Offloading Erasure Coding to CSD in Hyperledger Fabric

Junghyun Ryu[1], Hongsu Byun[1], Myungcheol Lee[2], Jinchun Choi[2] and Youngjae Kim[1]
[1]*Sogang University,* [2]*Smart Data Research Section, ETRI*

최근 블록체인 시스템에서 저장되는 데이터의 양이 기하급수적으로 증가함에 따라, 공산 효율성을 높이기 위한 방안으로 Erasure Coding (EC)이 채택 되고 있다[1]. EC 는 데이터 복제 방식과 비교하여 저장 공간 효율을 높이는 장점이 있는 반면, 행렬곱 기반의 계산집약적인 연산으로 노드의 CPU cycle 을 증가시켜 블록체인 시스템의 성능을 저하시키는 단점이 있다.

본 연구는 우선, 대표적인 EC 를 수행하는 하이퍼레저 패브릭[2]에서 EC 구동으로 인한 호스트/노드 CPU 사용율과 패브릭의 성능 저하를 분석한다. 다음으로, 계산 스토리지 디바이스 (예, SmartSSD, Newport CSD)에 EC 를 오프로딩을 하여 호스트의 CPU 사용률을 줄이고 패브릭의 성능 감소 가능성을 분석한다.

평가 및 분석(트랜잭션 처리 시간): 그림 1(a)는 패브릭에서 호스트 CPU 를 사용하여 EC 수행 시 (baseline) 시간에 따른 CPU 작업 스케줄을 보여준다. 먼저, 호스트는 전달 받은 블록을 검증(T_A)하고 그 이후 EC($T_{EC-Host}$)를 동기적으로 수행한다. 블록 검증이 끝나고 일정 대기시간 (T_B) 이후 새로운 블록이 도착하더라도 EC 를 수행하는 동안 CPU 는 다른 작업을 수행할 수 없다. 그림 1(b)는 CSD 를 사용하여 오프로드된 EC 를 비동기적으로 실행하였을 때 호스트 CPU 와 CSD CPU 의 작업 실행을 보여준다. CSD CPU(Newport CSD, ARM Cortex-A53, 1.0GHz)는 노드 CPU (AMD EPYC 7352, 2.3 GHz)보다 연산 능력이 낮다. 따라서, CSD 의 EC 수행 시간(T_{EC-CSD})은 $T_{EC-Host}$보다 크지만, EC 를 비동기적으로 실행하므로 호스트 CPU 는 EC 수행 완료 여부에 관계 없이 T_B 의 대기 시간 이후에 도착하는 새 블록을 바로 처리할 수 있다. EC 를 CSD 로 오프로딩한 노드에서 총 트랜잭션 처리 시간이 baseline 보다 더 빠르기 위한 조건들은 다음과 같다.

- CSD 에 $T_A + T_B$ 시간에 한번씩 EC 요청이 들어오게 되므로 $T_A + T_B > T_{EC-CSD}$ 를 만족해야 한다.
- Baseline 의 총 트랜잭션 처리 시간은 $\sum T_A + \sum T_{EC-Host}$ 이고, CSD 는 $\sum T_A + \sum T_B + T_{EC-CSD}$이므로 $\sum T_{EC-Host} > \sum T_B + T_{EC-CSD}$ 을 만족해야 한다.

그림 2 는 하이퍼레저 패브릭에서 사용율을 EC 수행 유무에 따른 노드의 CPU 사용율을 보여준다. 이때 노드의 CPU 부하를 관찰하기 위해 사용가능한 호스트 CPU 의 코어 수를 1 개로 제한하였다. 그리고, 16 개의 클라이언트에서 총 1 백만건 트랜잭션을 생성하여 우리의 실험 환경에서 가용 자원을 최대로 사용하게 설정했다. EC 를 수행했을 때(EC On)와 EC 를 수행하지 않았을 때(EC Off) 모두 150 초 까지 CPU(AMD EPYC 7352, 2.3 GHz)를 최대로(100%) 사용한 구간이 지속된다. 그 후 EC Off 상태에서 CPU 사용률이 60% 미만으로 감소한 반면, EC On 의 경우 CPU 를 최대로 사용하는 구간이 약 50 초 더 지속되었다(붉은색 음영 구간). 해당 구간 만큼 노드는 CPU clock cycle 을 추가로 소모한다.

결론 및 제안: 그림 2 에서 EC 수행으로 인한 추가적인 CPU clock cycle 소모량과 실행 시간의 증가는 CSD 에서 EC 를 비동기적으로 수행하여 그림 1 을 기반으로 도출한 수식을 만족한다면, 충분히 감춰질 수 있는 오버헤드이다. 따라서 우리는 이러한 관찰을 기반으로 하이퍼레저 패브릭[2]을 확장하여 CSD 에 EC 를 오프로딩하며 비동기/병렬 실행이 가능한 구조를 제안하고 구현을 미래 연구로 제안한다.

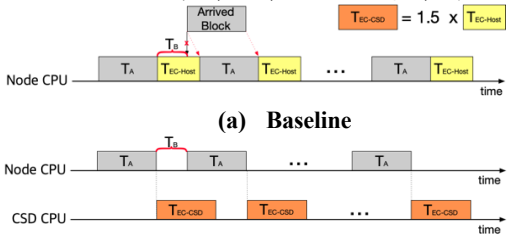

Fig1. Baseline vs. EC offloading to CSD

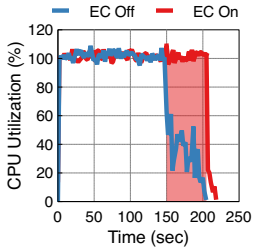

Fig2. CPU utilization

Acknowledgments This work was supported by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00136, Development of Big Blockchain Data Highly Scalable Distributed Storage Technology for Increased Applications in Various Industries).

References [1] X. Qi, Z. Zhang, C. Jin and A. Zhou, "BFT-Store: Storage Partition for Permissioned Blockchain via Erasure Coding," 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 2020, pp. 1926-1929.
[2] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, et al., "Hyperledger fabric: a distributed operating system for permissioned blockchains,"in Eurosys, pp. 1–15, 2018.