

TMAG-FED: 안전한 연합학습을 위한 2-패스 상호인증 가십 프로토콜

박상덕¹ · 장항배² · 이재우^{2*}

TMAG-FED: Two-pass Mutual Authenticatable Gossip Protocol for Secure Federated Learning

Sangdeok Park¹ · Hangbae Chang² · Jaewoo Lee^{2*}

¹Ph.D student, The Department of Security Convergence, Chung-Ang University, Seoul, 06974 Korea

^{2*}The Department of Industrial Security, Chung-Ang University, Seoul, 06974 Korea

요 약

인공지능(AI) 분야에서는 분산 저장된 데이터를 직접 공유하지 않으면서 서로 협력하여 인공지능 모델을 학습할 수 있는 연합학습(FL: Federated Learning)이라는 분산형 머신러닝 기법이 부각되고 있다. 특히, P2P(Peer to Peer) 네트워크에서 가십 프로토콜을 기반으로 연합학습을 수행하는 경우 트래픽 분산뿐만 아니라 각 클라이언트의 로컬 데이터를 이용하여 학습을 더 효율적으로 수행할 수 있다. 그러나, 연합학습은 일반적으로 신뢰성이 보장되지 않기 때문에 중요정보 유출 등과 같은 보안 위험이 발생할 수 있다. 본 논문은 P2P 연합학습 과정에서 정보보호 측면에서 발생할 수 있는 문제들을 해결하기 위해 2-패스 상호인증 기반의 안전한 푸시-풀 결합 가십(push-pull hybrid Gossip) 프로토콜을 제안한다. 이것은 한 번의 푸시(Push)와 풀(pull) 과정으로 송신자가 논스(NONCE) 값을 생성, 수신자와 공유함으로써 상호인증, 무결성 및 기밀성을 제공할 수 있어, P2P 네트워크 환경에서 각 클라이언트 간 신뢰 관계를 보장하게 되며 해당 클라이언트들은 수신 암호문을 복호화하여 추출한 모델 W들을 기반으로 집계된 정확한 모델 Wagg을 생성할 수 있다. 이 연구 결과는 향후 P2P 네트워크 환경에서 효율적인 통신뿐만 아니라 안전성과 정확성을 고려한 차세대 연합학습 전략으로 발전할 수 있다.

ABSTRACT

In the realm of artificial intelligence, Federated Learning(FL) has emerged as a promising technique. FL enables AI models to be trained collaboratively without compromising privacy by directly sharing sensitive data. Specifically, FL based on the Gossip protocol in peer-to-peer networks can enhance efficiency by leveraging the local data of each client and optimizing traffic distribution. However, the reliability of FL is not always guaranteed, leading to potential security risks such as data leakage. To address these challenges, this paper proposes a secure push-pull hybrid Gossip protocol based on two-pass mutual authentication. The protocol ensures mutual authentication, integrity, and confidentiality by generating and sharing NONCE values with the receiver in a single push and pull process. This fosters trust among clients in the peer-to-peer network environment, enabling them to generate accurate aggregated models Wagg based on decrypted model updates Ws. This research contributes to the development of advanced FL strategies that prioritize efficient communication, security, and accuracy in peer-to-peer network environments.

키워드 : 연합학습, 가십 프로토콜, 탈중앙학습, 상호인증

Keywords : Federated Learning, Gossip Protocol, Decentralized Learning, Mutual Authentication

Received 23 October 2024, Revised 7 November 2024, Accepted 25 November 2024

* Corresponding Author Jaewoo Lee (E-mail: jaewoolee@cau.ac.kr, Tel: +82-2-820-5935)

Associate Professor, The Department of Industrial Security, Chung-Ang University, Seoul, 06974 Korea

Open Access <http://doi.org/10.6109/jkiice.2024.28.12.1563>

pISSN:2234-4772

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론: 연구 배경 및 목표

최근 ICT 환경변화에 따른 클라우드 서비스 및 공급망 전반에서 비롯된 데이터의 폭발적인 증가와 그에 따른 유용한 정보공유는 국제사회의 중요한 이슈로 대두되고 있다. 이러한 디지털 기술의 발전과 ICT 환경변화는 다양한 분야에서 개인화된 서비스와 효율적인 데이터 분석의 필요성을 증가시켜왔다. 이러한 맥락에서 연합학습은 프라이버시를 보장하는 동시에 다양한 클라이언트 기기와 서버 간의 협력을 통해 지능적인 모델학습을 가능하게 하는 중요한 패러다임으로 부상했다[1, 2]. 즉, 연합학습을 활용하면 참여 클라이언트의 개인 훈련 데이터를 노출하지 않고 공동 머신러닝 모델을 구축할 수 있다.

연합학습은 각 참여 클라이언트가 로컬 데이터를 바탕으로 모델을 학습하고 학습된 모델의 업데이트된 데이터만을 중앙서버와 공유함으로써 프라이버시를 강화하고 데이터 전송에 따른 비용을 절감할 수 있다. 이러한 접근법은 의료, 금융, 스마트 시티 등과 같이 민감한 데이터를 다루는 산업 분야에서 그 가치가 높게 평가되고 있다.

그러나 연합학습의 효율성과 확장성은 여러 기술적 한계에 의해 제약받고 있다. 특히, 전통적인 서버-클라이언트 아키텍처에서 발생하는 통신비용, 데이터 동기화 문제, 서버의 병목현상, 그리고 확장성 제한은 연합학습이 다양한 응용 분야에서의 활용을 어렵게 만드는 주요 장애물이다[3]. 이와 함께, 다양한 기기와 복잡한 네트워크 환경에서의 동작은 성능저하, 통신 지연 및 불안정한 네트워크 연결 등의 문제를 야기할 수 있으며, 이는 전반적인 연합학습 과정의 신뢰성과 효율성을 저해하게 된다.

전통적인 연합학습 시스템은 중앙서버가 모델 업데이트를 집계하고 각 클라이언트는 서버로부터 모델을 받아 로컬 데이터에 기반한 학습을 수행하는 구조이다. 이러한 중앙집중식 접근법은 대규모 네트워크에서 과도한 통신비용과 트래픽 지연을 증가시키며, 서버의 과부하와 단일실패지점(single point of failure)을 초래할 수 있다. 뿐만 아니라, 모든 클라이언트가 동일한 업데이트를 받음으로써, 네트워크의 다양성과 개별 기기의 특성을 충분히 활용하지 못하는 문제가 있다. 이러한 문제는 대규모 네트워크와 다양한

기기로 구성된 시스템 환경에서 두드러지게 발생한다[4]. 또한, 연합학습은 보안에 취약할 수 있다. 최근 연구 결과[5]에 따르면 연합학습 과정에서 그래디언트를 통해 훈련 데이터까지 복원할 수 있다는 것이 증명되었다. 이에 대응하기 위해 연합학습 모델에서 적합한 보안기술 적용은 중요한 연구과제가 되었다.

이에, 최근 들어 연합학습의 새로운 통신 아키텍처로서 P2P 접근법이 제안되었다[6-8]. P2P 네트워크를 기반으로 한 연합학습은 중앙서버에 의존하지 않고 각 클라이언트가 동등한 위치에서 서로 직접 통신하며 학습을 진행하는 분산형 학습모델이다. 이 접근법은 네트워크 트래픽을 분산시키고 서버의 부하를 줄이며, 각 클라이언트의 로컬 데이터와 연계한 계산 능력을 더 효율적으로 활용할 수 있는 잠재력을 가진다. 그러나 P2P 연합학습은 클라이언트 간의 신뢰성 확보, 데이터 일관성 및 보안 문제해결 등 새로운 도전 과제를 수반한다[3, 4]. 특히, 네트워크상의 각 클라이언트가 서로 다른 데이터와 계산 능력을 가지고 있으며, 서로에 대한 사전 신뢰를 가지고 있지 않은 환경에서는 이러한 문제들이 더욱 심각해질 수 있다. 클라이언트 간에 직접 통신을 수행함으로써 발생할 수 있는 정보 유출 등의 보안 위험, 데이터조작 및 신뢰성 결여 등은 연합학습의 성능과 학습 데이터의 정확성에 심각한 영향을 미칠 수 있다. 이러한 문제를 해결하기 위해서는 클라이언트 간의 상호인증과 학습 데이터의 무결성 등 정보보호 서비스가 필수적으로 제공되어야 한다.

본 논문은 상기에서 언급한 데이터의 정확성과 보안 측면에서의 발생할 수 있는 문제들을 해결하기 위해 P2P 네트워크 기반 연합학습 과정에 적용할 수 있는 2-패스 상호인증 가십 프로토콜(Two-pass Mutual Authenticatable Gossip Protocol : TMAG-FED)을 제안한다.

TMAG-FED는 P2P 네트워크에서 각 클라이언트 간 안전하게 중요정보를 송수신하고 정확한 모델학습을 수행할 수 있도록 연합학습의 신뢰성과 보안성을 보장한다. 즉, 연합학습을 위한 풀과 푸시 통신 과정에서 기존 모델과 제안 모델의 비교(그림 1)하고 가십 프로토콜의 안전성을 증명한다. 가십 프로토콜 과정에서 악의적인 제3의 공격자가 P2P 네트워크 상에 존재한다고 가정하였을 때, TMAG-FED는 클라이언트 간 상호인증을 거쳐, 로컬의 학습 데이터를 보호함으로써 연합학습의 정확성 향상과 기존의 연합학습 과정에서

모델에 존재한 보안 취약점[5, 9]을 개선하는 연구 성과를 제시한다.

II. 관련 연구

2.1 P2P 네트워크에서의 연합학습

인공지능 분야의 분산형 기계학습 기술 중에서 연합 학습 환경에서의 가십 학습에 대한 탐구는 최근 국내외 연구진에게 상당한 주목을 받고 있다.

효율적인 연합학습 모델 구현 분야는 다음과 같이 연구되었다. Chenghao Hu[10]는 세분화된 가십 집계 메커니즘을 도입하여 중앙집중식 서버가 없을 때 대역폭을 최대한 활용할 수 있는 등 효율성과 모델 품질을 개선하는데 중점을 둔 연합학습 모델을 제시한다. 이 접근방식은 Peer의 세그먼트 모델을 활용하여 가속화된 전파를 촉진하고 Combo라는 분산형 연합학습 시스템을 통해 모델의 무결성을 보장한다. Combo의 설계는 연합학습 시스템에서 모바일 및 임베디드 장치의 고유한 불안정성 특성에도 불구하고 강력한 모델 수렴을 보장하여 작업자의 동적 참여에 적응하는 능력으로 특히 주목할 만하다.

연합학습 모델의 확장성 및 응용 분야는 다음과 같이 연구되었다. István Hegedűs[6]는 전통적인 연합학습과 가십 학습 프레임워크를 심층적으로 비교한다. 이 연구는 중앙 집계기가 필요하지 않아 분산학습 환경에서 개인정보보호 제공과 확장성을 향상시킬 수 있는 가십 학습의 잠재력을 강조한다. 또한, 가십 학습이 로컬 계산 리소스와 데이터를 효과적으로 활용하여 분산 노드에서 학습 프로세스를 균등 분배할 수 있는 방법 등을 제시한다.

그리고 Bal'azs Sz'or'enyi[8]는 분산 확률 밴디트 알고리즘을 최적화하는데 가십 프로토콜을 적용하는 방법을 제시한다. 이 연구는 가십 학습 패러다임을 불확실성 하에서 의사결정 영역으로 확장하여 가십 기반 알고리즘이 탐색-착취 딜레마를 효과적인 분산 방식으로 해결할 수 있는 방법을 보여준다. 이 논문의 연구 결과는 가십 학습이 전통적인 기계학습 작업을 넘어 응용 분야 확장 적용 가능성을 보여줌으로서 분산 의사결정 시스템의 연구를 위한 새로운 길을 열어준다.

연합학습의 정확성, 보안성 및 가용성 제고 분야는 다음과 같이 연구되었다. Róbert Ormándi[7]는 선형 모델로 가십 학습 프로토콜을 구현하는 기술적 복잡성을 탐구한다. 가십 알고리즘의 수학적 기반과 완전 분산 데이터 세트를 처리하는데 있어서 유의미한 효과를

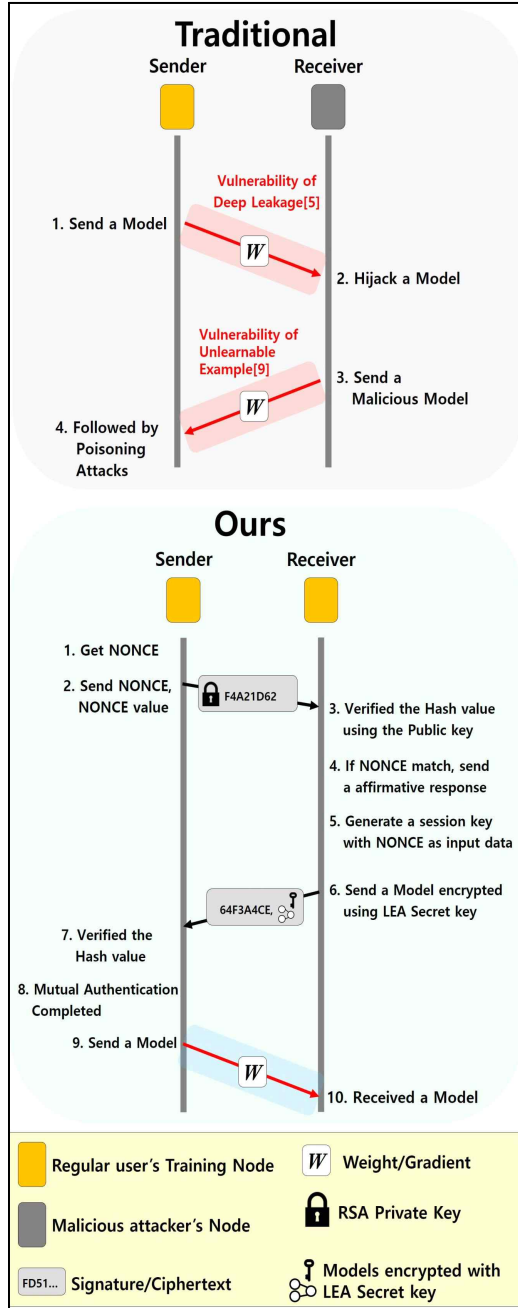


Fig. 1 Comparison between the traditional approach and ours

도출하기 위한 포괄적인 프레임워크를 제공한다. 이 연구는 중앙집중식 학습모델과 비슷한 정확도를 달성하는 동시에 단일장애지점에 대한 데이터 관련 개인정보보호 및 시스템 복원력을 크게 향상시키는 가십 학습의 가능성을 보여주는데 중요하다.

종합적으로 상기에 소개한 연구는 데이터 관련 개인정보보호, 확장성 및 동적 네트워크 적응을 포함한 분산 머신러닝의 가장 시급한 과제를 해결하기 위한 솔루션으로 가십 학습의 다양성과 잠재력을 강조한다.

III. 예비 연구(Preliminary Research)

3.1 연합학습

인공지능 관련 학습을 위한 데이터 수집 과정에서 발생할 수 있는 데이터 전송 비용 문제 및 개인정보 침해 문제를 해결하기 위해 연합학습이 제안되었다[1, 2]. 연합학습은 분산형 장치 또는 서버 전반에 걸쳐 기계학습 모델을 훈련하는 접근방식으로 통신비용 절감과 프라이버시 보장 및 보안성을 강화하는 목적으로 최근 국내외에서 활발히 연구되고 있는 분야이다. 특히, 연합학습은 데이터의 중앙집중화 없이 다양한 로컬 장치에서 모델 훈련을 수행함으로써 데이터 전송량을 최소화하는 장점을 제공한다. 또한, 프라이버시 침해 우려 없이 사용자 데이터를 효과적으로 보호하는 동시에 중앙서버에 대한 과도한 의존성을 감소함으로써 보안 위험을 낮출 수 있다.

연합학습 구조는 서버-클라이언트, 순차 모델, P2P 및 클러스터링 모델로 나누어 볼 수 있다. 대표적인 연합학습의 알고리즘은 FedSGD(Federated Stochastic Gradient Descent) 및 FedAVG(Federated Averaging) 등이 있다[11, 12].

FedSGD 알고리즘은 각 모바일 기기의 그래디언트를 서버로 전송하여 서버에서 전달받은 그래디언트의 평균 값을 구해 글로벌 모델을 업데이트한다. 또한, FedAvg 알고리즘은 FedSGD와 미니 배치를 함께 사용하여 각 모바일 기기가 직접 모델 업데이트를 진행한 후 서버는 가중치의 평균을 구해 글로벌 모델을 업데이트한다.

다음 그림 2는 서버-클라이언트 구조로 이루어진 일반적인 연합학습 과정을 도식화한 것이다.

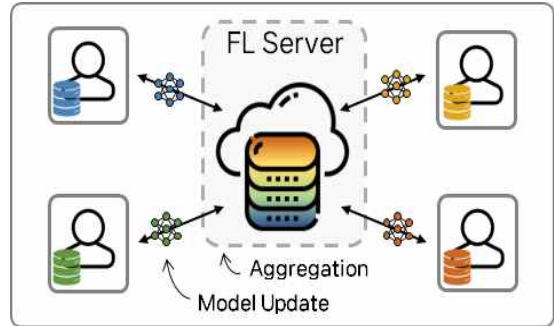


Fig. 2 Schematic of Federated learning[13]

모델 훈련은 각각의 클라이언트에서 수행되며, 각 클라이언트는 중앙서버로 개선된 모델 파라미터(또는 가중치)만을 전송, 개인정보 등 기밀 데이터의 외부 노출 없이 모델을 최적화할 수 있다. 네트워크 운용 환경에 따른 연합학습의 프로세스는 다음 그림 3과 같다.

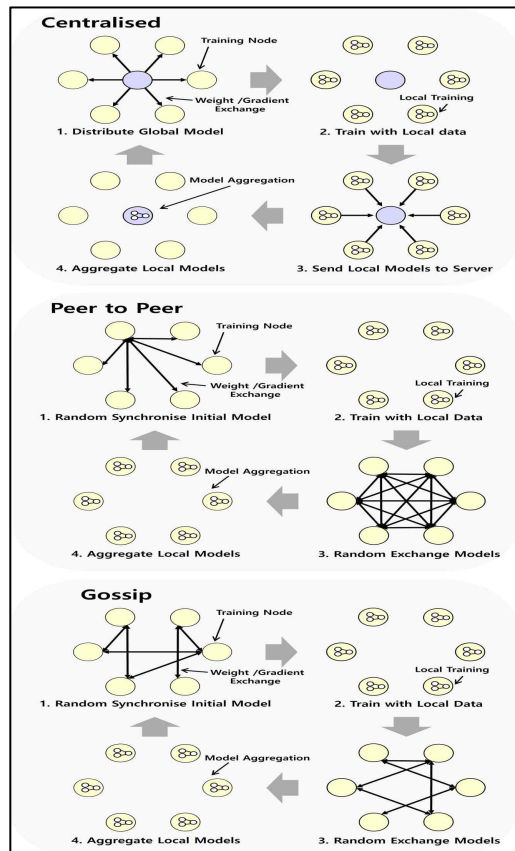


Fig. 3 Federated learning approaches

연합학습 프로세스는 다음과 같이 수행된다.

- ① 현재 시점에서 가장 최신 모델을 참여 클라이언트에 공유한다.
- ② 클라이언트 내 모델은 해당 기기에서만 존재하는 데이터를 활용하여 훈련한 후 업데이트된 미세한 부분(가중치)만 중앙서버로 암호화하여 전송한다. 이 과정에서 클라이언트 내 데이터는 해당 클라이언트에만 존재하고 중앙서버로 전송되지 않는다.
- ③ 중앙서버는 각 클라이언트로부터 받은 업데이트 값들을 집계한 후 글로벌 모델을 업데이트시킨다.
- ④ 중앙서버는 업데이트한 글로벌 모델을 다시 각 클라이언트에게 공유시키기 위해 ① 단계로 돌아가 이러한 과정을 반복하게 된다.

다만, 연합학습에서 외부와의 통신을 위해 참여 클라이언트의 네트워크 포트는 항상 열려있어야 하며 중앙서버가 단일 서버 모델일 경우 서론에서 언급한 바와 같이 단일장애지점 문제와 항상 서비스 가능 상태가 불가할 수 있다는 문제점이 내재되어 있다. 이러한 서버-클라이언트 구조의 연합학습 모델의 단점을 보완하기 위해 P2P 형태의 네트워크를 활용할 수 있지만 P2P 네트워크로 구성된 경우 네트워크의 작은 변화나 특정 노드의 불안정, 메시지 전달 지연 및 기기의 배터리 부족 등으로 연합학습이 정상적으로 진행되지 않을 수 있는 단점도 존재할 수 있다. 특히, P2P 네트워크 환경에서 연합학습을 구성할 경우 로컬 클라이언트의 참여 또는 이탈이 자유롭기 때문에 악의적인 공격자가 정당한 클라이언트로 가장하여 연합학습에 참여할 수 있다. 이로 인해 학습 참여자의 신뢰도에 따른 문제점도 존재하는데 인증받지 않은 비인가 참여자는 전달받은 모델을 기반으로 학습 데이터를 탈취하거나[5] 악의적으로 모델 업데이트를 방해하는 가짜 모델을[9, 14] 전송할 수 있다.

3.2 P2P 네트워크에서의 가십 프로토콜

가십 프로토콜은 전염병 프로토콜로도 알려져 있으며 소셜 네트워크에서 정보가 확산되는 과정에서 영감을 받은 통신 메커니즘이다. 이러한 프로토콜은 분산시스템을 위해 설계되었으며 특히, 대규모 P2P 네트워크에서 데이터를 전파하는데 적합하다. 가십 프로토콜에서 네트워크의 노드는 무작위로 선택된 Peer 세트와 정기적으로 데이터를 교환하며 사람들

사이에 소문이 퍼지는 방식과 유사하다. 이는 P2P 네트워크 전체에 데이터를 효율적으로 전파하여 참여 클라이언트에게 공유할 수 있다.

가십 프로토콜은 다음 4가지 특성을 갖는다.

① **분산화**: 가십 프로토콜은 정보의 배포를 제어하는 중앙기관 없이 탈중앙화된 방식으로 운영된다. 특히, 네트워크 결함에 영향을 거의 받지 않는다. 다시 말해 이 기능은 몇 개의 노드의 실패가 네트워크의 통신 능력에 큰 영향을 미치지 않기 때문에 본질적으로 확장 가능하고 장애에 견딜 수 있다.

② **확장성**: 가십 프로토콜의 효율성은 노드 수가 증가하더라도 크게 저하되지 않는다. 또한, 확장성은 로컬 노드 간 상호작용을 통해 달성되며, 각 노드는 임의로 선택된 작은 다른 노드의 서버 셋과만 통신하여 전체 네트워크 트래픽을 줄인다.

③ **내건고성과 내결함성**: 가십 프로토콜은 노드 장애와 네트워크 파티션에 대한 복원력이 매우 높다. 각 노드는 독립적으로 정보를 교환할 Peer를 선택하기 때문에 네트워크의 일부에 액세스할 수 없게 되더라도 가십 프로토콜은 여전히 작동할 수 있다.

④ **단순성과 유연성**: 가십 프로토콜은 메시지를 수신한 노드의 수는 지수적으로 증가하게 되고 선형적 성질을 갖게 되어 브로드 캐스팅하기 수월하다. 또한, 구현이 간단하며 메시지 배포, 데이터 집계 및 네트워크 토폴로지 관리 등 다양한 응용 분야에 적용이 가능하다.

가십 프로토콜은 많은 기계에 걸쳐 소프트웨어 또는 구성을 업데이트하기 위한 분산시스템과 네트워크에 걸쳐 데이터를 집계하기 위한 분산시스템 및 분산 데이터베이스의 일관성 유지 등 다양한 영역에서 성공적으로 적용되어 왔다. 연합학습의 측면에서 가십 프로토콜은 일반적인 분산뿐만 아니라 효율적이며 내결함성 방식으로 모델 업데이트의 분산 및 집계를 관리할 수 있는 유망한 접근방식을 제공한다. 그러나, 가십 프로토콜은 이러한 장점에도 불구하고 네트워크 트래픽 증가로 이어질 수 있는 메시지의 중복, 전파 속도와 네트워크 오버헤드 간의 상충 문제 등에 직면해 있다. 또한, 가십 프로토콜 기반 시스템은 데이터 일관성을 보장하고 분산 알고리즘의 수렴 등을 관리하기 위해 신중한 설계와 조정이 필요하다. 그럼에도 가십 프로토콜은 분산 네트워크에서 정보를 전파하기 위한 강력한 프레임워크를 제공한다.

본 논문에서는 개선된 연합학습 모델 구현을 위한 P2P 네트워크 환경에서의 안전한 가십 프로토콜 적용으로 기존의 전통적인 접근방식이 직면한 확장성의 어려움과 효율성 저하 및 정보보안 이슈의 해결을 보장하며 흥미로운 연구 방법을 제시한다.

3.3 상호 인증프로토콜

인증, 즉 개인식별(Identification) 또는 실체 인증(entity authentication)이란 특정 시스템 또는 사용자의 실제 신원을 확인하는 과정을 말한다[15]. 예를 들어, 목소리나 지문에 의해서 특정 인물의 신원을 파악할 수 있다. 또한, 신분증에 있는 사진과 신분증 소지자를 비교해봄으로써 그 소지자의 신분을 파악할 수 있게 된다. 통상적으로 클라이언트가 서버에게 자신의 신원을 증명하는 인증 방법은 크게 세 가지로 구분할 수 있다. 첫째는 사용자 고유의 물리적인 특성을 이용하여 개인의 신원을 확인할 수 있다. 음성, 지문 그리고 홍채의 혈관구조는 사람마다 상이한 유형을 띄고 있어 개개인을 식별하는데 사용되어질 수 있다. 둘째는 패스워드와 같이 사용자가 알고 있는 정보를 이용하는 방법이다. 사용자는 자신의 ID와 함께 패스워드를 제공함으로써 자신의 신원을 밝히고 시스템 서버에 로그인 할 수 있다. 셋째는 사용자가 소지하고 있는 것을 이용하는 것이다. 사용자에게 주어지는 메모리카드나 스마트카드와 같은 토큰 카드를 이용하여 클라이언트와 서버 간에 시도-응답 방식의 개인 식별 프로토콜이 수행된다. 토큰 카드에는 패스워드와는 달리 사용자들이 기억하기 힘든 2진 수열의 비밀키가 내장되어 서버와 함께 수행되는 개인식별 프로토콜의 보조수단으로 이용된다. 이러한 개인식별 프로토콜에는 대칭형 암호 또는 공개키 암호가 기반을 이루기 때문에 2진 수열의 비교적 큰 수의 비밀키가 사용된다. 개인식별 프로토콜이 컴퓨터와 컴퓨터 간에 수행된다면 별문제가 없지만 사용자와 컴퓨터간의 개인식별에 있어서는 패스워드와 달리 그러한 비밀키를 사람이 어떠한 보조수단 없이 기억하기는 거의 불가능하다. 인증은 대칭형 암호 및 공개키 암호에 기반을 둔 일방향 개인식별과 상호개인식별 프로토콜이 있다. 클라이언트가 서버에게 자신의 신원을 확인시키기 위해 주로 일방향 개인식별 프로토콜을 사용한다. 하지만, 공격자에 의한 가장 공격은 클라이언트뿐만 아니라 서버에 대해서도 동일하게 적용될 수가 있기

때문에 클라이언트의 입장에서는 자신의 신원을 확인시키는 동시에 서버의 신원도 확인할 수 있는 상호 개인식별 프로토콜이 요구되어진다. 특히, 시도와 응답이 프로토콜 진행 과정에서 임의의 공격자에게 관측될 수 있다고 할지라도 공격자가 정당한 실체를 가장하여 그것을 후속의 개인식별 프로토콜에서는 사용할 수 없어야 한다. 따라서, 시도-응답 방식의 개인식별 프로토콜에서 사용되는 시도는 매 세션의 프로토콜마다 그 값이 가변적인 난수, 순번 및 시각표 등을 이용하고 있는데, 오직 한 번만 사용된다는 의미에서 이런 것들을 ‘논스’라고 한다.

본 논문에서는 P2P 네트워크 환경에서 효율적으로 사용할 수 있는 해쉬함수, 대칭형 암호 및 공개키 암호에 기반을 둔 2-패스 상호 개인식별 프로토콜을 적용한다. 서버가 필요하지 않은 P2P 네트워크 환경에서 안전한 연합학습을 위해 참여 클라이언트 간 상호인증은 중요한 보안 요구사항이다. 왜냐하면, 정보 탈취 등을 목적으로 한 미상의 공격자가 정당한 클라이언트로 가장하여 P2P 네트워크 환경에서 연합학습에 참여할 경우 연합학습 모델 관련 데이터가 부정확할 수 있을 뿐만 아니라 중요 데이터가 비인가자에게 노출되거나 또 다른 공격 유형으로 악용될 수 있기 때문이다. 이에 따라, 다양한 네트워크 환경에서 안전한 연합학습을 통해 유의미한 학습모델 도출을 위해서는 참여 클라이언트 간 상호인증은 필수적으로 수행되어야 한다.

IV. 제안 알고리즘: TMAG-FED

4.1 2-패스 상호인증 가십 프로토콜

TMAG-FED는 클라이언트 간 상호인증을 통해 P2P 네트워크 기반 연합학습 과정에서 안전성을 제공한다. 즉, TMAG-FED는 푸시-풀 결합 가십 프로토콜 통신 과정으로 확장한다. P2P 네트워크 기반 연합학습에 참여한 각 클라이언트 $s, r \in C$ 에 대해 다음의 과정을 통해 안전한 통신 세션이 이루어진다.

① **로컬 학습(Local training)**: 클라이언트 s 는 자신이 보유한 데이터 D_s 를 이용해 로컬 모델을 학습한다. 이를 수식화하면 다음과 같다.

$$W_{t+1} \leftarrow W_t - \eta \nabla l(W_t; D_s)$$

② **푸시(push)**: 클라이언트 s 는 이웃 클라이언트들 중에서 사전 정의된 팬아웃(fanout : 해당 라운드에 통신할 클라이언트 수) f 만큼 클라이언트를 선택하고 난수 생성기를 사용하여 NONCEs 값을 생성한다. 이 NONCEs를 seed값으로 해쉬함수 $H()$ 에 적용하여 경량 블록암호 LEA(Lightweight Encryption Algorithm)의 대칭키 k 를 생성한다.

이후 $RSA(RSA(H(NONCEs))^d || NONCEs^e || LEAk(Ws))$ 을 계산하여 클라이언트 r 에게 전달한다. 여기서 d 는 송신 클라이언트 s 의 개인키를 의미하며, e 는 수신 클라이언트 r 의 공개키를 의미한다. 또한, $H()$ 는 해쉬함수(SHA-256)를 의미하고 LEA 블록 알고리즘은 부채널 공격에 대비, 마스킹 기법을 적용한다. 여기서 마스킹 기법은 암호화 알고리즘의 연산 과정에 나타난 중간값에 난수를 더하여 키를 감추는 것을 의미하는데, 전력·시차 분석 등 부채널 공격에 대응하기 위해 주로 사용한다.

③ **수신(Receive)**: 수신 클라이언트 r 은 전달받은 암호문을 r 의 개인키로 복호화하고 송신 클라이언트 s 의 공개키를 이용하여 서명값을 검증한다. 즉, 서명값을 통해 추출한 해쉬값과 NONCEs를 직접 해쉬 연산한 값이 동일하다면 정당한 클라이언트로부터 변조되지 않은 NONCEs를 수신한 것으로 판단하고 다르다면 NONCEs 값이 변조된 것으로 간주할 수 있다. 또한, 송신 클라이언트 s 는 수신 클라이언트 r 에게 NONCEs를 생성하여 보내지 않았다고 부인할 수 없게 된다. 다시 말해 송신 클라이언트 s 는 자신이 로컬 모델인 Ws 를 수신 클라이언트 r 에게 보낸 적이 없다고 부인할 수 없다는 것을 의미한다. 이후 수신 클라이언트 r 은 송신 클라이언트 s 와 안전하게 공유한 NONCEs를 seed 값으로 LEA의 대칭키 k 를 생성, 암호화된 로컬 모델 Ws 를 복호화하여 얻을 수 있다.

④ **풀(Pull)**: 클라이언트 r 은 수신 과정을 통해 송신 클라이언트 s 를 인증하였다면 자신이 생성한 대칭키 k 로 Wr 을 암호화하여 클라이언트 s 에게 전달한다.

⑤ **집계(Aggregation)**: 각 클라이언트는 정당한 송신 클라이언트로부터 암호문을 전달받았다면 복호화할 수 있고 그 과정에서 얻은 모델 W 들을 기반으로 집계된 모델 $Wagg$ 를 생성한다.

본 논문에서 제안한 TMAG-FED의 P2P 통신 과정에 참여하는 모든 클라이언트들은 제3의 신뢰할 수 있는 인증기관인 CA(Certificate Authority)로부터 RSA-2048

알고리즘에서 사용할 개인키와 공개키 쌍을 사전에 생성하여 제공받게 되고 이러한 클라이언트들은 최초로 모두 신뢰할 수 있는 클라이언트라는 가정을 기반으로 한다.

Algorithm 1 TMAG-FED Algorithm

Function: ClientUpdate(W^t)

```

1: initialize  $W^0$ 
2:  $\beta \leftarrow$  (split  $D_k$  into batches of size  $B$ )
3:
4: for each client epoch  $i=1,2,...,E$  do
5:   for each batch  $b \in \beta$  do
6:      $W^{t+1} \leftarrow W^t - \eta \nabla l(W^t; b)$ 
7:   end for
8: end for
9:  $W^{t+1} \leftarrow$  SecureCommunication( $W^{t+1}$ )
```

Fig. 4 Proposed Algorithm(TMAG-FED) process

V. 안전성 분석

5.1 암호 알고리즘

연합학습에서 클라이언트는 소형 디바이스로 계산량, 메모리 공간 등이 제한될 수밖에 없다. 이에 따라, LEA 경량암호알고리즘을 선택하여 소형 디바이스에 탑재, 암호 기능을 수행한다. 그러나, 이러한 사물인터넷 등에 활용하는 소형 디바이스는 알려진 부채널 공격에 쉽게 노출될 수 있다. 제안하는 2-패스 상호인증 가십 프로토콜에 적용한 LEA 경량암호알고리즘은 알려진 CPA(Correlation Power Analysis) 공격 등 부채널 공격에 대응 가능하다. 즉, TRNG(True Random Number Generator) 기반의 마스킹 방법을 적용한다. TRNG에서 생성되는 난수를 암호화 연산 중간값에 더하는 마스킹 기법을 적용한 결과, 최대 상관 계수가 매우 작아 잘못된 라운드키 값이 분석되어 CPA 공격 등 부채널 공격에 강인함을 확인할 수 있다[16]. 이렇게 난수를 이용한 마스킹 기법의 선형 연산 과정은 다음 그림 5와 같다.

Algorithm 2 Masked linear operation

Input: x', y' such that $x' = x \oplus m_x$, $y' = y \oplus m_y$, m_x and m_y are masked value for x and y , respectively.

Output: $z' (= (x * y) \oplus m_z)$, m_z such that $*$ is linear operation and m_z is the masking value for $z = x * y$.

```

1:  $z' \leftarrow x' * y'$ 
2:  $m_z \leftarrow (m_x \oplus m_y)$ 
3: return  $z', m_z$ 
```

Fig. 5 Masked operation process

또한, 논스를 입력 값으로 해쉬함수 $H()$ 를 활용, 매 세션마다 상이한 암호키 값(256비트)을 생성한다. 이 암호키는 LEA 경량암호알고리즘의 암호·복호화에 활용하게 되어 TMAG-FED에 적용한 암호 알고리즘은 중간자 공격 등 사이버 공격을 방어할 수 있다. 한편, LEA 경량암호알고리즘은 다양한 SW 환경에서 국제 표준 암호인 AES(Advanced Encryption Standard) 대비, 1.5배~2배 성능이 좋아, P2P 네트워크의 연합학습에서 자원이 한정된 다양한 소형 디바이스를 적용하는 클라이언트 환경에 적합하다.

제안한 TMAG-FED에 적용한 공개키 기반 암호 RSA-2048과 해쉬함수 SHA-256은 국제 표준으로 현재까지 보안 취약점이 발견되지 않는 등 안전성이 유지되고 있어, 세계 각국이 다양한 응용 분야에서 사용하고 있는 알고리즘이다.

TMAG-FED은 P2P 네트워크 상에서 클라이언트 간 안전한 데이터 공유를 위해 가십 프로토콜을 채택, RSA-2048과 SHA-256을 적용한다. 연합학습에 참여하는 클라이언트 간 안전한 키 공유와 데이터 W_s 보호 과정은 아래의 수식을 통해 확인할 수 있다. 즉, $RSA(RSA(H(NONCE_s))^d || NONCE_s) || LEA_k(W_s)$ 은 수신 클라이언트 r 이 송신 클라이언트 s 가 생성한 NONCEs라는 사실을 증명할 수 있다. 이것은 송수신 클라이언트 간 안전하게 세션 키를 공유하므로써 데이터 W_s 를 보호할 수 있다는 것을 의미한다.

5.2 2-패스 상호인증 가십 프로토콜

가십 프로토콜에는 푸시 가십과 풀 가십 방법이 있다. 본 논문에서는 두 가지 방법을 적절히 통합하여 네트워크 전체로의 메시지 전파 속도가 기존 FedAvg에 대비 향상되고, 모든 참여 클라이언트가 메시지를 수신할 수 있는 푸시-풀 결합 가십 프로토콜을 활용한 연합 학습을 수행한다[17]. 송신 클라이언트는 자신이 보유한 데이터 D_s 를 이용해 로컬 모델로 학습된 데이터 W 를 안전하게 푸시하기 위해 매 세션마다 상이한 난수를 통해 암호화 키를 생성한다. 이것은 난수 기반 인증프로토콜의 보안요구사항인 패스워드 추측 공격과 재생 공격 그리고 위조 공격 등 알려진 공격에 대해 안전성을 보장한다[18]. 더불어 제안한 가십 프로토콜 수행 과정에서 송신자의 서명 값을

수신자가 확인함으로써 수신자는 정당한 송신자로부터 위조되지 않은 학습 데이터를 제공받았다는 것을 확실하게 되고 송신자는 자신이 수행한 로컬 모델을 수신자에게 제공한 사실에 대해 부인할 수 없다(그림 6). 이것은 각 클라이언트들이 주기적으로 랜덤한 타켓을 골라 브로드 캐스팅하여 가십 메시지를 전송하는 네트워크 환경에서 필수적으로 수행되어야 하는 정보보호 요소이다.

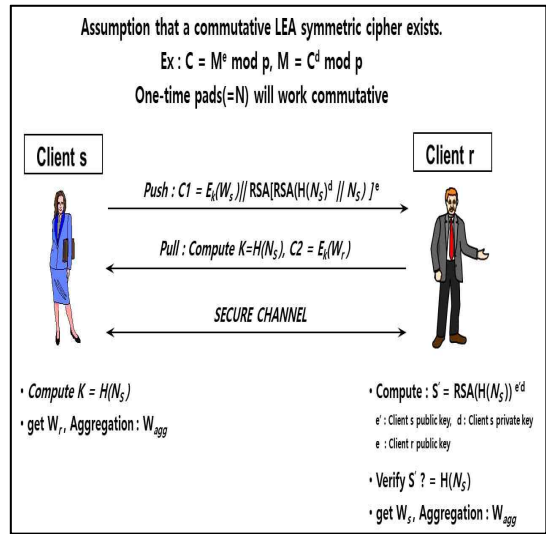


Fig. 6 Two-Pass Mutual Authentication Process

요약하자면, TMAG-FED는 P2P 네트워크에서 참여 클라이언트 간 매 세션마다 상호인증을 제공하며, 로컬 모델을 보호한다. 본 논문에서 제안한 TMAG-FED는 전체적인 P2P 네트워크에서 정확한 연합학습을 수행할 수 있도록 설계된 암호 프로토콜이다.

VI. 결 론

인공지능 등 첨단기술의 급속한 발달에 따라 정보보호 기술의 중요성과 수요는 지속적으로 증가하고 있다. 최근 ChatGPT의 상용화로 편리성과 함께 신뢰성이 사회적 이슈로 부각 된다. 즉, 빅데이터 시대를 맞아 넘쳐나는 정보의 홍수에 살고 있지만, 과연 사용자에게 신뢰할 수 있는 정확한 정보를 제공하는지에 대한 이슈는 지속적으로 대두되고 있다.

본 논문에서는 최근 첨단기술로 부각되고 있는 인공지능과 암호기술 그리고 가십 프로토콜을 접목한 TMAG-FED를 제시하였다. 제안한 알고리즘은 기존의 TLS1.3 프로토콜에서 사용하는 AES 암호보다 성능이 우수한 LEA 경량 암호를 적용하였다. 즉, P2P 네트워크 환경에서의 효율적인 연합학습 구현[17]뿐만 아니라 기밀성, 메시지 무결성, 인증 및 부인 봉쇄 등 정보보안 서비스 제공에 의한 로컬 모델 보호를 통해 정확한 학습모델을 제공할 수 있다는 것을 보여준다.

이러한 측면에서 본 논문의 연구 결과는 향후 인공지능 분야에서 개인정보보호 등을 제공하는 연합학습의 사이버보안 분야 연구 기반이 될 뿐만 아니라 광범위한 이기종 네트워크 환경에서의 연합학습 연구로 확장을 기대할 수 있다. 이것은 궁극적으로 국내 정보보안 분야 연구개발 및 ICT 산업 활성화에 기여할 것이다.

ACKNOWLEDGEMENTS

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Convergence security core talent training business support program (No. IITP-2024-RS-2023-00266605) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

- [1] B. M. Mahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Artificial intelligence and statistics*, Lauderdale: USA, pp. 1-10, 2017.
- [2] P. Kairouz, H. B. M. Mahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, and R. Cummings et al., "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no.1-2, pp.1-210, Jun. 2021.
- [3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50-60, May 2020. DOI: 10.1109/MSP.2020.2975749.
- [4] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. M. Mahan, T. V. Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," in *Proceedings of Machine Learning and Systems 1 (MLSys 2019)*, Palo Alto: USA, pp1-15, 2019.
- [5] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver: CA, pp. 1-11, 2019.
- [6] I. Hegedus, G. Danner, and M. Jelasity, "Gossip learning as a decentralized alternative to federated learning," in *Proceedings of the 19th IFIP WG 6.1 International Conference*, Kongens Lyngby: DK, pp. 17-21, 2019. DOI: 10.1007/978-3-030-22496-7_5.
- [7] R. Ormandi, I. Hegedus, and M. Jelasity, "Gossip learning with linear models on fully distributed data," *Concurrency and Computation: Practice and Experience*, vol. 25, no. 4, pp. 556-571, May 2012. DOI: 10.1002/cpe.2858.
- [8] B. Szorenyi, R. B. Fekete, I. Hegedus, R. Ormandi, M. Jelasity, and B. Kegl, "Gossip-based distributed stochastic bandit algorithms," in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta: USA, pp. 19-27, 2013.
- [9] H. Huang, X. Ma, S. M. Erfani, J. Bailey, and Y. Wang, "Unlearnable examples: Making personal data unexploitable," *arXiv preprint arXiv: 2101.04898*, Jan. 2021. DOI: 10.48550/arXiv.2101.04898.
- [10] C. Hu, J. Jiang, and Z. Wang, "Decentralized federated learning: A segmented gossip approach," *arXiv preprint arXiv: 1908.07782*, Aug. 2019. DOI: 10.48550/arXiv.1908.07782.
- [11] J. Konecny, H. B. M. Mahan, F. X. Yu, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv: 1610.05495*, Oct. 2016. DOI: 10.48550/arXiv.1610.05492.
- [12] H. B. M. Mahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," *arXiv preprint arXiv: 1602.05629*, Feb. 2016. DOI: 10.48550/arXiv.1602.05629.
- [13] S. Park, Y. Chang, and J. Lee, "FedWPSO: Federated Learning using PSO with added Weighted Aggregation," in *Proceedings of KIISE Computer System Society*, Gangwon: KR, pp. 1, 2024.

- [14] J. Zhang, X. Ma, Q. Yi, J. Sang, Y. G. Jiang, Y. Wang, and C. Xu, "Unlearnable clusters: Towards label-agnostic unlearnable examples," *arXiv preprint arXiv: 2301.01217*, Dec. 2022. DOI: 10.48550/arXiv.2301.01217.
- [15] C. Park, *Cryptographic Theory and Security*, 1st ed, Paju, KR: Dae Young Sa, 2001.
- [16] H. S. An and K. W. Shin, "Correlation Power Analysis Attack on Lightweight Block Cipher LEA and Countermeasures by Masking," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 7, pp. 1276-1284, Jul. 2017. DOI: 10.6109/JKIICE.2017.21.7.1276.
- [17] S. Park and J. Lee, "FedSnM: Score-and-Model based Communication-Efficient Federated Learning in Peer-to-Peer Network Environment," *Journal of the Korea Institute Of Information and Communication Engineering*, vol. 27, no. 2, pp. 192-198, Feb. 2023. DOI: 10.6109/jkiice.2023.27.2.192.
- [18] H. S. Kim and J. Y. Choi, "Security Analysis of RFID System using Nonce based Mutual Authentication Protocol," *Korean Institute of Information Scientists and Engineers*, vol. 36, no. 2D, pp. 7-12, Nov. 2009.



박상덕(Sangdeok Park)

고려대학교 수학과 이학사 (2001)
고려대학교 정보보호대학원 정보보호학과 공학석사 (2004)
중앙대학교 대학원 융합보안학과 박사 (2023~)
※관심분야 : 암호, 산업보안, 인공지능 보안, 연합학습



장항배(Hangbae Chang)

중앙대학교 컴퓨터공학과 공학사 (1997)
중앙대학교 대학원 컴퓨터공학과 공학석사 (1999)
연세대학교 대학원 정보시스템관리학과 공학박사 (2006)
중앙대학교 산업보안학과 정교수 (2014~)
※관심분야 : 개인정보보호, 산업보안, 보안 데이터 분석



이재우(Jaewoo Lee)

서울대학교 컴퓨터공학과 공학학사 (2006)
서울대학교 대학원 컴퓨터공학과 공학석사 (2008)
University of Pennsylvania Computer and Information Science, Ph.D (2017)
중앙대학교 산업보안학과 부교수 (2018~)
※관심분야 : 실시간 시스템, 사이버 물리 시스템 보안