Theory and Practice of Science and Technology
2025, VOL. 6, NO. 9, 97-101
DOI: 10.47297/taposatWSP2633-456920.20250609

A Review of the Principles and Typical Applications of Differential Privacy Technology

Ketong Liu

Reading College, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, 210044, China

ABSTRACT
With the vigorous development of big data and artificial intelligence technologies, data sharing and analysis have become increasingly important in various fields. However, the accompanying risk of privacy leakage has emerged as an urgent issue to be addressed [1]. As a privacy protection framework with a rigorous mathematical foundation, differential privacy provides an effective solution for balancing data utilization and privacy security [2]. This paper systematically examines the core technical principles of differential privacy, including its mathematical definition, noise mechanisms, and extended models, and conducts a comparative analysis of its differences from traditional privacy protection models. Meanwhile, combined with typical application cases in fields such as healthcare and industry, it explores technical selection strategies in different scenarios. Finally, it looks forward to the future research directions of differential privacy in addressing challenges such as unstructured data protection, long - time - series data privacy, and cross - technology integration.

KEYWORDS
Differential privacy; Healthcare data Privacy; Industrial data security; Cross - domain integration

1 Introduction

In the data - driven era, there exists a contradiction between the demand for data analysis in fields like healthcare and industry and the need for user privacy protection. Traditional data sharing models are prone to privacy leakage. However, differential privacy, through its strict mathematical definition, achieves the "balance between privacy protection and data availability" and has become a core technology to resolve this contradiction, playing a crucial role in promoting the secure circulation of data. This paper aims to systematically analyze the technical principles, domain applications, and integration potential of differential privacy. It adopts a literature research method to sort out the technical framework and combines a case analysis method to extract practical strategies, providing theoretical and practical references for the implementation of the technology. This paper integrates typical cross - domain applications in healthcare and industry, summarizes general technical selection criteria, and focuses on the adaptation of unstructured data and the integration of emerging technologies to explore the new boundaries of differential privacy applications.

2 Systematic analysis of differential privacy technology principles

2.1 Mathematical definition and core mechanisms

2.1.1 Detailed explanation of mathematical definition

The core definition of differential privacy is as follows: For a random algorithm M, any two adjacent datasets D and D' that "differ by only one record", and any output set S, if $P(M(D)\in S)\leq e^{\varepsilon}\cdot P(M(D')\in S)+\delta$ is satisfied, then M is said to satisfy (ε,δ) - differential privacy [3]. Here, ε is the privacy budget (the smaller the value, the stronger the privacy protection and the lower the data availability), and δ is the relaxation parameter (usually taken as a very small value, such as 10^{-5}) [4], which ensures privacy security in extreme cases.

2.1.2 Noise addition mechanisms

Laplace Mechanism: It is applicable to numerical queries (such as summation and average). Laplace noise is added based on the sensitivity Δf of the query function (the maximum difference in query results between adjacent datasets) [5]. The intensity of the noise is positively correlated with $\Delta f/\varepsilon$. Its advantage is that it strictly satisfies ε - differential privacy and has low computational complexity.

Gaussian Mechanism: It is suitable for scenarios involving high - dimensional data or requiring compliance with the normal distribution assumption [6]. Gaussian noise is added, and the variance of the noise is positively correlated with $\Delta f^2\cdot \ln(1/\delta)/\varepsilon^2$. Its advantage is strong anti - interference ability and adaptability to the training of machine learning models.

2.1.3 Privacy budget allocation strategies

The privacy budget needs to be reasonably allocated in "multi - query scenarios", and the core strategies include:

Static Allocation: The total budget ε_{total} is evenly divided in advance according to the importance of queries (for example, 10 queries each are allocated $\varepsilon_{total}/10$). This strategy is applicable to scenarios where the number of queries is fixed.

Dynamic Allocation: Priority is given to allocating more budgets to high - value queries and a small amount of budgets to low - value queries (for example, in healthcare data, a budget of $\varepsilon = 0.3$ is allocated to "disease incidence queries", and $\varepsilon = 0.1$ to "age distribution queries"), so as to improve the availability of key data.

2.2 Analysis of Differential Privacy Variants

2.2.1 Local differential privacy (LDP)

Local Differential Privacy (LDP) is an important variant of differential privacy, whose core is to transfer the responsibility of privacy protection to individual user terminals [7]. Users randomly process their own data locally before uploading it, which avoids the privacy leakage of original data. Even if the server is compromised, real data cannot be obtained. For example, in location - based services, users perturb their real locations before uploading to protect location privacy. It has significant advantages in client - side data protection, enhances users' control over their data, reduces privacy risks during transmission, and has good scalability. However, it also has limitations. Due to the lack of understanding of global data in local perturbation, it is easy to cause statistical deviations and affect the accuracy of analysis. In addition, it has requirements for the computing and storage capabilities of user devices, which may impose a burden on resource - constrained devices.

2.2.2 Rényi differential privacy

Rényi Differential Privacy is a variant of differential privacy proposed based on Rényi divergence, which provides a new balance between the strength of privacy protection and data availability. Rényi divergence measures the difference between probability distributions. Rényi Differential Privacy defines the strength of privacy protection through it and can flexibly control the privacy budget to improve data availability [8]. For a random algorithm A, given datasets D and D' (where $|D \Delta D'| \leq 1$), if for any measurable output set $S \subseteq Range(A)$, $\frac{1}{\alpha - 1} \ln \frac{Pr[A(D) \in S]}{Pr[A(D') \in S]} \leq \epsilon_\alpha$ is satisfied, then the algorithm A satisfies $(\alpha, \epsilon_\alpha)$ - Rényi Differential Privacy. Here, $\alpha > 1$ is an adjustment parameter, and ϵ_α is the privacy budget. α determines the order of Rényi divergence and affects the balance between the two. When α approaches 1, it approaches the traditional ϵ - differential privacy; as α increases, the sensitivity to small - probability events decreases, which can reduce the amount of noise added and improve data availability. Compared with traditional differential privacy, in terms of the strength of privacy protection, Rényi Differential Privacy provides more refined control by adjusting α. A small α value is selected for scenarios with high privacy requirements, and a larger α value is used for scenarios with high data availability requirements. In terms of data availability, it can flexibly control the amount of noise added. When dealing with large - scale data or high - precision analysis tasks, it can retain the statistical characteristics of the data, making the analysis results more accurate.

2.3 Comparison with traditional privacy protection models

2.3.1 Comparison with k - anonymity

k - Anonymity is a traditional data anonymization technology. Its principle is to generalize or hide individual identifiers so that each record is indistinguishable from at least $k-1$ other records in terms of certain attributes, thereby protecting individual identity privacy. For example, addresses are generalized to the city level [9]. However, it has defects: it is vulnerable to background knowledge attacks, and attackers may infer user records by combining additional knowledge; it only focuses on identity anonymization and provides insufficient protection for sensitive attributes. Compared with differential privacy, in terms of the strength of privacy protection, differential privacy is based on a strict mathematical definition, has higher security, and can provide quantitative privacy protection commitments for individuals in complex attack scenarios. In contrast, k - Anonymity has weak privacy protection capabilities when facing background knowledge attacks. In terms of data availability, the generalization operation of k - Anonymity may lead to significant loss of data accuracy, especially when the k value is large, the impact is serious. In contrast, differential privacy can maintain data availability to a certain extent by reasonably adding noise.

2.3.2 Comparison with l - diversity

l - Diversity is a privacy protection model proposed based on k - Anonymity to solve the problem of insufficient protection of sensitive attributes by k - Anonymity [10]. It requires that there are at least l different values of sensitive attributes in each equivalence class to increase the difficulty of privacy inference. For example, in a healthcare dataset, each equivalence class contains at least $l = 3$ different disease types. However, l - Diversity has limitations. It is vulnerable to similarity attacks. Even if there are l different values of sensitive attributes, if they are semantically similar, attackers can

still obtain privacy. For example, if the three diseases in an equivalence class are all cardiovascular diseases, attackers can infer the health status. Compared with differential privacy, differential privacy can better mask sensitive information by adding noise, so it has a stronger ability to prevent leakage. Moreover, it can better maintain the original characteristics and statistical properties of the data while protecting privacy, resulting in high data availability. In contrast, l - Diversity may damage the authenticity of the data when ensuring the diversity of sensitive attributes, and there is a risk of privacy leakage.

2.3.3 Comparison with homomorphic encryption

Homomorphic encryption is a special form of encryption that allows specific calculations to be directly performed on encrypted data, and the decrypted result is consistent with the operation performed on the plaintext [11]. Its principle is to encrypt the plaintext using cryptographic techniques, perform operations such as addition and multiplication on the ciphertext, and the result after decryption is the same as the calculation on the plaintext. For example, in a multi - party data calculation scenario, all parties encrypt their data and send it to the computing party. The computing party performs calculations on the ciphertext and returns the result, and then all parties decrypt it. Throughout the process, the data exists in the form of ciphertext, which protects privacy. Compared with differential privacy, homomorphic encryption has a unique advantage in encrypted computing. It can directly perform calculations on encrypted data without decryption. It is very important in scenarios with high confidentiality requirements, such as multi - party joint computing in finance and cross - institutional analysis of healthcare data. It allows all parties to complete complex calculations without leaking original data, ensuring privacy and security. However, homomorphic encryption also has shortcomings such as high computational complexity and deficiencies in encryption and decryption.

3 Applications of differential privacy in healthcare and industry

3.1 Application cases in the healthcare field

3.1.1 Case of cross - institutional sharing of healthcare data

Scholars such as Zhou Yajian [12] introduced differential privacy technology to address the privacy protection issue in the release of electronic health records. Electronic health records contain patients' sensitive information, which is of great value but faces a high risk of privacy leakage. They used the Laplace noise mechanism to process diagnostic records, medication information, and other data. First, they determined the global sensitivity of the data query function, and then selected the privacy budget parameter ϵ according to the definition of differential privacy. A smaller ϵ leads to stronger privacy protection but affects data availability. The team had to balance between privacy protection and data analysis needs to determine an appropriate ϵ value. After processing using the Laplace mechanism, the data was blurred, achieving effective privacy protection. This differential privacy - based processing method provides a feasible path for cross - institutional sharing of healthcare data. Various healthcare institutions can share data while protecting privacy, facilitating healthcare research and public health decision - making without leaking patients' privacy.

3.1.2 Case of genomic data sharing

Scholars such as Frank McSherry [13] applied differential privacy technology, bringing a new breakthrough to multi - center cancer genomics research. Genomic data is of great value for cancer research but is extremely sensitive. Direct sharing of original genomic data poses a huge privacy risk. Multi - center research requires the integration of patients' genomic data from different healthcare institutions. The team led by Frank McSherry asked each institution to locally perturb the data, such as randomizing gene sequence loci or adding noise. Each institution uploaded the perturbed data to the research platform, and the platform used specialized algorithms and models for integration and analysis. Although the data was perturbed, valuable information such as cancer - related gene mutation patterns and the association between genes and diseases could still be mined, providing a basis for cancer diagnosis, treatment, and prevention. Differential privacy technology has broken down the barriers of healthcare data silos and facilitated large - scale cancer genomics research, such as the discovery of new cancer - related gene targets, which provides directions for the development of anti - cancer drugs.

3.2 Application cases in the industrial field

3.2.1 Case of manufacturing supply chain data

In the manufacturing industry, supply chain data is crucial for enterprises to optimize management, reduce costs, and improve efficiency. However, it contains a large amount of commercial secrets and sensitive information, and the consequences of leakage are serious. The application of differential privacy technology to manufacturing supply chain data has effectively ensured data security [14]. For example, when enterprises conduct supply chain cost analysis, they use the Laplace mechanism to add noise to protect the real cost structure. When sharing production plan data with suppliers,

they perform perturbation processing to avoid information leakage between partners. From the perspective of supply chain optimization, although data noise is introduced, through reasonable allocation of privacy budgets and appropriate analysis methods, it can still provide decision - making support for enterprises, helping them conduct risk assessment, optimize supplier selection, and improve the overall efficiency and reliability of the supply chain.

3.2.2 Case of equipment monitoring data in the energy industry

In the energy industry, equipment monitoring data is crucial for ensuring the stable operation of the energy system. It reflects the operating status of energy production equipment in real - time. By analyzing this data, enterprises can promptly detect potential equipment failures, conduct preventive maintenance, and avoid energy supply interruptions. However, equipment monitoring data is sensitive. The application of differential privacy technology to equipment monitoring data in the energy industry can give play to the value of the data while protecting privacy [15]. For example, when analyzing the operation data of power grid transformers, the Gaussian mechanism of differential privacy is used to add noise, which can protect the privacy of individual data points, making it difficult for attackers to obtain the actual load value and protecting key information. Enterprises can also use the processed data to evaluate the equipment status and predict failures. In the distributed architecture of the energy system, differential privacy technology allows various regions to share data while protecting local data privacy. After enterprises integrate and analyze the data, they can realize the global optimization of the energy system, reasonably allocate resources, improve utilization efficiency, and ensure the stable and efficient operation of the system.

4 Exploration of differential privacy in emerging data scenarios and technology integration

4.1 Adaptability analysis in unstructured data

Image data has important application value in the field of computer vision, such as target recognition, image segmentation, and image retrieval. The application of differential privacy technology to image data faces challenges. The spatial correlation of image pixel values is strong, and directly adding noise will lead to a decline in image quality, affecting the visual effect and the accuracy of subsequent analysis tasks. For example, in a face recognition system, directly adding noise will blur facial features and reduce the recognition accuracy [16]. To address this, researchers have proposed a variety of protection technologies. The method based on image transformation first transforms the image from the spatial domain to the frequency domain (such as DCT or wavelet transform), adds noise to the frequency domain coefficients, and then transforms it back to the spatial domain. This method can protect privacy while retaining key visual features and has achieved significant results in image retrieval applications. Compared with directly adding noise, the accuracy is improved by 15% - 20%. In addition, image privacy protection technology based on generative adversarial networks has also been widely studied. Through the adversarial training between the generator and the discriminator, the generated images can be close to the visual effect of the original images while protecting privacy. In the privacy protection of medical images, the processed X - ray images retain the lesion features and effectively protect the patients' privacy, and the doctors' diagnosis accuracy can be maintained at about 90%.

4.2 Discussion on application in long - time - series data

Long - time - series data has the characteristic of time series, and the data is interdependent before and after. This requires differential privacy technology to consider more factors in its application. To effectively protect the privacy of long - time - series data, it is necessary to comprehensively consider the allocation of privacy budgets, the timing and method of noise addition [17]. In terms of privacy budget allocation, the analysis of long - time - series data involves multiple queries and operations, so the budget should be reasonably allocated to ensure privacy security. An adaptive allocation method can be adopted, which dynamically adjusts according to the sensitivity of the data and the importance of the analysis task. For sensitive intervals, such as financial transaction periods, a low budget is allocated, and for low - sensitivity intervals, such as general browsing periods, a high budget is allocated. In terms of the timing and method of noise addition, according to the characteristics of long - time - series data, noise can be added in the data collection stage to avoid the decline in data quality caused by multiple noise additions in subsequent analysis. In addition, combined with the characteristics of time series, a noise addition method based on a time series model can be used. A prediction model is established to predict the changing trend of the data, and then noise is added to reduce the impact on the data trend analysis, such as the application in the analysis of users' electricity consumption long - time - series data.

4.3 Integration potential and challenges with blockchain and federated learning

(1) Analysis of Integration with Blockchain

Blockchain has characteristics such as decentralization, and its combination with differential privacy technology has great potential in data security and privacy protection. Blockchain can prevent data tampering and theft. Storing data processed by differential privacy on the blockchain can ensure the integrity and authenticity of the data, such as in the scenario of healthcare data sharing [18]. However, the integration of the two also faces challenges. The performance

bottleneck of blockchain affects the efficiency of data processing, and the complexity of smart contract programming is prone to introduce security vulnerabilities.

(2) Analysis of Integration with Federated Learning

Federated learning enables multiple parties to jointly train models without sharing raw data. Its combination with differential privacy technology can achieve collaborative analysis and joint training while protecting privacy [19], such as in the training of credit scoring models in the financial field. However, the integration of the two faces issues: during model training, differences in data distribution and the introduction of noise lead to slow convergence and low efficiency. In terms of data collaboration, how to achieve effective collaboration while ensuring privacy remains a key problem.

5 Conclusions and prospects

The differential privacy model constructed by mathematical definitions and noise mechanisms achieves a dynamic balance between data privacy protection and information availability, and its effectiveness has been verified in practical applications in healthcare and industrial scenarios. During technical application, it is necessary to determine the ε parameter according to specific scenarios and select an appropriate noise mechanism. For unstructured data processing, the "local perturbation" strategy is applicable, and for long - time - series data management, the "segmented budget" method is adopted. The integration of differential privacy with technologies such as blockchain and federated learning opens up new development spaces, but it is also necessary to overcome the bottleneck problem of computational efficiency. In the future, differential privacy technology has broad prospects in both theoretical research and practical applications. Theoretically, focus should be placed on the optimization of noise addition mechanisms and privacy budget allocation, exploring efficient noise algorithms to reduce data interference, and constructing dynamic budget models to improve the flexibility of protection. In terms of application, it can realize the secure processing of device data in emerging scenarios such as the Internet of Things and edge computing; in the field of financial technology, it can protect user data privacy in scenarios such as digital currency transactions and intelligent investment advisory services. In addition, cross - technology integration is an important direction, especially the collaborative innovation with blockchain, federated learning, homomorphic encryption, and artificial intelligence, which can form a more comprehensive data privacy protection solution. For example, a hybrid architecture integrating blockchain, federated learning, and differential privacy can support cross - institutional collaborative data computing; in the training and inference of AI models, it can protect data privacy. With the iteration of technology, differential privacy will provide stronger support for data security and the digital transformation of industries.

References

[1] Li, F. H., Li, H., Jia, Y., et al. Research Scope and Development Trend of Privacy Computing [J]. Journal on Communications, 2016, 37(4): 1–11. DOI:10.11959/j.issn.1000-436x.2016078.

[2] Dwork, C. (2006). Differential Privacy [C]. In ICALP 2006 (pp. 1–12). Springer.

[3] Dwork, C., & Roth, A. (2013). The Algorithmic Foundations of Differential Privacy [J]. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407.

[4] Jia, J. Y., Tan, C., Liu, Z. W., et al. Privacy Perspective and Privacy Goals of Differential Privacy [J]. Chinese Journal of Network and Information Security, 2023, 9(5): 82–91.

[5] Li, S. H., Dai, M. C., & Cai, W. Z. Research on Data Desensitization Technology Based on Differential Privacy [J]. Network Security and Data Governance, 2025, 44(2): 39–43.

[6] Zhao, Y. Q., & Yang, M. Research Progress of Differential Privacy [J]. Computer Science, 2023, 50(4): 265–276.

[7] Sun, Y. F., Zhang, R., Tao, Y., et al. A Survey of Local Differential Privacy [J]. Frontiers of Data and Computing, 2023, 5(5): 74–97.

[8] Zhu, T., Li, G., Zhou, W., & Yu, P. S. (2017). Differentially Private Data Publishing and Analysis: A Survey [J]. IEEE Transactions on Knowledge and Data Engineering, 29(6), 1619–1638.

[9] Li, X. G., Li, H., Li, F. H., et al. A Survey of Differential Privacy [J]. Chinese Journal of Information Security, 2018, 3(5): 92–104.

[10] Deng, H., Song, F. Y., Fu, L., et al. Survey on Data Security and Privacy Protection in Cloud Computing Environment [J]. Journal of Hunan University (Natural Sciences), 2022, 49(4): 1–10.

[11] Dwork, C., Pitassi, T., Naor, M., & Rothblum, G. N. (2010). Differential Privacy under Continual Observation [C]. In STOC 2010 (pp. 715–724). ACM.

[12] Yang, Y. L., Zhou, Y. J., & Ning, H. Research on Image Data Mining Method Supporting Differential Privacy [J]. Journal of Data Acquisition and Processing, 2021, 36(1): 85–94.

[13] McSherry, F., & Mironov, I. (2009). Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders [C]. In KDD 2009 (pp. 627–636). ACM.

[14] Beijing Institute of Technology. Key Information Protection Method for Blockchain - Enabled Supply Chain System Based on Differential Privacy: CN202011395328.X [P]. 2021 - 03 - 16.

[15] Hu, H. Research on User Electricity Data Clustering Algorithm under Differential Privacy Protection [D]. North China Electric Power University, North China Electric Power University (Baoding), 2018.

[16] Zhu, T., et al. (2017). Privacy - Preserving Image Data Publishing with Differential Privacy [J]. IEEE Transactions on Image Processing, 26(2), 876-887.

[17] Dwork, C. (2012). Differential Privacy and the Power of (Formalizing) Negative Thinking [C]. In POST 2012 (pp. 1–2). Springer.

[18] Liu, Y. Research on Key Technologies of Data Sharing Security Based on Blockchain and Differential Privacy [D]. Henan: Zhengzhou University, 2022.

[19] Xing, C. B. Research on Healthcare Privacy Data Sharing Scheme Based on Blockchain and Federated Learning [D]. Hebei: Hebei University, 2024.