Asia-pacific Journal of Convergent Research Interchange
Vol.9, No.12, December 31 (2023), pp.33-42
http://dx.doi.org/10.47116/apjcri.2023.12.04

Application of SMPC (Secure Multiparty Computation) for Privacy Protection in MyData Environment

마이데이터 환경에서 프라이버시 보호를 위한 다자간 SMPC 적용에 관한 연구

Ji Yeon Lee[1], Soon Seok Kim[2]

이지연[1], 김순석[2]

[1] Professor, Department of AI Convergence Security, Halla University, Korea, jiyeon.lee@halla.ac.kr
[2] Professor, Department of AI Convergence Security, Halla University, Korea, sskim@halla.ac.kr

Corresponding author: Soon Seok Kim

Abstract: The SMPC (Secure Multiparty Computation) protocol is an encryption protocol that allows multiple parties to perform computations on their respective input values without revealing the inputs to each other, while obtaining a jointly computed result. Looking at the progress of My Data in the domestic health and medical field, the demand for medical services from consumers (citizens and individuals) is rapidly increasing. There is a growing interest in health-related matters, and with the rise of ICT-based healthcare services such as transmitting hospital medical records and test results as data, medical big data is being generated and stored. However, these data are not being properly utilized. With the amendment of the Data 3 Act, there is a significant increase in expectations regarding the utilization of healthcare big data. Additionally, with the advancement of artificial intelligence, there is a considerable focus and investment in the medical AI industry, which raises concerns about the protection of personal information. Concrete measures are required to address the scope of providing medical information and the security of sensitive personal information, including information processed from medical professionals' expertise. This paper aims to introduce the Multi-party PSI based on Multi-Point OPRF(Oblivious Pesudo Random Function) as a solution to address privacy protection in such a My Data environment and proposes methods for its application.

Keywords: Secure MultiParty Computation, My Data, Muti-Point OPRF, Pseudo Random Function

요약: 안전한 다자간 연산 (Secure Multiparty Computation: SMPC) 프로토콜이란 여러 사람이 자신이 가진 입력 값을 공개하지 않고도 입력 값을 연산하여 모두가 공동의 연산 결과를 가실 수 있도록 하는 암호화 프로토콜이다. 국내 보건의료분야 마이데이터 추진현황을 살펴보면 수요자(국민, 개인)의 의료서비스에 대한 요구가 급속도록 증가하고 있다. 건강에 대한 관심이 높아지고 있고, 병원간 병원 진료기록이나 검사결과를 데이터로 전송하는 등 ICT기반의 의료 서비스 증가에 따라 의료 빅데이터가 생성, 저장되고 있었으나 개인의 의료 정보 뿐만 아니라 의료진의 전문 지식으로 가공된 정보가 포함되어 있어 의료 정보 제공 범위 와 개인 민감 정보보안 등 문제로 인해 제대로 활용하지 못하고 있었다. 지난 데이터3법의 개정으로 보건의료 빅데이터 활용가능성에 대한 기대가 매우 커지고 있어 본 논문은 이러한 마이

Received: August 05, 2023; 1st Review Result: September 08, 2023; 2nd Review Result: October 12, 2023
Accepted: November 25, 2023

데이터 환경에서 프라이버시 보호를 위한 보안 이슈를 해결하기 위해 PSI(Private Set Intersection) SMPC 기반의 새로운 다중 시점(Multi-Point) OPRF 방법을 제안하였다.

핵심어: 안전한 다자간 연산, 마이 데이터, 다중 시점, 의사랜덤 함수

1. 서론

SMPC(Secure Multiparty Computation)는 개인정보 보호와 다자간 협업을 위한 보안 연산을 제공하는 암호화 프로토콜이다. 한편 SMPC의 장점은 아래와 같이 5가지로 요약해 볼 수 있다.

· 개인정보 보호 : 각 참여자의 입력값은 암호화 되어 다른 참여자에게 전송되므로 개인정보 노출을 방지하여 의료 데이터 등 민감 데이터 활용에 적용할 수 있다.

· 데이터 공유와 협업 : 다수의 참여자가 자신들의 데이터를 보호한 상태에서 연산을 수행하여 연산 결과를 공유함으로써 협업이 가능하다.

· 신뢰없는 환경에서의 안전한 연산 : 신뢰할 수 없는 환경에서도 안전한 연산을 가능하게 하여 결과의 무결성을 보장한다.

· 데이터 소유권 보장 : 참여자는 자신의 데이터를 보호하면서도 다른 참여자와 함께 연산을 수행할 수 있어 데이터의 소유권과 제어권을 유지할 수 있다.

· 다양한 응용 분야 : 의료, 금융, 클라우드 컴퓨팅, 인공지능 등 민감 데이터 활용 분야에서 다양하게 적용할 수 있다.

상기와 같은 장점이 있는 반면 통신 오버헤드의 문제점이 발생된다. 메시지의 크기나 네트워크 대역폭이 제한적일 경우 오버헤드의 문제가 발생할 수 있다.

전통적인 다자간 연산(MPC)[1]는 여러 참여자가 자신의 입력값을 공유하며 연산을 수행하는 참여자간의 신뢰를 기반으로 연산을 수행하는 암호화 기술이였다면, 본 논문에서 제안하는 안전한 다자간 연산(SMPC)은 여러 참여자가 자신이 가진 입력 값을 공개하지 않고도 참여자가 가진 입력 값을 연산하여 모두가 공동의 연산 결과를 가실 수 있도록 하는 안전한 암호화 프로토콜이다. 즉, 개인의 정보를 공개하지 않고 암호화하여 다른 참여자에게 전송하고, 연산 결과만 공개하는데 사용된다. 따라서 개인정보 보호와 데이터 공유의 중요성을 인식하고 있는 마이 데이터 환경에서 유용한 보안 연산 방법으로 인정받고 있다[2].

본 논문의 구성은 다음과 같다. 2장에서는 선행 연구에 대한 고찰과 의료 데이터 분야에서 연구사례를 살펴보고 3장에서는 기존의 SMPC프로토콜을 이루는 사전 지식 및 기술들을 설명한다. 그리고 4장에서는 제안 SMPC 프로토콜에 대해 기술하고 5장에서는 결론 및 향후 연구 방향에 대해 서술한다.

2. 연구 배경

2.1 마이 데이터 환경 특성 정의

마이 데이터란 데이터의 주체가 자신이라는 의미이다. 즉, 정보주체인 개인에게 개인데이터의 권한을 주고, 개인이 본인의 개인정보를 비롯한 모든 정보를 적극적으로 관리, 통제하는 실질적 권한을 부여하는 것이다. 과거 개인 데이터는 개인의 소유로

인정하지 않았는데 그 이유로 기업이나 서비스 제공자 등 사회가 공유하며, 공공 가치 창출을 중요하게 생각하였다[3]. 2012년 핀란드에서 개인정보호 및 개인 데이터에 관한 법률 개정이 이루어지며 마이 데이터란 개념이 확산되게 되었으며, 이후 전세계적으로 개념이 확산 되었다. 마이 데이터는 기본적으로 개인 데이터의 보호와 활용의 양면을 갖고 있으며, 균형과 조화를 모색하는 기본적인 특성을 갖고 있다[3].

2.2 팬더믹 이후 보건 의료 데이터 활용 동향

COVID-19 팬데믹으로 인해 미국 식품의약국(FDA)는 실제 의료 데이터의 활용을 가속화 해야 한다고 강조했다[4]. 이는 실제 보건의료 현장에서 생산되는 개인 의료 데이터를 의미하며, 제한된 범위에서 시행되는 임상시험 데이터와 대비되는 개념으로 일반 병원, 보험사, 국가 기관들이 갖고 있는 보건의료 빅데이터를 의미한다. 팬더믹 이후 국내에서도 이러한 보건 의료 의료빅데이터를 활용하여 의료 정책 수입 및 의료 개발에 활용해야 한다는 목소리가 높아지고 있다.

2.3 마이 데이터 환경에서의 보건 의료 데이터의 보안 이슈[5]

마이 데이터 환경에서 보건 의료 데이터 개방에서 안전한 공유를 위해 보안 이슈를 살펴보면 아래와 같다[5].

· 개인정보보호 : 보건의료 빅데이터에는 개인 식별 정보가 포함되어 있어, 개인정보 침해에 대한 우려의 목소리가 높아 이에 대한 적절한 보호 조치와 법적 규제가 필요하다.

· 불법 접근 및 해킹 위협 : 의료 데이터는 개인의 의료 정보 뿐만 아니라 의료진의 전문 지식으로 가공된 정보 등 민감하고 가치 있는 정보를 포함하고 있기 때문에 불법접근 및 해킹의 위험에 노출될 수 있다.

· 데이터의 무결성과 위조 방지 : 의료 데이터는 중요한 정보를 포함하고 있으므로 데이터의 무결성을 보장하기 위해 암호화, 디지털 서명 등의 보안 메커니즘을 도입해야 한다.

· 접근 권한 관리 : 보건 의료 데이터 개방시 접근할 수 있는 권한을 엄격하게 제한하고, 사용자의 신원 확인과 인증 절차를 강화해야 한다.

3. 관련 연구

다자간 연산(Multiparty Computation: MPC)프로토콜은 1982년 A. C. Yao[6]에 의해 처음 제안되었다. 제안된 MPC의 경우는 양자간의 경우로 두 백만장자가 자신의 재산을 공개하지 않고 자산의 총량을 계산하여 누구의 재산의 크기가 더 큰지 계산하는 시스템으로 백만 장자의 문제라 불린다. 예를 들어 세명의 어린이가 본인이 갖고 있는 돈을 알리지 않고 평균을 계산하고 싶은 경우 MPC의 방법으로 문제를 해결 할 수 있다.

1) 영이는 4000원을 갖고 있으나 친구들에게 알리지 않고 랜덤하게 3개로 나눈다. 4400원, -1100원, 700원으로 나누어 알려준다.

2) 철이는 5000원을 갖고 있으며, -600원, 3200원, 2400원으로 나누어 알려준다.

3) 혁이는 6000원을 갖고 있으며, 2000원, 0원, 4000원으로 나누어 알려준다.

	영이	철이	혁이	합계
	4500원	-1000원	500원	4000원
	-800원	3400원	2400원	5000원
	2000원	1800원	2200원	6000원
합계	5700원	4200원	5100원	

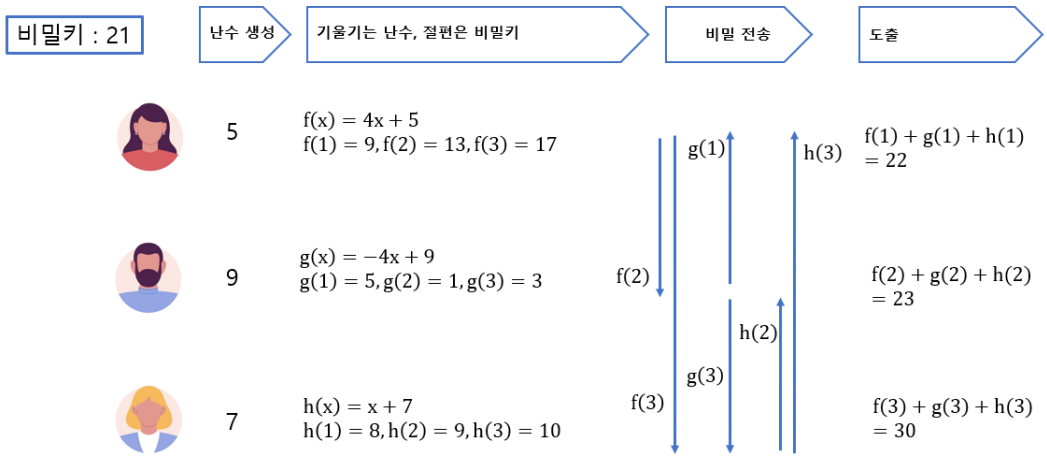
[그림 1] 다자간 연산 예시

[Fig. 1] Example of Muptparty Computation

4) 평균 : (5700 + 4200 +5200) /3 = 5000원

5) 비밀키에 의해 각자 갖고 있는 돈은 암호화되어 서로에게 알려지지 않는다.

3.1 Shamir의 Secret Sharing



[그림 2] Shamir의 비밀키 공유 프로토콜 예시

[Fig. 2] Example of Shamir's Secret Sharing Protocol

다자간 연산(Multiparty Computation: MPC) 프로토콜은 일반적으로 비밀 공유키를 기본 기술로 사용한다[6][7]. 비밀키 공유 프로토콜을 예를 들어 간단하게 설명하면, 개인키가 21인 경우 영이, 철이, 혁이가 각각 5, 9, 7을 나누어 가졌을 경우 외부 제3자가 이 중 한 명의 키를 탈취한다 해도 개인키가 21인 임을 알 수 없으므로 보안에 대한 방어력은 높아졌으나 분실 가능성도 결과적으로 높아졌다. 이에 대한 해법으로 Shamir가 아래와 같은 비밀키 공유 방법을 제안하였다[7].

1) n개의 참여자에 대해 $t < n$인 $t+1$ 개의 점 $(x_i, y_i), i = i, 2, \ldots, t+1$ 을 지나는 최대 t 차 다항식 $q(x)$에 대해 개인키 $s = q(0)$로 가정한다.

2) 이때 해커가 비밀키를 탈취하더라도 t개의 점으로 다항식을 구하여 s을 구할 수 있다.

3) 다항식은 랜덤하게 선택되며, 각 개인에게는 1개씩 주어지며, 개인에게는 하나의 점을 지나는 다항식은 무수히 많으므로 랜덤하게 보여진다.

4) 개인키 분배 과정은 [그림 2]와 같다.

결과적으로 $y = f(x) + g(x) + h(x) = x + 21$ 임을 알 수 있다.

3.2 OT(Oblivious Transfer) 프로토콜[8]

OT 프로토콜은 Rabin에 의해 1981년 처음 소개되었다[8]. 송신자가 수신자에게 메시지를 전송할 때, 송신자가 선택한 정보를 수신자에게 노출시키지 않으면서 전송하는 프로토콜이다. OT는 정보 전송에서 개인정보 보호와 기밀성을 보장하는 암호화 프로토콜로 당사자간의 안전하고 개인정보 보호를 위한 정보교환에 강력한 보안성을 가진 기술이다. 이후 OT는 SMPC 프로토콜의 기본 기술로 사용하기 위해 1-out-of-2 OT로 발전하게 되었다[9]. OT 프로토콜은 아래와 같다.

1) 송신자는 두 개의 메시지 m_0, m_1 을 가지고 있으며, 그 중 하나를 수신자에게 전송하려고 한다. 수신자는 송신자에게 받은 메시지를 공개하지 않는다.

2) 송신자는 두 개의 큰소수 p, q 를 선택하고 그 곱으로 $n(n = p * q)$을 선택한다. 그런 다음개인키 d 와 공개키 e 를 랜덤하게 선택한다.

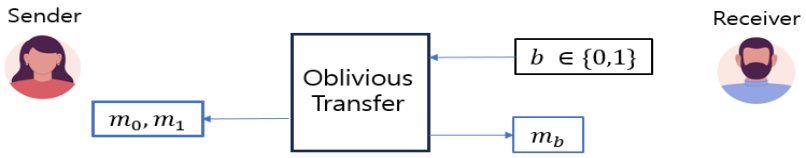
3) 송신자는 두 값 x_0, x_1 을 랜덤하게 선택하여 수신자에게 보내고 수신자는 b, $b \in \{0,1\}$와 r을 랜덤하게 선택한 후 그 중 x_b를 선택한다.

4) 수신자는 자신의 키 r 을 암호화한 v, $v = (x_b + r^e) mod\, n$을 송신자에게 보낸다.

5) 송신자는 다음과 같이 k_0, k_1 을 계산한다. $k_0 = (v - x_0)^d mod\, n$, $k_1 = (v - x_1)^d mod\, n$. 이때 k_0, k_1 둘 중 하나는 k 와 같을 것이나 남은 하나는 의미가 없을 것이다. 하지만 송신자는 선택된 b값을 모르기 때문에 k_0, k_1 중 어느것이 r과 같은지를 알 수 없다.

6) 송신자는 $m_0' = m_0 + k_0$, $m_1' = m_1 + k_1$을 계산하여 수신자에게 모두 보낸다.

7) 수신자는 r값을 알기 때문에 $m_b = m_b' + r$ 는 알 수 있으나 $k_{1-b} = (v - x_{1-b})^d mod\, n$ 값은 계산할 수 없으므로 m_{1-b}는 알 수 없다.



[그림 3] OT(Oblivious Transfer) 프로코톨 예시

[Fig. 3] Example of OT(Oblivious Transfer) Protocol

3.3 PSI(Private Set Intersection)from a Polynimoal based OPRF

Private Set Intersection(PSI) 기술은[10] 에서 각자의 정보를 가진 두 참여자가 상대방에게 자신의 정보를 제공하지 않고 공통된 정보를 찾는 프로토콜로이며, 보안검색, 중복제거, 인증 및 권한 부여 등에 활용될 수 있다. 이 기술은 아래와 같이 요약될 수 있다.

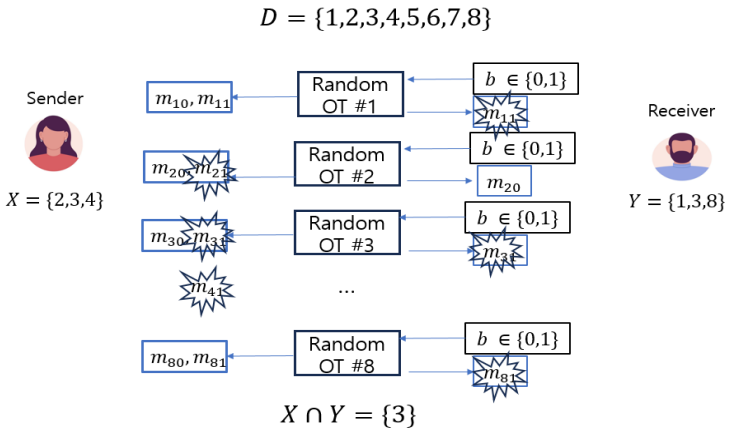
두 참여자는 각각의 집합을 가지고 있다고 가정한다. 송신자는 $x_1, x_2, \ldots, x_n$, 수신자는 $y_1, y_2, \ldots, y_n$ 로 간단하게 설명하기 위해 두 참여자의 집합의 원소의 개수는 같다고 가정하고 서로에 대한 정보는 공유하지 않는다. PSI 프로토콜의 단계는 아래와 같다. 여기서 $F_k(x)$는 Oblivious Pseudo Random Function $PRE_k(x)$이다.

1) 송신자는 PRF에 사용할 비밀키 k값을 선택한다.

2) 두 참여자는 n개의 OT를 실행한다. : i번째 실행에서 송신자는 k값을 입력하고 수신자는 본인이 갖고 있는 원소 y_i를 입력하여 $F_k(y_i)$를 계산한다.
3) 송신자는 $F_k(x_i)$를 계산하여 수신자에게 전송한다.
4) 수신자는 $F_k(y_1)$와 $F_k(x_1)$를 비교하여 교집합을 구한다.

OT기반 PSI의 계산 복잡도는 $X = \{x_1, x_2, \dots x_n\}, Y = \{y_1, y_2, \dots y_n\}$ 인 경우 각 원소의 하나마다 OT계산을 n번 반복해야 하므로 집합 X의 원소가 n인 경우 $O(n^2)$ 의 계산 복잡도가 나온다. 이미 만일 데이터가 대용량이라면 복잡도에 대한 개선이 필요하다 할 것이다.

한편 OT기반의 안전한 다자간 연산의 개선 방법으로 Kolesnikov 등[11]은 라그랑쥐 보간 다항식을 활용한 Sparse OTE를 바탕으로 한 PSI 프로토콜을 제안한 바 있다. 기존 OT 프로토콜의 문제점은 모든 메시지에 대해 OT 계산량이 필요해 여전히 많은 계산량을 필요로 한다는 것이다. 따라서 이러한 문제점을 개선하고자 한번에 다량의 메시지를 전달하면서 통신비용을 낮추는 방법으로 다항식을 활용하여 다항식의 계수만 전송하도록 하여 통신 비용을 효과적으로 줄였으며, 또한 기하급수적으로 커지는 데이터에 대해서도 일정량의 통신비용을 유지하기 위해 OPRF를 사용하였다. 실험결과 Sparse OTE 기반 PSI 프로토콜은 통신 비용측면과 계산 복잡도가 40% 개선된 것으로 밝혀졌다[10].



[그림 4] Oblivious Tranger기반 교차점 연산 프로토콜 예시

[Fig. 4] Example of Random Oblivious Transfer based Private Set Intersection

4. 안전한 다자간 연산 프로토콜

본 장에서 마이데이터 환경에서의 적용을 위해 Pinkas, Kolesnikov 등 [10][11] 이 제안한 Sparse OTE 기반 PSI 프로토콜을 기반으로 하여 Polynomial-based Oblivious Pseudo Randon Function (POPRF)를 제안한다. 기존 [10][11]에서 다중 시점 Oblivious Pseudo Random Function(OPRF)와 [11][12]에서 제안한 Oblivious PRF(OPRF) 을 확장하여 즉, 3명 이상 다자간 참여자(Multi-party)를 가정한 다중 시점(Multi-Point)을 갖는 다항식(Polynomial Interpolation) 기반 PSI 프로토콜을 제안한다. 제안하는 연구의 핵심기술은 [12]에서 제안한 3명 이상 다자간의 비밀키를 상호작용하는 zero sharing을 개선하여 3.1에서 설명한 Shamir sharing 을 적용하였고 Polynomial-based Oblivious Pseudo Randon Function (POPRF) 을 기반으로 한다. 제안하는 프로토콜은 [13]에서 소개한의 문제점인 공개키 연산의

계산량의 문제를 개선하였고, [14]에서 제안한 다중 참가자간 PSI 프로토콜을 개선하여 데이터의 크기에 영향을 받지 않는다.

4.1 Three-Party Private Set Intersection From a POPRF

먼저 3명의 참가자에 대해 PSI 프로토콜에 대해 설명한다. 3명의 참가자 $\boldsymbol{P_1, P_2, P_3}$ 가 각각의 집합 $\boldsymbol{X_i = \{x^i_1, x^i_2, \dots, x^i_m\} \subseteq \{0,1\}^*}$ 를 갖고 있다고 가정한다. 기본 이론은 간단하다. 참가자 $\boldsymbol{P_1, P_2}$ 의 교차점 $\boldsymbol{X_{12} = X_1 \cap X_2}$ 은 $\boldsymbol{P_2}$ 의 새로운 부분집합이 되고, $\boldsymbol{X_{123} = X_{12} \cap X_3}$ 은 $\boldsymbol{P_3}$의 새로운 부분집합이 되어 모두의 교차점을 구하게 된다.

1) P_1, P_2, P_3 에 대하여 모든 $k \in [m]$ 에 대하여, P_1는 PRF에 사용할 비밀키 $\{e^1{}_k \mid k \in [m]\}$ 인 k값을 무작위로 선택한다.
2) P_1와 P_2 는 OPRF $e^1{}_k = F_k(e^1{}_k)$ 를 실행한다.
   - P_1 송신자가 되어 $\{(x^1{}_k, F_k(es^1{}_k) \mid k \in [m])\}$ 으로 다항식을 구하여 다항식의 계수를 전송한다.
   - P_2 는 수신자가 되어 $X_2 = \{x^2{}_1, x^2{}_2, \dots, x^2{}_m\}$ 를 입력한다.
   - $x^2{}_k \in X_2 = x^1{}_k$ 에 대해서 $\hat{e}^2{}_k = F_k(e^2{}_k)$ 인 교차점을 찾고 교차점으로 이루어진 P_2 의 부분집합 $(x^2{}_k, \hat{e}^2{}_k) | k \epsilon [m]\}$ 구한다.
3) P_2 는 부분집합 $(x^2{}_k, \hat{e}^2{}_k) | k \epsilon [m]\}$ 으로 다항식을 구하여 다항식의 계수를 전송한다.
   - P_3 은 $X_3 = \{x^3{}_1, x^3{}_2, \dots, x^3{}_m\}$ 을 전송한다.
   - $x^3{}_k \in X_3 = x^2{}_k$ 에 대해서 ,$\hat{e}^3{}_k = F_k(e'^3{}_k)$ 인 교차점을 찾고 교차점으로 이루어진 P_3 의 부분집합 $(x^3{}_k, \hat{e}^3{}_k) | k \epsilon [m]\}$ 구한다.
4) P_1 은 마지막으로 리더가 되어 교차점을 찾고 교집합을 계산한다. ([그림 4] 참조).

4.2 Muti-Party Private Set Intersection From a POPRF

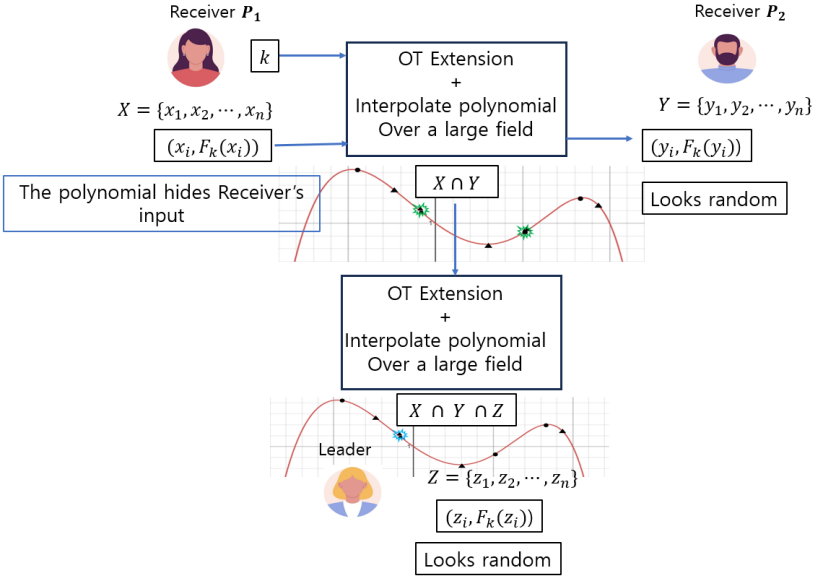
4.1 을 확장하여 단순 적용해 보면, $n > 3$ 인 참여자 $P_1, P_2, \dots, P_i, \dots, P_n$ 에 대하여 P_1, P_j 은 부분 교집합 $\boldsymbol{X_1 \cap X_2 \cap \dots \cap X_j}$ 을 구할 수 있다. 더 나아가 실제로 P_i, P_j 은 부분 교집합 $\boldsymbol{X_i \cap \dots \cap X_j}$ 을 구할 수 있다. 이러한 계산은 P_i 와 P_j 가 악의적으로 협력하는 참가자인 경우도 교차점을 계산을 할수 있다는 것이다. 두 참가자가 데이터의 교차점을 공유하지 않아도 교차점을 계산할 수 있어 악의적 협력자가 있어도 문제가 되지 않는다. 따라서 추가적인 정보가 노출되지 않음을 의미한다.

$n > 3$ 인 참여자 $P_1, P_2, \dots, P_i, \dots, P_n$ 에 대하여 각각의 참여자는 본인의 집합 $X_i = \{x^i{}_1, x^i{}_2, \dots, x^i{}_m\} \subseteq \{0,1\}^*$ 을 가지고 있다고 가정한다. 기본이론은 모든 참여자의 집합의 원소의 개수 m 는 같다고 가정하고 서로에 대한 정보는 공유하지 않는다. 이때 제안하는 다자간(Multi-party) 다중시점(Multi-Point OPRF) PSI 프로토콜의 단계는 아래와 같다([그림 5] 참조). 여기서 $F_k(x)$는 Oblivious Pseudo Random Function $PRE_k(x)$이다.

1) 모든 참여자 n명 $P_1, P_2, \dots, P_i, \dots, P_n$ 에 대하여 모든 $i \in [n]$, $k \in [m]$ 에 대하여, P_i는 PRF에 사용할 비밀키 $\{s^{i,j}{}_k \mid j \in [n]\}$ 인 k값을 무작위로 선택한다.
2) 모든 $i, j \in [n]$, P_i와 P_j 는 OPRF $F_k(s^{i,j}{}_k)$ 를 실행한다.
   - P_i 송신자가 되어 $\{(x^i{}_k, F_k(s^{i,j}{}_k) \mid k \in [m])\}$ 으로 다항식을 만들어 다항식의 계수를 전송한다.

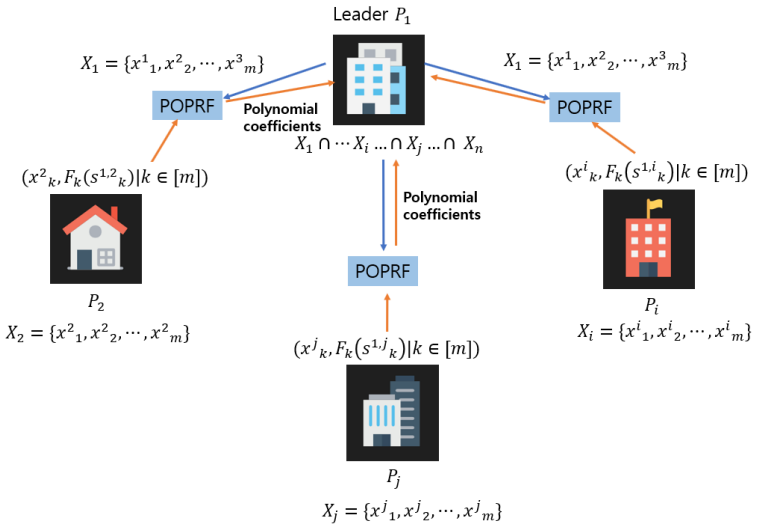
- P_j 는 수신자가 되어 $X_j = \{x^j{}_1, x^j{}_2, \ldots, x^j{}_m\}$ 를 입력한다.
- $(x^i{}_k = x^j{}_k, F_k(s^{i,j}{}_k) = F_k(s^{i,j}{}_k))$ 인 교차점을 계산하여 교집합을 구한다.

3) 이 때 서로 주고 받는 경우의 수를 줄이기 위해 리더인 P_1 를 결정한다.
4) P_1 가 리더가 되어 교차점을 찾고 교집합을 계산한다.



[그림 5] 3자간 다중 시점 의사 랜덤 함수를 사용한 교차점 연산 프로토콜(한명의 리더와 2명의 송신자와 두명의 수신자를 가정)

[Fig. 5] Three-party PSI Protocol Using the Proposed Oblivious Pseudorandom Function(assuming one sender and two receivers)



[그림 6] n 자간 의사 랜덤 함수를 사용한 교차점 연산 프로토콜(한명의 리더와 $n-1$명의 송신자를 가정)

[Fig. 6] Multi-party PSI Protocol Using the Proposed Oblivious Pseudorandom Function(assuming one leader)

5. 결론 및 향후 연구 방향

우리는 지금까지 다중 참가자를 가정한 안전한 다자간 연산(SMPC)을 위한 암호화 프로토콜에 대해 살펴보았다. 본 논문에서 제안하는 방식은 기존의 양자간 SMPC 알고리즘을 개선한 OT based PSI 프로토콜을 확장하여 마이 데이터 환경에서의 안전한 데이터 활용을 위해 3명 이상의 다중 참가자간 연산을 통한 안전한 다자간 연산(SMPC) 프로토콜을 소개하였다. 제안한 프로토콜과 선행 연구의 프로토콜 비교는 [표 1]과 같다[15].

[표 1] n 자간 PSI 프로토콜의 통신 복잡도와 모델 비교표

[Table 1] Comparison of Communication Complexity of Multi-party PSI Protocols Using[15]

Protocol	Communication complexity Party $P_i, i = 1, \dots, n,$	Security Model
[14]	$O(mn), m: the\ size\ of\ data\ set$	Semi-honest
[16]	$O(m)$	Semi-honest
[17]	$O(m)$	Semi-honest
Our Portocol	$O(m)$	Semi-honest

또한 마이 데이터 환경에서는 개인정보의 보호를 위해 본인이 갖고 있는 개인 정보인 원소 $\boldsymbol{y_i}$가 아닌 $\boldsymbol{F_k(y_i)}$를 계산하여 프로토콜에 적용하는데 이때 $\boldsymbol{F_k(y_i)}$ 값은 일종의 가명정보로서의 역할을 수행할 수 있을 것이다. 향후 연구방향으로는 본 논문에서 제시한 사미에르 세어링 공유키 방식은 데이터의 안전성은 매우 높아 의료 데이터, 금융 데이터등 민감 데이터에 적용하기에는 적절하나 계산 비용이나 통신 복장성이 높아 대역폭과 네트워크 리소스에 부하를 줄 수 있다는 단점이 있다. 공개키 공유 방식을 개선하고 계산능력은 줄이도록 알고리즘 보완하여 보다 구체화하여 실제 마이데이터 환경에 실험하여 그 결과를 확인해보고자 한다.

6. 감사의 글

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임(과제번호 HI23C0733).

References

[1] H. Eun, Ubaidullah, H. Oh, Efficient Outsourced Multiparty Computations Based onPartially Homomorphic Encryption, Journal of The Korea Institute of Information Security & Cryptology, (2017), Vol. 27, No.3, pp.477-487.
DOI: https://doi.org/10.13089/JKIISC.2017.27.3.477

[2] S. Y. Lee, K. K. Byun, S. W. Cho, A Study on the Impact of the Nasic Characteristic of MyData on the Its Setvices and

Government Policies-focusing in major countries.cases, The Journal of Korea Institute of Information, Electronics, and Communication Technology, (2023), Vol.6, No.2, pp.77-86.
DOI: https://doi.org/10.17661/jkiiect.2023.16.2.77

[3] K. H. Lee, Current Status of MyData Policy and Tasks in Health and Welfare, Korea Institute for Health and Social Affairs, (2021), Vol.11, pp.52-68.
DOI: https://doi.org/10.23062/2021.11.5

[4] Y. M. Kim, J. S.Jeong, I. Y. Park, Global biopharmaceutical industry R&D trends and implications examined through cases of COVID-19 vaccine and treatment development, Khidi Brief, (2021), Vol.332.
Available from: https://www.khidi.or.kr/fileDownload?titleId=453603&fileId=1

[5] Y. J. Song, K. Y. Park, Security/Privacy requirements for medical data sharing and utilization services, Journal of The Korea Institute of Information Security & Cryptology, (2010), Vol.20, No.3, pp.90-96.
Available from: http://www.riss.kr/link?id=A82352429

[6] A. C. Yao, Protocols for secure computations, Proceedings of the 23rd Annual Symposium on Foundations of Computer Science, IEEE, pp.160-164, (1982)
DOI: https://doi.org/10.1109/SFCS.1982.38

[7] A. Shamir, How to Share a Secret, Communications of the ACM, (1979), Vol.22, No.11, pp.612-613.
DOI: https://doi.org/10.1145/359168.359176

[8] M. Rabin, How To Exchange Secrets with Oblivious Transfer, IACR Cryptol, (2005), Vol.187.
Available from: https://www.researchgate.net/publication/220332997

[9] B. Pinkas, M. Rosulek, N. Trieu, A. Yanai, SpOT-Light: Lightweight Private Set Intersection from Sparse OT Extension, Advances in Cryptology–CRYPTO 2019, (2019), Vol.11694.
DOI: https://doi.org/10.1007/978-3-030-26954-8_13

[10] V. Kolesnikov, R. Kumaresan, M. Rosulek, N. Trieu, Efficient Batched Oblivious PRF with Applications to Private Set Intersection, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp.818-829, (2016)
DOI: https://doi.org/10.1145/2976749.2978381

[11] V. Kolesnikov, N. Matania, B. Pinkas, M. Rosulek, N. Trieu, Practical Multi-party Private Set Intersection from Symmetric-Key Techniques, Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, (2017)
DOI: https://doi.org/10.1145/3133956.3134065

[12] L. Kissner, D. Song, Privacy-Preserving Set Operations, Advances in Cryptology–CRYPTO 2005, (2005), pp.241-257.
DOI: https://doi.org/10.21236/ada457144

[13] C. Hazay, M. Venkitasubramaniam, Scalable Multi-party Private Set-Intersection, Public-Key Cryptography – PKC 2017, (2017), pp.175-203.
DOI: https://doi.org/10.1007/978-3-662-54365-8_8

[14] X. Yu, F. Li, W. Zhao, Z. Dai, D. Tang, Multiparty Threshold Private Set Intersection Protocol with Low Communication Complexity, Security and Communication Networks, (2022), Vol.2022, pp.1-12.
DOI: https://doi.org/10.1155/2022/9245516

[15] A. Miyaji, S. Nishida, A Scalable Multiparty Private Set Intersection, Lecture Notes in Computer Science, (2015), Vol.9408, pp.376-385.
DOI: https://doi.org/10.1007/978-3-319-25645-0_26

[16] C. Hazay, M. Venkitasubramaniam, Scalable Multi-party Private Set-Intersection, Public-Key Cryptography – PKC 2017, Lecture Notes in Computer Science, (2017), Vol.10174, pp.175-203.
DOI: https://doi.org/10.1007/978-3-662-54365-8_8

[17] R. Inbar, E. Omri, B. Pinkas, Efficient Scalable Multiparty Private Set-Intersection via Garbled Bloom Filters, Security and Cryptography for Networks, Lecture Notes in Computer Science, (2018), Vol.11035, pp.235-252.
DOI: https://doi.org/10.1007/978-3-319-98113-0_13