KIPS Trans. Comp. and Comm. Sys.
Vol.11, No.2 pp.51~58
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2022.11.2.51

Matrix Character Relocation Technique for Improving Data Privacy in Shard-Based Private Blockchain Environments

Yeol Kook Lee† · Jung Won Seo†† · Soo Young Park†††

ABSTRACT

Blockchain technology is a system in which data from users participating in blockchain networks is distributed and stored. Bitcoin and Ethereum are attracting global attention, and the utilization of blockchain is expected to be endless. However, the need for blockchain data privacy protection is emerging in various financial, medical, and real estate sectors that process personal information due to the transparency of disclosing all data in the blockchain to network participants. Although studies using smart contracts, homomorphic encryption, and cryptographic key methods have been mainly conducted to protect existing blockchain data privacy, this paper proposes data privacy using matrix character relocation techniques differentiated from existing papers. The approach proposed in this paper consists largely of two methods: how to relocate the original data to matrix characters, how to return the deployed data to the original. Through qualitative experiments, we evaluate the safety of the approach proposed in this paper, and demonstrate that matrix character relocation will be sufficiently applicable in private blockchain environments by measuring the time it takes to revert applied data to original data.

Keywords : Blockchain, Data Privacy, Hash Algorithm, Shard

샤드 기반 프라이빗 블록체인 환경에서 데이터 프라이버시 개선을 위한 매트릭스 문자 재배치 기법

이 열 국† · 서 중 원†† · 박 수 용†††

요 약

블록체인 기술은 블록체인 네트워크에 참여하는 사용자의 데이터가 분산 처리되어 저장되는 시스템이다. 비트코인과 이더리움을 필두로 세계적으로 관심을 받고 있으며, 블록체인의 활용성은 무궁무진한 것으로 예측되고 있다. 하지만 블록체인의 모든 데이터를 네트워크 참여자에게 공개하는 투명성으로 인해 블록체인 데이터 프라이버시 보호에 대한 필요성이 개인정보를 처리하는 각종 금융, 의료, 부동산 분야에서 떠오르고 있다. 기존 블록체인 데이터 프라이버시 보호를 위해서 스마트 컨트랙트, 동형암호화, 암호학 키 방식을 사용하는 연구들이 주를 이루었으나, 본 논문에서는 기존의 논문들과 차별화된 매트릭스 문자 재배치 기법을 사용한 데이터 프라이버시 보호를 제안한다. 본 논문에서 제안하는 접근방안은 원본 데이터를 매트릭스 문자 재배치 하는 방법, 배치된 데이터를 다시 원본으로 되돌리는 방법, 크게 두 가지로 구성이 되어있다. 정성적인 실험을 통해 본 논문에서 제안하는 접근방안의 안전성을 평가하였으며, 매트릭스 문자 재배치가 적용된 데이터를 원본 데이터로 되돌릴 때 걸리는 시간을 측정하여 프라이빗 블록체인 환경에서도 충분히 적용이 가능할 것이라는 것을 증명하였다.

키워드 : 블록체인, 데이터 프라이버시, 해시함수, 샤드

1. 서 론

블록체인은 데이터의 신뢰성과 효율성을 높일 수 있는 4차 산업혁명의 대표적인 기술 중 하나로서 2008년 사토시 나카모토의 비트코인 백서[1]가 공개되면서 세상에 알려지게 되었다. 블록체인 기술은 피투피 네트워크(Peer-to-Peer)를 기반으로 블록체인 네트워크에 참여하는 모든 참여자의 트랜잭션을 블록에 저장하고 각 노드가 동일하게 트랜잭션 내용을 해시 형태로 저장한다. 이렇게 저장된 트랜잭션들은 수정하거나 조작이 불가능하기 때문에 무결성과 투명성이 뛰어난 기술로 인정받아 상용화를 위한 다양한 연구가 진행되고 있다. 대표적인 블록체인이 상용화 사례는 IBM의 하이퍼레저 패브릭이다[2]. 월마트는 IBM 블록체인을 바탕으로 구축된 추적 시스템을 도입하여 납품하는 식품의 원산지, 운송 경로

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 지원사업의 연구결과로 수행되었음(IITP-2021-2017-0-01628*).
† 비 회 원 : 한영회계법인 사원
†† 비 회 원 : 서강대학교 컴퓨터공학과 박사과정
††† 정 회 원 : 서강대학교 컴퓨터공학과 교수 및 지능형 블록체인연구센터 센터장
Manuscript Received : June 23, 2021
Accepted : September 2, 2021
* Corresponding Author : Soo Young Park(sypark@sogang.ac.kr)

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

등 식품의 전반적인 정보를 실시간으로 확인 할 수 있게 하였다. 블록체인 기반의 유통 시스템의 도입한 후 식품의 원산지 추적 시간이 7일에서 2.2초로 단축되었다[3].

블록체인이 가지고 있는 투명성이라는 특징에 의해서 블록체인 네트워크에서 발생하는 트랜잭션에 대한 정보는 모든 참여자들에게 공개가 되어 어떤 거래가 언제 발생했는지 알 수 있다. 블록체인에 저장되는 트랜잭션은 블록체인 네트워크 안에서 송신자와 수신자의 주소, 데이터의 바이너리 값을 포함한 형태로 블록체인 네트워크 참여자들 간에 공유가 된다. 이러한 블록체인의 투명성은 블록체인 네트워크 참여자 간의 신뢰성을 보장할 수도 있지만, 특정 집단이나 기업에는 본인의 트랜잭션이 모두에게 공개가 되기 때문에 원하지 않는 방향일 수 있다. 예를 들어 개인의 금융 데이터 및 거래에 대한 상세한 내역, 병원에서 치료받은 환자의 의료데이터, 개인과 개인의 부동산 거래 데이터 등 민감한 개인정보가 포함된 데이터의 경우 프라이버시가 노출되어 개인정보가 유출되고 악용될 우려가 있다.

이러한 블록체인의 프라이버시 문제를 해결하기 위한 다양한 연구들이 존재하며[4-13], 크게 네 가지 형태로 연구들이 진행되고 있다. 첫 번째는 그룹 서명(Group-Signature)[4]을 적용하여 프라이버시를 보호하는 연구가 있다[5]. 그룹 서명은 특정 그룹에 속한 한 참여자가 서명하는 기법이며 서명을 누가 했는지 알 수 없다는 특징이 있어서 프라이버시를 보호할 수 있지만, 그룹 매니저의 권한이 크기 때문에 중앙화 문제가 있다는 한계가 존재한다. 두 번째는 링 서명(Ring-Signature)[6]을 활용하여 사용자의 공개키를 혼합하고 특정 사용자를 식별과 추적을 못하게 하는 연구도 있다[7,8]. 링 서명의 경우 어떤 참여자가 거래를 했는지 알 수 없다는 장점이 있지만 그룹 서명과 같이 특정 노드에 권한이 부여된다는 한계가 존재한다. 세 번째는 블라인드 서명(Blind-Signature)[9]을 활용한 프라이버시 기법[10,11]이 있는데 메시지를 보내는 송신자의 정체를 숨긴 채 대리인을 통해 서명하는 방식이다. 자신의 정보는 숨길 수 있지만, 악의적인 메시지의 전송을 막을 수 없고 송신자와 대리인의 신뢰를 전제로 한다는 한계가 있다. 마지막으로 zk-SNARKs(zero-knowledge Succinct Non-interactive Argument of Knowledge)를 활용한 블록체인 익명성 보호 기법[12,13]이 있으며 해당 연구는 특정 데이터를 공개하지 않지만 증명키와 검증키를 만드는 주체에 대한 의존도가 높기 때문에 중앙화 문제가 있다는 한계가 존재한다.

본 논문에서는 기존에 제안된 블록체인 프라이버시를 보호하는 방법과는 차별화된 블록체인 환경에서 개인정보와 같은 민감한 데이터를 블록체인에 저장하기 전에 암호화를 거친 후 블록체인의 각 샤드에 데이터를 저장하는 방법을 제안한다. 원본 데이터를 특정 인코딩 기법을 사용해 바이트 값으로 변환한 후 해당 값을 매트릭스를 사용한 블록화 과정을 거쳐 해시값을 추출한다. 암호화된 데이터를 블록체인의 샤드 수에 맞게 분배하고 원본 데이터의 복원을 진행할 때는 이전에 미리 추출해 놓은 해시값을 토대로 문자를 재배치하여 원본 데이터로 복원을 진행한다. 본 논문이 기여하는 바는 다음과 같다.

- 기존의 블록체인 프라이버시 기법과 다른 데이터 매트릭스 문자 재배치를 통해 프라이버시 보호를 제안
- 블록체인 네트워크에 저장된 데이터가 외부에 유출되더라도 원본 데이터가 어떤 형태인지 유추할 수 없음
- 데이터를 요청한 접근 권한자가 직접 원본을 검증하게 함으로써 검증 의존도를 낮춤

본 논문의 구성은 다음과 같다. 1장의 서론을 서술하고, 2장에는 블록체인에 대한 배경지식과 기존에 블록체인 프라이버시를 보호하기 위한 연구를 설명하고, 3장에서 본 논문의 접근 방안을 제시한다. 4장에서 연구에 대한 실험 결과를 확인하고 마지막으로 5장에서 결론과 향후 연구에 대해 알아본다.

2. 배경지식 및 관련 연구

2.1 배경지식

블록체인은 서론에서 언급했듯이 블록체인 네트워크를 통해 데이터를 분산 처리 하는 기술이다. 기존의 중앙서버가 존재하는 방식이 아닌 블록체인 네트워크에 참여하는 모든 노드가 동등하게 데이터를 저장 및 보관한다. 이러한 블록체인은 암호화와 분산처리를 동시에 적용하기 때문에 높은 보안성을 갖게 되고 데이터를 임의로 조작하거나 변경할 수 없다.

블록체인의 높은 보안성을 유지하기 위해서 해시 함수 함수를 사용하는데 해시 함수는 임의의 데이터를 일정하게 고정된 값으로 변환해주는 특징을 가지고 있으며 해시 함수의 동작 과정은 Fig. 1과 같이 표현할 수 있다. 해시 함수를 거친 데이터는 의미를 알아볼 수 없는 고정된 문자열로 변환이 된다. 해시 함수의 또 다른 특징으로는 해시 함수는 역방향으로 변환이 불가능한 단방향 함수의 특징을 가지고 있다. 단방향 함수의 특징 때문에 해시 함수의 결괏값으로부터 원본 데이터를 찾아내는 것은 불가능하다. 원본 데이터의 아주 작은 변화가 발생하면 해시함수의 결괏값은 완전히 다른 값으로 만들어지기 때문에, 이러한 해시 함수가 가진 특징을 사용하여 원본 데이터의 조작이나 변형을 해시 함수 비교를 통해서 확인할 수 있다.

2.2 관련 연구

블록체인의 프라이버시 보호를 위한 다양한 관련 연구들은 다음과 같다[4-13]. Zhang은 블록체인 기반 모바일 에지 컴퓨팅의 프라이버시를 보호하기 위해 그룹 서명(group-signature)[4]을 적용하여 서명의 유효성을 검증하고 프라이버시를 높이

Fig. 1. Hash Algorithm

는 기법을 연구했다[5]. 그룹 서명은 서명할 특정 그룹의 멤버 중 한 멤버가 서명을 하는 것이며 서명을 진행하면 검증자는 해당 그룹의 멤버 중 누가 서명을 했는지 알 수 없지만 1/n의 확률로 유추가 가능하고 검증하는 그룹 관리자의 권한이 절대적이라는 한계가 존재한다.

Li는 블록체인에 링 서명(ring-signature)[6]을 기반으로 익명성을 보장하는 연구를 제안한다[7,8]. 일반적인 거래는 송금자 한 명에 대한 본인의 디지털 서명을 통해 거래가 이루어지지만 Li의 연구에서는 링 서명을 사용하여 여러 사람의 공개키를 혼합하여 서명하여 프라이버시를 보호하는 방안을 제안하였다[7]. 링 서명은 암호화폐 모네로(Monero)[8]에도 적용되었고 링 서명을 적용하면 그룹 서명과 같이 그룹 내의 서명자가 누군지 알 수 없고 그룹 관리자가 존재하지 않는다. 따라서 그룹 서명보다 중앙화 문제는 줄어들지만, 완전히 중앙화 문제가 해결되는 것은 아니다.

Jian은 블라인드 서명(blind-signature)[9]을 적용하여 민감한 개인정보를 보호하는 연구를 진행하였다[10]. 블라인드 서명은 송신자와 대리 서명자가 존재하고 둘 사이의 한 쌍의 서명이 존재한다. 또한 Cai는 블록체인의 스마트 컨트랙트의 보안 성능을 향상하는 연구[11]를 진행했으며 이러한 블라인드 서명을 적용한 기법들은 블라인드 서명을 통해 메시지를 보내는 전송자의 정체를 숨길 수 있다. 하지만 악의적인 메시지의 전송을 막을 수 없으며 송신자와 대리 서명자 간에 신뢰해야 한다는 한계가 있다. 또한 대리 서명자에 대한 검증 의존도가 존재하기 때문에 검증의 중앙화 문제도 여전히 존재한다.

마지막으로 zk-SNARKs를 사용하여 프라이버시를 보호하는 연구가 존재한다[12,13]. Murtaza는 zk-SNARKs를 블록체인에 적용한 전자 투표 시스템에 대한 연구를 진행했다[12]. zk-SNARKs를 통해 유권자의 익명성을 보호하여 공정한 전자 투표와 검증을 가능하게 했고 Li는 블록체인 기반 사용자 인증 시스템에 zk-SNARKs를 적용하여 사용자의 아이덴티티와 속성에 대한 프라이버시 보호 시스템을 제안하였다[13]. zk-SNARKs를 활용하면 특정 정보를 공개하지 않고 증명이 가능하지만 증명키와 검증키를 생성하는데 주체 의존도가 높기 때문에 악의적으로 잘못된 키 생성을 할 가능성이 있고 중앙화 문제가 있다는 한계가 있다.

본 논문에서 제안하고자 하는 접근 방안과 기존 연구를 비교하면 Table 1과 같다. 표에서 n은 전체 참여자를 의미하고 g는 링 서명의 그룹 멤버를 의미한다. 본 논문에서는 기존 연구들이 가지고 있는 검증에 대한 의존도와 특정 참여자에게 권한이 집중되는 중앙화 문제를 해결하는 방안을 제안한다. 제안하는 기법은 기존 연구와 다르게 네트워크 참여하는 특정 인원에게 권한을 부여하여 검증하는 방식이 아닌 검증자가 따로 존재하지 않고 데이터 요청자가 직접 검증을 할 수 있기 때문에 탈중앙성이 1로 표현될 수 있다. 또한 데이터 요청자가 직접 검증을 할 수 있어서 검증 의존도는 없다고 표현될 수 있다. 본 연구의 기법은 검증 의존도가 낮기 때문에 블록체인 네트워크에 악의적인 노드가 존재하더라도 원본 데이터의 복원이 가능하다.

Table 1. Comparison with Existing Studies

Techniques	Decentralization Degree	Verification Dependency
Group-Signature	1/n	Group administrator
Ring-Signature	g/n	Group member
Blind-Signature	1/n	Representative signer
zk-SNARKs	1/n	Key constructor
This paper	1	None

3. 접근 방안

본 논문에서는 기존의 블록체인 프라이버시를 유지하기 위한 연구와 차별화된 접근 방안을 제시한다. 퍼블릭 블록체인이 아닌 미리 정해진 특정 집단이나 기업에서 사용할 수 있는 샤드 기반 프라이빗 블록체인 환경에서의 데이터 프라이버시에 대한 보호 기법을 다룬다.

샤딩이란 데이터베이스에서 사용되는 개념으로써 데이터를 분할 하여 저장하기 위해 사용되어 왔다[14]. 블록체인에서 일반적으로 샤딩은 트랜잭션을 병렬처리하는 확장성 기술로 정의되며 샤딩으로 나눠진 블록의 구간을 샤드라고 정의한다. 본 논문에서는 샤드를 블록체인 노드 혹은 네트워크 그룹으로 정의하고 해당 샤드 기반으로 트랜잭을 분할하여 검증하는 방법을 제안한다.

본 논문에서 제안하는 접근 방안은 블록체인 외부에서 사전에 전처리 과정을 거친 후 블록체인 네트워크에 데이터를 저장하는 형태로서, 기존 연구들과 달리 블록체인 네트워크의 특정 참가자에 대한 의존도 없이 데이터의 프라이버시를 보호하는 접근 방안을 제안한다. 제안하는 접근 방안은 Fig, 2와 같이 표현될 수 있다. 사용자는 프라이버시를 보호하고 싶은 특정 데이터를 데이터 랜덤 배치(Random placement) 과정을 통해서 데이터의 프라이버시를 보호할 수 있다. 또한, 사용자는 기존에 블록체인에 랜덤 배치를 통해 저장된 데이터들을 불러와 복원(Restore)과정을 거쳐 원본 데이터를 불러올 수 있다. 접근 방안의 전체적인 과정을 살펴보면 원본 데이터를 인코딩하고 샤드의 개수만큼 데이터를 분배하여 가로와 세로 순으로 데이터를 섞고 사용자가 데이터를 요청하면 원본 데이터로의 복원과 인증을 하는 방식이다. 데이터 랜덤 배치를 진행하는 이유는 원본 데이터를 유추 불가능하게 하여 프라이버시를 높이기 위함이다. 데이터를 랜덤 배치를 하기 위해서는 원본 데이터를 바이너리 데이터로 변환하고 블록체인의 샤드 개수에 맞게 분배한 후 Fig. 2와 같이 데이터를 가로와 세로 N x M형식으로 매트릭스화 한다. 매트릭스화가 완료된 상태에서 매트릭스의 가로줄과 세로줄에 배치

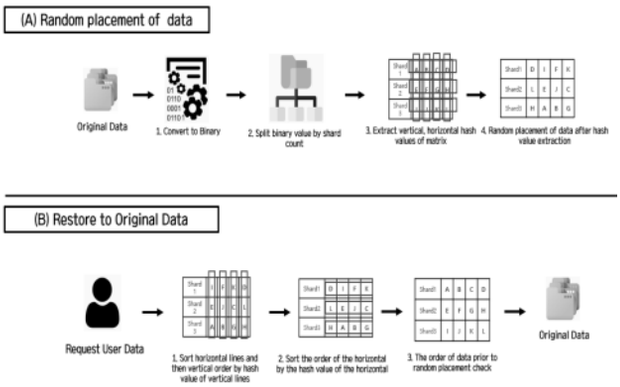


Fig. 2. Shard-based Matrix Character Relocation Technique Process

된 값들의 해시값을 추출함으로써, Fig. 2의 윗부분에 해당하는 원본 데이터 랜덤 배치 과정이 끝이 난다. 이렇게 랜덤으로 배치가 완료된 데이터를 사전에 정의된 샤드 기준으로 블록체인 네트워크에 저장한다. Fig. 2의 아랫부분에 해당하는 랜덤 배치된 데이터를 원본으로 복원할 때는 원본 데이터 랜덤 배치 과정의 역순으로 복원이 진행된다. 사용자가 블록체인에 저장된 데이터를 요청하고 미리 추출했던 해시값을 기반으로 랜덤으로 배치된 매트릭스를 원본의 형태로 재배치한다. 가로줄과 세로줄이 원본과 동일하게 재배치 되었는지 해시값으로 확인한다. 원본 데이터의 랜덤 배치와 저장 그리고 다시 데이터를 복원하는 과정을 다음 장에서 더 구체적으로 설명한다.

3.1 원본 데이터 랜덤 배치

본 장에서는 문자, 이미지, 영상 데이터 등의 특정 데이터를 인코딩 후 매트릭스 화를 통해서 블록체인 샤드에 배치하는 과정을 설명한다. 원본 데이터를 랜덤으로 배치하지 않은 상태에서 각 샤드에 분배 후 저장하는 경우 몇몇 인코딩 된 데이터들이 합쳐져서 원본 데이터의 특정 부분이 유출될 가능성이 존재한다. 따라서 본 논문에서는 샤드에 데이터를 분배하기 전에 전처리 과정을 통하여 데이터의 랜덤 배치를 실행한다. 예를 들어 특정 원본 데이터를 인코딩한 값이 'ABCDEFGHIJKL'이고 블록체인 내 샤드의 개수가 3개라고 가정했을 때 인코딩한 데이터를 매트릭스 화 하면 Fig. 3과 같다. 인코딩한 값을 샤드 1부터 샤드 3까지 분배하고 난 뒤 매트릭스의 가로줄에 속한 값들을 SHA-1 계열의 해시 알고리즘을 적용해 해시값을 얻어낸다. 해당 해시값들을 얻어낸 뒤, 매트릭스의 가로줄을 먼저 랜덤하게 배치한다. 블록체인에서 해시 알고리즘의 충돌 저항성이 높아 안전성이 좋은 SHA-256[15]을 사용하나 제안하는 접근방안에서는 랜덤 배치로 인해 충돌 저항성이 만족 되기 때문에 속도가 빠른 SHA-1 계열의 알고리즘을 사용한다.

매트릭스 가로줄을 랜덤 배치한 이후에, 매트릭스의 세로줄에 해당하는 해시값을 추출하는 과정을 진행한다. 해당 과정은 Fig. 4의 윗부분에 표현이 되어있다. Fig. 4의 윗부분을

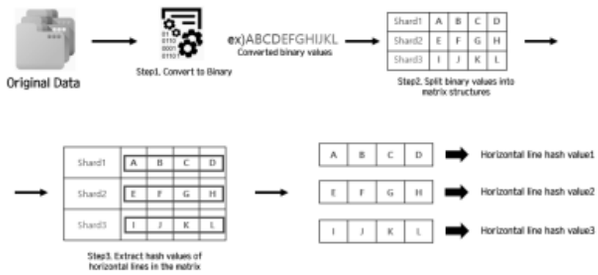


Fig. 3. Extract Hash Values from Matrix Horizontal Lines

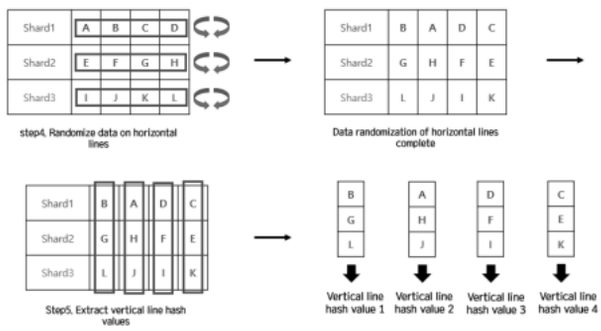


Fig. 4. Extract Hash Values from Matrix Vertical Lines

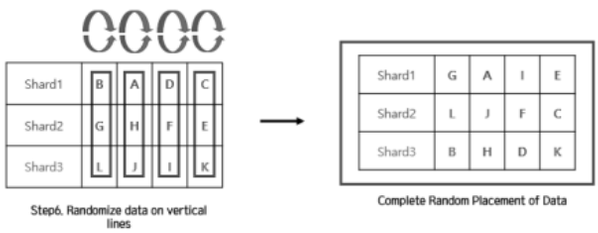


Fig. 5. Complete Random Placement of Matrix

보면 매트릭스의 가로줄을 랜덤하게 배치한 모습을 확인할 수 있으며, 가로줄을 랜덤하게 배치한 상태에서 매트릭스의 세로줄의 해시값을 추출한다. 세로줄의 해시값은 각각 'BGL', 'AGJ', 'DFI', 'CEK'이 생성되고 예시와 같은 4X3 매트릭스에서 총 7개의 해시값이 추출된다.

세로줄의 해시값을 추출하고 난 후에는 가로줄을 랜덤 배치 하였던 것처럼 세로줄을 다시 한번 랜덤 배치하며, 해당 과정은 Fig. 5와 같이 표현될 수 있다. 세로줄의 랜덤 배치까지 완료되면 Fig. 1에서 표현된 데이터 랜덤 배치 과정은 종료된다.

제안하는 접근 방안은 원본 데이터의 바이너리 값이 증가할수록 각 샤드에 할당되는 데이터 값이 많아지고 이에 따라 생성되는 매트릭스의 크기도 커지게 된다. 매트릭스 세로줄의 길이는 초기에 설정해 놓은 블록체인의 샤드 개수를 의미하기 때문에 고정값을 가진다. 하지만, 데이터의 크기에 비례하여 매트릭스의 가로줄의 길이가 길어지고 랜덤 배치된 데이터를 복원할 때의 경우의 수가 기하급수적으로 증가하기 때문에 이 문제를 해결하기 위해서 매트릭스의 블록화가 필요하다. 만약 매트릭스의 블록화를 하지 않았을 시 문자 데이

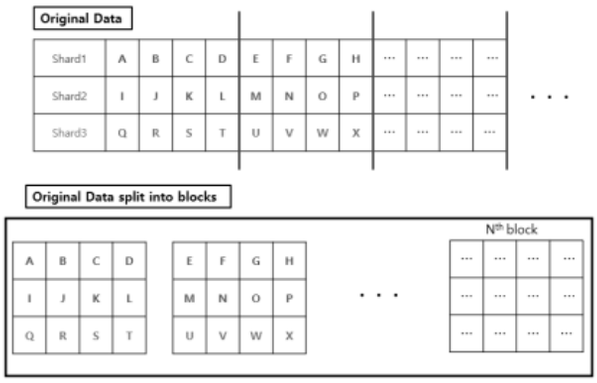


Fig. 6. Blocking the Data Matrix

터 33자리의 중복 값이 없이 없고 샤드가 3개라고 가정했을 때 개 이상의 경우의 수가 발생하게 되고 복원하는 데 엄청난 시간이 소요된다. 따라서 Fig. 6과 같은 원본 데이터 매트릭스의 블록화 과정을 거치게 되고 Fig. 6의 아랫부분의 각 블록을 본 논문에서는 데이터를 이루는 '최소블록'이라고 정의한다.

매트릭스 문자 재배치 기법을 통해 랜덤으로 섞인 원본 데이터를 복원하기 위해서는 앞선 과정에서 미리 추출해 두었던 가로줄과 세로줄의 해시값이 필요하다. 사용자가 블록체인에 저장된 데이터를 요청하면 데이터 매트릭스 블록을 랜덤 배치했을 때의 역순으로 세로줄의 값이 먼저 정렬된다. 해시값으로 원본 데이터를 찾는 과정은 Fig. 7과 같다.

3.2 원본 데이터 복원 과정

Fig. 7과 같이 원본 데이터의 가로줄 혹은 세로줄의 값이 'ABCD'라고 가정했을 때 원본'ABCD'의 해시값은'fb2f85c88567f3c8ce9b799c7c54642d0c7b41f6'이라고 가정한다. 원본의 값이 Fig. 7과 같이 'BACD', 'DABC', 'CADB' 등의 값처럼 배치되어 있을 때 원본의 배치인 'ABCD'로 복원하기 위해서 다시 한번 랜덤 함수가 사용된다. 'BACD', 'DABC', 'CADB' 등의 배치 해시값을 생성하고 원본값의 저장된 해시값 'fb2f85c88567f3c8ce9b799c7c54642d0c7b41f6'과 일일이 대조하는 작업을 거친다. 이 과정은 마치 작업 증명(Proof of Work)[16]와 유사한 방식이고 세로줄이 먼저 정렬된다.

세로줄이 정렬된 상태는 원본 데이터의 매트릭스에서 가로줄의 값만 랜덤으로 배치된 상태이기 때문에 가로줄의 해시값을 세로줄과 같은 방식으로 대조하는 과정을 거쳐 데이터의 원본값으로 복원이 완료된다.

4. 실험 및 안정성 평가

본 장에서는 제안한 접근 방안에 대한 안전성 및 탈중앙성에 대한 평가를 정성적으로 진행하며, 상황별로 매트릭스 문자 재배치된 데이터가 원본 데이터로 되돌아가는 시간을 측정하였다.

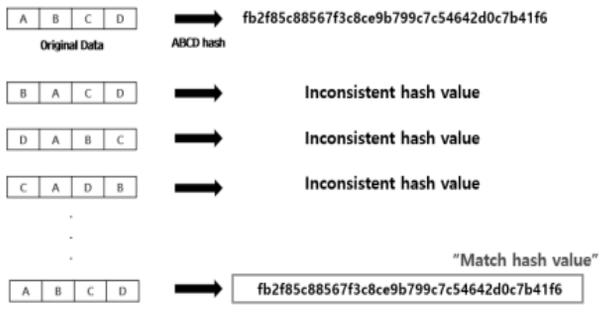


Fig. 7. The Process of Finding the Source Value using Hash Values

실험에서 일반 데이터를 바이너리 데이터로 바꾸기 위해서 블록체인에서 가장 널리 사용되는 Base64 방식을 사용하였다. 또한 바이너리 값을 해시화 시키기 위해서 접근 방안에서 제안하였던 SHA-1 방식을 적용하였다. 또한, 샤드의 개수는 2개, 3개, 4개로 가정하고 실험을 진행했다.

4.1 랜덤 배치 후 샤드에 분배된 데이터의 안정성과 탈중앙성

본 장에서는 제안하는 접근 방법의 안전성과 프라이버시, 인증에 대한 탈중앙성의 정성적인 평가를 진행한다. 각 샤드에 랜덤 배치된 원본 데이터가 분배되면 문자의 순서가 뒤섞여 있기 때문에 각 샤드에 속한 노드들은 블록체인에 저장된 데이터의 원본 데이터의 형태를 알 수 없다. 예를 들어 각 샤드에 저장된 데이터가 유출된다고 해도 블록체인의 내부 노드들뿐만 아니라 블록체인 네트워크 외부에서도 블록체인에 저장되어있는 데이터들의 원본을 알 수가 없다.

원본 데이터를 알기 위해서는 모든 샤드가 보유하고 있는 해시값이 필요하며, 해당 해시값이 가로를 의미하는지 세로를 의미하는지도 알고 있어야 한다. 만약 해시값이 악의적인 사용자에 의해 노출이 되더라도 원본이 어떠한 규칙으로 배치되어있는지 알 수 없고, 규칙을 모르는 상태에서 원본 데이터를 찾아내기 위한 조합법의 경우의 수는 32바이트 기준으로 번으로서, 산술적으로 매우 큰 수이다.

또한, 악의적인 사용자가 운 좋게 전체 데이터 블록의 일부분만 복원을 시킨다고 하더라도 나머지 데이터에 대한 정보를 알 수 없어서 전체 데이터의 원본은 완벽하게 알 수 없고, 본 논문의 접근 방안에 의해서 원본 데이터를 유추할 수 없어서 안전성이 높다고 평가할 수 있다. 또한, 본 논문에서는 원본 데이터를 요청한 사용자가 직접 데이터를 검증할 수 있어서 기존의 관련 연구보다 탈중앙성이 높고 검증 의존도가 낮다고 볼 수 있다.

4.2 샤드의 개수에 따른 최소블록 원본 복원시간

접근방안에서 언급했듯이, 본 논문에서의 샤드는 블록체인 내부의 그룹의 숫자를 의미함과 동시에 원본 데이터를 몇 등분으로 나눌 것인지를 의미한다. 본 실험에서 샤드의 개수에 따라 원본 데이터의 데이터 매트릭스를 이루는 최소한의 블록의 복원 시간을 측정했다.

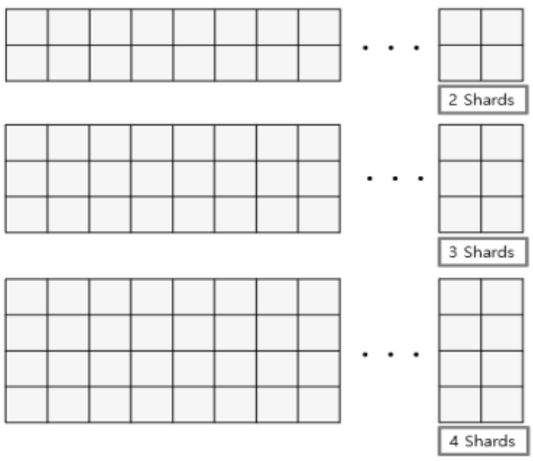
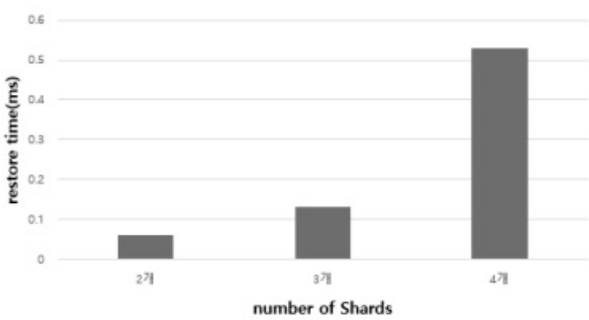


Fig. 8. Data Matrix by Number of Shards



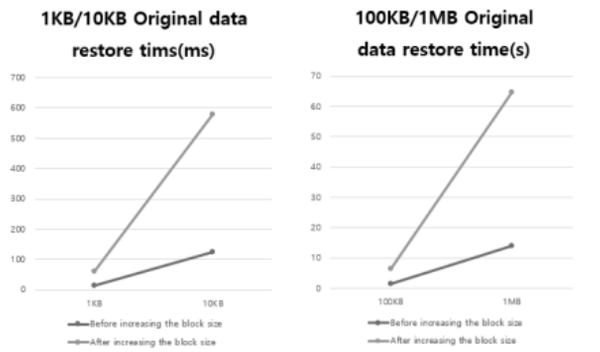
number of Shards	2 shards	3 shards	4 shards
restore time(ms)	0.06	0.13	0.53

Fig. 9. Minimum Block Restore Time According to Number of Shards

원본 데이터의 가로줄을 일정한 길이로 분할해서 블록화를 진행하는데 최소 블록의 가로 길이를 3으로 가정 했다. Fig. 8은 샤드의 개수에 따라서 원본 데이터가 어떻게 블록화가 진행되는지 도식화하여 표현한 그림이다. Fig. 8에 표현된 방식으로 최소 블록을 생성하고 최소 블록의 크기에 따른 원본 데이터 복원 시간은 Fig. 9와 같다. 샤드가 2개일 때 최소 블록의 복원 시간은 0.06ms이고 샤드가 3개, 4개 일 때의 최소 블록의 복원 시간은 각각 0.13ms, 0.53ms이다. 샤드의 개수가 늘어날수록 세로줄 랜덤 배치에 대한 경우의 수가 2부터 4팩토리얼까지 증가하고 그에 따른 세로줄의 원본 값과 해시값을 찾기 위한 대조과정이 많아지기 때문에 복원 시간이 늘어나는 것을 확인할 수 있다.

4.3 원본 데이터의 용량과 샤드의 개수에 따른 데이터 복원시간

본 장에서는 블록체인에 저장할 데이터의 크기에 따른 시간을 샤드의 개수 변화에 따라서 측정했다. 원본 데이터의 크기는 1KB, 10KB, 100KB, 1MB로 설정하여 실험을 진행했다. Fig. 10은 샤드의 개수와 원본 데이터의 크기에 따른 복원 시간을 의미한다. 샤드가 2개일 때 1KB의 복원 시간은 약



Data volume	1KB	10KB
Before increasing the block size	12.7ms	123.376ms
After increasing the block size	59.74ms	577.734ms

Data volume	100KB	1MB
Before increasing the block size	1.35s	13.854s
After increasing the block size	6.287s	64.785s

Fig. 10. Restore Time before and after Doubling the Data Matrix Block Size

12.7ms이고 10KB일 때는 약 복원 123ms이다. 데이터의 크기가 약 10배 차이 나는 것과 같이 복원 시간도 약 10배의 차이가 나는 것으로 확인되었다. 그래프를 통해 샤드가 늘어날수록 복원 시간이 늘어나고 샤드의 개수가 일정할 때 원본 데이터의 크기가 증가할수록 데이터의 복원시간이 일정하게 비례하여 증가한다. 이런 결과가 나온 이유는 결국 원본 데이터는 수많은 최소블록으로 이루어져 있기 때문에 샤드의 개수가 같을 때 데이터의 크기와 복원 시간은 비례한다.

4.4 본 데이터의 용량과 샤드의 개수에 따른 데이터 복원시간

본 논문에서 제안하는 기법은 원본 데이터 매트릭스를 분할할 때 블록의 가로줄의 값이 늘어날수록 원본 데이터의 총 블록의 개수는 줄어든다. 즉 가로줄의 길이가 길어지면 원본 데이터를 구성하는 최소 블록의 크기가 커지고 결국 복원하는 과정에서 해시값을 찾는 경우의 수가 늘어나고 프라이버시 보호가 더 강화된다. 따라서 최소 블록의 가로의 길이를 적절한 값으로 설정할 필요가 있다.

Fig. 11은 같은 크기의 데이터와 샤드의 수를 가질 때 최소 블록의 크기를 늘렸을 때의 복원시간을 보여준다. 1KB의 데

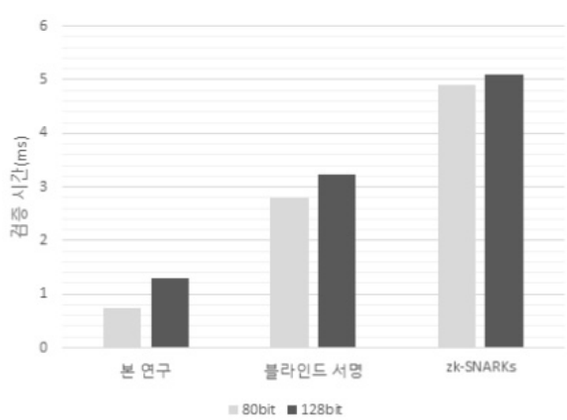


Fig. 11. Blind Signature, zk-SNARKS Verification Time Graph

이터를 샤드가 2개일 때 복원시간을 측정했다. 최소 블록의 크기를 2배로 늘리지 않았을 때의 복원 시간은 약 12.7ms지만 최소 블록의 크기를 2배로 늘렸을 때의 복원 시간은 약 59.7ms이다. 실험을 통해 데이터 최소 블록의 크기가 증가할수록 원본의 해시값을 찾는 경우의 수가 늘어나고 복원 시간이 늘어나는 것을 확인할 수 있다.

4.5 블라인드 서명, zk-SNARKs와 검증시간 비교

본 논문의 기법을 적용했을 때와 기존의 블라인드 서명, zk-SNAKRs의 검증시간을 비교하였다. 원본 데이터가 80bit일 때 블라인드 서명과 zk-SNARKs의 검증시간은 각각 2.79ms, 4.9ms이지만 본 논문의 기법을 적용했을 때의 검증시간은 0.73ms이고 원본 데이터가 128bit일 때는 각각 1.29ms, 3.23ms, 5.1ms의 검증시간이 측정되어 본 논문의 기법을 적용하였을 때 검증시간이 개선되었음을 확인할 수 있다.

5. 결론 및 향후 연구

블록체인의 안에 저장되는 데이터들의 프라이버시 보호를 위한 연구는 계속 진행됐으며 앞으로도 중요한 문제로 자리잡을 것이다. 본 논문에서는 기존에 블록체인 데이터의 프라이버시를 지키기 위한 연구들과 차별화된 접근 방안을 제안하였다.

본 논문에서 제안하는 방안은 크게 원본 데이터 랜덤 배치 과정과 랜덤 배치된 데이터를 원본으로 복원하는 두 과정으로 구분이 된다. 원본 데이터 랜덤 배치 과정에서는 원본 데이터를 해시화 하여 매트릭스에 랜덤 배치하고 각 샤드에 저장하는 과정을 제안하였다. 또한, 랜덤 배치된 데이터를 원본으로 복원하는 과정에서는 기존 블록체인에서 사용되는 PoW 방식을 적용하여 각 샤드에 배치된 데이터를 원본으로 되돌리는 접근 방안 또한 제안하였다.

또한, 실험을 통해서 접근방안에서 제안하는 각 과정에 대한 처리 속도를 측정하였으며, 정성적인 평가를 통해 본 논문에서 제안하는 방식이 충분히 안전한 방안이라는 것을 증명하였다. 하지만 본 논문에서 제안하는 접근방안은 프라이빗 블록체인 환경에서만 적용이 가능하고 블록체인 플랫폼에 따라서 인코딩 방식을 다르게 적용해야 한다는 한계점을 가지고 있다. 이러한 한계점을 넘어서 다양한 블록체인 플랫폼에서도 범용적으로 사용할 수 있도록 하는 향후 연구가 필요하다.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," [Internet], http://bitcoin.org/bitcoin.pdf.

[2] E. Androulaki, et al., "Hyperledger Fabric: A distributed operating system for permissioned blockchains," *EuroSys'18: Proceedings of the Thirteenth EuroSys Conference*, pp.1-15, 2018.

[3] R. Kamath, "Food traceability on blockchain: Walmart's pork and mango pilots," *The Journal of The British Blockchain Association*, Vol.1, Iss.1, pp.1-12, 2018.

[4] D. Chaum, "E. vanHeyst, Group siganture," in *Advances Cryptology*, pp.257-265, 1991.

[5] S., Zhang and J. H. Lee, "A group signature and authentication scheme for blockchain-based mobile-edge computing," *IEEE Internet of Things Journal*, Vol.7, No.5, pp.4557-4565, 2019.

[6] F. Zhang and K. Kim, "ID-based blind signature and ring signature from pairings. In *International Conference on the Theory and Application of Cryptology and Information Security*, Springer, Berlin, Heidelberg, pp.533-547, 2002.

[7] X. Li, Y. Mei, J. Gong, F. Xiang, and Z. Sun, "A blockchain privacy protection scheme based on ring signature," *IEEE Access*, Vol.8, pp.76765-76772, 2020.

[8] N. Shen, "Ring signature confidential transactions for monero," IACR Cryptology ePrint Archive, Vol.2015 pp.1098, 2015.

[9] D. Chaum, "Blind signatures for untraceable payments," In *Advances in Cryptology*, Springer, Boston, MA, pp.199-203, 1983.

[10] C. Li, Y. Tian, X. Chen, and J. Li, "An efficient anti-quantum lattice-based blind signature for blockchain-enabled systems," *Information Sciences*, Vol.546, pp.253-264, 2021.

[11] Z. Cai, J. Qu, P. Liu, and J. Yu, "A Blockchain smart contract based on light-weighted quantum blind signature," *IEEE Access*, Vol.7, pp.138657-138668, 2019.

[12] M. H. Murtaza, Z. A. Alizai, and Z. Iqbal, "Blockchain based anonymous voting system using zkSNARKs," In *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, pp.209-214, 2019.

[13] Q. Li and Z. Xue, "A Privacy-protecting authorization system based on blockchain and zk-SNARK," In *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies*, IEEE, pp.439-444, 2020.

[14] H. Dang, T. T. A. Dinh, D. Loghin, E. C. Chang, Q. Lin, and B. C. Ooi, "Towards scaling blockchain systems via sharding," In *Proceedings of the 2019 International Conference on Management of Data*, pp.123-140, 2019.

[15] D. Rachmawati, J. T. Tarigan, and A. B. C. Ginting, "A comparative study of Message Digest 5 (MD5) and SHA256 algorithm," In *Journal of Physics: Conference Series*, IOP Publishing, Vol.978, No.1, pp.012116, 2018.

[16] A. Gervais, G. O. Karame, K. Wust, V. Glykantzis, H. Ritzdorf, and S. Capkun, "On the security and performance of proof of work blockchains," In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp.3-16, 2016.

이 열 국

https://orcid.org/0000-0001-7450-7742
e-mail : hotgodj@hanmail.net
2018년 삼육대학교 컴퓨터시스템(학사)
2021년 서강대학교 컴퓨터공학과(석사)
2021년 ~ 현 재 한영회계법인 사원
관심분야 : 블록체인 프라이버시

서 중 원

https://orcid.org/0000-0002-3370-0551
e-mail : Jungwonrs@gmail.com.
2016년 State University New York at Buffalo (SUNNY) Management Information System(학사)
2020년 서강대학교 컴퓨터공학과(석사)
2020년 ~ 현 재 서강대학교 컴퓨터공학과 박사과정
관심분야 : 블록체인, 합의알고리즘

박 수 용

https://orcid.org/0000-0002-3979-0586
e-mail : sypark@sogang.ac.kr
1986년 서강대학교 컴퓨터공학과(학사)
1988년 Florida State University, Computer and Information Science(석사)
1995년 George Mason University, Information Technology (박사)
1998년 ~ 현 재 서강대학교 컴퓨터공학과 교수
2017년 ~ 현 재 지능형 블록체인연구센터 센터장
2018년 ~ 현 재 한국블록체인학회 학회장
관심분야 : 소프트웨어공학, 블록체인