오류 정정 부호를 활용한 고신뢰 차등 프라이버시 기법

지승하 [1], 전소은 [2], 이일구 [3]
[1] 성신여자대학교 융합보안공학과 학부생
[2] 성신여자대학교 미래융합기술공학과 박사과정
[3] 성신여자대학교 융합보안공학과, 미래융합기술공학과 부교수

{20210640, 220237020}@sungshin.ac.kr, iglee@sungshin.ac.kr

Highly Reliable Differential Privacy Technique Utilizing Error Correction Encoding

Seung-ha Ji[1], So-Eun Jeon[2], Il-Gu Lee[3]
[1]Dept. of Convergence Security Engineering, Sungshin Women's University
[2]Dept. of Future Convergence Technology Engineering, Sungshin Women's University
[3]Dept. of Convergence Security Engineering & Future Convergence Technology Engineering, Sungshin Women's University

요 약

IoT 장치의 개수가 급증함에 따라 네트워크 환경에서 송수신되는 데이터 양이 증가하였고, 이에 따라 데이터 전송과정의 보안 강화가 중요해지고 있다. 기존에는 데이터에 인공 노이즈를 추가하는 차등 프라이버시 기법(Differential Privacy, DP)을 적용하여 데이터를 보호하고 있다. 하지만 DP가 적용된 데이터를 수신하는 정상 사용자의 머신러닝 학습 정확도가 감소되는 문제가 있다. 본 논문에서는 고신뢰 데이터 전송을 위한 데이터 인코딩 기반의 DP 기법인 EN-DP (Encoding-based DP) 모델을 제안한다. 실험 결과에 따르면, EN-DP 를 통한 정상 사용자와 공격자 간의 학습 능력 정확도 간극을 종래 모델 대비 최대 17.16% 개선할 수 있음을 입증하였다.

1. 서론

네트워크에 연결되는 사물인터넷(Internet of Things, IoT) 장치의 개수가 증가함에 따라 네트워크 환경 상에서 송수신되는 데이터의 양이 급증하고 있으며, 데이터 전송 과정의 보안 강화가 중요해졌다. 또한 머신러닝 기술의 발전과 함께 학습 모델을 악용한 사이버 공격이 고도화되면서 데이터 보안의 중요성은 강조되고 있다. 이에 따라 데이터에 인공 노이즈를 추가하여 공격자가 사용자의 정보를 유추하지 못하게 하는 차등 프라이버시 기법 (Differential Privacy, DP)이 연구되고 있다 [1].

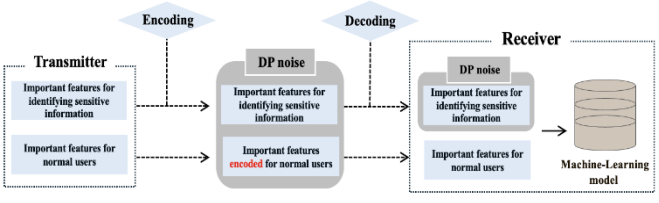
하지만, DP 는 데이터의 과도한 왜곡으로 인해 데이터의 유틸리티를 저하시키고 정상 사용자의 데이터 학습 결과가 부정확해지는 한계가 있다. 본 논문에서는 공격자와 사용자가 기계학습 모델을 활용하여 통계치를 도출하는 환경에서 인코딩 기법을 사용자의 통계치 도출에 유의미한 피쳐에만 적용함으로써 정상 사용자의 정확도 성능인 유틸리티 보장하면서도 공격자의 머신러닝 학습 정확도를 감소시켜 프라이버시를 함께 보장하는 EN-DP (Encoding-based DP) 모델을 제안한다.

2. 선행연구

Jung-Hwa Ryu [2]의 선행연구에서는 DP 의 세 가지 변수인 앱실론, 델타, 민감도를 결합 및 조정하여 프라이버시 보호와 사용자 효용성 사이의 최상의 균형을 찾아내며 입증했다. M.A.P. Chamikara [3]의 선행연구에서는 머신러닝을 활용한 얼굴 이미지를 학습할 때 생체 데이터와 연관된 개인의 민감 정보 유출을 방지하기 위해 학습 데이터에 DP를 적용하였으며 프라이버시 손실 방지를 입증했다. 그러나 두 선행연구 모두 DP 로 인한 데이터 왜곡이 사용자의 유틸리티를 감소시키는 문제는 다루지 않았다.

3. 오류 정정 부호를 활용한 고신뢰 차등 프라이버시 기법

본 장에서는 안전한 데이터 전송을 위한 데이터 인코딩 기반의 효율적인 DP 기법인 EN-DP 모델의 동작 방식을 서술한다. 그림 1 은 EN-DP 의 동작 메커니즘을 나타낸다.



(그림 1) EN-DP 의 동작 메커니즘

안전한 데이터 송수신을 위해 EN-DP 의 동작 방식은 모든 피쳐에 DP 를 적용하고, 프라이버시와 유틸리티의 트레이드오프 문제를 해결하기 위해 일부 피쳐 그룹에 인코딩 기법을 적용한다.

이 때 공격자의 타겟이 될 수 있는 민감정보를 식별하는데 중요도가 높은 피쳐에는 별도의 인코딩 절차 없이 DP 노이즈가 적용된 데이터를 송수신함으로써 공격자가 피쳐를 탈취하더라도 인공 노이즈로 인해 민감정보를 식별하기 어려워서 프라이버시를 개선할 수 있다. 반면에, 정상 사용자가 통계치를 도출하는데 유의미한 피쳐에는 사용자의 유틸리티 성능을 보장하기 위한 인코딩 값을 추가하여 송신하고, 수신자는 이를 다시 디코딩함으로써 원본 피쳐로 복구하여 정확한 통계치 도출이 가능하다.

피쳐 그룹은 사용자가 머신러닝 모델을 사용하여 통계치를 도출하는데 유의미한 피쳐 그룹과 머신러닝 모델로 민감 정보를 식별하는데 유의미한 피쳐 그룹으로 나뉘어진다. 각 사용자와 공격자가 타겟으로 하는 정보를 식별하기 위한 유의미한 피쳐는 라벨 별 피쳐의 중요도 순위를 도출하는 permutation importance 기법을 활용한다.

4. 성능 평가

4.1 성능 평가 환경

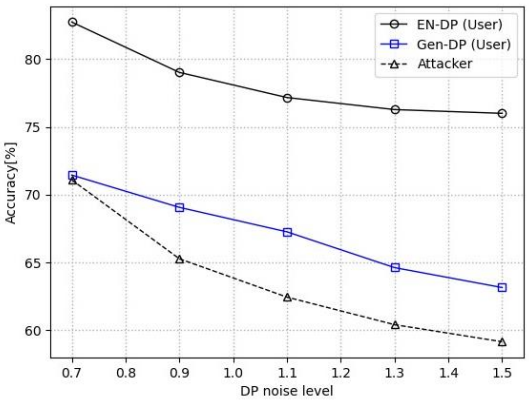
본 장에서는 제안하는 EN-DP 의 성능 평가 환경을 설명한다. 본 실험에서는 위조된 은행 계좌를 탐지하는 머신러닝 학습 데이터인 Bank Account Fraud Dataset Suite (NeurIPS 2022)을 활용하였다. 정상 사용자는 머신러닝 모델을 활용하여 위조 계좌를 탐지하고, 공격자는 머신러닝 모델로 금융 민감 정보를 유출한다고 가정했다.

정상적인 사용자와 공격자의 머신러닝 모델을 Decision Tree 방식으로 모델링하였다. 또한, 피쳐의 중요도 순위를 도출하기 위해 permutation importance 모델을 활용하였고, 인코딩은 Reedsolo 1.7.0 모듈을 사용하였다.

EN-DP 의 성능 비교 평가를 위해 인코딩을 적용하지 않고 DP 를 적용한 Con-DP (Conventional DP) [3]를 대표적인 종래의 방법으로 선정했다. EN-DP 와 Con-DP 의 정상 사용자의 정확도 성능과 민감정보를 식별하고자 하는 공격자의 정확도 성능 차인 보안 용량 (security capacity)를 비교 분석하였다. 인코딩 방식이 유한한 점을 악용하여 공격자가 브루트 포스 공격으로 인코딩 패턴을 추정 후 디코딩할 수 있는 현실적인 환경을 모델링하여 실험했다. 100 회 반복 시뮬레이션 했고, DP 노이즈 레벨에 따른 정확도를 산출하여 성능을 평가하였다.

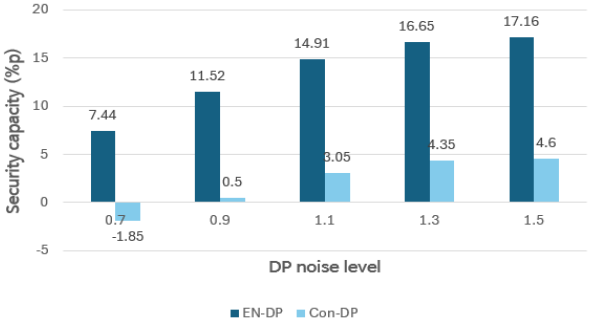
4.2 성능 평가 결과 및 분석

그림 2 는 본 논문에서 제안한 방식과 종래 방식의 위조 계좌 탐지 정확도와 공격자의 민감 정보 추정 정확도를 나타낸 그래프이다.



(그림 2) EN-DP 의 성능 평가 결과

그림 2 에 따라, DP 노이즈 레벨이 증가할수록 모든 모델의 정확도 성능이 낮아지는 결과를 보였다. 그에 반해, EN-DP 는 인코딩 기법을 통해 DP 가 적용된 데이터를 본래의 데이터로 복원이 가능하므로, DP 노이즈 레벨이 증가할수록 정확도가 떨어지더라도 Con-DP 과 공격자의 정확도 성능 대비 성능이 좋은 결과를 보였다. 반면에 공격자 모델과 종래 Con-DP 모델은 DP 노이즈 레벨이 증가할수록 데이터에 포함된 인공 노이즈로 인해 학습 정확도가 급격하게 떨어지는 경향을 보였다. 그림 3 은 EN-DP, Con-DP 의 보안 용량 비교 결과를 나타낸다.



(그림 3) EN-DP, Con-DP 의 보안 용량 결과

DP 노이즈 레벨이 높아질수록 EN-DP 와 Con-DP 의 보안 용량이 증가함을 보여준다. 특히 EN-DP 는 인코딩 기법을 통해 에러 정정함으로써 Con-DP 에 비해 보안 용량을 개선하여 프라이버시와 유틸리티의 트레이드오프 문제를 개선할 수 있음을 입증하였다.

5. 결론

종래의 DP 기법은 인공 노이즈를 민감 데이터에 섞어서 프라이버시를 개선했지만 정상 사용자의 성능과 유틸리티를 열화시키는 한계점을 극복하고자 고신뢰 데이터 전송을 위한 데이터 인코딩 기반의 DP 기법인 EN-DP 모델을 제안하였다. 실험 결과에 따르면 EN-DP 는 Con-DP 대비 최대 17.16%의 보안 용량을 개선시킬 수 있음을 입증하였다. 향후 연구에서는 공격자 노드의 인 · 디코딩 수행이 불가능한 메커니즘을 개발하여 보안 용량을 최적화하는 기법을 연구할 계획이다.

Acknowledgement

본 논문은 2024 년도 산업통상자원부 및 한국산업기술진흥원의 산업혁신인재성장지원사업 (RS-2024-00415520)과 과학기술정보통신부 및 정보통신기획평가원의 ICT 혁신인재 4.0 사업의 연구결과로 수행되었음 (No. IITP-2022-RS-2022-00156310)

참고문헌

[1] Bo_Ning, Yunhao Sun, Xiaoyu Tao, and Guanyu Li “Differential privacy protection on weighted graph in wireless networks,” Ad Hoc Networks, vol. 110, 102303, 2021.

[2] Jung-Hwa Ryu, Sun-Jin Lee, Hye-Yeon Shim, and Il-Gu Lee, "Parameter Optimization Techniques for Privacy and Utility Trade-off of Differential Privacy," The 24th World Conference on Information Security Applications, Jeju Island, South Korea, Aug. 2023, pp.3-4.

[3] M.A.P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, “Privacy Preserving Face Recognition Utilizing Differential Privacy,” Computer & Security, vol. 97, 101951, 2020.