 | 18,116 | 18,115 | 1.78 |

표 1. 실험 결과

표 1은 [1]에서 제공하는 코드를 기반으로 LRC를 적용한 실험 결과를 나타낸다. 입력 데이터는 10MB(=80,000,000비트)이며, [5]의 KZG 커밋먼트를 사용함을 전제로 하여, BLS12-381 곡선의 필드 원소 크기(384비트)에 맞춰 데이터를 분할하였다. 총 $\lceil \frac{80000000}{384} \rceil = 208334$개의 심볼이 생성된다. 이후, 2차원 Reed-Solomon 부호의 확장 방식을 그대로 따르기 위해, 행과 열 방향으로 각각 2배씩 확장하여 코드워드의 총 길이를 4배인 833,336으로 설정하였다.

LRC 구성에서는 이 코드워드를 정확히 46개의 로컬 그룹으로 나누었다. 이는 코드워드 길이에서 정확히 나누어떨어지는 값들 중, 로컬 그룹 크기(18,116비트)가 적절하게 설정되는 구조를 고려해 선택한 것이다.

본 실험은 전체 데이터가 아닌, 하나의 로컬 그룹을 복원하는 상황을 가정하여 수행되었다. 복원을 위한 최소 수신량은 RS 구성의 경우 로컬 그룹에 관계 없이 전체 데이터를 복원해야 하므로 원본 데이터의 크기에 해당하는 208,334인 것에 비해 LRC 구성에서는 로컬 그룹의 크기인 18,116이다.

샘플 수는 복원 최소 수신량을 확보하기 위해 샘플링하는 수를 의미한다.

RS 구성에서는 쿠폰 수집 문제(coupon collector's problem)에 기반하여 복원이 성공할 확률이 통계적으로 충분히 높아지도록 설정하였다. 그 결과 총 358,633개의 샘플이 필요했다. 반면, LRC 구성에서는 로컬 그룹을 알고 있다는 전제 하에 전역적인 무작위 복원 추출이 불필요하므로, 정확히 로컬 그룹의 크기보다 하나 작은 18,115개로 대폭 감소하였다. 이에 따라 전체 통신량 또한 35.31MB에서 1.78MB로 줄어들었으며, 약 94.96%의 감소율을 확인할 수 있다.

LRC 기반 DAS 구성은 전체 데이터를 복원하는 것만이 아닌, 특정 영역의 데이터만 복원 가능한 상황에서도 실질적인 효용을 제공할 수 있다. 이는 부호 구조의 다양성과 복원 전략의 유연성을 확대하며, 데이터 가용성 검증을 위한 새로운 설계 가능성을 여는 기반이 될 수 있다.

## Ⅲ. 결론

본 논문은 DAS 구조에서 기존 RS 부호 기반 설계의 한계를 지적하고, LRC의 개념적 적용 가능성을 고찰하였다. LRC는 로컬 그룹 단위의 복원을 지원함으로써, 전체가 아닌 부분 데이터 복원이 필요한 상황에서 보다 유연한 복원 전략을 제공할 수 있다. 이는 DAS 설계의 구조적 다양성을 확장할 수 있는 기반이 되며, 향후 효율적인 부호화 방식에 대한 연구 가능성을 제시한다.

## ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(RS-2024-00399401, 양자안전 보안인프라 전환 및 대양자 복합 안전성 검증기술 개발).

## 참 고 문 헌

[1] M. Hall-Andersen, M. Simkin, and B. Wagner, "Foundations of Data Availability Sampling," Cryptology ePrint Archive, Paper 2023/1079, 2023.

[2] A. Kate, G. M. Zaverucha, and I. Goldberg, "Constant-size commitments to polynomials and their applications," in Advances in Cryptology-ASIACRYPT 2010, pp. 177-194, Springer, 2010.

[3] M. Al-Bassam, A. Sonnino, and V. Buterin, "Fraud and Data Availability Proofs: Maximising Light Client Security and Scaling Blockchains with Dishonest Majorities," arXiv preprint arXiv:1809.09044, 2018.

[4] Kim, Y., Raman, R. K., Kim, Y.-S., Varshney, L. R., & Shanbhag, N. R., "Efficient Local Secret Sharing for Distributed Blockchain Systems," IEEE Communications Letters, vol. 23, no. 2, pp. 282-285, Feb. 2019.

[5] A. Kate, G. M. Zaverucha, and I. Goldberg, "Constant-size commitments to polynomials and their applications," Advances in Cryptology - ASIACRYPT 2010, Lecture Notes in Computer Science, vol. 6477, pp. 177-194, Springer, 2010.