데이터 프라이버시를 보존하는 내적 연산 프로토콜 제안

성유정 [1], 김민지 [1], 김가은 [1], 손용하 [2]
[1] 성신여자대학교 융합보안공학과 학부생
[2] 성신여자대학교 융합보안공학과 교수

20230869@sungshin.ac.kr, 20221331@sungshin.ac.kr, 20240927@sungshin.ac.kr, yongha.son@sungshin.ac.kr

A Privacy-Preserving Inner Product Protocol

Yujeong Sung[1], MinJi Kim[1], Gaeun Kim[1], Yongha Son[2]
[1, 2]Dept. of Convergence Security Engineering, Sungshin Women's University

요 약

최근 민감 데이터 보호의 필요성이 대두되면서, 서로가 가진 데이터를 숨기면서도 필요한 연산을 진행해 그 결과만 공유하는 프로토콜의 중요성이 부각되고 있다. 그러나 기존의 Decisional Diffie-Hellman (DDH) 기반 방식은 통신량 측면에서 효율적이지 않다. 본 논문은 Circuit-PSI 와 동형암호를 사용한 안전한 데이터 내적을 통해 통신량 문제를 해결하면서도 데이터의 보안을 유지하고 필요한 결론만을 공유할 수 있는 방식을 제안한다. 통신량 계산에 따르면 기존의 DH 기반 방식 대비 약 2.5 배에 달하는 통신량을 절감할 수 있다.

1. 서론

본 연구는 각자 데이터베이스를 가진 두 당사자가 서로의 데이터를 공개하지 않은 상태에서 데이터를 안전하게 조인한 후, 결합된 데이터베이스 위에서 특정한 함수를 계산하는 암호 프로토콜을 제안한다. 일반적으로, (식별자, 값) 쌍의 데이터를 가진 두 당사자가, 식별자를 기준으로 데이터베이스를 내부 조인한 뒤 해당하는 값들의 내적(Inner-Product)를 계산하는 것이 프로토콜의 목표이다.

X

식별자	값
aa@a.a	d_1
bb@b.b	d_2
cc@c.c	d_3
dd@d.d	d_4
ee@e.e	d_5

Y

식별자	값
xx@x.x	d'_1
bb@b.b	d'_2
ee@e.e	d'_3

$\Rightarrow \sum d_x \cdot d_y = d_2 \cdot d'_2 + d_5 \cdot d'_3$

(그림 1) 두 데이터베이스의 내적 연산

이러한 프로토콜은 정보 공유가 제한되어 있는 의료계에서 유용하게 사용될 수 있다. 예를 들어, 임상시험 참가자의 병원 방문 횟수를 알고 있는 제약회사와, 건강 지표에 따른 가중치를 갖고 있는 병원이 있을 때, 각 기관의 데이터를 공유하지 않고도 참여자 집단에 대한 임상적 효과를 도출할 수 있도록 해준다.

이러한 기술에 대한 수요가 생긴지 오랜 시간이 지나지 않았기 때문에, 알려진 기술들의 성능은 실용화 단계와는 거리가 먼 상태이다. 특히 많은 수의 프로토콜들이 타원곡선 공개키 암호 기술을 기반으로 설계되었으며, 이는 충분한 안전성을 보장할 수 있으나 타원 곡선의 내재적인 비효율성으로 인해 성능 문제가 발생하는 상황이다. 본 연구에서는 이러한 성능적 한계를 개선하는 신규 프로토콜 설계 방안을 제안하고자 한다.

2. 선행 연구

K. Chida.[1]은 타원곡선 군 상의 Decisional Diffie-Hellman(DDH)을 기반으로 안전하게 교집합의 내적을 계산하는 프로토콜을 제안했다. 그러나 DDH 기반 방식은 타원곡선 군 연산을 요구하기 때문에 계산 속도가 느리고, 통신을 주고받는 대상 역시 타원곡선 군의 점들이므로 통신량이 크다. 구체적으로, 타원곡선 그룹 G 의 한 원소는 $\mathbb{Z}_p$상의 두 점으로 표현되며, 보안을 위해 p 를 2^{256}으로 설정하기 때문에 점 하나가

32 바이트를 차지한다. 실제 프로토콜에서는 데이터 하나 당 타원곡선 원소 7 개의 통신이 필요하고, 결과적으로 450 바이트 가량의 통신량이 요구된다.

3. 제안 기법

본 장에서는 Circuit-PSI 와 동형암호(Homomorphic Encryption, HE)를 기반으로 한 새로운 데이터 내적 프로토콜을 제안한다. Circuit-PSI 는 Cuckoo hashing 및 Oblivious Key-Value Store 기반 OPRF 를 이용하여 빠른 속도와 통신량을 보이는 프로토콜이다. 이를 통해 두 참여자는 각 식별자에 대해 Boolean share 를 안전하게 생성할 수 있다. 각 당사자는 자신의 share 만을 보유하며, 두 share 의 XOR 결과는 두 식별자가 동일할 경우 1, 다를 경우 0 이 된다. 이때, XOR 연산이 수행되기 전에는 상대의 입력에 대한 정보를 유추할 수 없으므로 안전하다[3]. 따라서 Circuit-PSI 를 사용하여 식별자 정렬을 맞춘 뒤, 동형암호를 사용하여 데이터에 대한 내적 연산을 한다.

- 송신자와 수신자는 Cuckoo hashing 과 OKVS 를 통해 각 식별자들의 Boolean shares 를 생성한 뒤, 인덱스가 정렬된 상태에서 Boolean-to-Arithmetic 변환을 거쳐 a^R 의 Arithmetic shares 를 얻는다. 이때, a^R 는 식별자가 교집합의 원소일 경우 해당 식별자에 대응되는 수신자의 데이터를, 아닐 경우 0 의 값을 갖는 벡터이다.
- 송신자 측의 데이터를 a 라 하고, 수신자 측의 Arithmetic Share 을 b 라 하자.
- 송신자가 동형 암호의 키 쌍(pk, sk)를 생성하고, 공개키 pk 와 HEnc(pk, a)를 수신자에게 전송한다.
- 수신자는 HEnc(pk, ab)를 계산한 후, 자신의 출력값으로 랜덤한 값 r 을 선택하고, z = HEnc(pk, ab-r)를 송신자에게 전송한다.
- 최종적으로 송신자는 HDec(sk, z)를 계산하여 ab-r 을 출력값으로 얻는다.

이 방법에 따르면 송신자와 수신자는 서로의 입력값을 공유하지 않으면서도, 식별자의 교집합에 대응되는 데이터들에 대한 내적값을 얻을 수 있다.

4. 통신량 비교 분석

<표 1> DDH 기반 방식과의 통신량 비교

n	2^{14}	2^{16}	2^{18}	2^{20}
	Comm. (MB)			
DDH	4.3	17.2	68.8	275.2
OURS	1.7	6.7	27.4	111

프로토콜의 소요 시간은 실제 구현 없이는 개선 여부를 명확히 주장하기 어렵다. 그러나 통신량의 경우 주고받는 요소의 크기를 통해 이론적으로 예측이 가능하므로, 우리는 통신량 비교부터 수행하였다.

<표 1>은 데이터의 길이가 32bit 라고 가정하고, 두 통신 참여자의 데이터의 크기가 각각 $n = 2^{14}, 2^{16}, 2^{18}, 2^{20}$ 인 상황에서 통신량을 계산한 결과이다.

DDH 기반 방식과 비교했을 때, 모든 결과에서 동형암호를 사용한 방식은 약 2.5 배의 통신량을 절감할 수 있었다. 또한 두 프로토콜 모두 데이터의 크기에 선형적으로 비례하는 통신량을 보이므로, 임의의 데이터셋에 대해서도 유사한 수준의 개선을 보일 것으로 기대된다.

5. 결론 및 향후 연구 방향

전통적으로 안전한 데이터 통신에서는 통신 당사자 외의 제 3 의 공격자로 인한 공격만을 고려해왔다. 그러나, 데이터 자체가 하나의 자산이 되고 있는 현대에는 통신 당사자들 간에도 서로의 데이터를 보호할 필요가 생겼으며, 자연히 이에 대한 기술적 해결책에 대한 관심도 높아지고 있다. 이러한 배경에서, 본 논문은 서로 신뢰할 수 없는 두 기관이 보유하고 있는 데이터 간 연산을 지원하는 프로토콜의 신규 설계 방안을 제시하였다.

그 결과 많은 양의 통신을 필요로 한다는 DDH 기반 방식의 문제를 해결하기 위해 동형암호를 사용한 방식을 제안하였으며, 약 2.5 배의 통신량을 절감할 수 있음을 밝혔다.

또한, 본 연구에서 제안한 방법은 타원곡선 연산을 요구하지 않으므로 계산 속도 측면에서도 이점을 가질 것으로 기대된다. 추후 연구로 관련 기술에 대해 알려진 라이브러리들을 활용하여 프로토콜을 구현하고, 다양한 환경에서 성능을 비교·분석할 예정이다.

참고문헌

[1] K. Chida, K. Hamada, A. Ichikawa, M. Kii, and J. Tomida, “Communication-efficient inner product private join and compute with cardinality,” in Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security, 2023, pp. 678–688

[2] T. Lepoint, S. Patel, M. Raykova, K. Seth, and N. Trieu. Private join and compute from PIR with default. In M. Tibouchi and H. Wang, editors, ASIACRYPT 2021, Part II, volume 13091 of LNCS, pages 605–634. Springer, Cham, Dec. 2021. doi: 10.1007/978-3-030-92075-3_21.

[3] Srinivasan Raghuraman and Peter Rindal. 2022. Blazing Fast PSI from Improved OKVS and Subfield VOLE. In CCS 2022, Heng Yin, Angelos Stavrou, Cas Cremers, and Elaine Shi (Eds.). ACM, 2505–2517.