[논문리뷰] HETAL: Efficienct Privacy-preserving Transfer learning with homomorphic encryption

PMLC Lab meeting

2025.07.16 (Wed)

왜 이 논문을 리뷰하는 지?

Efficient and Straggler-Resistant Homomorphic Encryption for Heterogeneous Federated Learning

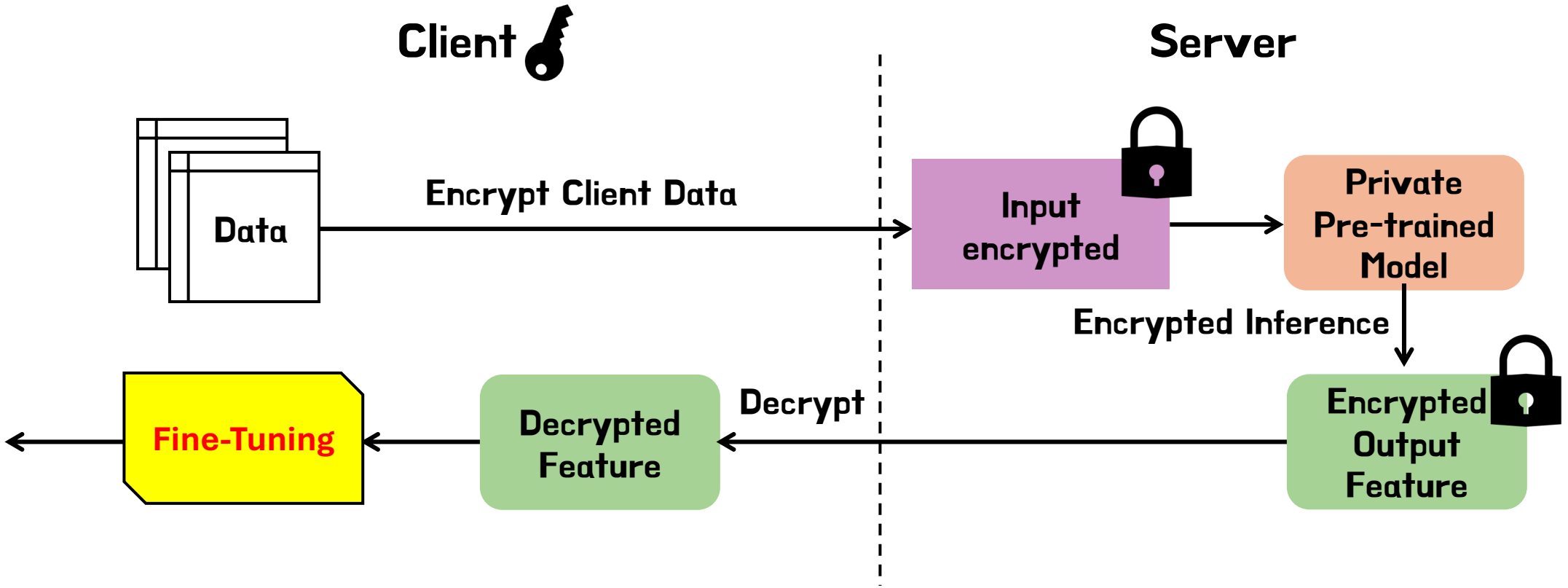
Nan Yan*, Yuqing Li*, Jing Chen*, Xiong Wang†, Jianan Hong‡, Kun He*, Wei Wang§
* School of Cyber Science and Engineering, Wuhan University, Wuhan, China
† School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
‡ School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
§ Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong
*{nanyan, li.yuqing, chenjing, hekun}@whu.edu.cn, †xiongwang@hust.edu.cn, ‡hongjn@sjtu.edu.cn, §weiwa@cse.ust.hk

[논문리뷰] HETAL: Efficienct Privacy-preserving Transfer learning with homomorphic encryption

PMLC Lab meeting

2025.07.16 (Wed)

기존 MLaaS 기반 Transfer Learning의 Privacy 한계



기존 연구의 한계: 클라이언트가 직접 fine-tuning을 해야 한다는 점에서 클라이언트에게

머신러닝에 대한 지식과 리소스를 요구

HETAL(Homomorphic Encryption based TrAnsfer Learning)

논문의 핵심 Research Question

Q. 클라이언트의 민감한 데이터를 보호하면서도,
암호화된 상태에서 Transfer Learning의 Training 단계까지도
효율적으로 처리할 수 있지 않을까?

A. HETAL은 Training 단계부터 암호화된 상태로 진행하면서도
정확도 거의 유지 및 학습 시간도 1시간 이내 완료

- 기존 연구의 한계: HE 기반 Transfer Learning 모델은 Inference 단계에 집중되어 있음

HETAL의 핵심 목표:

클라이언트 Privacy 보호를 위한 안전한 Fine-Tuning 아웃소싱

1. 클라이언트가 ‘ML 전문 지식 없이도’
   Transfer Learning을 활용할 수 있도록 지원

2. 서버가 Fine-tuning을 수행하지만,
   학습 데이터는 Encrypted 상태이므로
   ‘서버는 Final Model의 내용을 추론할 수 없음’

* 클라이언트 데이터는 클라이언트 키로 암호화돼서

서버는 복호화할 수 없지만, 공개 연산 키*로 계산 수행 가능

HETAL 프로토콜 프로세스

1. 클) Pre-trained 모델로 Feature (pt) 추출
2. 클) 암호화한 Feature (ct)를 서버에 전송
3. 서) 암호화된 Feature 사용하여
   Classification Layer를 Fine-tune
   * Early Stopping 지원
     1) 서) 검증용 데이터로 추론 logit 계산 (ct)
     2) 클) 복호화하여 loss 계산 (pt)
     3) 클) 과적합 방지 위해 stop 신호 서버로 보냄
4. 서) 암호화된 Classification Layer를
   클라이언트에 전송
5. 클) 이를 복호화해서 추론에 사용

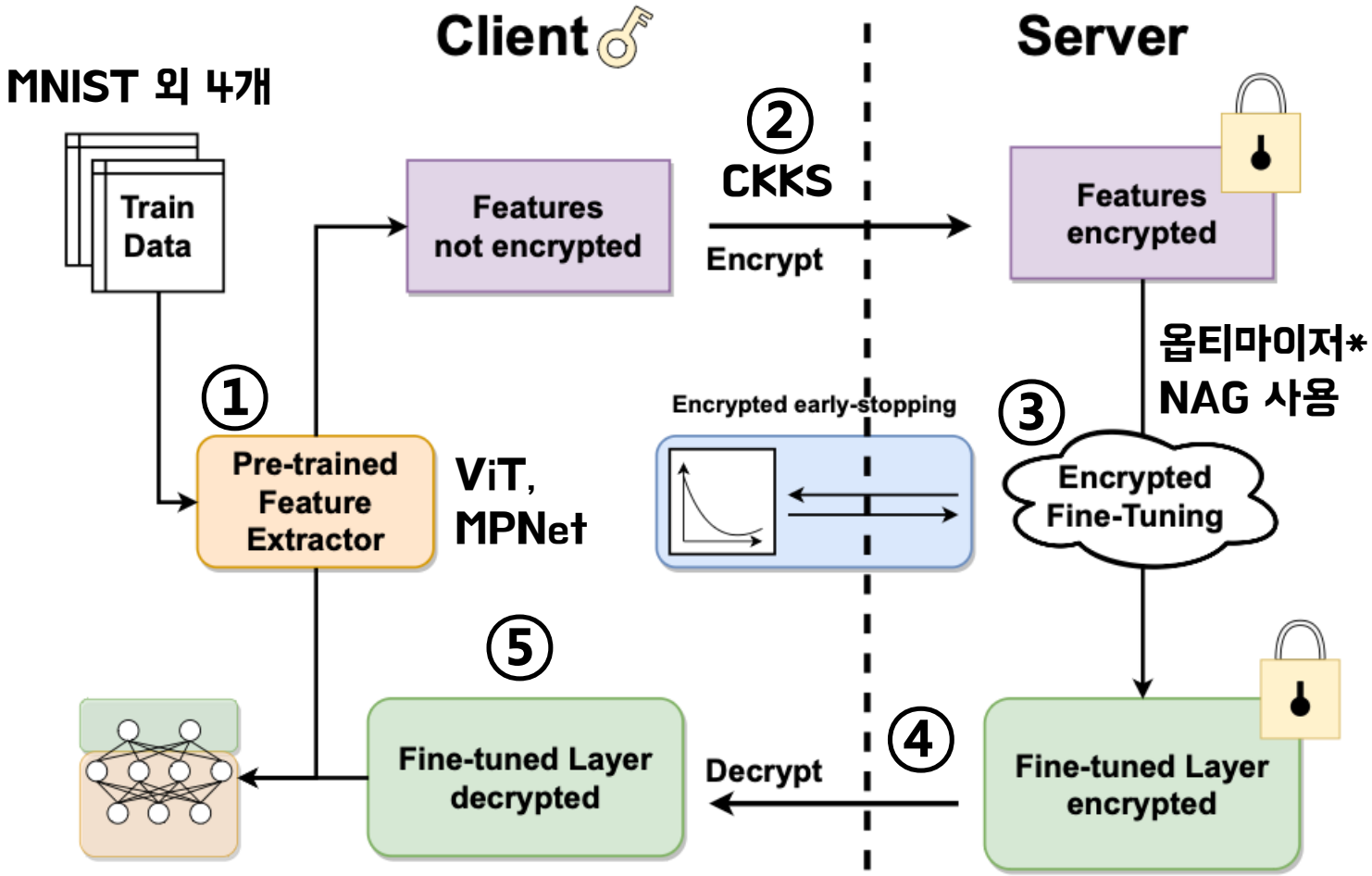


Figure 1. Our privacy-preserving transfer learning protocol (**HETAL**)

* 클라이언트 데이터는 클라이언트 키로 암호화돼서

서버는 복호화할 수 없지만, 공개 연산 키*로 계산 수행 가능

HETAL 핵심 개선점

1. 새로운 Softmax 근사 알고리즘 개발

   1) 기존 대비 넓은 범위를 높은 정밀도로 다룸

   2) 근사 도메인: [-128, 128]로 확장

      * 기존 연구 근사 도메인: [-8, 8]도 버거움

        - PrivGD (Jin et al., 2020)

   → HETAL 성과 1) 모델 학습 수백번도 가능

2. 행렬곱 계산의 최적화

   1) 암호화된 행렬 A와 B에 대한 AB^T, A^TB의

      계산 최적화 알고리즘 개발 (DiagABT, DiagATB)

   → HETAL 성과 2) 1.8~323배 속도 향상

   - 행렬곱: 전체 학습 시간의 최대 55% 차지

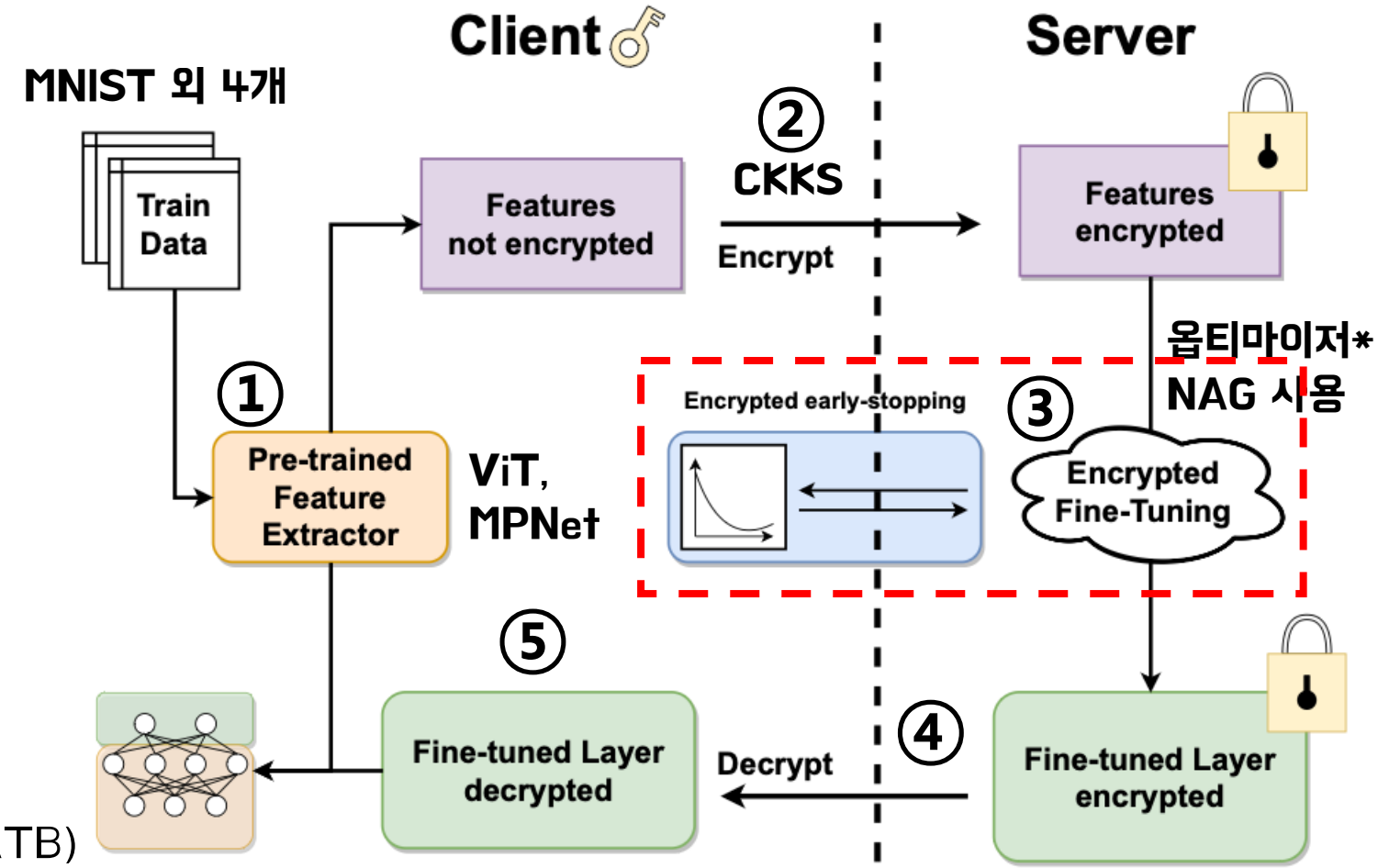


Figure 1. Our privacy-preserving transfer learning protocol (**HETAL**)

앞으로 해보고 싶은 연구

1. Data Privacy
   - 과거에 근무했던 병원, 은행에서의 경험
     + 데이터는 지금도 앞으로도 매우 중요

2. Privacy-preserving ML
   - 데이터가 안전하게 보호된 상태를 유지하며
     학습이 가능하다면 여러 산업에 100% 도움됨

3. Quantum-safe Cryptography
   - Quantum Computer가 상용화되더라도
     Lattice-based 암호(NP-hard에 가까운)인
     동형암호는 안전함