Differential Privacy is in Medical Images, Online Advertising, and Mobile Devices

Ick Namgung
University of Texas at San Antonio
San Antonio, Texas
ick.namgung@my.utsa.edu

***Abstract*—Data plays a central role in today's technology landscape, especially in emerging areas such as deep learning. However, this reliance on data raises concerns, particularly with respect to personal privacy. To address this, differential privacy has emerged as a safeguard against potential violations. This report explores the application of differential privacy in three different areas: generating synthetic medical images while preserving privacy, implementing measures in online advertising systems to protect user privacy, and using local differential privacy techniques such as HARMONY to ensure accurate analysis while preserving user privacy.**

***Index Terms*—differential privacy, synthetic medical images, local differential privacy**

I. INTRODUCTION

A traditional way to protect against data breaches is through anonymization or de-identification techniques. Anonymization involves removing personally identifiable information from data sets so that individuals cannot be re-identified. De-identification methods typically involve removing or modifying direct identifiers (such as names, SSNs, or addresses) and sometimes also suppressing or generalizing indirect identifiers (such as birthdates or zip codes) to minimize the risk of re-identification.

However, traditional anonymization methods have limitations, as they may not fully protect against re-identification, especially when dealing with high-dimensional data such as medical images. In addition, recent advances in data analysis techniques, particularly in the field of machine learning, have shown that even anonymized data can sometimes be re-identified through inference attacks or by combining multiple data sets, as happened to Netflix in 2006.

As a result, a newer approach, differential privacy, has emerged to provide stronger guarantees gurantees while still allowing useful analysis of the data. It provides a quantitative measure of how much the output of a computation would change if any single individual's data were added or removed from the dataset. The core idea of differential privacy is to add calibrated noise to data in such a way that preserves the privacy of individual data points is preserved, while still allowing accurate aggregate analysis. This noise ensures that the presence or absence of an individual's data does not significantly affect the output of the computation, thereby preventing attackers from inferring sensitive information about specific individuals. Differential privacy provides a strong privacy guarantee, even in the face of powerful adversaries who have access to auxiliary information or background knowledge. It has become increasingly important in data analysis, particularly in domains such as healthcare, advertising, and social science, where sensitive data must be analyzed while respecting individuals' privacy rights of individuals. Overall, differential privacy provides a principled approach to balancing privacy and utility in data analysis, allowing for meaningful insights to be derived from data sets while protecting individuals' sensitive information.

This report focuses on examining the applications of differential privacy in three domains: medical images, online advertising, and mobile device data collection. Preserving the confidentiality of medical information is critical to protecting individual rights, building trust in the healthcare system, and preventing the misuse of sensitive personal information. To address the handling of sensitive medical datasets, the report first investigates a differentially private generative model for synthesizing medical data without direct access to real datasets, focusing on privacy-preserving techniques. Faisal et al. aim to preserve privacy while achieving utility by training a generative model on pneumonia detection datasets and synthesizing X-ray images[1]. Second, online advertisers can build user trust, comply with regulations, and mitigate the risk of data breaches by prioritizing user privacy. Lindell et al. present a solution for preserving user privacy in online advertising campaigns, including introducing a minimum lower bound on the estimated reach of campaigns and using only public information for advertising [2]. Finally, mobile devices are highly personal and are often carried with users throughout their daily lives. The data collected from these devices is considered particularly sensitive and privacy protection is essential to avoid from raising significant privacy concerns among users. Nguyên et al. present an advanced data analysis tool that adheres to Local Differential Privacy (LDP), principles, HARMONY. HARMONY supports various data analysis tasks with improved accuracy and strong privacy guarantees, exemplified by its application in improving diagnostic information reporting in Samsung smartphones [3].

II. APPLICATIONS OF DIFFERENTIAL PRIVACY

A. Generating Privacy-Preserving Synthetic Medical Data

Privacy concerns associated with medical datasets, such as MRIs and X-rays underscore the reluctance of medical institutions to share sensitive data. This is where synthetic medical images come. Synthetic medical images are important

for privacy because they allow for the development and testing of medical imaging algorithms without the need to use real patient data, thereby protecting the confidentiality and security of sensitive medical information. Faisal et al. aim to address privacy issues by proposing a differentially private generative model for synthesizing medical data without direct access to real data sets. The approach involves training a generative model on pneumonia detection datasets, synthesizing X-ray images, and using them to train a predictive model. Faisal et al. also introduce a differentially private Generative Adversarial Network (GAN) architecture for generating synthetic X-ray images that supports both centralized and distributed radiology data generation processes. The primary focus is on adding noise exclusively to the generator while keeping the discriminator intact to ensure high quality image generation. The generator acts as a differentially private black-box model, ensuring user data privacy even when the discriminator is exposed. In such cases, each client must store and use its discriminator locally. The approach aims to prevent third parties from reconstructing source data by adding noise to learned weights. Faisal et al. claim that this approach preserves privacy. Faisal et al. also highlight its contribution of achieving 76% accuracy in generating private radiographic images, preserving utility through selective gradient sanitization, and eliminating the need for explicit clipping parameters using Wasserstein loss in the Wasserstein GAN (W-GAN).

Faisal et al. define some key concepts: a Generative Adversarial Network (GAN), Renyi Differential Privacy (RDP), and Gaussian noise. GAN is a deep learning approach for generative tasks, involving a generator and discriminator in an adversarial game to produce realistic data. The objective of GAN is to deceive the discriminator using iterative improvement of the generator. Renyi Differential Privacy (RDP) is introduced as a solution to the privacy budget restriction in DP because the use of Renyi divergence when measuring privacy loss provides more flexibility and reduces computation costs in comparison to traditional DP. Gaussian noise is proposed as a suitable choice to make the generator differentially private due to its additive property. The use of Gaussian noise in training iterations helps to maintain the privacy of in the generator. During the training epochs of old works, the noise upper bound increased exponentially, and such a loose upper bound led to higher privacy costs. However, the experiment in [1] uses the the earth mover distance between a real and a fake distribution to tighten the upper bound to exploit the Gaussian mechanism. In this way, 95% of the data stays within two standard deviations of the distribution, and this property would reduce the exponential parameter growth problem under iterations.

Moreover, it removes the overestimating privacy loss problem as well as a budget problem as Renyi differential privacy supports the composition of different mechanisms. Also, the proposed method focuses on privacy-preserving synthetic data generation, utilizing Wasserstein GAN (W-GAN) with Renyi differential privacy. In contrast to previous approaches lacking privacy guarantees, this method ensures higher utility. Unlike methods using DP-SGD with gradient clipping for both discriminator and generator, the approach exploits the gradient in the generator for privacy preservation. The choice of W-GAN is justified due to its effectiveness against mode collapse, leveraging the implicit 1-Lipschitz distance property to avoid hyperparameter tuning for gradient clipping. The 1-Lipschitz continuity in W-GAN maintains the gradient norm within a range of 1 during training. The combination of Renyi differential privacy and W-GAN's gradient penalty results in the generation of high-quality synthetic medical data. Additionally, Faisal et al. employ fake and real image-based comparative loss, enhancing the variation in trained data and enabling better generalization.

Faisal et al. also focus on prompting the exploration of differential privacy (DP) techniques to protect patient information. Previous works have introduced DP variations of stochastic gradient descent and the PATE mechanism, but there were issues such as tuning hyperparameters, causing bias, and privacy leakage during training. To overcome these challenges, a new approach inspired by the G-PATE mechanism and differential privacy using a Gaussian distribution has been introduced. DP-GAN is chosen as a solution for privacy leakage during real database training, and DT-GAN and conditional GANs are discussed for generating tabular synthetic data and providing partial privacy. Moreover, Wasserstein GAN (W-GAN) techniques are emphasized to eliminate the need for selecting a proper clipping parameter and addressing the challenges posed by high-dimensional radiology images. The proposed approach ensures privacy in both centralized and distributed settings, even under an untrusted server. This strategy involves noisy gradients to prevent the exploitation of real data by the server.

In the experiment, Faisal et al. discuss the challenges of directly using public medical data due to privacy concerns, prompting the exploration of synthetic data generated by Generative Adversarial Networks (GANs). Several studies have shown the efficacy of GANs in generating realistic medical data, such as cine-MRI, liver CT scans, and retinal images. These models have been applied to segmentation and image denoising, demonstrating satisfactory performance. However, previous works have failed because they often neglect the privacy aspect. In contrast, the proposed approach incorporates relaxed differential privacy by using Resnet18, offering stronger image recognition accuracy and enables the generation of high-fidelity, high-dimensional image data with a high noise multiplier. The use of Wasserstein loss helps to overcome mode collapse problems and ensures private data generation, allowing for the protection of patient-sensitive information.

Faisal et al. discuss challenges in ϵ- Differential Privacy (ϵ-DP) approaches related to noise accumulation and the need to minimize the privacy budget. ϵ indicates the upper bound of privacy loss due to change in data and choosing the proper ϵ value will maintain the utility-privacy trade-off. The iterative nature of deep learning processes, combined with subsampling and loose upper bounds, leads to a high privacy costs. To solve these problems, the Gaussian mechanism is introduced

Data	Algorithm	CNN (0.07)	CNN (1.02)	MLP
MNIST	Real	99	97	98
	G-PATE	51	49	25
	DP-SGD GAN	63	60	52
	Our approach	78.2	76	77.2
X-ray	Real	71.56	74.78	76
	DP-SGD GAN	60	58	40
	Our approach	76.172	76.245	74.484

Fig. 1. Comparison experiments on synthetic images of two data sets (MNIST and X-ray) about performance of NN(Neural Network) for accuracy % [9]

by emphasizing its use for maintaining a tighter privacy upper bound under the composition mechanism. The Gaussian mechanism, with a higher spread and lower peak, is preferred for noise balance, but traditional (ϵ, δ)- privacy does not permit its usage. Faisal et al. chose the mechanism that focuses on distance rather than the log ratio of probabilities, ensuring a strong guarantee of composition and suitability to the Gaussian mechanism. The use of Renyi Differential Privacy is beneficial to avoid overestimating privacy loss during multiple iterations. This relaxed privacy mechanism supports the composition of different mechanisms without exponential growth in the privacy budget.

In addition, the architecture and training process of a generative model is described, focusing on the generator (G) and the discriminator (D) in the context of the Wasserstein GAN. The confidence probability provides feedback on the realism of the generated data. During training, feedback from the discriminator is fed back to the generator. To prevent an overfitted discriminator, which hinders meaningful feedback, the discriminator is updated five times per generator iteration.

In summary, Faisal et al. discuss a privacy-preserving approach to generating synthetic data in machine learning models, particularly focusing in particular on sanitization through gradient clipping and noise addition. The strategy involves adding noise to the gradient of the generator to prevent the impact of individual examples on the overall learning process, following the principles of differential privacy. A selective sanitization approach is applied, where gradient clipping is performed on the initial layers of the generator, but not on the local layers that are not exposed to private data.

Noise is added only to the generator's gradient, not to the discriminator's gradient. This trade-off aims to ensure both image quality and privacy, crucial in the medical domain where image quality influences critical decisions related to diseases. By only sanitizing the gradient directly relevant to the noisy input, this selective noise addition helps preserve important gradient information, leading to high-quality synthetic data. Moreover, to reduce budget and stabilize training, a pre-trained discriminator approach is adopted, where discriminators are trained on client machines and sent to a central server for generator updates. The goal is to preserve important gradient information, leading to high-quality synthetic data.

Experiments are conducted using Kaggle Chest X-ray Images and MNIST datasets, with a focus on reliability and defense against mode collapse. To remove imbalance in data images, there is an equal number of images of normal patient data and Pneumonia patient data. Compared to the range of 50 to 60% accuracy of previous work, the introduced approach demonstrates comparable accuracy to real data, even with a low resolution. According to Fig. 1, the method is tested with varying noise levels, and results show that high noise multipliers still yield high-quality X-ray image data. Despite high noise levels, the generated images maintain quality compared to previous models. In X-ray images, the approach achieves 74.4% accuracy with MLP and 76.245% with CNN, close to the real image accuracy.

To conclude, Faisal et al. contribute to have a highly regularized GAN model to generate higher quality private X-ray images with a reasonable budget using W-GAN and selective noise addition. The authors aim to raise awareness about protecting medical data privacy and encourage further research in the field of medical image data.

B. A Practical Application of Differential Privacy to Personalized Online Advertising

Online advertisements often rely on collecting and analyzing vast amounts of user data, raising significant privacy concerns. Ensuring the ethical and secure handling of this sensitive information is crucial since Privacy breaches and misuse of user data can severely damage the reputation of both advertisers and platforms, with far-reaching consequences. Lindell et al. discuss the concept of differential privacy as a criterion for preserving privacy in analyses over a dataset. Differential privacy is interpreted as a mathematical treatment that limits the information gained on individual records when observing the output of a computation. The definition assumes a powerful attacker capable of determining all database entries except the one under attack. While this may seem stringent, Lindell et al. argue that it is the right notion of privacy, especially in the context of a recent attack on Facebook's advertising system. The attacker in [2] effectively has the same power as assumed by the definition of differential privacy, demonstrating its relevance to the given task. Lindell et al. emphasize that a weaker privacy notion would not be sufficient. Lindell et al. also introduce a differential privacy mechanism for releasing statistics in online Facebook advertising campaigns, aiming to preserve user privacy with reasonable accuracy, even for small campaigns. Lindell et al. believe that the simplicity and low cost of the proposed mechanism will incentivize adoption by large internet agencies, contributing to the preservation of web users' privacy. Lindell et al. illustrate the mechanism's flexibility, allowing control over the trade-off between the number of released statistics and their accuracy, tailored for different-sized advertising campaigns. Experimental evidence shown in Fig. 2. supports the claim that the mechanism causes only a reasonable decrease in the accuracy of the released data.

It defines privacy as a property of the mechanism, requiring that a change in a single entry minimally impacts the distribution of responses seen by a potential adversary. The concept of differential privacy is formalized using the Hamming distance

between databases, and Lindell et al. present the notion of neighboring pairs of databases. The approach of Lindell et al. involves a computational process applied to a database and a query, producing a randomized function. The definition of ϵ-differential privacy is provided, stating that the mechanism should be randomized and satisfy a specific probability inequality. The definition is explained through a mental game involving an adversary trying to distinguish between two possible values for a single entry. Moreover, Lindell et al. illustrate how the definition limits the adversary's advantage, emphasizing the small constant ϵ as the privacy parameter. The discussion concludes by showing that, for small values of ϵ, the probability of the adversary guessing correctly is constrained.

Lindell et al. introduce a basic technique for constructing mechanisms that preserve differential privacy, specifically in the context of releasing online advertising statistics on Facebook while safeguarding user privacy against attacks like those of Korolova [4]. Begin with discussing the seemingly stringent definition of differential privacy, which allows an adversary to select all entries in the database except the one under attack and choose two possible values for the targeted entry, Lindell et al. argue that despite the strictness, this strong definition is necessary and nothing weaker would suffice. The attack suggested by Korolova [4] is used to illustrate the point. The attack involves collecting auxiliary information about a user and running campaigns with specific criteria to infer restricted information about the user, such as sexual preferences. The success of such attacks demonstrates the need for a definition that assumes the adversary's capability to determine all entries in the database.

Lindell et al. outline the format of the database containing advertising campaign statistics presented to advertising companies on Facebook. There are four main statistics: IMPRESSIONS (number of times the ad was shown), UNIQUE IMPRESSIONS (number of different users shown the ad), CLICKS (number of times the ad was clicked), and UNIQUE CLICKS (number of different users who clicked the ad). Each entry in the database is a triple (U, I, C), where U is a user's identity, I is the number of times the user saw the ad, and C is the number of times the user clicked on the ad. Lindell et al. provide the relevant queries for the statistics, including IMPRESSIONS, CLICKS, UNIQUE IMPRESSIONS, and UNIQUE CLICKS, expressed as functions over the database entries.

The authors also proposed a privacy-preserving mechanism for releasing impression and click statistics that involves composing four differentially private mechanisms, denoted as S, each related to a specific query type q. The mechanisms calibrate the noise to the sensitivity, a magnitude of change in the output, of the respective query types using the method described for the sum queries. The privacy parameter ϵ is selected based on the sensitivity of the sum-query of each mechanism and the desired level of accuracy. The sensitivity of each sum query is determined: for unique impressions and clicks (q3 and q4), the sensitivity is 1; for non-unique impressions and clicks (q1 and q2), bounds are set at 20 and 3, respectively, representing the maximum changes attributed to a single person in a day. Each sub-mechanism S outputs a perturbed count using Laplacian noise. Mechanism S1 perturbs the number of impressions, S2 perturbs the number of clicks, S3 perturbs the number of unique impressions, and S4 perturbs the number of unique clicks. Negative values are rounded to 0 to ensure non-negativity, without affecting the privacy of the mechanism.

Lindell et al. discuss the instantiation of privacy parameters ($\epsilon1$, $\epsilon2$, $\epsilon3$, $\epsilon4$) for the proposed privacy-preserving mechanisms (S1, S2, S3, S4). The goal is to find privacy parameter ϵ that balance between privacy and utility. The overall privacy parameter ϵ for the mechanism is set to 0.2. The privacy parameters for sub-mechanisms are chosen based on the sensitivity of the respective queries and acceptable error magnitudes. Concrete privacy parameters are selected as follows: $\epsilon1$ (for impressions): Chosen as 0.03 to allow an error of a few thousand impressions. $\epsilon2$ (for clicks): Chosen as 0.11 to allow an error of a bit over a hundred clicks. $\epsilon3$ (for unique impressions): Chosen as 0.01 to allow an error magnitude of 500. $\epsilon4$ (for unique clicks): Chosen as 0.05 to allow an error magnitude of 70. The probabilities of error for different counts are analyzed, considering Laplace noise and the selected privacy parameters.

In the experiments, the probability of meaningful error is discussed for small advertising campaigns, emphasizing that noise's impact on correctness is more pronounced for small campaigns. The analysis includes the probability of error in impression count, click count, unique impression count, and unique click count. Lindell et al. discuss the concept of ϵ-differential privacy with $\epsilon = 0.2$, focusing on the privacy guarantees provided by a mechanism in the context of an attack attempting to reveal private information about Facebook users, particularly their sexual orientation. The ϵ-differential privacy ensures that the probability of categorizing a user as homosexual based on their Facebook profile information is bounded by $e^{\epsilon} = e^{0.2}$.

Their analysis emphasizes that running the attack multiple times is necessary to decrease the privacy parameter. Even with multiple attacks, the privacy guarantee has limitations, and the likelihood of correct categorization remains relatively low. Lindell et al. also suggest that the efficiency, time, and cost of such attacks should be considered, making it challenging for an attacker to obtain significant information. Additionally, combining privacy mechanisms and detecting attacks on user privacy can enhance overall privacy protection. Lindell et al. conclude by suggesting that online advertising service providers and platforms like Facebook should explore optimal security parameters and consider adjustments in pricing to support smaller privacy parameters without compromising user privacy.

The goal of the experiments on the raw data from Facebook online advertising reports is to evaluate the tolerability of accuracy degradation. The experiments involve measurements from four real-life campaigns, providing a practical assessment of the mechanism's performance. The focus is on experimental results, examining the impact of differentially private mech-

Impressions	Y1	Clicks	Y2	CTR	Unique impressions	Y3	Unique clicks	Y4	Unique CTR
177028		171		0.10	10709		161		1.50
10252.00		2.00		0.02	3055.00		2.00		0.07
36222.00		120.00		0.33	11735.00		19.00		0.16
212659.00		97.00		0.05	34263.00		97.00		0.28
Noisy Impressions		Noisy Clicks		Noisy CTR	Noisy Unique impressions		Noisy Unique clicks		Noisy Unique CTR
176334.35	-693.65	156.17	-14.83	0.09	10624.50	-84.50	195.05	34.05	1.84
10608.23	356.23	0.00	-4.80	0.00	3075.29	20.29	2.76	0.76	0.09
36301.08	79.08	127.48	7.48	0.35	11731.11	-3.89	18.44	-0.56	0.16
211845.50	-813.50	94.09	-2.91	0.04	34198.63	-64.37	102.26	5.26	0.30

Fig. 2. A table demonstrating the effect of applying Mechanism S to the raw data available in Christian Thurston's tutorial on Facebook online advertising reports [5].

anisms on the accuracy of released campaign statistics. The experiments involve repeated runs of the same data for the four campaigns, selecting fresh Laplacian random variables independently for each run.

In Fig. 2, the results of a single experiment are presented, detailing the selection of appropriate noise for each measurement in each campaign. The raw data and the perturbed data after applying Mechanism S are compared. Observed from Fig. 2, the click-through rate (CTR), representing the percentage of impressions resulting in a user clicking the ad, is highlighted as a crucial measurement for campaign evaluation due to its usage of Facebook whether an ad is presented to users matching the criteria. The findings indicate that the decrease in accuracy, as reflected in CTR values, is tolerable and very minor. The experimental results provide a practical validation of the differential privacy mechanism's performance, supporting the mathematical arguments presented regarding the reasonableness of accuracy degradation, even for small advertising campaigns.

The proposed approach in [2] addresses the challenge of extending privacy-preserving mechanisms to include additional count statistics on Facebook campaigns, such as social impressions and social clicks. Instead of applying separate mechanisms for each new statistic, the method involves predetermining a global privacy parameter ϵ and allowing marketers to choose subsets of statistics for a campaign. Marketers have control over the trade-off between the number of released sums and the accuracy of each sum, while maintaining a fixed level of privacy. This approach accommodates differences in the importance of counts for large and small campaigns, providing flexibility in privacy and accuracy trade-offs.

C. Collecting and Analyzing Data from Smart Device Users with Local Differential Privacy

The importance of privacy in mobile data collection is paramount, as mobile devices hold a vast trove of highly personal and sensitive information about users. Ensuring robust privacy safeguards for this data is critical to building user trust, complying with privacy regulations, and upholding the ethical standards essential for the responsible development and use of mobile technologies. Samsung collects mobile phone usage data through a diagnostic tool bundled with the Samsung Android OS, including settings, memory, battery usage, and log data. This data transmission to Samsung requires user consent but could potentially expose sensitive information, prompting the need for privacy protection mechanisms. Local differential privacy (LDP) offers strong privacy guarantees, as seen in Google's Rappor [6]. However, Rappor's LDP mechanism is limited to single categorical attributes, while Samsung's data includes both numeric and categorical attributes. In addition, Rappor doesn't support complex learning tasks crucial for Samsung's analytical models. Therefore, it is necessary to develop a new data collection technique based on LDP.

In a scenario where an aggregator (e.g., Samsung) collects data from users (e.g., smart device owners) to compute statistical models, the goal is to maximize model accuracy while safeguarding user privacy. Local differential privacy (LDP) is adopted, where users perturb their data before sending it to the aggregator. This perturbation, controlled by a privacy parameter ϵ, ensures that the aggregator cannot distinguish between users' true data with high confidence. Nguyên et al. outline the concept of ϵ-LDP, where each user can have personalized privacy protection based on their privacy requirements. Two types of analytics tasks under ϵ-LDP are discussed: mean value and frequency estimation, and empirical risk minimization, including tasks like linear regression, logistic regression, and support vector machines (SVM).

Nguyên et al. explain the general privacy-preserving data aggregation approach, where users perturb their data using a function f before sending it to an aggregator. The perturbation function determines the trade-off between privacy and utility and the key is to design the perturbation function f to strike the right balance between protecting user privacy and enabling accurate aggregation of the data. Nguyên et al. also explore the design of a perturbation function f to enable accurate estimation of mean values for numeric attributes and frequencies for categorical attributes, first considering the case where all attributes have a numeric domain of [-1, 1]. The perturbation function f applies the Laplace Mechanism to a tuple containing numeric attributes. The resulting perturbed tuples are then averaged by the aggregator to estimate the mean of each attribute. However, this approach can lead to significant error, especially when the number of attributes is large. A method proposed by Duchi et al. [3] claims to address this issue, but upon evaluation, it was found to be biased and could violate differential privacy when d numeric attributes are even. However, an algorithm using HARMONY is presented in [3], and this proposed algorithm achieves similar privacy guarantees and error bounds but with better efficiency because a user only needs to send a single bit to the aggregator which removes a communication overhead between a user and the aggregator.

When dealing with both numeric and categorical attributes in user data records, HARMONY aims to estimate the mean value for numeric attributes and the frequency distribution for categorical ones while maintaining privacy. For binary attributes, it employs the classic randomized response method to estimate distributions accurately. For multiple categorical attributes, it extends Bassily and Smith's method, using ran-

dom projection to handle each attribute individually. Hence, HARMONY introduces an alternative solution to Bassily and Smith's method to improve accuracy, especially for smaller categorical domains. Bassily and Smith's method tends to be unstable with small categorical domains due to the introduction of huge noise when a random matrix gets larger. HARMONY simplifies the process by estimating mean values for numerical attributes and frequency distributions for categorical ones separately. This solution ensures differential privacy and provides accurate estimates for both types of attributes.

Nguyên et al. investigate designing machine learning models that can be expressed as empirical risk minimization under ϵ-local differential privacy. The authors focus on three common learning tasks - linear regression, logistic regression, and SVM classification. Nguyên et al. also discuss using stochastic gradient descent (SGD) to compute the optimal model parameter, which is a common approach. However, under the privacy-preserving setting, the gradient is not directly available to the aggregator and needs to be collected privately. Prior work [7,8] has suggested using the Laplace mechanism or Duchi et al.'s method to obtain a noisy version of the gradient. However, this is still insufficient for the authors' specific use case, which is explored in more detail in the next paragraph.

To address the inaccuracy issue caused by noisy gradients in stochastic gradient descent (SGD) for parameter optimization, mini-batch gradient descent is proposed instead. Mini batch gradient updates the weights of the model based on a small random subset of the data set instead of a single example or the entire data set. In mini-batch gradient descent, a group of users submits noisy gradients, and the parameter vector is updated with the mean of these gradients. However, when the dimensionality of the data is large, the acceptable mini-batch size becomes too large, leading to premature termination of the algorithm. To mitigate this problem, dimension reduction is proposed. Each user's data is projected into a lower-dimensional subspace using a random matrix, reducing the dimensionality of the gradients. This reduces the error in the average noisy gradient obtained from a mini-batch of users. Consequently, the acceptable mini-batch size is reduced, allowing for more iterations of the algorithm. Above Algorithm 4 of [3] outlines the mini-batch gradient descent method with dimension reduction. The aggregator generates a random matrix and maintains a lower-dimensional parameter vector. Users project their data into the reduced space, compute noisy gradients, and submit them to the aggregator for parameter updates. The algorithm terminates when the parameter update is sufficiently small or when enough users have participated. Nguyên et al. employ two public datasets, US and Brazil, sourced from the Integrated Public Use Microdata Series in Fig. 3. The US dataset comprises 9 million tuples with 23 attributes, including 6 numeric attributes (e.g., age) and 17 categorical attributes (e.g., gender). The BR dataset contains 4 million records with 18 attributes, including 6 numeric and 12 categorical attributes. Both datasets feature a numeric attribute named "total income," used as the dependent variable in linear regression, logistic regression, and SVM models. Additionally,

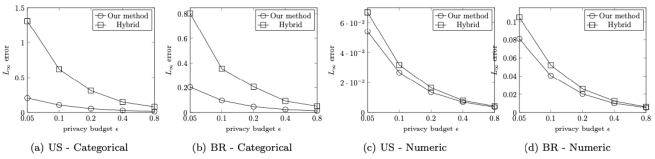

Fig. 3. Experiments of applying Hybrid and Harmony to generate noisy tuples which then are used to estimate means and frequencies of each value in US and BR datasets extracted from the *Integrated Public Use Microdata Series* [10].

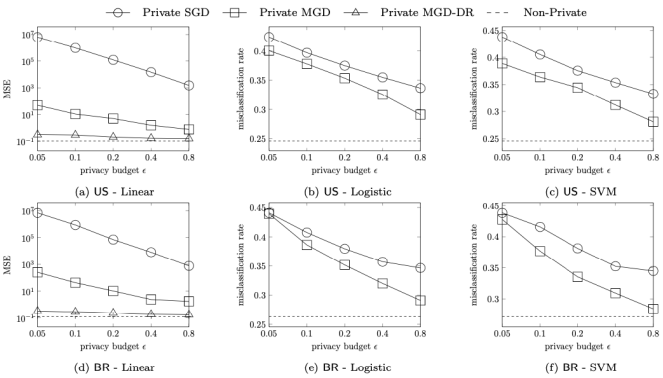

Fig. 4. Experimental results of the performance of each method for empirical risk minimization [11].

the numeric attributes are normalized to the range [-1, 1] for consistency in analysis.

In the experiments, a method called HYBRID is devised to address the challenge of collecting noisy tuples from users to estimate both numeric attribute means and categorical value frequencies according to Fig. 3. Hybrid combines existing solutions for numeric and categorical attributes: Duchi et al.'s method for numeric attributes and Bassily and Smith's method for categorical attributes. By employing these methods with appropriate privacy budgets, Hybrid ensures ϵ-local differential privacy (LDP).

In experiments, when applied to the US and BR datasets, HYBRID is compared against HARMONY. Based on Fig. 3, the results demonstrate that the proposed solution consistently outperforms Hybrid in terms of accuracy for estimating categorical value frequencies. Additionally, for numeric attribute mean estimation, the proposed solution slightly outperforms Hybrid across both datasets and various privacy budget settings. Notably, HARMONY incurs lower communication costs per user compared to Hybrid, as it requires only the transfer of 1 bit per user and is simpler to implement.

In the second set of experiments, linear regression, logistic regression, and SVM classification are performed on the US and BR datasets, with "total income" as the dependent variable and other attributes as independent variables. Categorical attributes are transformed into binary ones. Four methods are evaluated: private stochastic gradient descent (SGD), mini-batch gradient descent (MGD), MGD with dimension reduction (MGD-DR), and non-private SGD.

For linear regression, according to Fig. 4, MGD suffers from unsatisfactory accuracy due to the large mini-batch size required when the dimensionality is large. MGD-DR, which incorporates dimension reduction, achieves accuracy close to non-private SGD. In Fig. 4, for logistic regression and SVM, MGD and MGD-DR demonstrate effectiveness in reducing misclassification rates compared to private SGD. Overall, the experiments highlight the effectiveness of mini-batches and dimension reduction in empirical risk minimization under local differential privacy.

Nguyên et al. introduce HARMONY, a solution for collecting and analyzing users' personal data under ϵ-local differential privacy. HARMONY can handle multiple numeric and categorical attributes and compute accurate statistics, ranging from simple metrics like mean and frequency to machine learning models such as linear regression, logistic regression, and SVM classification. HARMONY achieves optimal asymptotic error bounds and demonstrates high accuracy in practical scenarios. Additionally, it is efficient in terms of communication and computational overhead. Extensive experiments on real data validate its effectiveness. Future work in [3] will explore the application of HARMONY in real-world scenarios, such as Samsung's diagnostic info report app.

CONCLUSION

In this report, I investigate how differential privacy techniques have been applied to strengthen personal privacy in multiple fields. The report highlights the effectiveness of employing W-GAN and selective noise addition to enhance the quality of generated private X-ray images contributing significantly to the field of medical data privacy. Additionally, the proposed approach for extending privacy-preserving mechanisms to include additional count statistics in Facebook campaigns offers marketers flexibility in balancing privacy and accuracy trade-offs, catering to differences in the importance of counts for campaigns of varying sizes. Furthermore, the introduction of HARMONY as a solution for collecting and analyzing users' personal data under ϵ-local differential privacy demonstrates its efficacy in handling multiple attributes and computing accurate statistics, with promising applications in real-world scenarios.

REFERENCES

[1] Faisal, F., Mohammed, N., Leung, C. K., and Wang, Y. (2023, February 8). Generating Privacy Preserving Synthetic Medical Data. IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10032429

[2] Lindell, Y., and Omri, E. (2011, March 29). A practical application of differential privacy to personalized online advertising. Cryptology ePrint Archive. https://eprint.iacr.org/2011/152

[3] Nguyên, T. T., Xiao, X., Yang, Y., Hui, S. C., Shin, H., and Shin, J. (2016, June 16). Collecting and analyzing data from smart device users with local differential privacy. arXiv.org. https://arxiv.org/abs/1606.05053

[4] A. Korolova. Privacy violations using microtargeted ads: A case study. IEEE International Workshop on Privacy Aspects of Data Mining (PADM 2010), 2010. See: http://theory.stanford.edu/ korolova/Privacy_violations_using_microtargeted_ads.pdf.

[5] C. Thurston. An explanation of facebook advertising reports. http://www.internetmarketingsolution.com.au/facebook-ad-performance-reports.html, 2010.

[6] U. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In CCS, pages 1054–1067, 2014.

[7] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Privacy aware learning. J. ACM, 61(6):38:1–38:57, 2014.

[8] J. Hamm, A. C. Champion, G. Chen, M. Belkin, and D. Xuan. Crowd-ml: A privacy-preserving learning framework for a crowd of smart devices. In ICDCS, pages 11–20, 2015.

[9] D.S.Kermany,etal.Identifyingmedicaldiagnosesandtreatablediseases by image-based deep learning. Cell 172, 2018, 1122–1131.e9.

[10] IPUMS. Integrated public use microdata series. https://www.ipums.org.

[11] C. Cortes and V. Vapnik. Support-vector networks. Machine Learning, 20(3):273–297, 1995.