Privacy Preservation in Data Lifecycle

Devna Ramesh

***Abstract*—In the era of increasing data collection through mobile devices and data centers, the preservation of privacy throughout the data lifecycle has become a paramount concern. This research paper delves into the multifaceted aspects of privacy preservation in the four distinct phases of the data lifecycle: Data Collection, Data Publication, Data Analysis, and Data Consumption. Drawing insights from recent work presented by Zhikun Zhang, a research group leader at CISPA, the paper navigates through innovative approaches and methodologies aimed at safeguarding user privacy.**

The focal points of discussion include "PrivSyn: Differentially Private Data Synthesis" for preserving privacy in the synthesis of tabular data, as well as the vulnerabilities associated with Graph Neural Networks and the intricacies of privacy risk assessment in machine learning models. The paper also explores the concept of "machine unlearning" in the context of the "Right to be Forgotten," shedding light on the challenges and potential research areas in efficiently removing user data from machine learning models.

I. Introduction

THE relentless growth of data, facilitated by mobile devices and data centers, raises significant privacy concerns, particularly considering the sensitivity of the information collected. The four distinct phases of the data lifecycle—Data Collection, Data Publication, Data Analysis, and Data Consumption—demand meticulous attention to privacy preservation. Zhikun Zhang's recent research, as presented in his talk, provides a comprehensive framework for understanding and addressing privacy challenges across these stages. In the Data Collection phase, the focus lies on local differential privacy, a mechanism that introduces perturbations to raw data before uploading it to the data center. This perturbation ensures that statistical information remains intact while shielding individual records' privacy. Moving to the Data Publication phase, the discussion centers on differentially private synthetic datasets, leveraging differential privacy methods to extract information from sensitive data and synthesize new datasets with robust privacy features. Data Analysis, a critical phase where datasets inform machine learning models, introduces the concept of risk assessment. Specific attacks targeting various model architectures, such as membership inference, property inference, and model inversion, highlight the need for robust privacy measures. In the Data Consumption phase, the exploration of "machine unlearning" intersects with the "Right to be Forgotten," presenting challenges in efficiently removing user data from machine learning models. As data regulation entities establish privacy laws, including Europe's GDPR and California's CCPA, compliance becomes integral to data center design. The subsequent sections of this research paper will delve deeper into each phase, examining current research, methodologies, and potential avenues for further exploration in privacy preservation.

II. Data Publication: Generation of Tabular Data

Naïve Method:- In the generation of synthetic tabular data, the naïve method involves maintaining differential privacy by calculating the joint distribution of all attributes and subsequently introducing noise. While effective for small datasets with few attributes, this poses challenges for high-dimensional datasets due to significant memory constraints.

Graphical Method:- The graphical method utilizes the original dataset to construct a graphical model ensuring differential privacy. Each node in the model represents an attribute, and edges signify high correlations between nodes. Once the graphical model is established, the synthetic dataset can be created by sampling the edge one by one from the conditional distribution. The two existing approaches, Bayesian network and Markov random field, suffer from limitations such as sparse representation, loss of crucial correlation information, and high storage costs.

PrivSyn:- PrivSyn presents an improved approach by extracting marginal tables from the data instead of constructing the entire graph. Marginals of low degree, such as two-way marginals, are used to represent the dataset, encapsulating correlation with a weak assumption of conditional independence. This is discussed in detail in the paper "PrivSyn: Differentially Private Data Synthesis" [1]
There are two main challenges with this approach - marginal selection capturing high correlation while avoiding excessive noise, and synthetic dataset creation considering noise and the separate nature of marginals.
Marginal Selection: When it comes to marginal selection, there are two kinds of error involved:-

- Noise Error – Error incurred by Differential privacy when adding marginals
- Dependency Error – Error incurred as a result of information lost while leaving out marginals

An optimization problem is proposed that minimizes noise and dependency errors as shown in Fig. 1, with a proposed greedy algorithm for selection called Marginal Selection Algorithm. Another approach involves combining small two-way marginals into larger margins. This is called Marginal Combine Algorithm.

Dataset Generation: Gradually Update Method (GUM): Given the randomly initiated dataset ds , for each noisy marginal, ds is updated to make it consistent with the marginal. If we change ds to make it completely consistent with the current marginal in each step it increases the error for other marginals even though it achieves low error of the target marginal. This is the approach for Min-Cost Flow (MCF). To avoid this situation, the idea of multiplicative update is

Optimization Problem Formulation:

$$\text{minimize} \sum_{i=1}^{m} [\psi_i x_i + \phi_i (1 - x_i)]$$

$$\text{subject to } x_i \in \{0, 1\}$$

Fig. 1. Optimization Problem Formulation

adopted in the GUM method. GUM updates a randomly initiated dataset for each noisy marginal, maintaining consistency without fully adhering to the marginal in each step. This multiplicative update strategy prevents excessive errors for other marginals. Fig. 2 shows the experimental setup done to test the utility of synthetic dataset generated with the above method. The perfomance is compared amongst three competitors -PrivBayes (Bayesian Network), PGM (Markov Random Field), PrivSyn. PrivSyn significantly outperforms competitors in preserving high-dimensional correlation . PrivSyn demonstrates superior performance compared to PrivBayes and PGM. Evaluation Metrics: The utility of synthetic datasets is evaluated using three metrics:

1) Pair-wise Marginals (Average L1 Error)
2) Range Query (Average L1 Error)
3) Classification (Misclassification Rate)

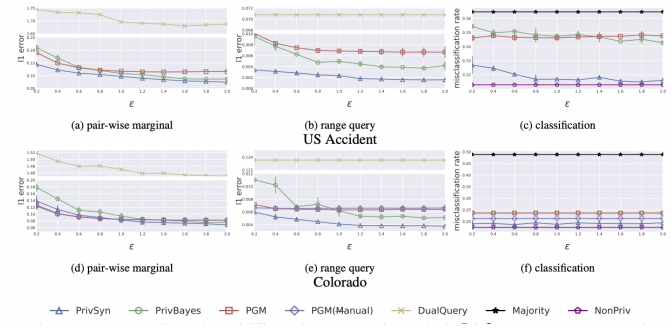

Fig. 2. Experimental Setup- Synthetic Dataset Generation

While PrivBayes and PGM effectively capture low-dimensional correlation, PrivSyn excels in preserving high-dimensional correlation, particularly in scenarios involving range queries and classification. The proposed approach addresses challenges in marginal selection and synthetic dataset creation, contributing to advancements in differentially private data synthesis for tabular data. Further exploration into the practical implications of these methods and their scalability is warranted.

III. Data Publication: Privacy Preservation in Graph Data

The distinction between node classification and graph classification was explored, emphasizing the determination of labels for individual nodes or the entire graph, respectively. An illustrative example was provided, highlighting applications such as predicting the toxicity of molecules.

Node Classification: ML models like MLP, utilizing node features, were discussed as the naïve method for node classification. The introduction of Graph Neural Networks (GNNs) improved accuracy by considering graph correlation. GNNs aggregate information from each node's neighbors through aggregators like GCN, GIN, and GAT, producing node embeddings for downstream tasks.

Graph Classification: For graph classification, node embeddings must be converted to graph embeddings. Common methods used are mean pooling and max pooling. Hierarchical pooling is an improvised technique that improves the accuracy of the embedding conversion by summarizing the node to graph embedding hierarchically. This captures the structural information in an effective manner compared to min/max pooling. Mean cut is one technique to achieve this by partitioning the graph in the most efficient way possible. The graph paradigm of target graph embedding used for downstream tasks is considered to preserve the privacy of the target graph. The paper 'Inference Attacks Against Graph Neural Networks' [2] discusses how the graph embeddings could be a window to attacking the target graph despite its abstract nature.

Inference Attacks Against Graph Neural Networks

The paper on "Inference Attacks Against Graph Neural Networks" delves into potential vulnerabilities in the use of graph embeddings. Three inference attacks were discussed:

1. **Property Inference Attack:**

- Objective: Infer basic properties of the target graph (e.g., density, number of nodes, edges) using graph embeddings.
- Methodology: Training multi-output ML models on an auxiliary dataset with varied distributions from the target graph.
- Results: Demonstrated superior performance compared to baseline, with sensitivity to bucketization methods. From Fig. 3, it is clear that in general, Meanpool has a lower accuracy compared to DiffPool and MincutPool.

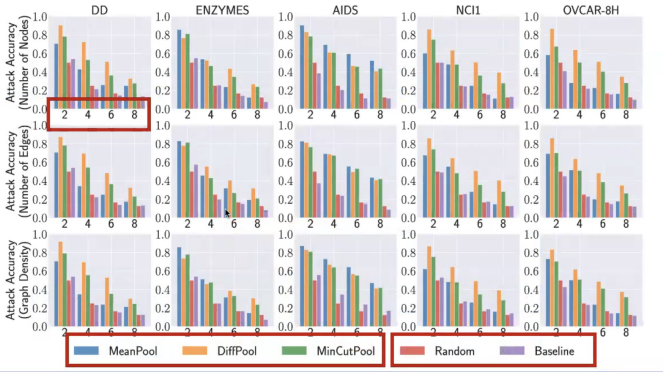

Fig. 3. Property Inference Attack

2. **Subgraph Inference Attack:**

- Objective: Infer whether a subgraph is part of the original graph using the subgraph and target graph embeddings.
- Methodology: Transforming subgraph to embedding, aggregating with target graph embedding, and then fed to an MLP or Deep Learning Model. The embedding extractor and the aggregator are trained simultaneously, and is called as the attack model.
- Results: Experimentally validated with consideration of positive and negative pairs as shown in Fig. 4.

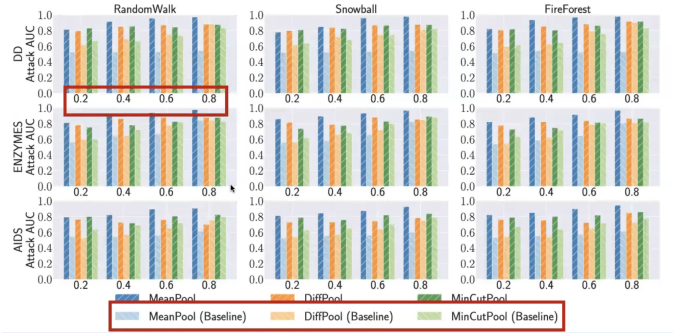

Fig. 4. Subgraph Inference Attack

3. **Graph Reconstruction Attack:**

- Objective: Reconstruct a graph with similar structural statistics as the target graph using target graph embeddings.
- Methodology: Training a graph autoencoder which is similar to an image encoder with an additional graph matching component. Once this is trained, the decoder is used as an attack model- which has the target graph embedding as input and the graph as output
- Results: Experimental results showcased reconstruction using the attack model. Fig. 5 shows that the proposed attack achieves good reconstruction performance regarding graph isomorphism and macro-level graph statistics.

Graph isomorphism

Dataset	DiffPool	MeanPool	MinCutPool
AIDS	0.875 ± 0.003	0.794 ± 0.003	0.869 ± 0.002
ENZYMES	0.670 ± 0.019	0.653 ± 0.022	0.704 ± 0.012
NCI1	0.752 ± 0.005	0.771 ± 0.010	0.693 ± 0.007

Macro-level graph statistics (cosine similarity)

Dataset	Target Model	Degree Dist.	LCC Dist.	BC Dist.	CC Dist.
AIDS	MeanPool	0.651 ± 0.001	0.999 ± 0.001	0.987 ± 0.001	0.876 ± 0.002
	DiffPool	0.894 ± 0.001	0.999 ± 0.001	0.983 ± 0.001	0.787 ± 0.002
	MinCutPool	0.888 ± 0.003	0.999 ± 0.001	0.983 ± 0.001	0.785 ± 0.006
ENZYMES	MeanPool	0.450 ± 0.070	0.646 ± 0.005	0.959 ± 0.001	0.516 ± 0.037
	DiffPool	0.519 ± 0.007	0.661 ± 0.008	0.958 ± 0.001	0.504 ± 0.005
	MinCutPool	0.467 ± 0.019	0.490 ± 0.009	0.916 ± 0.001	0.414 ± 0.009
NCI1	MeanPool	0.736 ± 0.003	0.999 ± 0.001	0.877 ± 0.001	0.402 ± 0.001
	DiffPool	0.633 ± 0.002	0.999 ± 0.001	0.877 ± 0.001	0.495 ± 0.002
	MinCutPool	0.570 ± 0.002	0.999 ± 0.001	0.877 ± 0.001	0.496 ± 0.001

Fig. 5. Graphical Reconstruction Attack

Defense Mechanism:- Random Noise Addition: A defense strategy was proposed involving the addition of random noise to graph embeddings. Experimental results demonstrated the effectiveness of this defense mechanism in mitigating inference attacks while maintaining acceptable accuracy for normal graph classification. The results can be seen in Fig. 6

The exploration of privacy preservation in the data publication phase, particularly in graph data, highlighted the

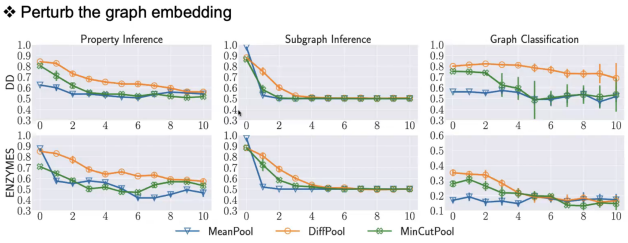

Fig. 6. Defence Mechanism to Inference Attacks Against GNNs

vulnerabilities associated with graph embeddings. The paper's findings on inference attacks underscore the need for robust defense mechanisms. The proposed addition of random noise emerges as a promising strategy to enhance the security of graph data during the publication phase. Further investigations into advanced defense mechanisms and their implications on real-world applications are warranted.

IV. Privacy Risk Assessment of Machine Learning Algorithms

The right to be forgotten, a fundamental privacy concept, mandates the removal of an individual's data from any data center. The application of this right to the training set of a machine learning (ML) model is known as machine unlearning. The paper, "Holistic Evaluation of Differentially Private Machine Learning"[3] explores the unintended privacy risks associated with machine unlearning and introduces a novel membership inference attack. Contrary to its intended purpose of enhancing privacy, machine unlearning may inadvertently leak information. The study investigates the potential privacy risks, proposing effective mitigations.

Unintended Privacy Risks:
Machine unlearning, designed to protect data owner privacy, is scrutinized for unintended information leakage. A novel membership inference attack leverages differences between the outputs of the original and unlearned ML models to infer whether a target sample was part of the original model's training set. The study reveals that, in multiple cases, this attack outperforms classical membership inference on the original ML model, especially for well-generalized models.

Attack Methodology:
The attack pipeline involves passing the target sample through both the original and unlearned models. The aggregated posterior from both models is then used by the attack model to predict membership status. The shadow model paradigm, employing a shadow dataset, is used for feature construction, with positive and negative datasets created using the original and unlearned ML models, as depicted in Fig. 7 Experimental Setup: The experimental results demonstrate the effectiveness of the proposed membership inference attack. Overfitting impact is explored, revealing that difference-based methods perform well on well-generalized models, while concatenation methods excel on overfitted models. Sorting the posteriors further enhances attack performance. Transferability Attack:

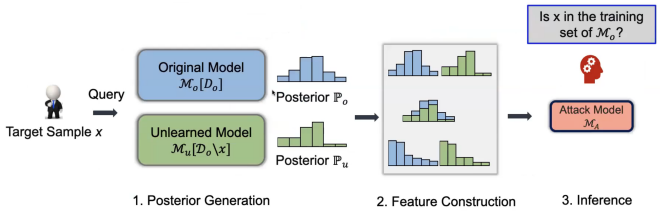

Fig. 7. Attack pipeline - ML Membership Inference Attck

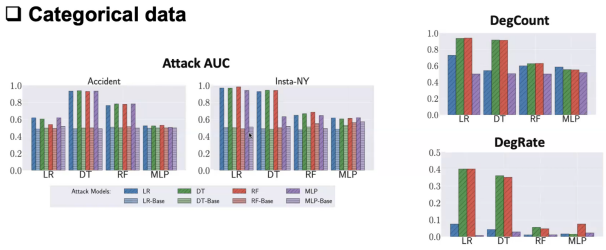

Fig. 8. Experimental Results- ML Membership Attack

Transferability attack, exploring model and dataset transferability, demonstrates that strong performance can be achieved even with a different model structure and dataset.

Defense Mechanisms: Several defense mechanisms are investigated(Fig. 9):

- Naïve Method: Using top-k instead of the full posterior proves less efficient, and it is also observed that temperature scaling is effective only for neural network models.
- Differential Privacy (DP): While this seems to be effective against the attack, DP comes at the cost of model accuracy loss.

$\mathcal{D}_o$ ($\mathcal{M}_T$)	$\mathcal{M}_A$	ND	Top-1	Top-2	Top-3	Label	TS	DP[ϵ_1]	DP[ϵ_2]
Insta-NY (LR)	RF	0.976	0.947	0.965	0.965	0.546	0.635	0.519	0.477
	DT	0.972	0.946	0.961	0.961	0.546	0.654	0.524	0.500
	LR	0.969	0.948	0.960	0.962	0.546	0.610	0.519	0.500
	MLP	0.970	0.948	0.960	0.966	0.453	0.653	0.506	0.504

Fig. 9. Defence Mechanism Exploration

This study is the first to explore unintended privacy risks caused by machine unlearning. The proposed membership inference attack reveals vulnerabilities that counteract the intended privacy protection. The findings underscore the importance of refining machine unlearning processes and implementing effective defense mechanisms, such as releasing only predicted labels, temperature scaling, and differential privacy, to ensure robust privacy protection in practical ML implementations.

V. Conclusion

In this comprehensive exploration of privacy preservation within the data lifecycle, we have delved into crucial aspects of privacy preservation in data collection, publication, analysis, and consumption. Drawing insights from the research presented by Zhikun Zhang, a leading expert in the field, we have scrutinized recent advancements in privacy-preserving techniques, spanning from tabular data synthesis to graph neural networks and machine unlearning. Considering the entire data lifecycle, it becomes evident that safeguarding privacy is a multifaceted challenge. From the inception of data collection to its ultimate consumption in machine learning models, each phase presents unique privacy concerns. The intersection of differential privacy, synthetic data generation, and defense mechanisms demonstrates the evolving landscape of privacy preservation. As data regulation entities continue to establish privacy-related laws, the need for innovative solutions to comply with these regulations becomes paramount. Future research should focus on refining existing methodologies, exploring practical implications, and addressing scalability concerns. Furthermore, investigations into the real-world applications of these privacy-preserving techniques are crucial to bridge the gap between theoretical advancements and practical implementations. This research paper has been crafted with the intention of shedding light on the intricacies involved in privacy preservation throughout the data lifecycle. The insights gained from recent research work, as presented by Zhikun Zhang, pave the way for advancing privacy-preserving techniques and ensuring data security in an era of escalating data collection and utilization. By addressing challenges at each phase of the data lifecycle, we can work towards a more secure and privacy-aware data ecosystem.

References

[1] Zhikun Zhang, Tianhao Wang, Ninghui Li, Jean Honorio, Michael Backes, Shibo He, Jiming Chen, and Yang Zhang. Privsyn: Differentially private data synthesis. In 30th USENIX Security Symposium (USENIX Security 21), 2021.

[2] Zhikun Zhang, Min Chen, Michael Backes, Yun Shen, and Yang Zhang. 2022. Inference attacks against graph neural networks. In Proceedings of the 31st USENIX Security Symposium (USENIX Security'22). USENIX Association.

[3] DPMLBench: Holistic Evaluation of Differentially Private Machine Learning C Wei, M Zhao, Z Zhang, M Chen, W Meng, B Liu, Yuan Fan, Wenzhi Chen - arXiv preprint arXiv:2305.05900, 2023

VI. Other Work

Translational AI in Medicine - Dr. Fei Wang

In the talk on 'Translational AI in Medicine,' Dr. Fei Wang provided a comprehensive examination of the essential facets contributing to a robust model within the clinical domain. He talked about six pillars including accuracy, actionability, transportability, transparency, privacy, and actionability. Dr. Wang explained that achieving an optimal model requires a balance among these six interrelated concepts, emphasizing their non-exclusivity.

The presentation featured insightful examples to explain the above mentioned aspects, such as a machine learning model designed to predict the necessity of a PTHrP test for detecting

tumor presence and another model predicting SARS-Cov 2 infection based on blood test results. The inspiration behind the development of these models to be very thoughtful — employing machine learning to enhance clinical medicine's efficiency and improve patient experiences.

The other two notable concepts that stood out to me in the talk were the imperative of privacy preservation through federated learning and the implementation of a decentralized swarm network for updating model parameters without disclosing private patient information. These aspects not only underscored the attractiveness of Dr. Wang's research but also emphasized their critical importance in the evolving landscape of medical AI.

Security Foundations for Cloud-based IoT Systems - Luyi Xing

In his presentation on "Security Foundations for Cloud-based IoT Systems," Luyi Xing delves into recent research on vulnerabilities within IoT cloud-based infrastructure. Through an advanced exploration of the state of the art, Dr. Xing systematically analyzes and addresses security challenges, aiming to prevent potential attacks. Dr. Xing's research spans critical areas, including the IoT supply chain, security models and policies, and emerging IoT paradigms. Within each domain, he investigates novel attacks, fundamental design issues, and develops formal methods for systematic problem detection. A particular focus of his inquiry lies in the MQTT protocol, widely used in App-to-Device communication. Recognized for its lightweight nature, adaptability to resource-restrained devices, and functionality in unreliable network conditions, MQTT serves as a cornerstone in Dr. Xing's examination. Dr. Xing identifies and discusses three significant attacks on the MQTT protocol: the last will attack, client ID and session hijack, and subscribing to MQTT wildcard hashtags. Recognizing the vulnerabilities inherent in MQTT, Dr. Xing proposes a comprehensive approach that combines robust security practices with innovative tools, notably introducing the P-Verifier. Positioned as a substantial advancement, P-Verifier emerges as a pivotal tool for fortifying IoT systems against sophisticated threats.