차등 프라이버시를 활용한 IoT 네트워크 합성 데이터 생성 파이프라인

김재철[1], 박승운[1], 차재식[2], 손은영[3], 손윤식[4]
[1]동국대학교 컴퓨터·AI학과 석사과정
[2]동국대학교 컴퓨터공학과 학사과정
[3]동국대학교 영어통번역학과 학사과정
[4]동국대학교 컴퓨터·AI학과 교수
2017111790@dgu.ac.kr, cystem@dgu.ac.kr, ckwotlr@dongguk.edu, youkise@dgu.ac.kr, sonbug@dgu.ac.kr

An IoT Network Synthetic Data Generation Pipeline Based on Differential Privacy

Jaecheol Kim[1], Seungun Park[1], Jaesik Cha[2], Eunyoung Son[3], Yunsik Son[4]
[1]Dept. of Computer Science and Artificial Intelligence, Dongguk University
[2]Dept. of Computer Science and Engineering, Dongguk University
[3]Dept. of English Lunguistics, Interpreatation and Transloation, Dongguk University

요 약

사물 인터넷(IoT) 네트워크의 발전은 실시간 데이터 수집과 자동화를 가능하게 했지만, 데이터 공유 및 배포 과정에서 민감 정보가 유출될 수 있는 보안 위협을 내포한다. 본 논문은 이를 해결하기 위해 합성 데이터 생성과 Differential Privacy(DP)를 결합한 파이프라인을 제안한다. 특히, 표 형식 데이터 생성에 특화된 디퓨전 모델인 TabDiff를 사용하여 IoT 데이터를 합성하고, 여기에 유틸리티 보존에 중점을 둔 Utility Privacy(UP-DP)를 적용한다. 제안된 파이프라인은 최종적으로 생성된 데이터가 원본 데이터와 갖는 통계적 유사성 및 개인정보 보호 수준을 종합적으로 평가한다. 본 연구의 핵심 기여는 민감 정보 추론 공격을 효과적으로 방어하면서도 데이터의 유용성을 높은 수준으로 유지하는 방법론을 정립하여, 안전하고 신뢰할 수 있는 IoT 데이터 활용의 기반을 마련했다는 데 있다.

1. 서론

사물 인터넷(IoT)은 스마트 시티, 의료 등 다양한 분야에서 실시간 데이터 수집과 자동화를 통해 큰 발전을 이루었다. 하지만 이는 IoT 네트워크의 민감한 정보를 탈취하기 위한 공격자들의 표적이 되었고, 악의적인 행동을 탐지하는 침입 탐지 시스템(IDS)이 효과적인 대안으로 떠올랐다. 특히 딥러닝 기반 IDS는 IoT 네트워크의 풍부한 데이터를 사용하여 공격 탐지에 좋은 성능을 보였다. 그러나 이는 모델 학습을 위해 데이터셋 배포, 공유 과정에서 민감한 정보를 추론하려는 새로운 위협으로 떠올랐다.

본 논문은 이를 해결하고자, IoT 네트워크의 합성 데이터를 생성하고 Differential Privacy(DP)[1]를 적용해 평가하는 파이프라인을 제안한다. 디퓨전 모델을 사용해 원본 데이터와 유사한 고품질의 합성 데이터를 생성한다. 이는 희귀한 공격 데이터 증강뿐 아니라 원본 데이터의 근사를 통해 본질적 보호 효과를 제공한다. 또한 DP로 정교한 노이즈를 부가하여 유틸리티 - 프라이버시 트레이드오프를 균형 있게 유지한다.

본 연구의 핵심 기여는 IoT 네트워크에서 디퓨전 모델과 DP를 결합해, 통계적 유용성을 높게 유지하면서 민감 정보 추론 공격에 견고한 실용적 파이프라인을 제시한 데 있다.

2. 관련 연구

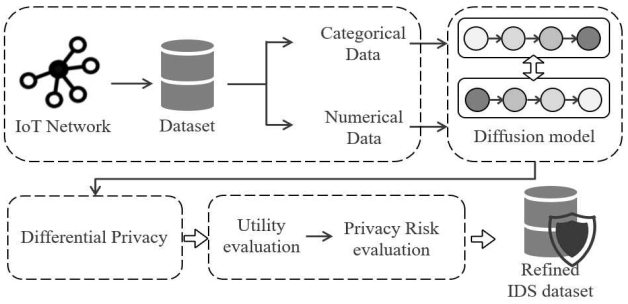
초기 합성 데이터 생성은 생성자와 판별자로 이루어진 GAN을 통해 이루어졌으나, 학습 불안정 및 모드 붕괴와 같은 한계가 있다.[2] 이를 극복하기 위해, 본 연구는 노이즈를 점진적으로 추가하고 제거하는 디퓨전 모델을 채택했다. 특히 TabDDPM[3], TabSyn[4]과 같은 표 형식 데이터 생성 디퓨전 모델은 데이터의 유형마다 서로 다른 노이즈를 추가하여 원본 데이터와 유사한 고품질의 합성 데이터를 생성한다.

DP는 정교하게 보정된 노이즈를 추가하여 개인 포함 여부를 통계적으로 식별 불가능하게 만드는 수학적 보장이다. 이는 데이터 분석을 허용하면서도

개개인의 프라이버시를 보장한다. 하지만 기존 DP 메커니즘은 데이터 유틸리티를 고려하지 않은 한계점이 존재한다. 따라서 본 연구는 데이터 유틸리티와 프라이버시 위협의 트레이드오프를 고려한 Utility Preserving DP(UP-DP)[5]를 사용하여 이를 해소하고자 한다.

3. 제안하는 파이프라인

본 연구는 디퓨전 모델로 합성 데이터를 생성하고 UP-DP를 적용해 유틸리티와 프라이버시 리스크를 평가하는 파이프라인을 제안한다. 전체 과정은 데이터 준비 - 모델 학습 - 평가의 3단계로, 준비 단계에서 수치형/범주형을 구분한 뒤 디퓨전 모델로 점진적 노이즈 주입, 제거로 합성 데이터를 만들고 UP-DP로 노이즈를 추가한다. 마지막으로 원본과의 유사성과 프라이버시 리스크를 측정해 데이터의 유용성과 프라이버시 견고성을 검증한다.



(그림 1) 제안하는 파이프라인의 전체 흐름도

3.1. 데이터 샘플링

본 연구는 MQTT-IoT-IDS2020 데이터셋[6]을 사용하여 수행된다. 해당 데이터셋은 IoT 네트워크 공격인 scanning, 브루트 포스와 같은 다양한 공격 유형과 패킷 데이터의 길이, 수 등이 포함되어 있다. 학습을 위해 공격 비율과 정상 데이터 비율 기준을 1대1로 나눈다. 이후, 각 열이 수치형 데이터인지 범주형 데이터인지 구분하여 데이터 유형에 맞게 디퓨전 모델이 처리할 수 있도록 한다. 표1은 데이터셋의 유형을 나타낸다.

<표 1> MQTT-IoT-IDS2020 데이터셋의 공격 유형

Class	Number
Scan_A	46,473
MQTT_bruteforce	31,311
Sparta	118,374
Scan_sU	42,520
Normal	119,339
Total	238,678

3.2. 표 기반 합성 데이터 생성

표 형식 데이터는 수치에 의미가 없는 범주형과 의미가 있는 수치형 데이터가 혼합된 이질적인 구조로, 피처 간 복잡한 종속성을 포함한다. 이를 다루기 위해 표 형 데이터의 복잡한 분포를 안정적으로 학습할 수 있는 TabDiff[7]가 제안되었다. TabDiff는 범주형에는 점진적인 마스킹을 적용하여 마스킹을 예측, 복원하는 과정을 반복해 분포를 모델링한다. 수치형은 단계적인 Gaussian 노이즈 주입 및 제거를 통해 분포를 학습한다. 특히 TabDiff는 노이즈 스케줄을 고정하지 않고, 훈련 단계에서 피처별로 최적의 스케줄을 학습한다. 이는 각 피처 분포의 특성에 맞춰 노이즈를 추가하여 고품질 데이터 생성이 가능하다.

3.3. UP-DP 적용

생성한 합성 데이터는 실질적 보호를 위해 DP를 적용한다. DP는 노이즈가 강할수록 프라이버시 보호 효과는 상승하지만, 유틸리티는 감소한다. 반대로 노이즈가 약할수록 프라이버시 보호 효과는 감소하지만, 유틸리티는 상승한다. 따라서 DP 적용 시, 유틸리티와 프라이버시 간 트레이드오프가 중요하다.

UP-DP는 이를 고려해 설계된 메커니즘이다. 표형 데이터에 맞게 범주형에는 Laplace 메커니즘을, 수치형에는 Gaussian 메커니즘을 적용한다. 이후, 노이즈 주입으로 발생하는 분포의 왜곡을 분위수 정합으로 구조적 복원을 수행하고, 분포의 자연스러움을 위해 Jitter를 추가한다. 최종적으로 분포의 정교함을 위해 동적 Kolmgorov-Smirnov(KS) 조정을 계산해, 구간화된 분포를 KS 통계량 기반으로 분포 차이를 최소화한다. 또한, 데이터의 유효성 유지를 위해 조정된 값이 해당 구간 경계 내에 있도록 Clipping을 수행해 DP를 적용한다.

3.4. 데이터 유틸리티 평가

본 연구는 합성 데이터에 DP 적용 시, 유사성 평가를 위해 Wasserstein distance, Jensen-Shanon Divergence (JSD), KS Test로 통계적 유사성을 측정한다. 하지만 이러한 지표는 표 형 데이터를 반영하지 못하는 단점이 존재하기에, Synthetic Data Vault(SDV) Fidelity[8]를 사용한다. 이는 표 형 데이터가 실제 데이터의 통계적 속성과 패턴을 얼마나 충실히 재현했는지 확인하기 위해 수치형에는 KS, 범주형은 Total Variance Distance를 통해 계산한다.

3.5. 프라이버시 리스크 평가

UP-DP가 적용된 데이터의 프라이버시 리스크 평가를 위해 속성 추론 공격과 Membership Inference Attack(MIA)을 사용하였다.

속성 추론 공격은 알려진 속성을 사용해 정보를 추론하는 공격으로, k-Neareset Neighbor (kNN)을 사용하여 예측 정확도를 계산한다.

MIA는 특정 데이터가 모델 훈련에 사용되었는지 추론하는 공격이다. 훈련에 사용된 멤버 데이터는 높은 예측 확률이나 낮은 손실값을 보이지만, 훈련에 사용되지 않은 비 멤버 데이터는 상대적으로 낮은 예측 확률과 높은 손실값을 보인다. 이러한 차이를 공격 모델에 학습시키고, 해당 데이터를 질의해 멤버십 여부를 추론한다. 따라서 MIA의 성공률이 높을수록 멤버 데이터의 유출 위험이 커진다.

4. 실험 결과 및 분석

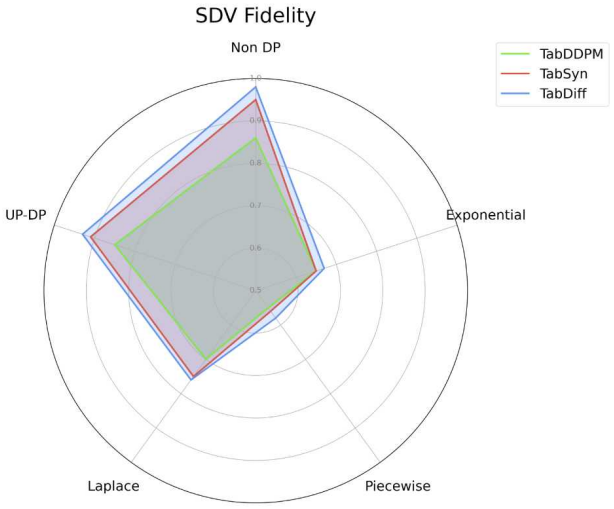
제안하는 파이프라인의 성능 평가를 위해, 본 연구에서는 TabDDPM, TabSyn, TabDiff를 사용하여 비교한다.

표 2는 각 모델에 Laplace, Exponentialm Piecewise 그리고 UP-DP를 적용하고, 통계적 지표를 사용하여 나타낸 결과이다. Wasserstein과 JSD는 값이 낮을수록 좋고 KS Test는 높을수록 우수하다. 모든 모델에서 Laplace와 비교 시, UP-DP가 더 우수한 유틸리티 보존을 보였고, TabDiff의 경우 가장 우수함을 확인하였다.

<표 2> 각 모델의 통계적 지표

Diffusion model	DP	Utility Metrics		
		Wasserstein	JSD	KS Test
TabDDPM	Without DP	0.04	0.15	0.95
	Laplace	0.08	0.24	0.72
	Exponential	0.09	0.34	0.38
	Piecewise	0.59	0.60	0.38
	UP-DP	0.08	0.28	0.80
TabSyn	Without DP	0.04	0.15	0.95
	Laplace	0.17	0.51	0.68
	Exponential	0.13	0.35	0.39
	Piecewise	0.61	0.61	0.39
	UP-DP	0.02	0.11	0.83
TabDiff	Without DP	0.01	0.05	0.99
	Laplace	0.01	0.18	0.73
	Exponential	0.10	0.36	0.44
	Piecewise	0.61	0.63	0.39
	UP-DP	0.01	0.07	0.98

그림 2의 SDV Fidelity에서는 각 DP 메커니즘을 적용했을 때, 모두 비슷한 양상을 보였으며, UP-DP가 가장 좋은 성능을 보였다. 특히, TabDiff와 UP-DP의 조합이 가장 높은 수치를 기록하며, 실제 데이터와 가장 유사하였다.

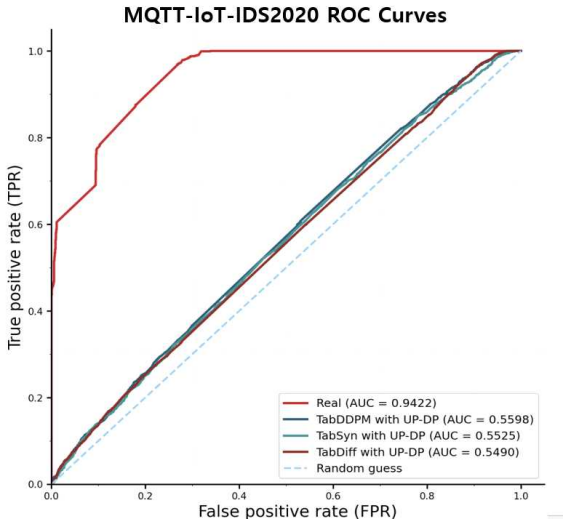


(그림 2) 모델별 DP 메커니즘 적용 시 SDV Fielity 수치

속성 추론 공격은 실제 데이터에서 모두 0.98로 심각한 누출을 보였다. UP-DP 적용 시, 모든 모델에서 유의미한 성능 개선을 보였고, 특히 TabDiff와 UP-DP 조합은 0.33으로 가장 좋은 성능을 보였다.

<표 3> 속성 추론 공격 평가

DP Mechanism	Diffusion model		
	TabDiff	TabSyn	TabDDPM
Without DP	0.98	0.98	0.98
UP-DP	0.33	0.34	0.40



(그림 3) 각 모델 별 멤버십 추론 공격의 ROC 커브

MIA는 AUC-ROC 커브를 통해 평가된다. 1.0에 가까울수록 멤버와 비 멤버를 잘 구별하며, 실제 데

이터는 0.95로 개인 정보 유출 위험이 매우 크다. 반면 합성 데이터에 UP-DP 적용 시, 모두 0.5에 근접한 수치를 보였다. 이는 멤버와 비 멤버를 구별하는 능력이 무작위 추측에 가까운 수치로, UP-DP가 효과적으로 데이터를 방어하고 있음을 의미한다. 특히 TabDiff와 UP-DP의 조합은 0.51로 가장 우수한 성능을 보였다.

5. 결론

본 논문은 IoT 네트워크라는 도메인에서 안전한 데이터 활용을 위해 디퓨전 모델과 UP-DP를 적용하고, 유틸리티와 프라이버시 리스크를 평가하는 파이프라인을 제안하였다. 제안한 파이프라인은 표형 데이터를 생성하는 디퓨전 모델인 TabDiff를 사용하여 합성 데이터를 생성하고, UP-DP를 적용해 유틸리티를 보존하며 프라이버시 리스크에 강건한 데이터를 생성하였다. 유틸리티는 SDV Fidelity와 통계적 지표, 프라이버시 리스크는 속성 추론 공격과 MIA로 평가하였다. 특히 TabDiff와 UP-DP의 조합은 유틸리티 지표에서 기존 모델 대비 우수성을 입증했을 뿐만 아니라 ROC 커브 값을 0.5 수준으로 현저히 낮추며 강력한 프라이버시 보호 효과를 확인하였다. 본 연구의 가장 큰 의의는 확산 모델을 UP-DP와 통합하여 IoT 네트워크 도메인에 적용하고, 개인 정보 보호 측면에서의 실질적인 성능을 입증했다는 점이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임 (과제번호 RS-2025-02412990). 이 성과는 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임(2710086167). 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2025-2020-0-01789)

참고문헌

[1] Dwork, C. Differential Privacy. In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP 2006), Part II., Venice, Italy, 10 – 14 July 2006; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg; Volume 4052, pp. 1 – 12, 2006

[2] Mescheder, L.; Geiger, A.; Nowozin, S. Which Training Methods for GANs do Actually Converge? International conference on machine learning. PMLR, Stockholm, pp. 3481 – 3490, 2018.

[3] Kotelnikov, A.; Baranchuk, D.; Rubachev, I.; Babenko, A. TabDDPM: Modelling Tabular Data with Diffusion Models. In Pro-ceedings of the 40th International Conference on Machine Learning (ICML 2023), pp. 17564-17579 ; PMLR: Honolulu, HI, USA, 2023.

[4] Zhang, H.; Zhang, J.; Srinivasan, B. Mixed-Type Tabular Data Synthesis with Score-based Diffusion in Latent Space. In Pro-ceedings of the International Conference on Learning Representations (ICLR 2024), pp. 1 – 21; 2024.

[5] Alabdulwahab, S.; Kim, Y.-T.; Son, Y. Privacy-Preserving Synthetic Data Generation Method for IoT-Sensor Network IDS Using CTGAN. Sensors 2024, 24, 7389, 2024.

[6] Hindy, H.; Tachtatzis, C.; Atkinson, R.; Belle, E.; Bures, M.; Tair, A.; Shojafar, M.; Loukas, G. Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset). In Proceedings of the International Networking Conference (INC 2020); pp. 297 – 308; 2020.

[7] Shi, J.; Xu, M.; Hua, H.; Zhang, H.; Ermon, S.; Leskovec, J. TabDiff: A Mixed-type Diffusion Model for Tabular Data Generation. In Proceedings of the International Conference on Learning Representations (ICLR 2025), 2025.

[8] DataCebo, Inc. SDMetrics: A Synthetic Data Evaluation Library. Available online: https://docs.sdv.dev/sdmetrics/ (accessed on 22 August 2025).