사일로 데이터 신뢰 협업을 위한 다중 식별자 동의 기반의 데이터 공유 모델에 관한 연구

라경진, 김태훈, 이임영
순천향대학교 소프트웨어융합학과

rababi@sch.ac.kr, 20134101@sch.ac.kr, imylee@sch.ac.kr

A Study on Multi-Identifier Agreement based Data Sharing Model for Trustworthy Cross-silo Data Collaboration

Gyeongjin Ra, Taehoon Kim, Imyeong Lee
Department of Software Convergence Soonchunhyang University

요 약

최근 4 차산업혁명의 도래와 ICT 기술의 발전은 거대한 양의 데이터를 생성, 수집, 가공, 분석하여 사용자에게 가치 있는 정보를 제공함에 따라, 가용성과 보안 및 프라이버시 사이의 적절한 균형점이 요구되고 있다. 이를 위해 안전한 데이터 공유 모델은 인증 서버 IDP (Identity Provider)가 사용자의 신원을 보증하고 암호화통신의 기본 설정을 중재하여, 당사자의 신원 증명과 데이터 접근의 다양한 프로비저닝을 제공함으로써 당사자간에 암호화 세션을 통해 보안공격에 안전한 신뢰 채널을 형성한다. 하지만 기존의 중앙집중식 공유 모델은 호환성 문제와 단일오류지점을 야기하고, 분산 데이터 공유 모델은 편파 된 데이터사일로 무결성 문제와 서버간 신뢰성 문제가 발생한다. 또한 데이터통합시 사용자 재식별, 추론 등의 프라이버시 문제를 고려하여야 한다. 본 논문에서는 사일로 데이터협업시 가용과 보안을 향상시키기 위해 탈 중앙화 된 신뢰 데이터 공유 모델을 제안한다. 이를 위해 IPDL (InterPlanetary Distributed Ledger) 프레임워크는 DID (Distributed Identity) 기반의 검증 가능한 자격증명과 분산파일시스템을 통해 데이터 자격증명 발급 및 분산 저장, 조회를 제공하여 변조검증을 보증한다. 이때 MAP (Multi-identity agreement processing)은 Private set intersection 암호학 도구를 통해 당사자간의 비밀 계산계약체계를 형성하여 프라이버시를 보호하고, 다중 식별자가 하나의 원자성을 갖도록 동적 업데이트를 수행한다. 기존의 신뢰 데이터 모델을 분석하여 본 연구의 타당성을 제시한다. 이후 프라이버시와 기밀성, 가용성 및 확장성을 고려한 신뢰 데이터 모델을 설계하고 보안 분석을 통해 효율적이고 안전한 데이터 모델 설계를 입증하였다.

Ⅰ. 서 론

최근 4 차산업혁명의 도래와 ICT 기술의 발전은 거대한 양의 데이터를 생성, 수집, 가공, 분석하여 사용자에게 가치 있는 정보를 제공한다. 정보의 가용성과 정확성을 위해 다양한 데이터 공유는 필수적이며, 보안과 프라이버시 또한 중요한 고려사항으로 둘 사이의 적절한 균형이 필요하다. 이를 위해 안전한 데이터 공유는 통신 당사자간에 인증 서버 IDP (Identity Provider)가 사용자의 신원을 보증하고 암호화통신의 기본 설정을 중재하여, 당사자의 신원 증명과 데이터 접근의 다양한 프로비저닝을 제공함으로써 당사자간에 암호화 세션경로를 형성하여 보안공격에 안전한 신뢰 채널이 설정된다 [1]. 하지만 기존의 중앙집중식 공유 모델은 호환성 문제와 단일오류지점을 야기하고, 분산 데이터 공유 모델은 편파 된 데이터사일로 무결성 문제와 서버간 신뢰성 문제가 발생한다. 따라서 최근 DID (Distributed Identity)의 분산식별자와 Verifiable Credential Data Model 은 데이터의 소유권을 변조 검증이 가능한 불변의 블록체인 원장에 기록함 함에 따라, 자격증명을 주장함으로써 탈중앙 신뢰 데이터 공유 모델이 제안되었다. 이후 블록체인의 오버헤드와 프라이버시 문제를 해결하기 위해 분산파일시스템 등의 하이브리드 블록체인이 연구되고 있지만 암호학 프로토콜와 결합된 구체적인 연구는 미흡하다. 따라서 본 논문에서는 사일로 데이터 협업 시 가용과 보안을 향상시키기 위해 탈 중앙화 된 신뢰 데이터 공유 모델을 제안한다. 기존 DID 의 확장 모델로, IPDL (InterPlanetary Distributed Ledger) 프레임워크는 검증 가능한 자격증명과 분산파일시스템을 통해 데이터 자격증명 발급 및 분산 저장, 조회를 제공하여 변조검증을 보증한다. 또한 MAP (Multi-identity agreement processing)의 Private set intersection query processing 은 암호학 도구를 통해 당사자간의 비밀 계산계약체계를 형성하여 프라이버시를 보호하고, 다중 식별자가 하나의 원자성을 갖도록 동적 업데이트를 수행한다 [1].

Ⅱ. 관련연구

최근 IDP 패러다임은 탈중앙형 DID 로 사용자의 디지털식별자 URL 과 디지털식별자를 정의한 문서를 매핑하고 이를 블록체인 원장에 기록하는 검증 가능한 자격증명 모델이 연구되었다. 따라서 사일로 데이터 협업 환경에서 TTP 서버의 신뢰 문제를 해결하고, 위조와 변조의 내성을 가진다. 이후 원장의 프라이버시를 위해 데이터 공유 시 영지식 증명 등의 비밀계산 공유 모델이 연구되었다 [2]. 이러한 솔루션은 상호 운용성 및 공동 작업을 위해 서로 다른 서버를 페더레이션으로 구축하고 개인 정보 소유자가 개인 정보 보호 기술을 통해 익명화

프로세스를 제어하고 서비스 보안을 강화할 수 있도록 요구된다. 따라서 본 논문에서는 단일블록체인 방식과 달리 계산 오버헤드와 프라이버시를 고려하여, 분산파일시스템 기반의 하이브리드 블록체인과 암호학 도구의 결합으로 신뢰 데이터 협업을 위한 프라이버시 데이터 공유 프로토콜을 제안한다.

Ⅲ. 제안방식

블록체인 내부 시스템에 따라, 사용자는 공개키 인증서를 통해 암호화 데이터의 자격증명을 요청한다. 제안방식은 전체 2 개의 데이터 생성 및 분산저장 IPDL Phase, 동적업데이트 MAP Phase 로 구성되며 각각 하위 Step 을 가진다. 사용자와 연결된 로컬서버는 사용자의 데이터를 수집하고, Elgamal 암호화 연산을 적용한다. 이때 암호화 계산을 위해 비트 단위로 암호화하여 XOR 동형을 유도한다. 다음의 알고리즘을 수행한다.

- ElGl.Enc (pk, m), 공개키를 통한 메시지의 ElGamal 암호화를 수행한다.
- ElGl.Dec 비밀키와 암호문 ElGl.Enc (pk, m)이 주어지면 ElGl.Dec 을 실행하여 일반 텍스트 m 을 복구한다.
- ElGl.Mul 공개 키 및 일련의 암호문 ElGl.Enc (pk, mi) 암호화 메시지 mi 가 주어지면 기본 메시지의 추출형태를 암호화하는 암호문을 동형으로 계산한다.
- ElGl.Enc (pk, ∏mi) = ElGl.Mul ({ElGl.Enc (pk, mi)} i)을 통해 암호화의 랜덤 순열조합을 계산한다.

1. IPDL Phase

Step 1. 사용자는 자신의 데이터를 ElGl.Enc (pk, m)하여 핸들링서버에 전달한다.

Step 2. 핸들링서버는 메타데이터를 생성하고 해시를 통해 CID (Contents Identifier)을 생성하여 글로벌 서버 간 블록체인을 통해 커밋한다.

2. MAP Phase

Step 1. 참가자 pi 로부터 형식 (query, CID, f)의 메시지를 수신할 때, 여기서 f = (f1, f2)는 두 데이터베이스에 대한 두 당사자의 비밀계산을 위한 함수이다. F 는 Enc (pk, mi)의 암호화 및 복호화 등의 계산 수행을 포함한다. 여기서 D 는 당사자의 데이터베이스이고 kenc 는 암호화 키이다.

Step 2. 각 당사자 pi 는 튜플 입력 = (G, kenc, ρ)입력은 여기서 G = (V, E)는 그래프로 매핑되며, kenc 는 암호화 키이다.

Step 3. 전체 map Λ: Dp1∪ Dp2 → U 를 계산한다. 여기서 U = {0,1}은 사용자 ID 를 의미한다. 사용자 Λ (S1) = Λ (S2)의 교집합인 S1 /~ S2 인 행 Λ (S)를 rowS.Foreachrow S ∈D 에 할당된 CID 를 의미한다.

Step 4. 각 pi 에서 (submit, CID, input_pi)를 수신하면 U = {0,1}에서 임의의 요소를 샘플링하고 mapΛ를 계산한다. Vp1 ∪Vp2 → {0,1}에서 v1, v2 ∈Vp1 ∪Vp2 두 꼭지점에 대해 Λ (v1) = Λ (v2) v1 과 v2 가 G1 ∪G2 의 동일한 연결된 구성 요소에있는 경우에만 메세지를 참가자 pi 로 보낸다.

(a) 양쪽 모두 CID 에 대한 설정 메시지를 보내지 않은 경우 메시지를 무시한다.

(b) 그렇지 않다면 (query, CID, (f1, f2), i)를 기록한다. (query, CID, (f1, f2), 3 − i)를 이미 기록한 경우 (response, CID, f1 (Dp1, Dp2))를 p1 로 보내고 (response, sid, f2 (Dp1, Dp2))를 p2 로 보낸다.

Ⅳ. 제안방식 분석

1. Security

(1) Authentication/ Confidentiality: DDH(Decisional Diffie-Hellman)의 확률적 다항식 시간 알고리즘으로 구성된 암호화 체계에 기반한다. ElGl.Gen 보안 매개 변수 λ 가 주어지면 ElGl.Gen (λ)은 공개-개인 키 쌍 (pk, sk)을 반환하고 메시지 공간 M 을 지정한다. 암호화 프로토콜의 끝에서 각 당사자는 상대방에게만 알려진 키로 암호화된 정점 레이블을 수신함으로써 정당한 사용자만이 안전한 메세지를 획득하여 인증과 기밀성을 보증한다.

(2) Integrity: CID 는 해시함수의 비가역성과 암호화에서 랜덤 오라클은 출력 도메인에서 균일하게 선택된 랜덤 응답으로 모든 고유한 쿼리에 응답하는 오라클 모델로 데이터 협업 시 데이터의 무결성을 보증한다.

(3) Reliability: 탈중앙 분산원장을 통한 공모공격과 단일지점장애로 내성을 통해 중간자공격과 재전송공격 등의 안전한 통신을 보증한다.

(4) Privacy: CPA(ciphertext indistinguishability)보안 암호화 체계에 따라 어느 쪽도 m 을 학습하지 않고 당사자에게만 알려진 키 k 로 암호화 되어야하는 회로 내에서 계산된 메시지 m 을 수행한다. 그룹의 식별 요소를 사용하여 ElGl.Mu을 사용하여 암호문을 다시 무작위화하여 동일한 길이의 평문 두 개를 임의로 선택하여 암호화한 것이 공격자 A 에게 주어졌을 때, A 는 둘 중에서 어떤 암호문이 자신이 원하는 암호문인지 선택할 확률은 매우 희박하다.

2. Efficiency

하이브리드 블록체인으로 암호문을 외부서버에 저장하고 해시로 계산된 데이터식별자와 자격증명 값을 동적 업데이트하여 원장의 기록을 중복제거함에 따라 효율적이다. 또한 ElGl.Mul 을 사용하여 비트 단위로 암호화하여 XOR 동형을 유도한다. 이를 통해 프로토콜의 왜곡 된 회로 부분에서 그룹 작업을 수행하는 것을 피하여 CPU 통신 비용을 줄인다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음 (2021-0-01516)

참 고 문 헌

[1] Kreuter, B., Patel, S., & Terner, B. (2020, September). Private Identity Agreement for Private Set Functionalities. In International Conference on Security and Cryptography for Networks (pp. 172-191). Springer, Cham.

[2] Gouru, N., & Vadlamani, N. (2021). Cops-cooperative provenance system with zkp using ethereum blockchain smart contracts. In Research Anthology on Blockchain Technology in Business, Healthcare, Education, and Government (pp. 572-586). IGI Global.