Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 12, No. 2, 2025

A Review of Business Data Security Based on Differential Privacy Protection

Fanjing Dong *

College of Transportation Engineering, Chang'an University, Xi'an, Shaanxi, 710018, China
*** Corresponding author Email:** Fanjing0607@outlook.com

Abstract: This paper summarizes the current status of the application of differential privacy technology in commercial data protection, and discusses its theoretical foundation, application methods and future development direction. By introducing a random noise mechanism, differential privacy ensures that an attacker cannot infer individual privacy information through statistical analysis during data release or query, while maintaining the analytical value of the data. This paper summarizes the main research results of differential privacy, analyzes its application prospects in the big data environment, and proposes that differential privacy technology has a wide range of application potentials in commercial data protection as well as future research directions.

Keywords: Differential Privacy; Big Data; Privacy Protection; Commercial Data Security.

1. Introduction

With the rapid development of artificial intelligence, a large amount of commercial information is becoming increasingly transparent in the data network, especially in the booming e-commerce, it is particularly urgent and crucial to protect commercial privacy data and prevent privacy invasion, which is largely related to the survival and development of related enterprises. Differential privacy, as a means of cryptography that consists of a rigorous mathematical foundation and a quantitative formal proof of the risk of privacy leakage, can effectively resist differential attacks and become a new privacy protection model. Differential privacy algorithms formed by combining differential privacy protection with the topic of data security protection in the commercial field can make the protected data widely used for accurate data analysis [1]. The purpose of this paper is to overview the current status of the application of differential privacy technology in commercial data protection, analyze its advantages and shortcomings, and explore the future research direction.

2. Differential Privacy Data Security

2.1. Definition and Theoretical Foundations of Differential Privacy

Differential privacy, as a cutting-edge privacy-preserving technique, effectively prevents an attacker from inferring the existence or non-existence of individual data records based on query results by introducing random noise into the original data. Its core definition is that for any two datasets differing by only one data point, the probability distribution of the output results produced by the differential privacy mechanism is nearly the same. In contrast to traditional privacy-preserving schemes, the differential privacy model assumes that the attacker has maximum background knowledge, i.e., the attacker is able to obtain information about all other records except the target record. Differential privacy techniques are based on rigorous mathematical formulae derivation, which can effectively solve the problems of over-reliance on the attacker's background knowledge and lack of quantitative analysis and evaluation in traditional privacy protection techniques [2].

2.2. Differential Privacy in Big Data Privacy Protection

Differential privacy technology achieves a balance between privacy protection and data utility by adding noise (e.g., Laplace noise or Gaussian noise) to the data, which not only effectively protects privacy, but also ensures the accuracy of the statistical analysis results. Google's RAPPOR project, for example, randomizes user data with the help of differential privacy technology, which protects the user's privacy and allows Google to obtain valuable information at the level of data aggregation, even if an attacker has auxiliary information that cannot be deduced from individual user details. This allows Google to obtain valuable information at the level of data aggregation, so that even if an attacker gets hold of auxiliary information, he or she will not be able to deduce the details of a single user.

2.2.1. Data Release and Query Protection

Differential privacy techniques protect the privacy of a query by adding noise to the query results before the data is released, while maintaining the statistical properties of the data, such as in the case of SQL queries. Whether it is a simple SELECT, COUNT, SUM query or a complex multi-table JOIN query, differential privacy techniques can provide accurate query results without compromising privacy [3].

2.2.2. Machine Learning and Model Training

Differential privacy techniques in the field of machine learning, on the other hand, protect data privacy by adding noise to the training process, such as the differential privacy optimized gradient descent algorithm (DP-SGD), which adds noise to the gradient update to protect the privacy of the training data while maintaining the performance and utility of the model. Differential privacy techniques can also protect the privacy of model parameters at model release time [1].

2.2.3. Real-time Data Processing and Dynamic Updates

Differential privacy techniques are not only applicable to static data, but also support real-time data stream processing and dynamic data updates. For example, by adding noise in

statistical queries and dynamic updates of real-time data streams, differential privacy techniques can provide real-time data processing capabilities without compromising privacy.

2.2.4. Data Visualization and Statistical Analysis

In data visualization and statistical analysis, differential privacy techniques protect privacy by adding noise when plotting heat maps, histograms, or calculating statistics such as mean and median. This approach provides accurate statistical analysis results without compromising privacy.

2.2.5. Financial Data Analysis

In the financial field, differential privacy techniques also play an important role. For example, in statistical analysis of financial data and querying of transaction data, by adding noise, differential privacy techniques can provide powerful analysis and processing capabilities without compromising privacy.

2.3. Variations and Improvements in Differential Privacy Adaptation to Business Data

2.3.1. Privacy Budget Management for Differential Privacy

Privacy budget management refers to the reasonable allocation of privacy budget in multiple queries or data releases to balance privacy protection and data utility. By reasonably allocating the privacy budget, the utility of data is maximized, resource optimization is achieved, and the continuity of privacy protection can be protected for a long period of time in multiple data releases, which is applicable to scenarios requiring multiple data releases, such as data analytics platforms, data sharing systems, and so on.

2.3.2. Adaptive Differential Privacy

Adaptive differential privacy allows privacy-preserving mechanisms to dynamically adjust the amount of noise based on the sensitivity of the data. For example, increasing noise when data sensitivity is high and decreasing noise when sensitivity is low. Dynamically adjusting the privacy protection intensity according to the sensitivity of the data improves data utility and can adapt to different data environments and privacy requirements. is suitable for scenarios that require dynamic adjustment of the privacy protection intensity, such as personalized recommender systems, dynamic data distribution, and so on.

2.3.3. Continuous Differential Privacy

Continuous differential privacy mostly protects the privacy of real-time data streams by continuously adding noise, which can better adapt to the dynamic changes of data and ensure a consistent level of privacy protection during continuous data release. It is suitable for data streams that require real-time processing, such as financial transaction data, network traffic, etc. It can adapt to the dynamic changes of data and maintain the continuity of privacy protection, and can be applied to real-time data analysis, IoT data streams, and online learning.

2.3.4. Local Differential Privacy

Localized differential privacy is a method to protect privacy directly at the data source (client), ensuring that the data has been privacy-protected before it is uploaded to the server, so that even if the data collector is not trustworthy, the user's privacy cannot be inferred. It is particularly suitable for distributed scenarios such as Internet of Things (IoT) and mobile devices, and is widely used in privacy-preserving search engines, advertising systems, and mobile application data analysis.

The definition of LDP can be formalized as follows: for any two adjacent datasets D and D′ and any possible output y, an algorithm A is called a (ε,δ)-local differential privacy algorithm if it satisfies the following condition: $pr[A(x)\in y] \leq e^{\varepsilon}Pr[A(x') \in y] + \delta$

where x and x′ are individual data points in D and D′ , respectively.

3. Analysis of the Current State of Research

3.1. Chronological Research Development

The concept of differential privacy was first formally introduced by Cynthia Dwork in 2006, by adding noise to the data, making it impossible for an attacker to infer the existence or non-existence of individual data records from the query results [4]. In 2007, Dwork et al. further proposed a Laplace noise-based differential privacy mechanism, which preserves the statistical properties of the data while protecting privacy.

Subsequently, differential privacy techniques have made important advances in several areas. In 2010, research began to focus on the application of differential privacy to real datasets [5]. In 2011, the (ε, δ)-differential privacy variant was proposed for scenarios that require trade-offs between privacy and utility [6, 7]. In 2012, the gradient descent algorithm for differential privacy optimization (DP -SGD) was proposed to protect the privacy of machine learning training data.

In 2014, Google introduced RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response), which is a technique based on Local Differential Privacy (LDP) that enables the privacy of data directly on the client side [8]. In 2015, the research began to explore the application of differential privacy to Internet of Things (IoT) data, proposing differential privacy mechanisms applicable to IoT devices to protect sensor data privacy.

In 2016, breakthroughs were made in the application of differential privacy in deep learning, and optimized deep learning algorithms were proposed, which were widely used in tasks such as image classification and natural language processing [9]. In 2017, a privacy budget management mechanism was proposed, which reasonably allocates the privacy budget to maximize the utility of the data, and the differential privacy technique has been widely used in the fields of data mining, data publishing, and machine learning [10]. In 2018, adaptive differential privacy mechanism was proposed to maximize data output, dynamically adjusting the amount of noise according to data sensitivity, which improves the adaptability in dynamic data environments [11].

In 2020, privacy-preserving model release mechanisms are proposed to protect the privacy of model parameters, which are widely used in healthcare data analysis and financial risk prediction [12]. In 2021, research begins to focus on the convergence of differential privacy and blockchain technology, and explores how to further enhance the privacy protection of data through blockchain technology [13]. In 2022, differential privacy technology makes advances in the processing of real-time data streams, and proposes mechanisms applicable to real-time data streams to protect the privacy of dynamic data, which are applied to network traffic analysis and financial transaction monitoring [14]. processing, and propose mechanisms applicable to real-time data streams to protect dynamic data privacy, which can be applied to

network traffic analysis and financial transaction monitoring [14].

In 2023, the noise addition mechanism of differential privacy is optimized to reduce the impact of noise on data utility and improve the applicability in large-scale data analysis [15]. In 2024, differential privacy makes a breakthrough in distributed machine learning, and a mechanism applicable to distributed machine learning is proposed to protect the training data privacy, which is widely used in federated learning [16].

Differential privacy technology has made breakthroughs in several key areas such as machine learning, federated learning, and real-time data stream processing. With the continuous optimization and expansion of related technologies, differential privacy is expected to play a more important role in the privacy protection field.

3.2. Domestic and International Research Status

3.2.1. Status of Domestic Research

In China, the core objective of protecting commercial data is to fully explore its value while realizing the balance of interests of multiple parties and promoting the formation of a win-win situation. The formulation of a legal framework that suits the development law of commercial data is crucial to promoting the healthy circulation of commercial data in the market and maintaining a market environment of fair competition. At present, policymaking in the field of anti-unfair competition is relatively more centralized. From the informatics perspective, a study has proposed a malicious behavior detection scheme based on machine learning to effectively deal with the problem of data risk identification; from the management perspective, a comprehensive scheme for commercial data risk assessment and management has been proposed. In the field of mobile Internet, differential privacy techniques have been widely recognized. Data collection, sharing and analysis techniques based on local differential privacy models have been widely used, and large Internet companies, such as Google, Apple and Microsoft, have applied localized differential privacy techniques to their actual products. In addition, Uber and Volcano Engine use centralized differential privacy technology in their data platforms to ensure the security of users' private data.

Zhao Peipei pointed out from the perspective of jurisprudence that the protection of commercial data faces many dilemmas. Due to the diverse forms of commercial data, some of the data are difficult to be recognized as trade secrets, thus failing to obtain corresponding protection [17]. Qingqing Ye et al. provide an in-depth analysis of the research status of localized differential privacy in China. In the untrustworthy third-party data collector scenario, the localized differential privacy technique emerges. The technique has strong privacy protection capability, which not only can fully consider the background knowledge of the attacker and quantify the degree of privacy protection, but also can perturb the data locally to effectively resist the privacy attacks of untrustworthy third parties and provide all-round protection for sensitive information [18].

3.2.2. Current Status of Foreign Research

Internationally, protection measures for commercial data focus on key areas such as privacy protection, data control, data transmission and government regulation. In Europe and the United States, in the era of big data, research on data protection for trade secrets and other data has been relatively perfect and mature.

Arijit Sarkar and other scholars delve into data protection mechanisms based on differential privacy through case studies of many foreign industries, and they propose that many large international companies such as Google and Apple are widely adopting the centralized model approach of Federated Learning, which usually collects data from users' mobile devices and processes the data in the cloud to build centralized machine learning models, and differential privacy techniques are used to quantify the level of privacy protection in databases. Meanwhile, the future optimization direction of differential privacy mentioned in the paper is to minimize the fluctuation of data accuracy due to noise protection, and utilize other data algorithms such as Laplace, so as to achieve a better balance between privacy protection and data utility [19].

Scholars such as Krishna Mohan Pd Shrivastva have pointed out that Big Data is a huge collection of structured, unstructured and semi-structured data, and its complexity requires a privacy-preserving mechanism that is independent of the underlying data structure and record type. Differential privacy techniques have attracted attention for their compatibility with various types of datasets. Incorporating differential privacy into commercial data security protection is a practical strategy, but there are limitations to this approach, i.e., it may have some impact on the integrity of the data in some cases. When commercial data does not require high accuracy, differential privacy techniques can be directly applied; however, if data accuracy is critical, existing algorithms need to be optimized or other algorithms need to be introduced to combine with them to achieve better privacy protection [20].

4. Business Data Protection

4.1. Theoretical Research Protection

Differential privacy builds a strong privacy protection framework that makes it difficult to restore the privacy details of individual users even if the attacker has additional background information. This is highly compatible with the three core features of big data (large data volume, fast processing speed, and diverse data types), and can effectively address various privacy challenges in big data scenarios. Differential privacy technology not only protects privacy but also preserves the analytic value of data, which is widely used in the fields of commercial data analytics and machine learning, and helps enterprises meet the requirements of data protection regulations, as exemplified by the European Union's General Data Protection Regulation (GDPR) and the U.S.'s California Consumer Privacy Act (CCPA) [21] [22].

4.2. Business Data Security Case Studies

To balance user data privacy protection and model performance in machine learning model training, Microsoft developed the differential privacy optimized gradient descent algorithm (DP-SGD).

By adding noise (e.g., Laplace noise or Gaussian noise) during the gradient update process and accurately regulating the noise intensity, the algorithm effectively prevents the attacker from inferring individual user's private information from the noisy gradient, successfully realizing the balance between privacy protection and model performance and utility, so that the user's sensitive information can be adequately protected, even in the event of a data leakage. This

privacy protection mechanism is widely used in Microsoft's Azure machine learning platform and Cortana voice assistant.

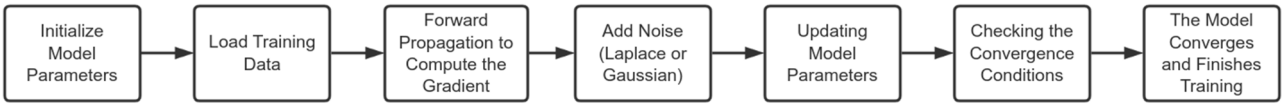

Figure 1. DP-SGD algorithm implementation flow

Microsoft has developed the SmartNoise platform, an open-source differential privacy tool designed to help organizations and researchers perform data analysis and machine learning while protecting privacy.

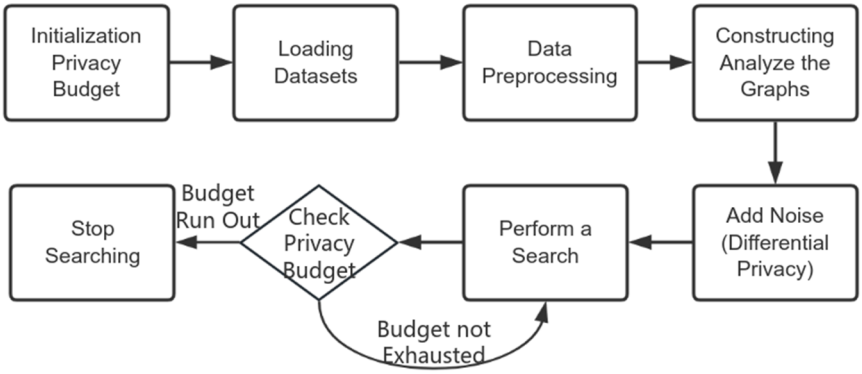

Figure 2. SmartNoise Platform Privacy Protection Flow

The SmartNoise platform includes core libraries that provide differential privacy algorithms and mechanisms, support for multiple programming languages (e.g., Python and R), and ensure memory security; a SQL data access layer that allows users to build analytic graphs through the SQL language, and SQL data access support for multiple database engines (e.g., SQL Server, PostgreSQL, Spark, etc.) Layer can ensure privacy budget management for each query privacy protection through privacy budget (epsilon) management. Microsoft supports the generation of differential privacy synthetic data for machine learning and deep learning tasks through the SmartNoise platform, such as the use of techniques such as Differential Privacy Generative Adversarial Networks (DP-GANs) and improved PATE-GANs, which can be used to generating high-quality synthetic data while preserving privacy [23].

5. Summary and Outlook

5.1. Discussion and Analysis

Although differential privacy performs well in balancing privacy protection and data utility, it still faces many challenges in practical applications. While adding noise protects privacy, it may also significantly degrade data utility, especially in high-precision analysis tasks where differential privacy struggles to meet data accuracy requirements. The allocation and management of privacy budgets also need to be precisely computed, otherwise it may lead to insufficient privacy protection or data utility degradation. The management of privacy budgets becomes especially complex in multiple query or data distribution scenarios, and needs to be dynamically adjusted to accommodate different query requirements.

The implementation of differential privacy techniques requires some computational overhead, especially in terms of noise addition and privacy budget management. For large-scale datasets and complex queries, this computational overhead may significantly increase and thus affect the efficiency of the system. Differential privacy is mainly suitable for statistical analysis and machine learning model training, while its applicability may be limited for certain application scenarios that require precise data, such as financial transactions.

Although differential privacy provides strong privacy protection, it cannot completely prevent privacy leakage in some cases especially when the attacker has a large amount of auxiliary information. In summary, differential privacy techniques still need to be further optimized to cope with more complex attack scenarios and higher data utility requirements.

5.2. Future Research Directions

Differential privacy protects privacy by adding noise, but this process may weaken the utility of the data. Current research has struck a balance between privacy and utility but there is still room for improvement. Future research directions can focus more on developing smarter noise addition mechanisms that dynamically adjust the amount of noise according to the sensitivity of data and the complexity of the query, so as to further optimize the balance between privacy protection and data utility, establish more accurate privacy-utility trade-off models, and find the optimal privacy parameters through mathematical optimization methods in order to maximize the utility of data [24].

Although differential privacy technology provides strong privacy protection in theory, users' trust and acceptance of privacy protection technology is still low. Enterprise companies may consider following up with more user privacy education activities to improve users' understanding and trust in differential privacy technology, and relevant technicians to develop more transparent differential privacy mechanisms to show users the specific process and effect of privacy protection.

Significant progress has been made in the application of

differential privacy in business data protection, but the application in some specific domains (e.g., healthcare, finance) still needs to be further explored Differential privacy in financial risk assessment and transactional data analytics to protect users' financial privacy.

References

[1] Zhao, Y., & Yang, M. (2023). A review of advances in differential privacy research. Computer Science, 50(4), 265-276.

[2] Wang, Y., Yu, Y., Yuan, Q., et al. (2024). Application of differential privacy in data protection. ZTE Technology. Retrieved from https://link.cnki.net/ urlid/34.1228. TN. 2024 1012.1529.002.

[3] Li, X., Hui, L., Fenghua, L., et al. (2018). An overview of differential privacy. Journal of Information Security, 3(5), 92-104.

[4] Dwork, C. (2006). Differential privacy. In International Colloquium on Automata, Languages, and Programming (pp. 1-12). Springer, Berlin, Heidelberg.

[5] Dwork, C. (2010). Differential privacy in new settings. In ACM-SIAM Symposium on Discrete Algorithms (pp. 174-183). SODA 2010, Austin, Texas, USA.

[6] Dwork, C. (2011). A firm foundation for private data analysis. ACM.

[7] Dwork, C. (2011). The promise of differential privacy: A tutorial on algorithmic techniques. In Foundations of Computer Science (pp. 1-2). IEEE.

[8] Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. arXiv.org. Ithaca, NY: Cornell University Library.

[9] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. arXiv.org. Ithaca, NY: Cornell University Library.

[10] McMahan, H. B., Moore, E., Ramage, D., et al. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics.

[11] Wang, Y. X., Fienberg, S. E., & Smola, A. (2018). Privacy-preserving data sharing via probabilistic modeling. arXiv preprint arXiv:1802.04918.

[12] Wei, K., Li, J., Ding, M., et al. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security.

[13] Liu, R., Cao, Y., Chen, H., et al. (2021). Flame: Differentially private federated learning in the shuffle model. In AAAI Conference on Artificial Intelligence.

[14] Cheng, A., Wang, P., Jian, C., et al. (2022). Differentially private federated learning with local regularization and sparsification. IEEE Transactions on Mobile Computing.

[15] Shi, Y., Wei, K., Li, S., et al. (2023). Towards the flatter landscape and better generalization in federated learning under client-level differential privacy. arXiv preprint arXiv:2305. 00873.

[16] Zhang, X., Chen, X., Hong, M., et al. (2024). Understanding clipping for federated learning: Convergence and client-level differential privacy. In International Conference on Machine Learning.

[17] Zhao, P. (2024). Research on commercial data protection model in the era of big data. Henan Science and Technology, 51(05), 132-136.

[18] Ye, Q., Meng, X., Zhu, M., et al. (2018). A review of localized differential privacy research. Journal of Software, 29(07), 1981-2005.

[19] Sarkar, A., Sharma, A., Gill, A., & Thakur, P. (2023). A differential privacy-based system for efficiently protecting data privacy. IEEE, 1399.

[20] Shrivastva, K. M. P., Rizvi, M. A., & Singh, S. (2014). Big data privacy based on differential privacy: A hope for big data. IEEE, 776.

[21] Zhao, J., Yue, X., Feng, C., et al. (2022). An overview of data privacy security based on the General Data Protection Regulation. Computer Research and Development, 59(10), 2130-2163.

[22] Wu, S. (2018). Personal information protection in the California Consumer Privacy Act of 2018. Information Security and Communications Privacy, 12, 83-100.

[23] Ji, X. (2022). Data protection impact assessment highlights - Compilation and analysis based on Microsoft's DPIA report. WeChat Public Platform. Retrieved from http://mp. weixin. qq.com/s?__biz=MzA4MjAyNzk0NQ==&mid=2649472358 &idx=1&sn=07cdd6e27193573a3dc35ff872ce84a7.

[24] CSDN. (2021). Problematic code implementation of local sensitivity. Differential Privacy Code Implementation Series (VIII). Retrieved from https://blog.csdn.net/ qq_41691212/ article/ details/121772608.