완전동형암호를 이용한 의료 데이터 분석 시스템 구현

엄유진, 이윤영, 양희훈, 유호영
충남대학교 공과대학 전자공학과

Dependable Medical Data Analysis System using Fully Homomorphic Encryption

Yujin Eom, Yunyoung Lee, Hyhoon Yang, and Hoyoung Yoo
Department of Electronic Engineering, Chungnam National University
E-mail : yjeom.cas@gmail.com, yunyoung@o.cnu.ac.kr, hhyang.cas@gmail.com, hyyoo.cnu@gmail.com

Abstract

Medical consultation and diagnosis increasingly rely on artificial intelligence (AI) chatbots, raising critical concerns about the privacy of sensitive medical data during processing and transmission. This necessitates a system that guarantees the confidentiality of personal medical information while restricting access to analysis results exclusively to the data owner. To address this issue, this paper presents a secure medical data analysis system leveraging homomorphic encryption (HE). By utilizing the HEAAN (Homomorphic Encryption for Arithmetic of Approximate Numbers) library, the proposed system ensures the protection of sensitive data while enabling efficient and privacy-preserving data analysis.

I. 서론

의료 상담 및 진단 분야에서 인공지능(AI) 챗봇의 활용이 증가함에 따라 민감한 의료 데이터의 처리와 전송 과정에서 개인정보 보호 문제가 중요한 과제로 부상하고 있다. 따라서 의료 진단 및 데이터 분석 중 개인정보의 기밀성을 철저히 보장할 수 있는 시스템의 필요성이 절실히 제기되고 있다. 해당 시스템은 개인의 민감한 의료 데이터를 안전하게 보호하면서, 분석 결과에 접근할 권한을 데이터 소유자인 개인에게만 부여하는 것을 핵심 목표로 해야 한다. 본 논문에서는 이러한 요구를 충족하기 위해 완전동형암호 scheme 중 HEAAN(Homomorphic Encryption for Arithmetic of Approximate Numbers) 라이브러리를 활용한 안전한 의료 데이터 분석 시스템을 제안한다.

This research was supported by National Research Foundation of Korea(NRF) funded by the Korea government(MSIT) (IITP-2024-RS-2024-00436406 (50%), 2020M3H2A1078119) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2022R1A5A8026986)

II. 배경 지식

2.1 동형 암호

동형 암호는 데이터의 복호화 없이 암호화된 데이터에 대해 연산을 수행할 수 있는 암호화 방식이다. 식 1 은 동형암호의 원리를 보여준다.

$$Enc(p_1)\ \Delta\ Enc(p_2) = Enc(p_1\ \Delta\ p_2) \qquad (1)$$

식 1 에서 p_1, p_2 는 평문 상태의 데이터를 의미하고 Δ 는 연산을 의미한다. 즉, 동형암호에서는 암호화된 상태에서의 연산 결과가 평문에서의 계산 결과를 암호화한 결과와 같다. 동형 암호화는 특정 연산만 무제한 수행 가능한 부분 동형 암호화(PHE)에서 덧셈과 곱셈이 제한적으로 가능한 일부 동형 암호화(SHE)로 발전하였고, 현재는 덧셈과 곱셈이 무제한 가능한 완전 동형 암호화(FHE)까지 개발되었다. 완전 동형 암호 scheme 중 하나인 HEAAN 은 실수와 복소수 연산에 특화된 알고리즘으로, 근사값 연산을 지원하며 bootstrapping 기법을 통해 오류를 감소시켜 높은 실용성을 제공한다[1].

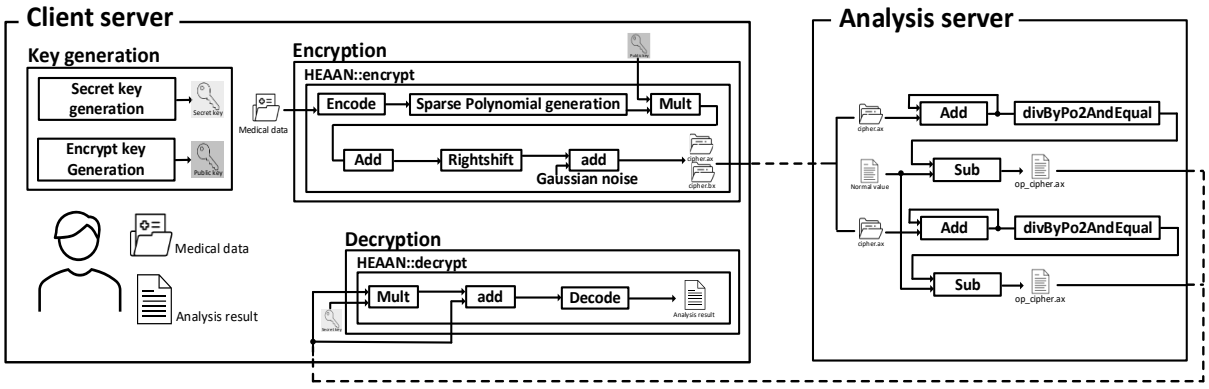

그림 1. 시스템 블록 다이어그램

표 1. 시스템 동작 결과

	Python SW	Proposed System	HIV infection
CD+T cell count	-31.875	-31.875 + *1.343e^{-8}i	High risk
HIV-1 viral load	-5912.5	-5912.5 + *3.467e^{-11}i	High risk
HIV antibody response strength	-31.865	-31.875 - *1.582e^{-9}i	High risk

*error for encryption/decryption

III. 제안하는 의료 분석 시스템

그림 1 은 제안하는 의료 분석 시스템의 블록 다이어그램을 보여준다. 전체 시스템은 개인의 의료 데이터를 가지고 있는 클라이언트 서버와 데이터 연산을 수행하는 분석 서버로 분리된다. 서버는 Linux OS 환경에서 Python 의 Flask 프레임워크를 사용하여 구현되었고, 내부 동형암호의 동작은 Linux OS 환경에서 HEAAN 라이브러리를 적용하여 구현되었다.

시스템 동작을 위해 가장 먼저 암호화 및 복호화를 위한 키를 생성한다. 이후 암호화 키를 이용해 데이터를 암호화한다. 암호화된 데이터는 분석 서버로 전송되어 데이터의 연산을 수행한다. 연산된 데이터는 다시 클라이언트 서버로 전송되어 복호화 키를 이용해 복호화되어 개인이 확인할 수 있다. 이러한 구성을 통해 암호화된 상태에서 데이터를 연산하므로 높은 보안성을 가지며. 키를 개인만 가지고 있으므로 분석 결과는 개인만이 확인할 수 있게 된다.

IV. 실험 결과 및 결론

제안하는 시스템은 다양한 의료 분석에 활용될 수 있으며, 이를 보이기 위해 HIV 진단을 예시로 하였다. HIV 감염 진단을 위해 CD+T cell, HIV-1 viral load, HIV 항체 반응도를 진단 기준으로 하였으며 이는 HIV 감염 진단에 사용되는 수치이다[2]. 분석 소프트웨어인 Python SW 에서는 여러 날의 데이터를 수집한 뒤, 각 진단 항목별로 평균 값을 계산한다. 이 값을 정상 기준치와 비교하여 차이를 계산한다. 계산된 값이 0 보다 작으면 감염 가능성이 높고, 0 보다 크면 감염 가능성이 낮은 것으로 평가한다. 제안하는 시스템의 분석 서버 또한 같은 방식으로 연산하지만 암호화된 데이터로 연산한다는 점에서 차이가 있다.

표 1 은 Python SW 와 제안하는 시스템의 실험 결과를 보여준다. Python SW 결과와 제안하는 시스템의 동작 결과가 암호화 및 복호화 과정으로 인한 허수부의 작은 오차를 제외하면 같은 것을 확인할 수 있다. 모든 진단 기준에 대한 항목이 0 보다 낮은 값을 가지므로 HIV 감염 가능성이 높은 것으로 평가할 수 있다. 이를 통해 본 논문에서 제안된 시스템이 민감한 의료 데이터의 안전한 분석을 가능하게 함을 입증하고, 의료 데이터 보호의 새로운 방향성을 제시한다.

참고문헌

[1] Kim, J, Song, Y, Lee, J, Cheon, J. H, & Kim, D. “Homomorphic Encryption for Approximate Arithmetic Numbers”, Cryptology ePrint Archive, 2018

[2] Doe, J, Smith, A., & Brown, B. (2023). CD4+ T-cell percentage is an independent predictor of clinical progression in AIDS-free antiretroviral-naive patients with CD4+ T-cell counts > 200 cells/mm³. Journal of Clinical Immunology, 45(3), 123–135.