동형암호기반 데이터 유출 방지 표준 동향

나재훈 ITU-T SG17(정보보호 분야) WP4(어플리케이션 보안) 의장, 한국전자통신연구원 전문위원

1. 머리말

현대 문화는 ICT(Information and Communication Technology)의 영향을 많이 받고 있으며, 기술이 발전하는 속도가 빠르다 보니 기술 개발 초기 단계에 미리 어떤 산업에 어떻게 응용할 것인지를 염두에 두고 연구를 추진하기도 한다.

이러한 변화는 디지털 전환(Digital transformation)의 한 면이다. ICT가 발전하면서 데이터가 사이버 공간에 집중되고, 그 양이 방대해지면서 단순히 통계적 처리만 해도 가치 있는 정보를 획득할 수 있다. 그러나 이러한 처리 과정에서 개인 정보가 유출되거나 역공학(Reverse engineering)을 통해 민감한 개인 정보를 비즈니스에 악용하는 사례가 발생하고 있다. 이에 유럽에서는 유럽연합 일반 데이터 보호 규칙(GDPR)을, 한국에서는 데이터 3법을 시행한다. 이러한 추세에 따라 데이터를 처리할 때 데이터를 안전하게 획득·처리·파기하는 메커니즘과 제도가 필요하다. 본고에서는 암호학적 관점에서 데이터 처리에 안전성과 협업성을 제공하는 완전동형암호(FHE, Fully Homomorphic Encryption)에 대해 살펴본다.

2. 동형암호(Homomorphic encryption)

디지털 전환과 암호학이 만나는 분야가 있다. 암호학에서는 오래전부터 데이터에 대한 암호처리 메커니즘을 연구해 왔다. 암호를 푸는 키를 사용할 경우, 키가 노출되면 해킹에 매우 취약해지기 때문에 공개키를 사용하지 않는 암호화가 특히 주목받았다. 암호화된 정보를 복호화하지 않은 채 전산 처리하고, 최종 사용자만 그 내용을 복호화할 수 있게 하는 동형암호 알고리즘이 대표적이다.

암호화는 일반적으로 데이터의 기밀성과 무결성을 보장한다. 그러나 동형암호는 데이터의 무결성을 보장하지 않고 가단성(Malleability)을 제공하여 암호화된 데이터를 연산 처리할 수 있다. 동형암호 스킴(Scheme)은 오랫동안 연구됐으나 크게 관심을 받지는 못하다가 RSA, ElGamal, Pailier 등의 부분동형암호 스킴이 개발됐으며 Pre-FHE(Pre-Fully homomorphic encryption) 시대를 거쳐 1~4세대로 구분된다.

2.1 동형암호의 유형

동형암호는 비밀키에 액세스하지 않고 암호화된 데이터를 연산처리하게 한 암호화의 한 형태다. 동형의 의미는 대수학의 동형(Homomorphism)이며 암호화 및 복호화 기능이 일반 텍스트

와 암호문 사이에서 동형으로 처리되는 것이다[1]. 즉, 암호화된 데이터를 복호화하지 않고 암호화된 상태에서 직접 연산처리를 해도 그 결과가 원문 처리 결과와 동일한 암호화 방법인것이다. 그 유형은 다음과 같이 분류된다.

- Partially Homomorphic Encryption(PHE): 주어진 데이터 세트에 대해 무제한의 시간 동안 한 가지 유형의 수학 연산(예를 들어 곱셈)만 허용한다.
- Somewhat Homomorphic Encryption(SHE): PHE에 비해 허용 범위가 넓지만 여전히 제한적이어서, 주어진 데이터 집합에 대해 덧셈과 곱셈을 몇 차례만 허용한다.
- Fully Homomorphic Encryption(FHE): 최선의 방법으로, 데이터에 대해 횟수에 제한 없이 다양한 유형의 연산을 허용하지만 대신 성능 측면에서 현저히 불리하다.

3. 유스케이스

3.1 스토리지 아웃소싱

데이터 스토리지 아웃소싱은 조직 내 스토리지 운영에 따른 공간, 운영, 기술, 인력 낭비를 줄이고 유지보수 및 업그레이드에 따른 문제를 해결할 수 있는 효율적인 방안이다. 속지의 법과 규제가 상충하는 경우에도 이를 해결하고자 스토리지를 해외로 아웃소싱할 수 있다.
동형암호는 플랫폼 엔지니어가 권한을 악용해 사용자를 스토킹하는 것과 같은 사고를 미연에 차단할 수 있다. 또한 동형암호를 사용하면 데이터를 클라우드에 안전하게 저장할 수 있다. 동시에 암호화된 데이터를 연산이나 검색에 사용할 수 있다.

3.2 헬스케어

헬스케어 시스템은 의료정보에 ICT를 접목해의료기관에서 일어나는 제반 업무, 즉 기록과 계산, 업무현황, 작업지시, 환자의 건강정보를 다루는 시스템이다. 하지만 헬스케어 산업은 개인정보보호에 특별히 신경을 써야 한다. 치료 중 발견되는 환자의 병명과 병력은 민감한 정보다. 또한 이러한 정보를 통계적으로 처리한 결과물도 어느 정도 지역과 문화적 정보가 포함돼 있어서 정보 활용에 규제가 필요하다.
동형암호는 이러한 데이터를 전처리해 암호화 된 형태로 저장한다. 필요에 따라 암호화된 데이터의 내용을 알지 않아도 안전하게 연산처리도 할 수 있다.

3.3 DNA 분석

개인의 유전정보를 이용하는 개인 맞춤형 정밀 치료 기술이 활발히 개발되고 있다. 그러나 DNA는 매우 민감한 정보로서 필연적으로 프라이버시 문제를 일으킨다. 동형암호는 이러한 이슈를 해결하는 대안이다. DNA 정보를 동형암호화해 저장하고 처리하면 DNA 정보의 유출을 방지하는 효과가 있다.

3.4 가상물리시스템(Cyber physical system)

스마트 팩토리, 스마트 시티, 디지털 트윈과 같은 인프라는 가상물리 시스템에 기반을 둔다.

이러한 시스템은 센서와 제어기, 엑츄레이터라는 구조를 포함하는데, 현재 센서 데이터를 임의조작하는 일이 가능하다. 동형암호를 이용하면 센서 데이터를 암호화해 제어기에서 복호화되지 않은 상태에서 연산처리가 가능해 가상물리시스템의 안전성을 높일 수 있다.

3.5 기계학습(Machine learning)

머신러닝은 방대한 데이터를 분석해 미래를 예측하는 기술이다. 즉, 컴퓨터가 스스로 학습해 입력되지 않은 정보를 습득한 뒤 문제를 해결한다. 그러나 개인정보보호가 적용되는 환경에서는 머신러닝이 얻는 데이터의 정확성이 문제가 된다. 비식별화 처리가 된 데이터는 개인정보의 결합도가 낮아 데이터의 가치가 떨어지기 때문이다. 이러한 상황을 극복하는 방법으로 순수 데이터를 그대로 암호화하고, 암호화된 데이터를 개인정보 유출 없이 기계학습해 관련 패턴을 찾아 산업에 적용할 수 있다.

3.6 양자컴퓨팅 내성 암호

인터넷뱅킹, 전자상거래 같은 암호체계는 풀기가 거의 불가능한 수학 문제에 기반을 둔 국제 표준 공개키 암호인 'RSA(Rivest Shamir Adleman)'와 'ECDSA(Elliptic Curve Digital Signature Algorithm)'를 사용한다. RSA는 소인수분해 대상 숫자 단위가 무한히 커지면 이를 풀 수 없다는 수학적 난제로 잠금장치를 걸어놓은 것이다. 즉 공개키로 암호화하고 개인키로 복호화하는 방식으로 정보를 잠근다. 그런데 이들 암호는 양자컴퓨터가 출현하면 폐기될 전망이다. 양자컴퓨터가 사용하는 '쇼어 알고리즘'으로 실시간 해독이 가능하기 때문에 전자상거래에서 사용하는 암호화 통신이 무용지물이 될 수 있다. 양자컴퓨터에 대응할 수 있는 암호가 동형 암호다. 동형암호를 포함한 격자기반암호는 양자컴퓨터가 도입돼도 깨지지 않는 차세대 암호체계라는 것이 암호학계의 중론이다. 격자 문제는 '현재로서는 풀 수 있음이 증명되지 않은 문제'인 NP 완전 문제(NP complete problem)로 분류된다.

3.7 금융 협업

이상거래 탐지(Fraud detection)나 개인신용 평가(Credit scoring)뿐만 아니라 고객의 프라이버시를 보호하면서 맞춤 서비스에 동형암호를 적용하면, 데이터 기밀성을 보장하면서 다자간 협업 서비스가 가능하다. 알리바바의 자회사 앤트 파이낸셜(ANT financial)은 신용분석, 마케팅 분석 및 은행데이터 결합분석에 동형암호기술을 적용한다. SAP는 2018년 'SAP's Guiding Principles for AI'를 발표할 때 동형암호를 핵심 요소기술로 소개했다.

4. 산업 및 표준 동향

4.1 산업 동향

동형암호 개념은 Rivest , Adleman과 Dertouzos가 1978년에 처음 제시했다. 이후 IBM의 연구원인 Craig Gentry가 2009년에 격자 기반 암호화를 사용하는 완전한 동형암호(Fully FHE)를 위한 구조가 최초로 제시했다. 동형암호스킴과 관련해 여러 오픈소스 구현물이 존재하며, 다음과 같은 목록을 참조할 수 있다[2].

- Microsoft SEAL: BFV 및 CKKS 스킴을 지원하는 마이크로소프트의 오픈소스 라이브러리.
- PALISADE: BGV, BFV, CKKS, TFHE 및 FHEW와 같은 여러 동형 암호화 체계를 지원하며 다자간 지원을 제공하는 방위 계약 업체 컨소시엄(DARPA 자금을 지원받는)의 오픈소스 라이브러리.
- HELib: CKKS 및 BGV 체계와 부트스트랩을 지원하는, IBM의 초기에 널리 사용되는 라이브러리.
- FHEW/TFHE: TFHE 스킴을 지원하며, TFHE는 FHEW에서 설계되었지만, FHEW는 더 이상 활발하게 개발되지는 않는다.
- HeaAn: 고정 소수점 근사 산술을 기본적으로 지원하는 CKKS 스킴을 구현한 라이브러리.
- Λ ○ λ("LOL"이라고 발음): FHE를 지원하는, 링 기반 격자 암호화를 위한 Haskell 라이브러리.
- NFLlib: 저수준 프로세서 프리미티브를 사용하여 고성능 동형 암호화를 탐색하기 위한 유럽 HEAT 프로젝트의 파생 결과인 라이브러리.
- cuHE: 이 라이브러리는 동형 암호화를 가속화하기 위한 GPGPU 사용에 관한 연구.
- Lattigo: Go로 작성된 격자 기반 암호화 라이브러리.
- Concrete: TFHE 스킴의 사용자 맞춤을 지원하는 라이브러리.

4.2 표준 동향

개인정보보호가 규제화되면서, 전통적 암호스킴을 제치고 동형암호가 대안으로 대두됐다. 2009년 IBM에서 제안한 동형암호는 산업계에서 자발적으로 표준화의 필요성을 인식해 컨소시엄 형태의 표준화가 진행되고 있다[2]. 또한 이를 글로벌 환경에 보급하는 것을 목표로 공적표준화 기구에서도 표준화 작업을 2020년에 시작했다. 관련 표준화 기구로 Homomorphic Encryption Standardization 컨소시엄, ITU-T, ISO/IEC JTC 1 등이 있으며 관련 표준화 활동을 소개한다.

가. Homomorphic Encryption Standardization[2]

많은 기업과 개인이 클라우드 스토리지 및 컴퓨팅으로 전환함에 따라 이를 쉽게 사용하기 위한 기준이 필요하다. 이에 API를 균일화하고 단순화하며 애플리케이션 개발자에게 API를 사용하도록 표준화하고 있다. 참여자로 산업계에서는 Microsoft, Samsung SDS, Intel, Duality Technologies, IBM, Google, SAP 등, 기관으로는 NIH, NIST, NSF, UN/ITU 등, 학계는 서울대, Boston Univ., Columbia, EPFL, MIT, UCSD 등이 있다.

본 콘소시엄은 동형암호의 보안, API 및 애플리케이션 등 세 가지 백서를 기반으로 동형암호에 대한 표준을 개발하고 있다. 커뮤니티의 주요 구성원 검토를 거쳐 공개 의견 수렴 기간이 지난 후 보안 백서는 두 번째 표준화 워크숍(March 15-16 2018, MIT, Cambridge MA, USA)에서 공개적으로 승인돼 동형암호 표준의 첫 번째 버전을 제정했다. 이 표준은 스킴 설명, 보안 속성에 대한 설명, 보안 매개 변수에 대한표를 제공한다. 표준의 향후 버전에서는 동형암

호를 위한 표준 API 및 프로그래밍 모델을 기술할 예정이다.

나. ITU-T SG17[3]

ITU-T SG17에서는 동형암호를 이용해 산업에 적용할 수 있는 분야 중 하나로 기계학습 분야를 정했다. 그리하여 동형암호의 이해와 데이터 분석에 있어서 개인정보를 보호하기 위한 처리구조, 절차와 특성에 관한 지침을 개발하는 중이다. 삼성SDS, 서울대, ETRI가 에디터로 참여해 활동을 하고 있으며, 2020년 3월에 신규아이템(TR.sgfdm, FHE-based data collaboration in machine learning)을 채택해 개발에 착수했다.

주요 내용은 완전동형암호 기술을 사용해 기계학습의 보안 추론 서비스 및 데이터 집계에 대한 보안 지침을 제공한다. 또한 데이터 소유자가 기계학습 모델 공급자의 추론 서비스를 사용하는 반면 각 당사자는 자신의 데이터를 공개하지 않는 구조와 절차를 제공한다.

다. ISO/IEC JTC 1/SC27[4]

ISO/IEC JTC 1/SC27 WG2에서는 기존에 IS 18033-6, Encryption algorithms — Part 6: Homomorphic encryption 표준이 있고 2019년 개정판을 제정했다. 이 표준은 부분 동형 암호를 위한 지수 ElGamal 암호와 Paillier 암호, 두 가지 메커니즘으로 구성되어 있다. 부분 동형암호는 하나의 유형 연산만, 다시 말해 덧셈(Paillier 암호 경우) 또는 곱하기(지수 ElGamal 암호 경우)를 지원하는 스킴을 기술한다.

반면 2021년 1월부터 표준아이템(PWI 15150 — Fully homomorphic encryption)을 발굴하기 위한 사전 모임이 있었다. 여기서는 완전 동형암호(FHE)에 대해 임의 작업을 지원하고 암호화 데이터에 대한 임의 계산을 허용하는 동형암호 스킴 표준을 위한 작업을 수행한다. 또한 이 모임에서는 SP Suitability of standardization of fully homomorphic encryption(FHE)에대한 협의를 통해 표준제안 개선 작업과 이를 활용하기 위한 유스케이스 'Approzimate HE in analyzing encrypted data(서울대 천정희 교수)'의 발표가 있었다. 향후 회의를 거쳐 그 결과를 근거로 WG2 회의에 신규아이템 제안을 계획하고 있으며 한국에서는 서울대와 삼성SDS에서 참여하고 있다.

5. 결론

유럽의 GDPR과 한국의 데이터 3법에 적절히 대처하고, 4차 산업을 육성하려면 프라이버시를 온전히 보호할 수 있는 암호가 필요하다. 완전동형암호는 이러한 역할에 적합하다는 평가를 받는다. 이에 따라 금융권이나 기계학습과 같은 분야에서 완전동형암호에 대한 관심이 고조되고 있는 상황에서 한국(서울대)이 그 핵심기술 개발에 선두주자로 활동을 하고 있다는 것이 매우 반갑다.

완전동형암호는 암호학적으로도 진일보한 이론이다. 이론적 검증도 완성됐으며, 시스템 구축과 그 연산처리 성능에서는 아직 개선해야 할 숙제가 있지만 많은 노력을 통해 커다란 진전이 있었다. 암복호화 과정은 RSA 처리속도에 근사하며, 연산처리 과정에서 속도를 개선하고자 하드웨어 기반의 연산 가속기가 연구 중이다.

완전동형암호 이론을 산업에 효과적으로 적용하려면 표준화가 필연적이라 사료된다. 현재 기술과 병행하여 표준이 개발되고 있으므로 국내와 국제, 사실표준화와 공적표준화, 학계와 산업계, 정책적인 부분에서 개발에 대한 경쟁과 조율이 필요하다. 향후 동형암호에 대한 전략적 대응을 기대해 본다.

[참고문헌]

[1] Homomorphic Encryption, https://en.wikipedia.org/wiki/Homomorphic_encryption

[2] Homomorphic Encryption Standardization, https://homomorphicencryption.org/

[3] ITU-T SG17: Security,https://www.itu.int/en/ITU-T/studygroups/2017-2020/17/Pages/default.aspx

[4] ISO/IEC JTC 1/SC 27,https://isotc.iso.org/livelink/livelink?func=ll&objId=8916258&objAction=browse 2015.

※ 출처: TTA 저널 제193호
(코로나 이슈로 각 표준화기구의 표준화회의가 연기・취소됨에 따라 TTA 저널로 대체합니다)