Journal of Electronics and Information Science (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2024.090209
ISSN 2371-9524 Vol. 9 Num. 2

Privacy Protection in Information and Communication Technology Applications Based on Big Data

Dongyuan Ge[1,a,*], Qianyi Fang[2,b], Qi Han[1,c]

[1]*State Grid Heilongjiang Information & Telecommunication Company Ltd, Harbin, Heilongjiang, China*
[2]*Faculty of Science, Northeast Forrstry University, Harbin, Heilongjiang, China*
[a]*2914756@qq.com*, [b]*2834300729@qq.com*, [c]*389025758@qq.com*
[*]*Corresponding author*

Keywords: Big Data, Information and Communication Technology Applications, Privacy Protection, Data Encryption

Abstract: With the advent of the information age, big data has become an important force driving the rapid development of information and communication technology, while also bringing huge challenges to user privacy and security. This article is based on various privacy protection methods such as data anonymity, differential privacy, and ciphertext. Through quantitative and empirical research, the application of big data in information communication was deeply explored. Therefore, this article focused on how to efficiently apply data in network security while ensuring its availability and accuracy. A new data anonymity model was designed to address the issue of data anonymity, which can maximize the protection of user privacy and the availability of user data; the impact of different noise addition strategies on the accuracy of data analysis in differential security systems was studied; on this basis, a study was conducted on the combination of multiple encryption methods to improve their security in storage and transmission. In the latest data security assessment, data encryption led with a high score of 9.5, demonstrating its outstanding performance in protecting data security. This study provided a new approach and method for China's privacy protection in the field of information and communication, which can enhance the country's credibility in the field of public information and communication.

1. Introduction

In the context of rapid development in today's society, big data has become the main driving force for promoting innovation in information and communication technology. With the collection and processing of personal and corporate data on the Internet, data privacy and security issues under the Internet environment have received more and more attention. Although various methods and approaches have been proposed to address the above issues, there are still many problems and deficiencies that urgently require in-depth research and improvement. At the same time, this article also evaluates and optimizes existing algorithms, providing new ideas for privacy protection research in the big data environment, in order to improve the effectiveness of privacy protection in

big data, and enhance the credibility and acceptability of big data.

This article adopts a combination of qualitative and quantitative research methods to explore the impact of big data on privacy in the ICT (Information and Communication Technology) environment, and evaluates existing privacy protection measures. This article conducts research on privacy and security issues in data collection, processing, and storage, and explores their impact on user trust and technology adoption. By comparing relevant research results, it is hoped to discover the advantages and disadvantages of various research methods and provide reference for future research. On this basis, this article also explores technological innovations in encryption, anonymous processing, and user permission management.

The architecture of this article consists of three parts: firstly, by systematically sorting and analyzing the privacy protection issues involved in existing big data and information communication technologies, the research purpose and requirements of this article are determined. Secondly, empirical analysis and evaluation of existing privacy protection methods are conducted to identify key technical and policy deficiencies. Finally, based on the research results, new ideas and methods are provided for user privacy protection in the network environment to address increasingly complex information security issues. This article conducts in-depth and detailed research on the above-mentioned issues, and combines them with practical applications to ensure the scientific and practical nature of the research results, thereby proposing practical and feasible solutions for privacy protection issues in China's ICT environment.

2. Related Work

With the rapid development of big data and information communication technology, privacy protection in the application of information communication technology provides new ideas and methods to ensure the security and credibility of user data. Chen Qiqi explored the research progress and inspiration of information and communication technology in children's geography [1]. Wang Guidong studied that the Internet has improved the total factor productivity of manufacturing enterprises from the perspective of the ICT productivity paradox [2]. Zhang Yan studied the anti-interference grounding design scheme for electronic information and communication engineering [3]. Chen Congbo studied the impact of information and communication technology on urban innovation output [4]. Zhang Yurong studied the response of the information and communication technology industry to necessary patent litigation risks based on fuzzy set qualitative comparative analysis [5]. Although various privacy protection methods have been proposed in existing research, there is a lack of systematic evaluation of their practical application effectiveness and applicability in real data environments. In addition, previous studies have overlooked the active participation and feedback of users on privacy protection, which limits the comprehensiveness and effectiveness of this method.

The research results on privacy protection in the application of information and communication technology can provide users with a reliable and highly adaptive operable privacy protection mechanism. Han Guoying explored the impact of information and communication technology on promoting the upgrading of rural industrial structure [6]. Shukla S studied the role of information and communication technology in reducing agricultural supply chain risks in India [7]. Kramer V studied when and how the orientation of information and communication technology affects the role pressure of sales personnel [8]. Abraham J studied tablets as a social innovation technology that promotes development through information and communication technology [9]. Naik B J studied the attitudes of farmers towards information and communication technology tools [10]. There is still a lack of practical and effective evaluation of privacy protection in existing research, especially in large-scale practical applications where there is a lack of sufficient experimental and data support,

resulting in many theories and methods being difficult to implement.

3. Methods

3.1 Design and Data Collection

This article conducts research from the following aspects: identifying the main data types and sources suitable for big data privacy protection under information and communication technology, and conducting in-depth research with multiple major communication companies and network service providers. The collected data mainly includes: network usage records, location information, interaction information, transaction information, etc. Before collection, the data is anonymous to ensure the security of the study.

Information entropy is used to measure the uncertainty or amount of information in a dataset and is a way to assess the risk of privacy breaches:

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) \tag{1}$$

Among them, $H(X)$ is the entropy of the random variable X, and $P(x_i)$ is the probability of the variable taking a certain value.

3.2 Data Processing and Privacy Protection Technology

To ensure the privacy of personal information involved in the research process, a data desensitization based method is adopted. Firstly, the user's identity is anonymized by replacing or eliminating their identity information such as name, address, and phone number. The use of differential privacy algorithm to add noise to the data ensures that the results of data analysis cannot be used for inferring user identity. A role-based data access control method is adopted, and strict restrictions are imposed to allow only authorized researchers to use it.

Differential privacy protects personal privacy by adding noise, which is defined as:

$$\Pr[\mathcal{A}(D) \in S] \leq e^{\epsilon}\Pr[\mathcal{A}(D') \in S] + \delta \tag{2}$$

Among them, $\mathcal{A}$ is a random algorithm; D and D' are adjacent datasets; S is all possible subsets of the algorithm output; ϵ (privacy loss parameter) and δ (upper bound of small probability events) control the probability of privacy leakage.

3.3 Construction of Big Data Analysis Framework

This article uses methods such as machine learning and data mining to construct a big data analysis framework and evaluate privacy protection behavior in network environments [11]. Its main content includes: data preprocessing, feature extraction, model training, and evaluation. Due to its ability to efficiently process massive amounts of data and perform real-time analysis, Apache Spark is used as a processing platform. In the learning process, methods such as decision trees, random forests, and neural networks are selected to effectively handle the characteristics of nonlinear big data.

3.4 Experimental Design and Evaluation

This article evaluates various privacy protection methods through a series of tests, including two parts: simulation environment and real applications. In a simulation environment, a virtual user database is constructed and simulated to verify its security performance. In practical applications,

several partners are selected for experimentation, and the actual application effect and user feedback of this method are monitored. The evaluation criteria mainly include: privacy protection intensity, data processing efficiency, and user satisfaction [12].

If a dataset satisfies the k-anonymity condition, then at least $k-1$ other data records on all identifiable attributes are indistinguishable. The following formula can be used to represent:

$$\forall i, |\{j \mid \text{rec}(i) = \text{rec}(j)\}| \geq k \tag{3}$$

Among them, $\text{rec}(i)$ represents the identifiable attribute of the i-th record in the dataset.

3.5 Optimization and Implementation of Privacy Protection Policies

On this basis, this article designs a series of privacy protection strategies, including improvements in data encryption methods, new anonymous technologies, and dynamic control of data access. The above technologies strictly comply with relevant regulations and industrial standards. On this basis, this article also tracks the results of policy implementation to ensure that user privacy protection can better respond to the ever-changing technology and market.

For an encryption algorithm, its security strength is usually related to the length n of the key, and its anti cracking ability can be approximately described using the following formula:

$$S = 2^n \tag{4}$$

Among them, S is the average number of attempts required to perform a successful attack, and n is the key length.

4. Results and Discussion

4.1 Practical Application Conditions of Privacy Protection Policies

This study aims to conduct in-depth evaluations of privacy protection strategies in information and communication technology, especially in big data applications. Therefore, multiple typical environments are selected for experiments, including commercial data centers, cloud computing platforms, and mobile communication networks. These environments are chosen because they are widely used for processing and storing large amounts of user data, such as location information, browsing history, purchasing behavior, and social media activities. Before conducting any experimental operations, all data must undergo strict desensitization processing to ensure the safety of the experiment and comply with current privacy protection regulations.

In the evaluation of data security and efficiency, the following four indicators are crucial:

Privacy leakage risk: Information entropy is a method of measuring data uncertainty, and a decrease in entropy after leakage indicates an increase in data predictability, thus increasing the risk of leakage.

Processing efficiency: This indicator is evaluated by measuring the time and data throughput required for the system to complete the same data processing task. Time efficiency and high throughput are key factors in measuring the superiority of system performance.

Data integrity: Whether the data has been damaged or distorted during processing is verified by comparing the consistency of the data before and after processing. This comparison helps confirm the security of data during transmission or storage.

User satisfaction: Through regular survey questionnaires, feedback on the acceptance and satisfaction of privacy protection measures is collected from users, in order to evaluate the effectiveness of the measures and the trust of users.

4.2 Experimental Results

(1) Processing time and throughput under different experimental numbers

Experiment numbers 1-2, 3-4, 5-6 have data volumes of 1GB, 5GB, and 10GB, respectively. Experiment numbers 1, 3, and 5 represent strategy A, while 2, 4, and 6 represent strategy B. Strategy A supports higher parallelism, while strategy B has poorer parallelism. The processing time and throughput under different experimental numbers are shown in Figure 1.

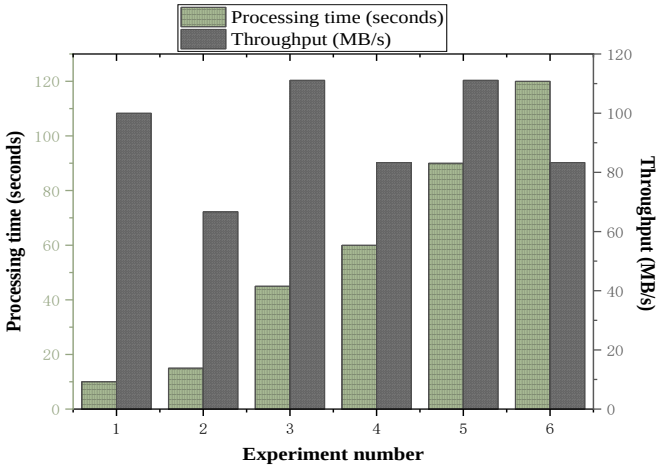

Figure 1: Processing time and throughput under different experimental numbers

In the first and second experiments, it is observed that when the processing time is extended from 10 seconds to 15 seconds, the data throughput rate decreases from 100 MB/s to 66.67 MB/s, revealing a possible trend of reduced throughput rate due to prolonged processing time. However, this trend is not without exceptions. From the third to the sixth experiment, although the processing time has increased, the changes in throughput rate are not very consistent. Especially in the third experiment, although the processing time increases to 45 seconds, the throughput unexpectedly increases to 111.11 MB/s. In contrast, the processing time of the fourth experiment is further extended to 60 seconds, but the throughput rate decreases to 83.33 MB/s. The fifth and sixth experiments also show a trend of this fluctuation.

(2) Data integrity

In order to evaluate data integrity, it is determined whether the data has been damaged during processing by comparing the degree of consistency of the data before and after processing. The raw data is shown in Table 1.

Table 1: Raw data

Serial number	Name	Age	Salary (yuan)
1	Zhang San	28	5000
2	Li Si	35	7000
3	Wang Wu	42	8500
4	Zhao Liu	25	5500

It is assumed that a simple salary adjustment project is initiated with the goal of increasing the salary of each employee by 10%. The adjusted new salary data is detailed in the table below. This

increase not only reflects the company's recognition of employees' efforts, but also rewards their contributions. The processed data is shown in Table 2.

Table 2: Processed data

Serial number	Name	Age	Salary (yuan)
1	Zhang San	28	5500
2	Li Si	35	7700
3	Wang Wu	42	9350
4	Zhao Liu	25	6050

To ensure the integrity of data processing, a detailed comparison is made between the data before and after processing. The research results of Tables 1 and 2 are as follows:

Employee Zhang San's serial number is 1, and his original salary is 5000 yuan. After a 10% increase, the salary has been adjusted to 5500 yuan, accurately matching the expected growth rate.

Employee Li Si's serial number is 2. The original salary is 7000 yuan, but after processing, it increases to 7700 yuan, which also meets the 10% promotion standard.

Employee Wang Wu, serial number 3, is adjusted from 8500 yuan to 9350 yuan, with an accurate increase.

Finally, employee Zhao Liu, with serial number 4, has his salary increased from 5500 yuan to 6050 yuan, with a consistent increase.

The consistency check of each data item shows that no data damage or distortion occurred during this processing. Therefore, it can be considered that the integrity of the data has been fully maintained in this update.

(3) Encryption and decryption times under different encryption techniques

The encryption and decryption times under different encryption techniques are shown in Figure 2.

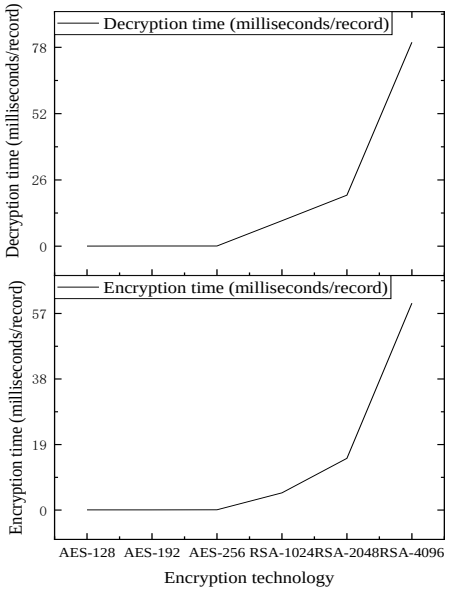

Figure 2: Encryption and decryption times under different encryption techniques

In the field of encryption technology, AES (Advanced Encryption Standard) and RSA (Rivest Shamir Adleman) exhibit significant performance differences (the following numbers represent key length). AES, as a symmetric encryption method, only takes milliseconds to encrypt and decrypt, which is much faster than RSA. Compared to the rapid execution of AES, RSA, as an asymmetric encryption algorithm, extends the encryption and decryption time from a few milliseconds to tens of milliseconds. The fundamental reason for this difference lies in their working mechanism: AES uses the same key for encryption and decryption, with a simple and efficient process; RSA relies on a pair of keys - public and private - and its encryption and decryption processes involve more complex mathematical operations, resulting in longer processing times.

When discussing encryption algorithms AES and RSA, the impact of key length on performance is particularly significant. The AES algorithm has improved the encryption and decryption speed from 128 bits to 256 bits, but the growth rate is small, indicating that AES still has good computational efficiency under different key length conditions.

(4) Evaluation of privacy protection effectiveness

The evaluation results of privacy protection effectiveness are shown in Figure 3.

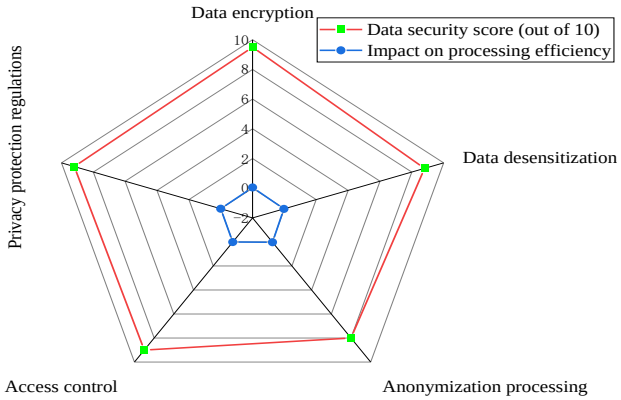

Figure 3: Evaluation results of privacy protection effectiveness

Figure 3 shows the evaluation of the impact on privacy protection. The latest data security evaluation shows that data encryption performs very well in data protection, with a score of 9.5. However, due to the need to replace or remove sensitive fields, the processing efficiency of this method has decreased by 3%. As for privacy protection regulations, as they mainly involve the standardization of policies and processes rather than direct technical operations, they have no direct impact on processing efficiency, and efficiency changes remain at 0%.

5. Conclusion

This article conducted in-depth research on privacy protection in information and communication technology in the big data environment, focusing on privacy protection technologies such as data anonymity, differential privacy, and ciphertext. Meanwhile, the performance of this method in practical applications was evaluated through a series of experiments. Experimental results have shown that combining anonymity with differential privacy can not only ensure data security but also reduce the risk of privacy leakage. In addition, a satisfaction survey on big data found that if users clearly recognize its existence and functions, their trust in big data can significantly increase. The limitation of this study is that the current experimental scale and scope cannot fully cover all possible scenarios. In addition, some advanced privacy protection technologies have higher requirements for computing resources and more complex management

methods in certain application scenarios, so their applications are limited in resource limited environments. On this basis, the scope and depth of the experiment should be further expanded, especially in the fields of mobile computing and cloud computing. At the same time, new technologies such as machine learning and artificial intelligence can be utilized to automatically adjust and optimize privacy protection mechanisms to better meet the growing data security requirements. In addition, in subsequent research, the subjective perception and acceptance of users should be fully utilized to ensure that the provided technological products conform to human nature and practicality.

References

[1] Chen Qiqi, Feng Jian. The research progress and inspiration of information and communication technology (ICT) in children's geography [J]. Progress in Geographic Science, 2024, 43 (1): 147-162.

[2] Wang Guidong, Yang Delin. Does the Internet improve the total factor productivity of manufacturing enterprises—Further exploration of the paradox of information and communication technology productivity in China [J]. Statistical Research, 2023, 40 (6): 63-76.

[3] Zhang Yan. Anti interference grounding design for electronic information and communication engineering [J]. Information Record Materials, 2023, 24 (6): 185-187.

[4] Chen Congbo, Ye Azhong, Chen Juan. The impact of information and communication technology on urban innovation output [J]. Economic Geography, 2022, 42 (10): 92-99.

[5] Zhang Yurong, Zhu Jingyu. A Study on the Risk of Necessary Patent Litigation for NPE Standards in the Information and Communication Technology Industry - Based on Fuzzy Set Qualitative Comparative Analysis [J]. Intelligence Journal, 2022, 41 (2): 104-111.

[6] Han Guoying, Liu Tongshan. Can information and communication technology promote the upgrading of rural industrial structure? Evidence from the Third National Agricultural Census in Beijing. Research on Agricultural Modernization, 2023, 44 (1): 108-118.

[7] Shukla S , Kapoor R , Gupta N ,et al.Role of information and communication technology in mitigating risks in Indian agricultural supply chains[J].Supply Chain Management: An International Journal, 2023, 28(3):544-558.

[8] Kramer V, Krafft M.When and how information and communication technology orientation affects salespeople's role stress: the interplay of salesperson characteristics and environmental complexity[J].European Journal of Marketing, 2023, 57(3):659-682.

[9] Abraham J, Geobey S. Digital elixir? The Aakash tablet as a social innovation in information and communication technology for development (ICT4D)[J].Social Enterprise Journal, 2022, 18(2):288-305.

[10] Naik B J , Rao B M , Rambabu P ,et al.Attitude of Farmers towards Information and Communication Technology (Ict) Tools[J].Current Journal of Applied Science and Technology, 2021, 39(43):72-81.

[11] Gupta M K, Chandra P. A comprehensive survey of data mining[J]. International Journal of Information Technology, 2020, 12(4): 1243-1257.

[12] Ni, Z. & Ouyang, T. On the Use of ICT by Enterprises to Activate Industrial Heritage: A Case Study of Taoxichuan in China. Journal of Organizational and End User Computing, 2023, 35(3), 1-21.