symmetry

MDPI

Article

Sensitive Data Privacy Protection of Carrier in Intelligent Logistics System

Zhengyi Yao [1], Liang Tan [1,2,3,*], Junhao Yi [4], Luxia Fu [1], Zhuang Zhang [1], Xinghong Tan [1], Jingxue Xie [1], Kun She [5], Peng Yang [1], Wanjing Wu [1], Danlian Ye [1] and Ziyuan Yu [1]

[1] College of Computer Science, Sichuan Normal University, Chengdu 610066, China; 20201301016@stu.sicnu.edu.cn (Z.Y.); 20191301007@stu.sicnu.edu.cn (L.F.); 20201303010@stu.sicnu.edu.cn (Z.Z.); 20201301018@stu.sicnu.edu.cn (X.T.); 20201391019@stu.sicnu.edu.cn (J.X.); 20231393001@stu.sicnu.edu.cn (P.Y.); 20231303010@stu.sicnu.edu.cn (W.W.); yedaly@stu.sicnu.edu.cn (D.Y.); 20231393041@stu.sicnu.edu.cn (Z.Y.)
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100864, China
[3] Institute of Cyberspace Security, University of Electronic Science and Technology of China, Chengdu 610054, China
[4] Software Engineering Department, Chengdu Jincheng College, Chengdu 611731, China; yijunhao@cdjcc.edu.cn
[5] College of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; kun@uestc.edu.cn
* Correspondence: jkxy_tl@sicnu.edu.cn

Abstract: An intelligent logistics system is a production system based on the Internet of Things (IoT), and the logistics information of humans has a high degree of privacy. However, the current intelligent logistics system only protects the privacy of shippers and consignees, without any privacy protection for carriers, which will not only cause carriers' privacy leakage but also indirectly or directly affect the logistics efficiency. It is particularly worth noting that solving this problem requires one to consider the balance between privacy protection and operational visibility. So, the local privacy protection algorithm ϵ-L_LDP for carriers' multidimensional numerical sensitive data and ϵ-LT_LDP for carrier location sensitive data are proposed. For ϵ-L_LDP, firstly, a personalized and locally differentiated privacy budgeting approach is used. Then, the multidimensional data personalization perturbation mechanism algorithm L-PM is designed. Finally, the multidimensional data are perturbed using L-PM. For ϵ-LT_LDP, firstly, the location area is matrix-partitioned and quadtree indexed, and the location data are indexed according to the quadtree to obtain the geographic location code in which it is located. Secondly, the personalized random response perturbation algorithm L-RR for location trajectory data is also designed. Finally, the L-RR algorithm is used to implement the perturbation of geolocation-encoded data. Experiments are conducted using real and simulated datasets, the results show that the ϵ-L_LDP algorithm and ϵ-LT_LDP algorithm can better protect the privacy information of carriers and ensure the availability of carrier data during the logistics process. This effectively meets the balance between the privacy protection and operational visibility of the intelligent logistics system.

Keywords: data privacy; location privacy; smart logistics platform; privacy protection

Citation: Yao, Z.; Tan, L.; Yi, J.; Fu, L.; Zhang, Z.; Tan, X.; Xie, J.; She, K.; Yang, P.; Wu, W.; Ye, D.; Yu, Z. Sensitive Data Privacy Protection of Carrier in Intelligent Logistics System. *Symmetry* **2024**, *16*, 68. https://doi.org/10.3390/sym16010068

Academic Editor: Antonio Palacios

Received: 22 October 2023
Revised: 1 December 2023
Accepted: 14 December 2023
Published: 4 January 2024

1. Introduction

Intelligent logistics is a modern logistics model based on the Internet of Things (IoT) technology [1,2], which is an important branch of the development and innovation of IoT technology [3]. Intelligent logistics is characterized by informationization, intelligence, and automation [1,2,4,5], and is an important link in the development of the digital economy of enterprises. Intelligent logistics has become one of the most important directions in the development of the logistics industry [1–3] and has become a focus of attention in academia

and industry. The intelligent logistics platform is the core and carrier of intelligent logistics, which is built based on an information-sharing mechanism and has the characteristics of reasonable distribution of a series of logistics resources such as transportation, storage, packaging, unloading, and distribution, etc. It can break the information barrier between consumers, logistics enterprises (carriers), and suppliers, and provides better and more stable logistics services for users of all levels. At the same time, the flexible allocation of logistics resources according to market demand in the intelligent logistics platform can solve the problem of resource tension in the peak season and resource waste in the off-season of logistics to a certain extent, to truly realize the intelligence, automation, and informationization of logistics [6]. M. Yang et al. [7] designed an intelligent logistics platform containing major applications such as e-commerce, self-service collection, and delivery, big data analysis, path location, and distribution optimization to accelerate the construction of a symbiotic and win–win logistics ecosystem and the benign development of the ICT industry. E. Mathew [8] proposed a conceptual model of an intelligent logistics distribution platform, analyzed the value creation process of the stakeholders of the intelligent logistics platform, and pointed out that it would be a new solution for logistics and distribution problems.

There are four object entities in the intelligent logistics platform, which are people, objects, fields, and transportation equipment, among which people–object entities include shippers, consignees, and carriers. A shipper, also known as a "cargo owner", is entrusted to be the carrier to transport goods (luggage or parcels) and pay the freight costs of social organizations or individuals. The consignee is the social organization or individual who has the right to pick up the goods (luggage or parcels). The carrier is the person or company who enters into a contract of the carriage of goods with the shipper in their own name or entrusts others in one's own name. In the process of logistics transportation, the main responsibility of the carrier is to ensure that the transported goods reach their destination on time and safely and are finally delivered to the consignee.

In the intelligent logistics platform, the sensitive information of the human and the object entity has a high degree of privacy. Currently, a large number of research works exist in academia and industry on logistics information privacy protection schemes for the sensitive data of shippers and recipients, but without any privacy protection for carrier logistics process data. For example, S. Liu and J. Wang [9] proposed a courier system based on (near-field communication) technology and encryption protocol, which uses the secure NFC tags instead of traditional orders to protect the personal information of consignees and shippers. X. Lin et al. [10] proposed a blockchain-based logistics privacy protection scheme to ensure the auditability and traceability of personal and logistic information. Among them, personal information is mainly targeted at shippers and consignees, and the privacy protection of logistics process data of carriers is relatively lacking. Therefore, this will not only cause the carrier's personal privacy information to be leaked but also indirectly or directly affect the logistics efficiency of the whole intelligent logistics platform. D. Waters [11] pointed out that the leakage of precise information about the logistics transportation process of logistics transporters (carriers) will increase the threat to the security of logistics goods, as well as cause a decrease in the quality of logistics services. For example, malicious personnel, competing companies, etc., obtain detailed logistics process data of carriers, and through techniques such as machine data mining and data analysis, obtain private information about the driving habits, physical health, and home address of carriers, which can be used to obstruct and hijack the designated logistics activity services with precision. T. Sativell and R. Sabar [12] studied the threats to high-value goods faced by logistics companies where the information leakage of carrier equipment or carriers is one of the reasons for the precise hijacking of cargoes. An elaborate logistics truck robbery scheme reported by the National Broadcasting Company (NBC) [13] hijacked a United Parcel Service Inc. (USP) truck for six hours, resulting in the loss of a quarter of the packages. One of the important reasons why the hijackers were able to carry out the robbery plan was that they gained access to the logistics truck driver's transportation routes and driving habits.

It is worth noting that an intelligent logistics system is a production system, and solving this problem requires considering the balance between privacy protection and operational visibility. If traditional privacy protection technologies such as encryption [14], multi-party secure computing [15], or access control [16,17] are adopted, this will inevitably affect the availability of carrier data in the logistics system. Encryption is the most commonly used privacy protection technology. Carrier information is transmitted, stored, and shared after encryption, and only an entity with a key can be decrypted and accessed. Although encryption protects privacy, data cannot be directly counted, processed, or processed, it increases the complexity of data usage. The goal of multi-party secure computing is a computing model designed by the participating parties to protect their privacy. Its application scenario does not match the protection of carrier information privacy in logistics systems, so multi-party secure computing cannot be applied to protect carrier information privacy in logistics systems. Access control is the most important way to achieve privacy protection. The essence of privacy protection is to share private information with authorized entities at appropriate times and in appropriate ways. In traditional access control, permissions are formulated and implemented by system managers, and ensuring that permissions are reasonably allocated and not tampered with is the most critical issue. However, in traditional access control, not only unreasonable authorization can cause privacy leakage, but the system manager's permissions are too large and the permission data are easily tampered with, which can also lead to illegal data access.

Therefore, to address the problem of the lack of privacy protection for carrier logistics to process sensitive data, this paper proposes a local differential privacy protection scheme for carrier logistics to process sensitive data in an intelligent logistics platform. The scheme can protect the privacy of carrier users while also satisfying the personalized privacy needs of carrier users, as well as ensuring the availability of logistics process data. The research content and contributions of this paper are as follows.

1. A local differential privacy protection algorithm ϵ-L_LDP (ϵ-Logistic Local Differential Privacy, ϵ-L_LDP) for the carrier's multidimensional numerical sensitive data in the logistics process is proposed. First, a personalized local differential privacy budget method is used to introduce a carrier user's personalized privacy budget, which ensures that each carrier can modify the privacy protection budget of sensitive data according to individual or carrier company requirements. Then, the data are normalized to $[-1, 1]$ using each attribute data security domain value, introducing [0,1] uniform random variables, and designing the multidimensional data personalized perturbation mechanism algorithm L-PM on segmented perturbation mechanism PM. Then, the carrier's multidimensional numerical data are perturbed using the L-PM algorithm.
2. The local differential privacy protection algorithm ϵ-LT_LDP (ϵ-Logistic Trajectory Location Differential Privacy, ϵ-LT_LDP) for the location-based data of carriers is put forward. First, the location region is matrix partitioned and quadtree indexed, and the location data are indexed according to the quadtree to obtain the geographic location code (including area code and inner code) where it is located. Then, this paper adopts a personalized local differential privacy budget method, which introduces a personalized privacy budget for carrier users and ensures that each carrier can modify the privacy protection budget of sensitive data according to personal or company requirements. Then, the geographic location coding vector is normalized to $[-1, 1]$, and a certain probabilistic Bernoulli variable is introduced, which is determined by the privacy budget and the specific geographic location coding vector value, to realize the personalized random response algorithm L-RR for location trajectory data. Finally, the L-RR algorithm is used to realize the perturbation of geographic location coding data.
3. In this paper, the privacy of the ϵ-L_LDP algorithm and the ϵ-LT_LDP algorithm are analyzed and proved. In addition to that, simulation experiments are conducted to verify the usability of the algorithms. The evaluation criteria in the experiments

are the standard mean square error (MSE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The experimental datasets include: the simulation dataset, which is the GAUSS dataset and UNIFORM dataset; the multidimensional numerical real dataset, which comprises the BR dataset and MX dataset; and the real trajectory (location) dataset, which comprises the GPS data of more than 14,000 cabs selected from 3 August 2014 to 30 August 2014, in Chengdu city. Among them, the ϵ-L_LDP algorithm is experimentally compared with three existing algorithms, and the ϵ-LT_LDP algorithm is experimentally compared with two existing algorithms for implementation. The experimental results show that the data processed by the scheme in this paper have higher usability under the same privacy budget. The privacy protection scheme proposed in this paper not only ensures the privacy protection of the carrier logistics process but also meets the personalized privacy needs of carrier users, which will help the intelligent logistics platform provide a better and more stable logistics services.

The rest of this paper is organized as follows: Section 2 of this paper introduces the research results of the scholars in the logistics data privacy protection; Section 3 carries out the introduction of relevant privacy protection algorithm knowledge, including differential privacy, local differential privacy and location privacy, etc.; Section 4 describes the scenarios and problems of intelligent logistics platforms, as well as elaborates the design of carrier sensitive data privacy protection schemes; Section 5 reports the experimental environment of the privacy algorithm and experiments on the algorithm on multiple datasets, followed by analysis and comparison; and Section 6 concludes the full paper.

2. Related Work

In recent years, with the development of intelligent logistics, domestic and foreign researchers have made significant progress in terms of protecting the privacy of sensitive data of shippers and receivers in intelligent logistics systems. We will introduce the research results in detail from the following three perspectives: firstly, the privacy protection of sensitive data in logistics systems; secondly, the privacy protection of facial data in logistics; and thirdly, the integration of logistics systems and blockchain to enhance data privacy.

In terms of the privacy protection of the sensitive data in logistics systems, X. Liu et al. [18] combined the web technologies of cloud computing and QR codes to build a secure and trusted logistics system with the complete protection of personal privacy from shipper to consignee. The paper [19] optimizes the logistics costs while protecting user privacy information by using artificial intelligence methods, formalizing the problem as a distributed constraint optimization problem (DCOP), and using various cryptographic encryption techniques to process data in the logistics system. F. Xu et al. [20] proposed a unified privacy-preserving mechanism for participants to register and cancel strategic and business relationships; set and clear job information for cargo transportation under predefined partnerships; and update and clear tracking and tracing data associated with a given job or partnership. Q. Gao [21] proposed a secure logistics information scheme, LIP-PA, using attribute-based encryption and location-based key exchange that implemented a logistics information privacy protection scheme with location and attribute-based access control for mobile devices to provide privacy protection for personal and logistics information. All these solutions start from logistics systems and use cloud computing technology [18], artificial intelligence technology [19], privacy protection mechanisms [20], and cryptography [19,21] to improve the privacy of logistics data and ensure the privacy of sensitive user data.

In terms of the privacy protection of sensitive data on face-sheet data in logistics, W. Qian et al. [22] proposed a k-anonymity model to protect logistics information. However, the name and phone number of the consignee are printed directly on the logistics manifest, and this scheme only protects some personal information among users. To solve the problem of the paper [22], H. Qi et al. [23] proposed a logistics courier management system based on encrypted QR codes; in this system, encrypted QR codes are used to store all

information about goods, and real-time logistics information about goods is automatically updated via General Packet Radio Service (GPRS) or Wi-Fi, and an improved genetic algorithm is used to provide carriers with an optimal delivery route. X. Zhan et al. [24] proposed a logistics information privacy protection system (LIPPS) based on encrypted QR codes, which stores the cipher text in the QR codes through a segmented encryption method, thus realizing a different level of authorization mechanism to decrypt the corresponding information to complete logistics business operations. W. Yan et al. [25] proposed a QR code and information-hiding-based logistics system privacy protection scheme, which uses information-hiding technology to embed user privacy information into QR codes on courier face slips and designs a JPEG image steganography algorithm for QR codes to complete the access permission control of privacy information. Both the paper [22] with a k-anonymity model and the papers [23–25] based on QR code technology aimed to solve the plain-text privacy problem of paper logistics sheets, and the paper [23] used cryptographic encryption technology to encrypt the e-logistics sheets within the logistics system, which avoid the user's personal privacy from sensitive data leakage due to internal employees.

In terms of improving the privacy protection by combining logistics systems with blockchain, the paper [26] presents the current research results of blockchain technology in supply chain, logistics, and transport management and suggests that the application of blockchain technology in the logistics industry is one of the main themes of future research. E. Tijan et al. [27] studied decentralized data storage represented by blockchain technology and its use in sustainable logistics and supply chain management as a way to improve the security and privacy of data storage in logistics systems. N. Rožman et al. [28] proposed an approach to integrate blockchain and IoT technologies into modern supply chains, where blockchain technology is used not only to write down agreements and conduct transactions but also as a trusted public list of services and information, ensuring their security and trustworthiness. H. Yi [29] proposed a secure logistics technology using blockchain to protect the privacy of individuals. The main contribution is the proposed blockchain model for logistics, which uses the security and anonymity of blockchain technology to achieve the effect of logistics data privacy protection. H. Duan et al. [30] proposed a privacy-preserving scheme to improve the logistics business, using a combination of blockchain and anonymous authentication to achieve the control and management of users' access rights to private data, and to divide and store data from different services, and the blockchain nodes receiving transactions are used as transit nodes to synchronize data within the chain, ultimately achieving the effect of data privacy protection. H. Li et al. [31] proposed a blockchain-assisted secure storage scheme for logistics data, in which data generation and aggregation, session establishment, record encryption, and storage, and an efficient consensus mechanism are introduced to improve the efficiency of the consensus process as a way to prevent the logistics data privacy leakage. Papers [27,31] used the decentralized and distributed storage characteristics of blockchain to ensure the security and privacy of logistics data, while the papers [28–30] used the trustworthiness, tamper-evident, and anonymity of blockchain to ensure the security and privacy of logistics data.

In summary, an overall analysis of the research work on the privacy protection of sensitive logistics data is shown in Table 1. From Table 1, it can be obtained that the existing research on the privacy protection of logistics data mainly focuses on shippers and consignees, but there is less research on the privacy protection of sensitive data for protecting carriers. The carrier is an important part of the human–object entity in the intelligent logistics platform. The security of sensitive data privacy information of carriers directly or indirectly affects the efficiency of the whole intelligent logistics platform. Therefore, the privacy protection scheme applicable to the sensitive data of carriers proposed in this paper has certain practical significance.

Table 1. Comparison table of related work.

Researches	Shippers	Consignees	Carriers	Characteristics
Paper [15]	✓	✓	×	Optimizing logistics costs with the use of artificial intelligence and encryption technology to protect users' private information.
Paper [16]	✓	✓	×	Update and clear tracking and tracing data associated with a given job or partnership under a predefined partnership.
Paper [17]	✓	✓	×	Attribute-based encryption and location-based key exchange for mobile device access control.
Paper [18]	✓	×	×	The k-anonymity model protects the logistic information, but the consignee's information is still printed on the logistic face-sheet.
Paper [19]	✓	✓	×	Privacy protection for e-logistics sheets within the logistics system using QR code technology and encryption.
Papers [20,21]	✓	✓	×	The use of QR code technology to solve the problem of privacy leakage caused by the explicit printing of paper logistics sheets.
Papers [23,27]	✓	✓	×	Ensure the security and privacy of logistics data with the decentralized and distributed storage features of blockchain.
Papers [24,25]	✓	✓	×	Ensure the security and privacy of logistics data using the trustworthiness, tamper-evident, and anonymity of blockchain.
Paper [26]	✓	✓	×	The combination of blockchain and anonymous authentication enables users to control and manage access to private data.

The table summarizes the characteristics of the relevant research work and its focus on the privacy and security of shippers, consignees, and carriers.

3. Preliminaries

This chapter introduces the relevant basic knowledge used in the scheme, providing a theoretical basis for the algorithms in the scheme. Among them, the definitions of privacy, differential privacy, local differential privacy, and location trajectory privacy are provided, as well as detailed introductions including noise mechanisms and random response mechanisms.

3.1. Differential Privacy

3.1.1. Definition of Differential Privacy

Privacy refers to sensitive information that entities such as individuals and organizations do not wish to be known externally [32] and thus express themselves selectively. Examples include personal financial information, medical records, travel records, shopping order data, etc. In recent years, many privacy-preserving approaches based on k-anonymity [33] and division (e.g., l-diversity [34], t-closeness [35], $(\alpha, k)-$anonymity [36]) have been proposed, but new attack models (e.g., composition attacks [37], foreground knowledge attacks [38], etc.) pose a serious threat to the effectiveness of these privacy-preserving methods. In 2006, Dwork [39] proposed a new definition of privacy protection for addressing privacy breaches in databases, which is called differential privacy (DP). The differential privacy as a privacy-preserving model is strictly defined in terms of the strength of privacy protection, i.e., deleting and adding any one record will not affect the query result. Also, an extremely strict attack model is defined in the differential privacy algorithm, which does not care how much background knowledge the attacker possesses and proves its privacy mathematically rigorously.

Definition 1. *Differential privacy [39,40]. D and D' are two adjacent datasets that differ by at most one tuple, i.e., $\Delta(D, D') = 1$, the randomized algorithm M: $D \to R^d$, where $Ran(M)$ is all possible outputs of M on D and D', and any subset S of $Ran(M)$; M satisfies ϵ-differential privacy if it satisfies the following inequality (1).*

$$Pr[M(D) \in s] \leq e^{\epsilon} \times Pr[M(D') \in s] \tag{1}$$

In inequality (1), Pr denotes the probability of the risk of privacy being disclosed; ϵ is the privacy budget, which defines the level of privacy protection and reflects the level of privacy protection that the algorithm M can provide. The smaller the value of ϵ, the higher the level of privacy protection. From Definition 1, it can be seen that the differential privacy protection mechanism limits the effect of any one record on the output of the algorithm M. This definition theoretically ensures that algorithm M satisfies ϵ-differential privacy, whilst the implementation a differential privacy algorithm requires the introduction of a noise mechanism.

3.1.2. Noise Mechanisms

There are many ways to implement differential privacy protection mechanisms, and one of the most common methods is to use noise addition mechanisms to achieve differential privacy. There are two representative noise mechanisms, namely the Laplace mechanism [41] for numerical types of data and the exponential mechanism [42] for discrete types of data. In differential privacy-preserving algorithms, different noise mechanisms are chosen depending on the type of data, and the required noise size is closely related to the global sensitive.

Definition 2. *Sensitivity [41]. For any function $f : D \rightarrow R^d$, the global sensitivity of the function f can be expressed by Equation (2).*

$$\Delta f = \max_{D,D'} \|f(D) - f(D')\|_p \tag{2}$$

In Equation (2), D and D' differ by at most one record; R denotes the real space of the mapping; d denotes the query dimension of the function f; and p denotes the metric. Δf uses the L_p distance, which is usually measured using the 1-order paradigm distance (L1).

Theorem 1. *Laplace mechanism [41]. For any function $f : D \rightarrow R^d$, an algorithm M satisfies ϵ-differential privacy if the output of the algorithm M satisfies the following Equation (3).*

$$M(D) = f(D) + < Lap_1(\Delta f/\epsilon), \ldots, Lap_d(\Delta f/\epsilon) > \tag{3}$$

In Equation (3): $Lap_i(\Delta f/\epsilon)(1 \leq i \leq d)$ are mutually independent Laplace variables, where the amount of noise is proportional to Δf and inversely proportional to ϵ. The larger the added noise, the greater the global sensitivity of the algorithm M.

It easily follows from Equation (3) that the $i(1 \leq i \leq d)$ element in $M(D)$ is caused by the Laplace noise, the standard absolute error is shown by Equation (4), and the variance is shown by Equation (5).

$$error^i_{abs} = E|M(D)_i - f(D)_i| = E|Lap(\frac{\Delta f}{\epsilon}) = \frac{\sqrt{2}\Delta f}{\epsilon} \tag{4}$$

$$error^i_{var} = E(M(D)_i - f(D)_i)^2 = \frac{2(\Delta f)^2}{\epsilon^2} \tag{5}$$

The exponential mechanism [42] is commonly used in non-numerical types of data to randomly select data to achieve differential privacy protection. The key in the exponential mechanism is designing a reasonable score function $u(D, t)(t \in O)$, where D denotes the input dataset and t denotes the output term selected in the output domain O. Different score functions need to be designed for different statistical queries or statistical tasks.

Theorem 2. *Exponential mechanism [42]. Let $u(D,t)$ be the score function corresponding to the input dataset D and the output t. The following Equation (6) is satisfied by the algorithm M. The algorithm M satisfies ϵ-differential privacy.*

$$M(D,u) = \{t : |Pr[t \in O] \propto exp(\frac{\epsilon u(D,t)}{2\Delta u})\} \tag{6}$$

In Equation (6), Δu denotes the global sensitivity of the score function $u(D,t)$. From Equation (6), it can be seen that the higher the score of the score function, the higher the probability of being selected for output.

3.2. Local Differential Privacy

3.2.1. Definition of Local Differential Privacy

The differential privacy introduced in the previous section is centralized differential privacy, which is a protection model that assumes a trusted third party; however, in real-world application scenarios, absolutely trusted third parties do not exist. With the rapid development of crowdsourcing technology, IoT application devices are widely used. A large number of mobile devices perform the function of data collection, uploading data to untrustworthy third parties, posing a serious threat to the privacy of individuals. At this time, local differential privacy (LDP) [43,44] came into being. Localized differential privacy is a proposed data privacy protection framework based on differential privacy protection techniques, and local differential privacy provides stronger privacy guarantees than centralized differential privacy. LDP does not require a trusted third party and directly adds noise to private data locally to protect personal information from disclosure. At the same time, third-party collectors can analyze and infer the statistical value of the group's data.

Under the local differential privacy protection model [45], the possibility of third-party data collectors stealing and disclosing user privacy is fully considered. In the local differential privacy protection model, each user (collector) first perturbs their sensitive data to ensure the privacy of their data, then transmits the perturbed data to the third-party data collector, and finally, the data collector performs statistical analysis through the data to obtain the value of the data information of the data collector community. In [44], a local differential privacy protection model is defined as follows.

Definition 3. *Local differential privacy [43,44]. Given n users, each corresponding to a record, and given a privacy algorithm M, if the following inequality (7) is satisfied, then M satisfies ϵ- local differential privacy.*

$$Pr[M(t) = t^*] \leq e^{\epsilon} \times Pr[M(t') = t^*] \tag{7}$$

In inequality (7), Pr denotes the probability of the risk of privacy being disclosed; algorithm M has a definition domain of $Dom(M)$ and a value domain of $Ran(M)$; t and $t'(t,t' \in Dom(M))$ are any two records, and $t^*(t^* \in Ran(M))$ is the same output of any two records t and t' on algorithm M. ϵ is the privacy budget: the smaller the value of ϵ, the higher the degree of privacy protection.

In the local differential privacy preservation model, users perform data perturbation in their local area to accomplish privacy preservation. However, different users have different privacy requirements, promoting the emergence of personalized local differential privacy [46].

Definition 4. *Personalized local differential privacy [46]. The privacy setting preference of the user $u_i(1 \leq i \leq n)$ is (τ, ϵ_i), τ is the security domain, and ϵ_i is the personalized privacy budget of the user u_i, for any two inputs t, $t'(t,t' \in \tau)$, and any outputs $t^*(t^* \in Dom(M))$; Algorithm M satisfies the ϵ-personalized local differential privacy if it satisfies the following Equation (8):*

$$Pr[M(t) = t^*] \leq MAX(e^{\epsilon_i}) \times Pr[M(t') = t^*] \tag{8}$$

F.D. McSherry [47] proposed that differential privacy has composition properties, namely sequential composition [47] and parallel composition [47]. Sequential composition ensures that the privacy budget can be distributed in different steps in the privacy method, while parallel composition ensures that the privacy of the differential privacy algorithm is satisfied over the disjoint subsets of the dataset. From the differential privacy of Definition 1 and the local differential privacy of Definition 3, it follows that differential privacy is defined on adjacent datasets, and the local differential privacy is defined on two of the records, which do not undergo any transformation in their privacy guaranteed form. Thus, local differential privacy also satisfies the sequential composition and parallel composition [45,47]. The specific sequential composition and parallel composition are described below.

Nature 1: Sequential composition [45,47]. Given a dataset D and n algorithms $M_i(1 \leq i \leq n)$, assuming that algorithm $M_i(1 \leq i \leq n)$ satisfies ϵ-local differential privacy, then the sequence combination of algorithms $M_i(1 \leq i \leq n)$ on D satisfies ϵ-local differential privacy, where $\epsilon = \Sigma_{i=1}^{n} \epsilon_i$.

Nature 2: Parallel composition [45,47]. Given a dataset D, partitioned into n mutually disjoint subsets $D = \{D_1, D_2, \ldots, D_n\}$, and algorithm $M_i(1 \leq i \leq n)$ satisfying ϵ_i-local differential privacy on the dataset D_i, the composition operation of algorithm $M_i(1 \leq i \leq n)$ on $D = \{D_1, D_2, \ldots, D_n\}$ still satisfies ϵ_i-local differential privacy.

3.2.2. Randomized Response Mechanism

The dominant mechanism for local differential privacy implementation is the randomized response (RR) mechanism [43,44]. In 1965, S. L. Warner [48] proposed the randomized response (RR) technique. Later, we called it W-RR. The main idea of W-RR is to use uncertain responses to sensitive questions to achieve the effect of privacy protection for the original data. The randomized response technique requires two main steps, namely the perturbation statistics of the data and the data correction.

Because of the importance of the randomized response technique for local differential privacy, this subsection details the data perturbations and statistical corrections for the randomized response technique. To introduce the randomized response technique in concrete terms, we introduce a problematic application scenario in which the proportion of people who smoke is surveyed. However, for most people, the question of whether or not they smoke is a sensitive one, and many users are reluctant to reveal private information about whether or not they smoke. Suppose there are n_1 users and the true percentage of smokers is π, but we do not know that and need to count the percentage $\widehat{\pi}$. So a sensitive question is asked: "Do you smoke?". Each user responds to this, and the ith user answers with "Yes" or "No", assuming that with the help of a random device (e.g., a non-uniform coin toss, Bernoulli distribution, etc.), their probability of answering the true answer is p, with a $1 - p$ probability of answering the opposite answer. First, perturbation statistics are performed. The statistical value of the number of smokers can be obtained by counting the responses of n users with the help of the perturbation method provided by the random device. In the hypothesis that the number of people who answered "Yes" is n_2, the number of people who answered "No" is $n_1 - n_2$. The proportion of users who answered "Yes" is shown in Equation (9), and the proportion of users who answered "No" is shown in Equation (10).

$$Pr(X_i = \text{“Yes”}) = \pi p + (1 - \pi)(1 - p) \tag{9}$$

$$Pr(X_i = \text{“No”}) = (1 - \pi)p + \pi(1 - p) \tag{10}$$

Obviously, the proportions expressed in Equations (9) and (10) are not unbiased estimates of the true proportions. Therefore, the next step is to correct the statistical results by constructing a likelihood function, as in the following Equation (11):

$$L = [\pi + (1 - \pi)(1 - p)]^{n_1}[(1 - \pi)p + \pi(1 - p)]^{n - n_1} \tag{11}$$

The maximum likelihood estimate of $\widehat{\pi}$ for π is calculated as shown in Equation (12) below:

$$\widehat{\pi} = \frac{p-1}{2p-1} + \frac{n_2}{(2p-1)n_1} \tag{12}$$

Calculating the mathematical expectation of $\widehat{\pi}$ yields that $\widehat{\pi}$ is an unbiased estimate of the true distribution π. The result of the calculation is shown in Equation (13).

$$\begin{aligned} E(\widehat{\pi}) &= \frac{1}{2p-1}[p-1+\frac{1}{n}\sum_{i=1}^{n} X_i] \\ &= \frac{1}{2p-1}[p-1+\pi p+(1-\pi)(1-p)] \\ &= \pi \end{aligned} \tag{13}$$

In summary, the proportion of true smokers π can be obtained based on the total number of people n with a perturbation probability of p. If it is made to satisfy ϵ-local differential privacy, its privacy budget can be set [43,44], as shown in (14). For example, p = 0.75, and its privacy budget ϵ is the level of privacy protection of ln3.

$$\epsilon = \ln\frac{\mathrm{p}}{1-\mathrm{p}} \tag{14}$$

The randomized response technique W-RR [48] is only applicable to respond to discrete data with only two values, but not with other data types. Therefore, a large number of scholars have improved on W-RR. For example, the stochastic response methods include RAPPOR [49], S-Hist [50], k-RR [51], and O-RR [52] for multi-valued discrete data, and MeanEst [53,54] and Harmony-mean [55] for continuous data.

3.3. Location and Trajectory Privacy

Definition 5. *Location points. The trajectory position point is represented by $l_i =< X_i, Y_i, timestamp >$, l_i denotes the position point of user u_i at the timestamp moment, X_i denotes the longitude of the position point, and Y_i denotes the latitude of the position point.*

Definition 6. *Trajectory. A trajectory t is a sequence formed by a series of position points in chronological order, which is denoted as Equation (15):*

$$t = l_1 \rightarrow l_2 \rightarrow \ldots \rightarrow l_a \tag{15}$$

The a in Equation (15) represents the number of position points in trajectory t.

Definition 7. *Trajectory dataset. A trajectory dataset T is a collection of combinations of a series of trajectories, which is represented in Equation (16):*

$$T = \{t_1, t_2, \ldots, t_b\} \tag{16}$$

where b denotes the number of location points in the trajectory dataset.

Definition 8. *Geo-indistinguishability (GI) [56]. A mapping mechanism K satisfies the definition of geographical indistinguishability determined by the parameter on the set of locations X when and only when the two location points x and x' satisfy the following Equation (17):*

$$K(x,z) \leq e^{\epsilon d(x,x')} K(x',z) \tag{17}$$

In Equation (17), x represents the set of locations in the region, $x, x', z \in X$, and $d(\,)$ represents the Euclidean distance between two location points. In the case of the privacy protection of location data, the risk of privacy leakage is limited to a defined range, which in

turn is determined by the Euclidean distance $d(\ )$ and parameters ϵ. This privacy protection mechanism is the geo-indistinguishability algorithm.

Definition 9. *Location local differential privacy. Derived from Definition 3, assume that there are n location points, each of which has a corresponding record. Given a privacy algorithm M, i.e., a definition domain Dom(M) and a value domain Ran(M), if any two positional records l and* l' $(l, l' \in Dom(M))$ *satisfy the same output result* $l^*(l^* \in Ran(M))$*, and satisfy the following inequality (18):*

$$Pr[M(l) = l^*] \leq e^{\epsilon} \times Pr[M(l' = l^*)] \tag{18}$$

where ϵ *is the privacy budget. Satisfying inequality (18), the algorithm M satisfies the* ϵ*-location local differential privacy.*

4. Local Differential Privacy Protection Solutions for Carrier Sensitive Data

4.1. Scenario and Problem Description

The smart logistics scenario in this paper is shown in Figure 1. Figure 1 includes a platform and three processes, one of which refers to the smart logistics platform, which is the carrier and core of the smart logistics with the characteristics of informatization, automation intelligence, etc. The smart logistics platform collects and integrates all the logistics data as a way of providing users with better quality logistics services. Therefore, the smart logistics platform needs to collect real-time logistics process data, with shippers and recipients communicating using mobile devices (smartphones, ipad, etc.), logistics stores and warehouses communicating using the Internet, and collection devices on carrier equipment transmitting data using wireless communications (4G, 5G). An intelligent logistics platform using logistics process data can realize the intelligent and automated scheduling of logistics resources. The three processes refer to "consignment process, transportation process, receiving process", in the lower part of Figure 1 from left to right to correspond to this, the core of the three processes of logistics transportation. Among them, the consignment process is the shipper, which will need to consign the goods submitted to the logistics company, the logistics company for unified distribution and scheduling to the warehouse transit. The carrying process refers to the logistics company by the transport requirements of the goods, scheduling the corresponding logistics transportation resources for logistics transportation, and the consignment of goods transported to the destination of the warehouse. The receiving process means that the logistics company distributes the goods from the warehouse at the destination and delivers the consigned goods safely to the consignee through the logistics distribution personnel. In Figure 1, the state data of the transportation process is extremely important; in the transportation process, one or more intelligent sensors are installed on each carrier equipment, and the state data of the carrier equipment, such as load, position, speed, and temperature, are collected by the intelligent sensors and transmitted to the cloud platform through the wireless network.

This paper analyzes several smart logistics systems, including the open source smart logistics system and the smart logistics platform system of Chengdu Fankonghui Technology Co., Ltd. (Chengdu, China) which both collect and monitor real-time data on the status of the logistics process, especially the status of the transportation process. In the TCS cooperative transportation system provided by Fankonghui Technology Co., LTD., the status data of the logistics process is tested in real-time, as shown in Figure 2. Figure 2 shows a screenshot of a real logistics monitoring system, where the Chinese (non English) expressions cannot be modified. The meanings of these expressions include carrier vehicle information and logistics status information. In Figure 2a,b, the status data of the logistics process contain the basic information of the carrier and the transportation equipment, as well as the status data of the logistics process (load, position, speed, temperature, light, etc.). In the TCS cooperative transportation system, these logistics process state data are utilized to provide intelligent functions for logistics services, such as real-time monitoring, historical track view, and alarm management.

In the logistics process data, basic information about the carrier and carrier equipment, including vehicle owner information, license plate number, length, axle, vehicle fuel type, etc., is protected for data privacy using symmetric encryption algorithms. Among them, symmetric encryption algorithms can choose encryption algorithms such as SM4 and AES. The intelligent logistics platform binds the symmetric encryption key to the computers of the carrier equipment. The symmetric encryption key is generated by the intelligent logistics platform and embedded in the computers of the carrier equipment during production. Therefore, this plan will focus on the real-time data privacy protection issues collected by intelligent sensors in logistics process status data, including numerical and positional data. The analysis is as follows.

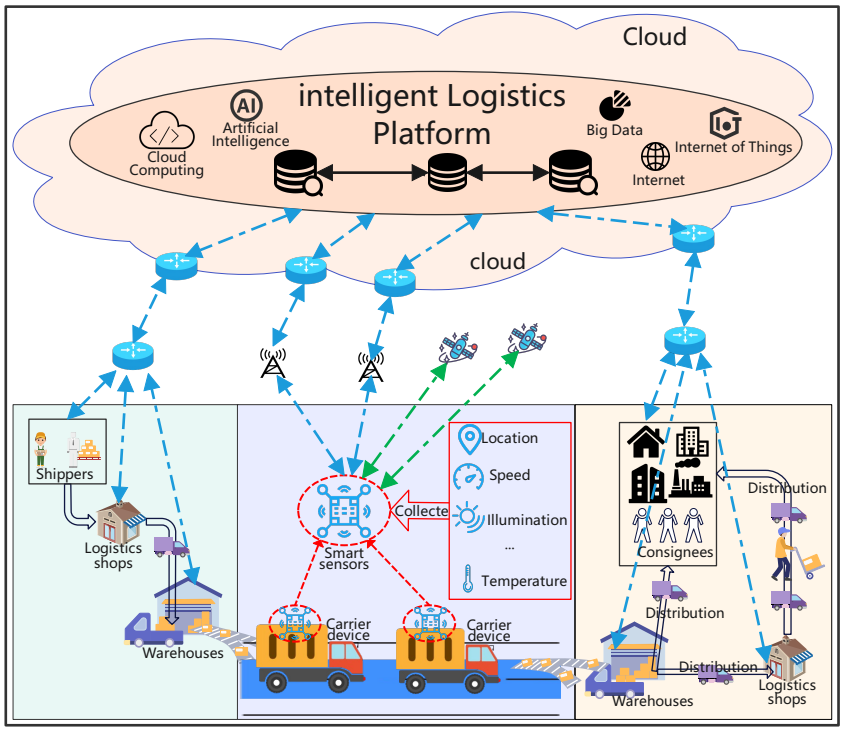

Figure 1. Model of the logistics process and its data collection in an intelligent logistics platform.

Figure 2. Diagram of the logistics process state collection data of the TCS collaborative transport system.

Question 1: Numerical data leakage problems for carriers. The intelligent logistics platform collects the state data of the logistics process in real-time, which includes data on speed, load, temperature, humidity, light, etc. The intelligent collection device installed on the carrier equipment collects these data and transmits the data to the intelligent logistics platform through the wireless communication network. However, the data such as speed,

load, temperature, humidity, light, etc. were directly transmitted to the smart logistics platform without any privacy protection processing, as shown in Figure 2 these data were collected in the smart logistics system.

Question 2: Location data leakage problems for carriers. The smart logistics platform collects real-time data on the status of the logistics process, including location data, which are collected by the intelligent collection device installed on the carrier equipment and transmitted to the smart logistics platform through the wireless communication network. However, these location data are directly transmitted to the smart logistics platform without any privacy protection processing, as shown in Figure 2, and the location data are collected in the smart logistics system.

As can be summarized from questions 1 and 2, it can be concluded that the intelligent logistics platform does not provide privacy protection for the carrier's status data (such as location, speed, load, temperature, humidity, light, etc.), which poses a risk of private data leakage for carriers. Malicious people obtaining logistics process data through untrustworthy third-party data servers, using data mining, big data, machine learning, and other methods, can obtain sensitive information such as the carrier's driving habits, behavioral preferences, home address, religious beliefs, and even health status, which can be used to precisely hinder or hijack the specified logistics activity services, which will directly or indirectly affect the logistics efficiency of the entire intelligent logistics platform, as well as cause the leakage of sensitive information about the carrier. Therefore, the privacy protection objectives are not only protecting the carrier's sensitive information based on the privacy requirements of the carrier company, but also satisfying the privacy preferences of the individual carrier, the statistical characteristics of the multidimensional numerical data available to the intelligent logistics system, as well as the availability of the carrier's trajectory routes. In non-absolutely trusted third-party data servers, the central differential privacy model is collected in a way that is vulnerable to attacks. Therefore, this paper proposes a local differential privacy protection solution for carrier-sensitive data, including a local differential privacy protection algorithm for a carrier's multidimensional numerical data and a local differential privacy protection algorithm for carrier location-based data.

In the logistics industry, transport modes include land, sea, and air transport, of which land transport is one of the most common and dominant modes in the global transport industry. Therefore, the local differential privacy protection solution for carrier-sensitive data designed in this paper focuses on the logistics process of land transport modes. First, we define the carrier set in detail as follows.

Definition 10. *Carrier. The carrier is the person who enters into a contract of the transport of goods with the shipper in their own name or entrusts others in one's own name. In the process of logistics transportation, the main responsibility of the carrier is to ensure that the transported goods reach their destination on time and safely and are finally delivered to the consignee. Carrier users are made up of individual transporters and carrier companies registered on the intelligent logistics platform. The basic information of individual transport operators includes name, ID number, contact details, vehicle information (license plate number, vehicle fuel type, vehicle load), and length of employment. The basic information about the carrier company includes the company name, the company address, the type of vehicle and its number, the number of registered transporters, and the geographical area of the company. Basic information about a carrier's transport staff includes the carrier to which they belong, name, ID number, contact details, vehicle information (license plate number, vehicle fuel type, vehicle load), and length of employment. Assuming a total of n carrier users are registered on the intelligent logistics platform, a carrier set* $U = (u_1, u_2, \ldots, u_n)$*, where* $u_i (1 \leq i \leq n)$ *can be either an individual in the transport industry or a carrier company.*

4.2. Local Differential Privacy Preservation Algorithm for Carrier Multidimensional Numerical Data ϵ-L_LDP

4.2.1. The General Idea of ϵ-L_LDP

To design a local differential privacy protection algorithm for the carrier multidimensional numerical data, it is first necessary to analyze the logistics process data in the data collection device, and then extract the numerical data related to carrier privacy to obtain the definition of the carrier multidimensional numerical data as follows.

Definition 11. *Multidimensional numerical data of carriers. In the multidimensional set of numerical attributes $A = \{A_1, A_2, \ldots, A_{10}\}$, numerical attribute A_1 represents the speed of the carrier; numerical attribute A_2 represents the load of the carrier; numerical attribute A_3 represents the temperature in the carrier's load compartment; numerical attribute A_4 represents the light in the carrier's load compartment; numerical attribute A_5 represents the number of times that the carrier's load compartment has been hit; numerical attribute A_6 represents the humidity in the carrier's device; numerical attribute A_7 represents the age of the carrier; numerical attribute A_8 represents the length of time the carrier has been in employment; and numerical attributes A_9 and A_{10} are reserved for the addition of numerical data at a later date, where the attributes are independent of each other. The set of attribute safety domains $\Gamma = \{\tau_1, \tau_2, \ldots, \tau_{10}\}$, $\tau_j (1 \leq j \leq 10)$ is the numerical safety range of attribute $A_j (1 \leq j \leq 10)$ that can be disclosed by the carrier u_i, e.g., speed attribute A_1 has a safety domain of [0,130] in km/h.*

According to the characteristics of carrier multidimensional numerical data, this paper designs a local differential privacy protection algorithm model for carrier multidimensional numerical data, which is shown in Figure 3. The local differential algorithm ϵ-L_LDP is formed in the local differential privacy protection model.

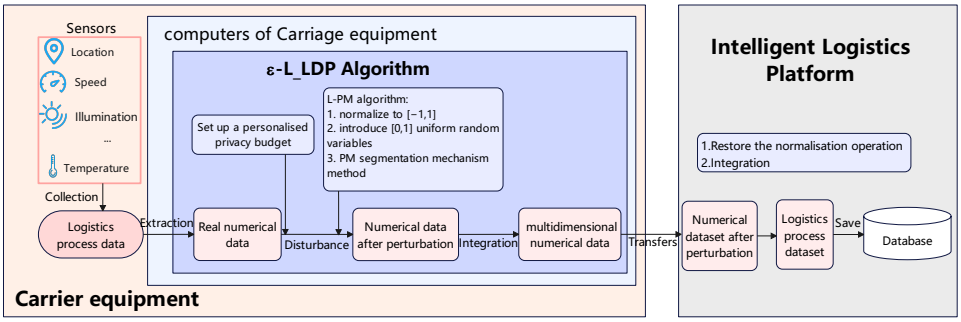

Figure 3. Diagram of the local differential privacy protection algorithm model for multidimensional numerical data for carriers.

The overall idea of the ϵ-L_LDP algorithm can be derived from Figure 3 as follows. First, a personalized local differential privacy budget approach is introduced, allowing carriers to set a personalized privacy budget for sensitive data based on their own or their carrier's privacy needs. Then, the data are normalized to $[-1, 1]$, using the individual attribute data security domain values to introduce a uniform random variable x of $[0, 1]$. Next, the PM segmentation mechanism is used to set the interval of data perturbation response values, compare the uniform random variable x with $\frac{e^{\epsilon_i/2}}{e^{\epsilon_i/2}-1}$, and use the result of this comparison to determine the interval from which the response values returned after data perturbation, yielding the numerical data personalization perturbation mechanism algorithm L-PM based on the PM segmentation mechanism. Finally, the L-PM perturbation algorithm is used to perturb the carrier's multidimensional data type data, and the perturbed data are collated and aggregated to obtain the perturbed numerical dataset, which is the local differential privacy protection algorithm for the carrier's multidimensional numerical data ϵ-L_LDP. The intelligent logistics platform receives the perturbed multidimensional numerical data, then performs the data reduction and normalization operation, and finally collates and aggregates the data into the logistics process data. Finally, the data

are collated and aggregated into a logistics process dataset. Therefore, the next section will focus on the L-PM perturbation algorithm and the ϵ-L_LDP local differencing algorithm.

4.2.2. Perturbation Algorithm L-PM

The most important thing in the local differential privacy protection model is the algorithm that provides random perturbation to the data. In this scheme, the segmented perturbation algorithm L-PM consists of three processes, as follows.

The first process is the normalization of the numerical data, where the data are normalized to $[-1, 1]$ using the individual attribute data safety domain values. The value of attribute $A_j(1 \leq j \leq 10)$ of the carrier $u_i(1 \leq i \leq n)$ is denoted by $t_i[A_j]$ and the range of the security domain of the value of attribute $A_j(1 \leq j \leq 10)$ is denoted by by τ_j. We calculate the normalization calculation factor k based on the maximum value τ_{j_max} and the minimum value τ_{j_min} of the security domain τ_j interval, and K is shown in the Equation (19).

$$k = [1 - (-1)]/(\tau_{j\_max} - \tau_{j\_min}) \tag{19}$$

The data are then normalized to the interval $[-1, 1]$ using the normalization calculation factor k; the specific steps are shown in Equation (20).

$$\begin{aligned} &Nort_i[A_j] \\ &= -1 + k(t_i[A_j] - \tau_{j_min}) \\ &= 2(t_i[A_j] - \tau_{j_min})/(\tau_{j_max} - \tau_{j\_min}) - 1. \end{aligned} \tag{20}$$

The normalization reduction operation can be deduced from the normalization process, and the specific steps are shown in the Formula (21).

$$(z_j + 1) \times (\tau_{j_max} - \tau_{j_min})/2 + \tau_{j_min} \tag{21}$$

In Formula (21), $1 \leq j \leq 10$.

The second process is the PM mechanism of the numerical data. The PM segmentation mechanism was introduced because it is more data-worthy to count the mean value of the carrier's multidimensional numerical data of the logistics process in the intelligent logistics platform. In academia, studies on local differential privacy for numerical data have focused on local differential privacy for frequency estimation. However, it is more data-worthy to count the mean value of the multidimensional numerical data of the logistics process of carriers in an intelligent logistics platform. The mean estimation algorithms used in local differential privacy protection studies are the MeanEst algorithm [53,54], Harmony-mean algorithm [55], Duchi algorithm [57], and PM algorithm [58]. The MeanEst algorithm is higher in terms of communication cost, release error, and time complexity. The harmony algorithm is a partial improvement on the MeanEst method but still suffers from the problem of large deviations of individual data from the original data. The Duchi method has an absolute value greater than 1 for both values of its perturbed output, making the variance of the perturbed values always greater than 1. In 2019, N. Wang et al. [58] proposed a PM segmentation mechanism to improve the drawbacks of the Duchi method obtain very accurate results when the privacy budget is large, so we design a personalized perturbation mechanism algorithm called L-PM for carrier multidimensional numerical data based on the PM segmentation mechanism. This is achieved by introducing a uniform random variable x of $[0, 1]$, comparing the uniform random variable x with $\frac{e^{\epsilon_i/2}}{e^{\epsilon_i/2}+1}$, and then perturbing it according to the result of the comparison to obtain the perturbed response value as a way to satisfy the privacy of the local differential privacy-preserving model.

The third process is data perturbation. The output of the data perturbation results in a response value which is a segment of continuous values $[-C, C]$, where $C = \frac{e^{\epsilon_i/2}+1}{e^{\epsilon_i/2}-1}$. Here,

the perturbed value $t_i^*[A_j]$ has a higher probability of occurring in the middle segment of the value domain and a lower probability of occurring in the values at the ends. Equation (22) is its probability density function.

$$pdf(t_i^*[A_j] = x | t_i[A_j]) = \begin{cases} p, \ x \in [l(T_i[A_j]), r(T_i[A_j])]; \\ \frac{p}{e^{\epsilon_i}}, \ x \in [-C, l(T_i[A_j])) \cup (r(T_i[A_j]), C]; \end{cases} \tag{22}$$

In Equation (22), $p = \frac{e^{\epsilon_i} - e^{\epsilon_i/2}}{2e^{\epsilon_i}+2}$, $l(T_i[A_j]) = \frac{C+1}{2} \times \mathrm{Nort_i[A_j]} - \frac{C-1}{2}$,$r(T_i[A_j]) = (l(T_i[A_j]) + C - 1$.

The perturbation mechanism algorithm L-PM for the carrier multidimensional numerical data is described in Algorithm 1.

Algorithm 1 Algorithm L-PM

Input: $t_i[A_j], \epsilon_i, \tau_j$;
Output: $t_i^*[A_j]$;
1: Use τ_j to normalize $t_i[A_j]$ to [−1,1] and obtain Nor$t_i[A_j]$;
2: Draw x uniformly at random from [0,1];
3: **if** $x < \frac{e^{\epsilon_i/2}}{(e^{\epsilon_i}+1)}$ **then**
4: Randomly draw $t_i^*[A_j]$ from $[l(T_i[A_j]), r(T_i[A_j])]$;
5: **else**
6: Randomly draw $t_i^*[A_j]$ from $[-C, l(T_i[A_j])) \bigcup (r(T_i[A_j]), C]$;
7: **end if**
8: **return** $t_i^*[A_j]$;

In Algorithm 1, there are three input parameters in algorithm L-PM, $t_i[A_j]$ is the value of attribute $A_j (1 \le j \le 10)$ for carrier $u_i (1 \le i \le n)$; ϵ_i is a restriction on the attacker's lack of ability to discriminate any two values within the security domain range τ_i, and is the personalized privacy budget of carrier $u_i (1 \le i \le n)$; τ_j is the security domain range for the value of attribute $A_j (1 \le j \le 10)$. The output of the L-PM algorithm is a data-perturbed response value $t_i^*[A_j]$. In algorithm L-PM, line 1 normalizes $t_i[A_j]$ to the interval $[-1, 1]$, lines 2–6 use the extraction of [0,1] uniform random variables x compared with $\frac{e^{\epsilon_i/2}}{e^{\epsilon_i/2}+1}$ for data segmentation perturbation purposes.

4.2.3. Local Differential Privacy Protection Algorithm ϵ-L_LDP

There are three input parameters in Algorithm 2, $T = \{t_1, t_2, \ldots, t_{10}\}$ is the set of carrier data tuples, e.g., data tuple $t_i = \{t_i[A_1], t_i[A-2], \ldots, t_i[A_{10}] (1 \le i \le n)\}$ of carrier u_i; $\Gamma = \{\tau_i, \tau_i, \ldots, \tau_{10}\}$ is the set of security domain value ranges for each attribute; and ϵ_i is the personalized privacy budget of the carrier $u_i (1 \le i \le n)$.

Algorithm 2 Algorithm ϵ-L_LDP

Input: $T = \{t_1, t_2, \ldots, t_n\}, \Gamma = \{\tau_i, \tau_2, \ldots, \tau_d\}, \epsilon_i$;
Output: $z_j (1 \le j \le d)$;
1: **for** $i = 1$ to n **do**
2: **for** $j = 1$ to d **do**
3: $t_i^*[A_j] = \mathrm{L-MP(t_i[A_j], \epsilon_i, \tau_j)}$;
4: **end for**
5: send $t_i^* = \{ t_i^*[A_1], t_i^*[A_2], \ldots, t_i^*[A_d]\}$ to Server;
6: **end for**
7: Aggregate data, calculate mean $z_j = \frac{1}{n} \sum_{i=1}^{n} t_j^*[A_j]$, and do a normalized reduction of the mean;

The local privacy protection algorithm for carrier multidimensional numerical sensitive data ϵ-L_LDP is described in Algorithm 2.

From Algorithm 2, the algorithm ϵ-L_LDP can be divided into three overall steps as below, where lines 1–6 are executed on the computer on the carrier's device, and line 7 is executed on the data server of the intelligent logistics platform.

S-1: In lines 2–4, the multidimensional numerical attributes of the carrier's logistics process are subjected to data perturbation by the L-PM algorithm.

S-2: Line 5 is where the carrier sends the perturbation values for the 10 attributes to the data server of the intelligent logistics platform.

S-3: Line 7 is where the server aggregates the perturbation data sent by all the carriers, counts the mean value of each attribute data for each carrier separately, and performs a normalized reduction operation on the mean value to finally obtain the mean value estimation result of the attribute data.

4.3. Local Differential Privacy Protection Model for Location Data of Carriers

4.3.1. The General Idea of ϵ-LT_LDP

To design a local differential privacy protection algorithm for carrier location-based data, it is first necessary to analyze the logistics process data in the data collection device and then extract the location-based data to obtain the definition of carrier location-based data as follows.

Definition 12. *Location data of the carriers. The carrier's location dataset* $L = \{< X_i, Y_i, timestamp >\}$ $(1 \leq i \leq n)$, X_i *denotes the longitude of the location data,* Y_i *denotes the latitude of the location data, timestamp denotes an instantaneous moment, and the location data security field* τ.

According to the characteristics of the carrier location-based data, we design a local differential privacy protection model for the carrier location-based data, as shown in Figure 4, in which the local differential algorithm ϵ-LT_LDP is the core.

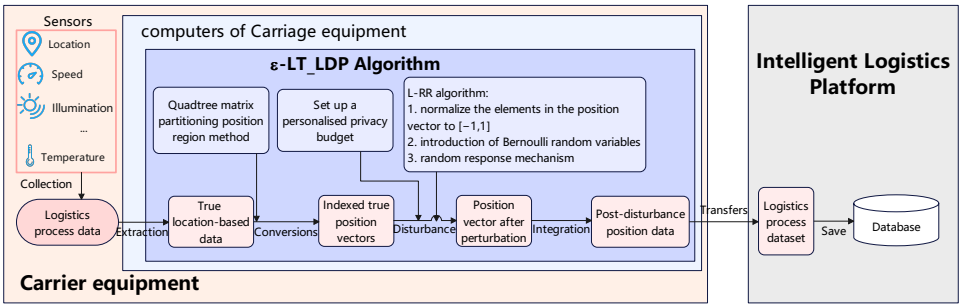

Figure 4. Diagram of the local differential privacy protection algorithm model for carrier location-based data.

The overall idea of the ϵ-LT_LDP algorithm can be obtained from Figure 4. Firstly, this paper chooses the location area index partitioning method of the quadtree matrix partitioning to implement the encoding of geolocation data and obtain the location data vector. Secondly, a personalized local differential privacy budget approach is introduced, allowing carriers to set personalized privacy budgets for sensitive data according to their own or their carrier's privacy needs. Thirdly, the elements of the geographic location encoding are normalized to $[-1, 1]$ and a Bernoulli variable with a certain probability is introduced, which perturbs the elements of the location encoding vector based on the vector values of the privacy budget and the geographic location encoding, resulting in a personalized perturbation mechanism algorithm L-RR for the location vector based on the RR random perturbation mechanism. Finally, the L-RR perturbation algorithm is invoked to perturb the carrier's geographic location encoding vector data and return the area code and response vector, i.e., the perturbed location data are obtained, and the whole process above

is the personalized local differential privacy protection algorithm for the carrier location-based data ϵ-LT_LDP. Then, the perturbed location data are received by the intelligent logistics platform, collated, and put into the logistics process dataset. Subsequently, this subsection will detail the quadtree index construction method for location data, the L-RR algorithm, and the ϵ-LT_LDP algorithm.

4.3.2. Quadtree Index Construction Method for Location Data

When performing local differential privacy on location data, the direct data perturbation of the longitude and latitude of the location data would greatly affect the usability of the location data. Therefore, we design a local differential privacy protection algorithm for carrier location data using a location region segmentation encoding followed by the data perturbation of the encoded vector. There are many studies on the geographic region segmentation in academia, the uniform grid partitioning method (UG method) [59], the adaptive grid partitioning method (AG method) [59], the Kd-tree partitioning method [60], the Quad-tree partitioning method [61], and the QLP method [62]. Among them, the method of Z. Yang et al. [62] using quad-trees and matrices can encode the geographic location of geographic regions on location data with high efficiency and rationality, and its segmentation of geographic regions is also applicable to the local differential privacy model, which uses matrices to segment the location regions, and then indexes the location regions after the matrix segmentation using quad-trees. Therefore, the geographic location coding of the location data in this paper uses the method in the paper [62]. The intelligent logistics platform performs the matrix partitioning of the geographical area as widely as possible and establishes a quad-tree index structure, which is then sent to all the smart collection devices of the logistics process. In the logistics process, real-time locations are obtained in the smart collection devices through DBS, GPS, etc. The location data are indexed according to the location area after the quad-tree matrix partitioning to obtain the code of the generalization unit in which it is located. The matrix partitioning and indexing of the quad-tree is constructed as shown in Figure 5, with the process and method proved by the paper [62].

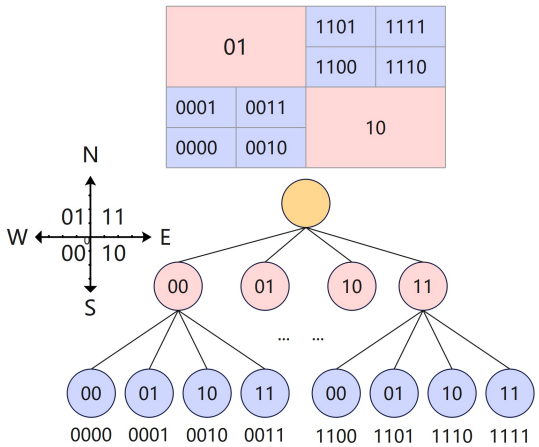

Figure 5. Grid matrix partitioning and quadtree indexing.

4.3.3. Perturbation Algorithm L-RR

The implementation of local differential privacy algorithms requires a perturbation mechanism algorithm to perturb the data in order to obtain the effect of privacy protection for a single dataset while ensuring that the statistical properties of some or all of the datasets remain largely unchanged. As seen in the previous section, it is not possible to directly perturb the latitude and longitude data, but rather encode the location dataset using grid matrix partitioning and quadtree indexing to obtain the encoded vector of location data before perturbation.

We designed the L-RR algorithm which is divided into three steps. The first step is performing a $[-1, 1]$ normalization operation on the elements of the encoded vector of location data. The second step is introducing a Bernoulli variable with a certain probability which is based on the privacy budget and the vector value of the geolocation encoding. The third step is that each element of the coded vector is perturbed according to the value of the resulting Bernoulli variable u.

The personalized perturbation mechanism algorithm L-RR for location-based data is shown in Algorithm 3.

Algorithm 3 Algorithm L-RR

Input: **B**, ϵ_i;
Output: **S**;
Initialize vector **S**;
for $j = 1$ to length(**B**) **do**
    $z = -1 + 2 \times \mathbf{B}_j$;
    Introduction of Bernoulli variables, $Pr[u = 1] = \frac{z \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1}{2e^{\epsilon_i} + 2}$;
    if $u = 1$ **then**
        $\mathbf{S}_j \leftarrow 1$;
    else
        $\mathbf{S}_j \leftarrow 0$;
    end if
end for
return S;

In Algorithm 3, there are two input parameters, one is the carrier's position vector **B**, and the other is the privacy budget ϵ_i. The output parameter is the perturbed position vector **S**.

4.3.4. Local Differential Privacy Protection Algorithm ϵ-LT_LDP

In the algorithm ϵ-LT_LDP, the location data are firstly encoded using grid matrix partitioning and quadtree indexing to obtain a location data encoding vector; then, the L-RR perturbation mechanism algorithm is invoked to perturb each element of the location data vector, and the perturbed location data (i.e., the area code vector and the perturbed response vector) are obtained.

In algorithm ϵ-LT_LDP, there are five input parameters: $< X_i, Y_i, timestamp >$ is the location data of the carrier $u_i (1 \leq i \leq i)$ at a given moment; τ is the security domain of the location data; ϵ_i is the privacy budget of the carrier $u_i (1 \leq i \leq i)$; m is the number of perturbation layers; and l is the number of generalization layers. The output is a vector of zone codes H and a vector of instantaneous randomized response S. The ϵ-LT_LDP algorithm is shown in Algorithm 4.

From Algorithm 4, the ϵ-LT_LDP algorithm can be divided into four steps, as follows:

S-1: Initialize the vectors **M** and **H** and use the latitude and longitude of the carrier location data to generate the area code **H** and the inner code **C**, where **H** and **C** form the unit code **M**.

S-2: Convert **C** into a $2h$-bit vector **B**, such that $\mathbf{B}[t + 1] = 1$ and $\mathbf{B}[i] = 0 (i`t + 1)$, where t is the decimal number corresponding to the binary sequence **C**.

S-3: Randomized response. Execute the algorithm L-RR to perturb the inner code **B** into a response vector **S** of equal length.

S-4: The area code **H** and the perturbed position vector **S** are returned and sent to the data server in the intelligent logistics platform.

Algorithm 4 Algorithm ϵ-LT_LDP

Input: $< X_i, Y_i, timestamp >, \epsilon_i, \tau, m, l$;
Output: **H**, **S**;

Initialize $2(l-1)$-bit vector $\mathbf{M} \leftarrow 0$;
Initialize 2^h-bit vector $\mathbf{H} \leftarrow 0$;
$\mathbf{M}_x \leftarrow min(2^{l-1}(x - \tau_{x\_min})/(\tau_{x\_max} - \tau_{x\_min}), 2^{l-1} - 1)$;
$\mathbf{M}_y \leftarrow min(2^{l-1}(y - \tau_{y\_min})/(\tau_{y\_max} - \tau_{y\_min}), 2^{l-1} - 1)$;
$M[oddbits] \leftarrow$ convert $\mathbf{M}_x$ to a $2(l-1)$-bit vector;
$M[evenbits] \leftarrow$ Convert $\mathbf{M}_y$ to a $2(l-1)$-bit vector;
$\mathbf{H} \leftarrow$ First, $2(m-1)$bits of $\mathbf{M}$;
$\mathbf{C} \leftarrow$ Last, $2(l-m)$bits of $\mathbf{M}$;
$t \leftarrow$ Convert $\mathbf{C}$ to decimal number;
$\mathbf{B}[t+1] \leftarrow 1$;
$\mathbf{S} \leftarrow$ L-RR($\mathbf{B}$,ϵ_i);
return S, **H**;

4.4. Model Algorithmic Analysis

In the carrier's multidimensional numerical data local differential privacy protection algorithm ϵ-L_LDP, according to the definition of local differential privacy (LDP), attributes $t_i[A_j](1 \le i \le n, 1 \le j \le 10)$ are independent of each other, $t_i[A_j], t_i^{'}[A_j] \in \tau_i$, the perturbed data $t_i^*[A_j] \in [-C, C]$, and the privacy budget $\epsilon_i(\epsilon/\sqrt{n} \le \epsilon_i \le \epsilon)$ is set by the carrier user u_i. The L-PM uses a security domain τ to normalize the original data to $[-1, 1]$. From the local differential privacy definition shown in Equations (7) and (23) can be obtained from it, as follows.

$$\frac{Pr[L - PM(t_i[A_j], \epsilon_i, \tau_i) = t_i^*[A_j]]}{Pr[L - PM(t_i^{'}[A_j], \epsilon_i, \tau_i) = t_i^*[A_j]]} = e^{\epsilon_i} \tag{23}$$

The derivation from Equation (23) leads to Equation (24) as follows.

$$\frac{Pr[L - PM(t_i[A_j], \epsilon_i, \tau_i) = t_i^*[A_j]]}{Pr[L - PM(t_i^{'}[A_j], \epsilon_i, \tau_i) = t_i^*[A_j]]} \le e^{\epsilon} \tag{24}$$

Thus, the ϵ-L_LDP algorithm satisfies the protection model of the local differential privacy algorithm.

From the paper [58], in the PM algorithm, the maximum absolute error bound is $O(\sqrt{d\log(1/\beta)}/\epsilon\sqrt{n})$, and the maximum absolute value error under individual attribute perturbation is $O(\sqrt{\log(1/\beta)}/\epsilon\sqrt{n})$. Since the carrier's privacy budget is set by itself, the maximum absolute value error of the ϵ-L_LDP algorithm is $O\left(\sqrt{d\log(d/\beta)}/\left(\min_{i\in[n]} \epsilon_i\sqrt{n}\right)\right)$. When $\min_{i\in[n]} \epsilon_i = \epsilon/\sqrt{n}$, the maximum absolute value error of the ϵ-L_LDP algorithm is $O(\sqrt{d\log(d/\beta)}/\epsilon\sqrt{n})$. The maximum absolute value error under individual attribute perturbation is $O(\sqrt{\log(1/\beta)}/\epsilon\sqrt{n})$.

In the carrier's location data local differential privacy protection algorithm ϵ-LT_LDP, according to the definition of local differential privacy (LDP), the number of carriers is n, the attributes $< X_i, Y_i, timestap > (1 \le i \le n)$ are independent of each other, and any $j(1 \le j \le n), < X_j, Y_j, timestamp >, < X_j, Y_j, timestamp >' \in \tau$, $\mathbf{H}^*$, $\mathbf{S}^*$ is the output of the algorithm ϵ-LT_LDP. If the algorithm is M, as proved and given in the paper [62], then $\mathbf{S}^i = M(\mathbf{B}^i)$. It can be obtained that Equation (25), as follows.

$$\frac{Pr[\epsilon - LT\_LDP(l_i, \epsilon_i, \tau, m, l) = (\mathbf{H}^*, \mathbf{S}^*)]}{Pr[\epsilon - LT\_LDP(l_i^{'}, \epsilon_i, \tau, m, l) = (\mathbf{H}^*, \mathbf{S}^*)]} \le e^{\epsilon} \tag{25}$$

In Equation (25), $l_i = \langle X_i, Y_i, timestamp \rangle$, $l_i^{'} = \langle X_i, Y_i, timestamp \rangle^{'}$, and the privacy budget $\epsilon_i (\epsilon/\sqrt{n} \leq \epsilon_i \leq \epsilon)$ set by the carrier u_i. The proof proceeds as shown in Equation (26) below. So, the ϵ-LT_LDP algorithm satisfies the protection model of the local differential privacy algorithm.

$$\begin{aligned}
&\frac{Pr[\epsilon - LT_LDP(\langle X_i, Y_i, timestamp \rangle, \epsilon_i, \tau, m, l) = (\mathbf{H}^*, \mathbf{S}^*)]}{Pr[\epsilon - LT_LDP(\langle X_i, Y_i, timestamp \rangle^{'}, \epsilon_i, \tau, m, l) = (\mathbf{H}^*, \mathbf{S}^*)]} \\
&= \frac{Pr[L - RR(\mathbf{B}, \epsilon_i) = \mathbf{S}^*]}{Pr[L - RR(\mathbf{B}^{'}, \epsilon_i)] = \mathbf{S}^*} \\
&= \frac{Pr[L - RR(\mathbf{B}_j, \epsilon_i) = \mathbf{S}_j^*]}{Pr[L - RR(\mathbf{B}_j^{'}, \epsilon_i) = \mathbf{S}_j^*]} \\
&= \frac{z \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1}{z^{'} \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1} \\
&\leq \frac{max(z \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1)}{min(z^{'} \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1)} \\
&= \frac{1 \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1}{-1 \times (e^{\epsilon_i} - 1) + e^{\epsilon_i} + 1} \\
&= e^{\epsilon_i} \\
&\leq e^{\epsilon}
\end{aligned} \tag{26}$$

5. Experimental Results and Analysis

5.1. Experimental Environment

The computer parameters on the carrier equipment are CPU, Intel Core(TM) i5-10500CPU×6 @ 3.1GHz (Intel, USA), 32G RAM (KingSton, USA), 500G storage (KingSton, USA), Windows 10 (Microsoft, USA). The experimental data include both open source datasets and simulated datasets. The experimental programming software is MATLAB R2020b and PyCharm Community Edition 2021.2. The algorithms are implemented by Python 3.6, and the logistics platform of Fankong Hui Network Technology Co., Ltd (Chengdu, China) is used as the intelligent logistic platform.

5.2. Experimental Analysis of ϵ-L_LDP Algorithms

In this section, we experimentally study the effect of the number of attributes and privacy budget on the usability of the ϵ-L_LDP algorithm, and the evaluation criterion used in the experiment is the mean square error (MSE), which is defined as follows (27).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (t_i - \widehat{t_i})^2 \tag{27}$$

where t_i is the real data and $\widehat{t_i}$ is the estimated data. The Harmony mean algorithm [55], the Duchi algorithm [57], and the PM multidimensional data mean estimation algorithm [58] are chosen for the comparison experiments. Since the concept of attribute safety domain is not defined in the algorithms [55,57,58], the maximum range of all possible values of the attributes in the dataset was taken as the safety domain of the attributes, and the safety domain τ in the ϵ-L_LDP algorithm was used for the normalization process. In the mean estimation experiments, each method was repeated 100 times to take its mean value in order to eliminate the effect of randomly generated errors.

5.2.1. The Impact of the Number of Attributes on Usability

In order to investigate the effect of the number of attributes on the usability of the algorithm, a randomly generated simulation dataset with 5000 carriers and a range of attributes [2, 4, 6, 8, 10] was used. Let $\epsilon = 0.5$, and each carrier generates a random value

in the interval $[\epsilon/\sqrt{d}, \epsilon]$ as their personalized privacy budget. The randomly generated simulation datasets are the GAUSS dataset and the UNIFORM dataset. The GAUSS dataset is a dataset that follows a Gaussian distribution with a mean of 60 and a standard deviation of 1. The UNIFORM dataset is a dataset that follows a uniform distribution with a mean in the range of [30, 70]. The experimental results are shown in Figure 6. From the experimental results, it can be seen that the MSE of the Harmony mean algorithm, the Duchi algorithm, the PM algorithm, and the ϵ-L_LDP algorithm gradually increase with the increase in the number of attributes for both the GUASS dataset and the UNIFORM dataset. The Harmony-mean algorithm shows a more significant increase in MSE with the increase in the number of attributes. When the number of attributes is 2, the MSE of the PM algorithm, the Duchi algorithm, and the ϵ-L_LDP algorithms is not significantly different, and when the number of attributes is greater than 2, the MSE of the ϵ-L_LDP algorithm is lower, indicating that the ϵ-L_LDP algorithm is better in terms of usability.

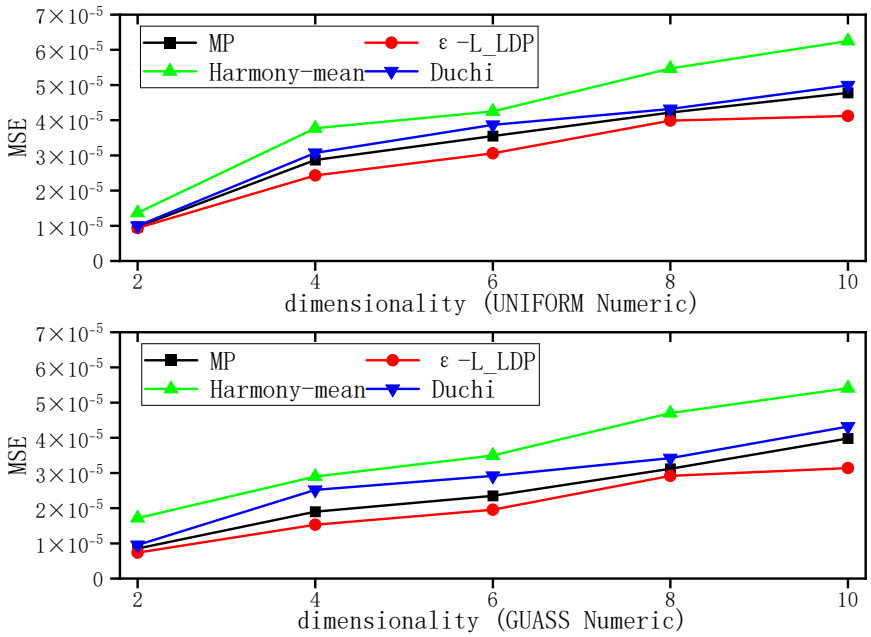

Figure 6. Experimental results for the UNIFORM dataset and the GUASS dataset.

5.2.2. The Impact of Privacy Budgets on Usability

Real datasets were used for the experimental data in studying the impact of privacy budget on algorithm usability. The reason why we chose the numerical data from public datasets to replace numerical data from carriers for our experiment is to facilitate the comparison between our proposed algorithm and other algorithms. Two public datasets, BR and MX, i.e., the Brazilian and Mexican census survey records, were extracted from IPUMS [63]. In total, 16 attributes are included in BR, among which 6 are numerical attributes, and 19 attributes are included in MX, among which 5 are numerical attributes. The data of the 5 numerical attributes in the BR dataset and MX dataset are selected for the experiment, and the range of values of the privacy budget was set to [0.5, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]. Personalization is applied in the ϵ-L_LDP algorithm designed to set their personalized privacy budget ϵ_i, with ϵ_i randomly drawn values in the interval $[\epsilon/\sqrt{d}, \epsilon]$. Experiments with these algorithms were carried out on data with numerical attributes in the BR dataset and the MX dataset, and the results are shown in Figure 7. In the experimental results of the MX public dataset, the MSE of the ϵ-L_LDP algorithm is slightly lower than that of the PM algorithm at a privacy budget of 0.5. At a privacy budget of 1.0, the MSE of the ϵ-L_LDP algorithm and the PM algorithm are very close to each other. At a privacy budget of 1.5, the MSE of the ϵ-L_LDP algorithm is significantly lower than that of the PM algorithm. At such privacy budgets, the MSE of the harmony-mean algorithm is

significantly higher than that of the ϵ-L_LDP algorithm, and the MSE of the Duchi and PM algorithms is slightly higher than that of the ϵ-L_LDP algorithm. At privacy budgets greater than 2.0, the MSE of the Duchi algorithm, PM algorithm, and ϵ-L_LDP algorithm gradually approached each other. In the results for the BR dataset, it is obtained that the MSE of ϵ-L_LDP is the smallest for any privacy budget case, and the MSE of the harmony-mean algorithm is larger for both, especially for the privacy budget in the [0.5, 2.0] interval class, the MSE of the harmony-mean algorithm is a bit too large, resulting in a lower usability of the harmony-mean algorithm low. Overall, the MSE of the harmony-mean algorithm, the PM algorithm, the Duchi algorithm, and the algorithm all gradually decreased as the privacy budget increased, and the MSE of the four algorithms differed less when the privacy budget was greater than 3.0.

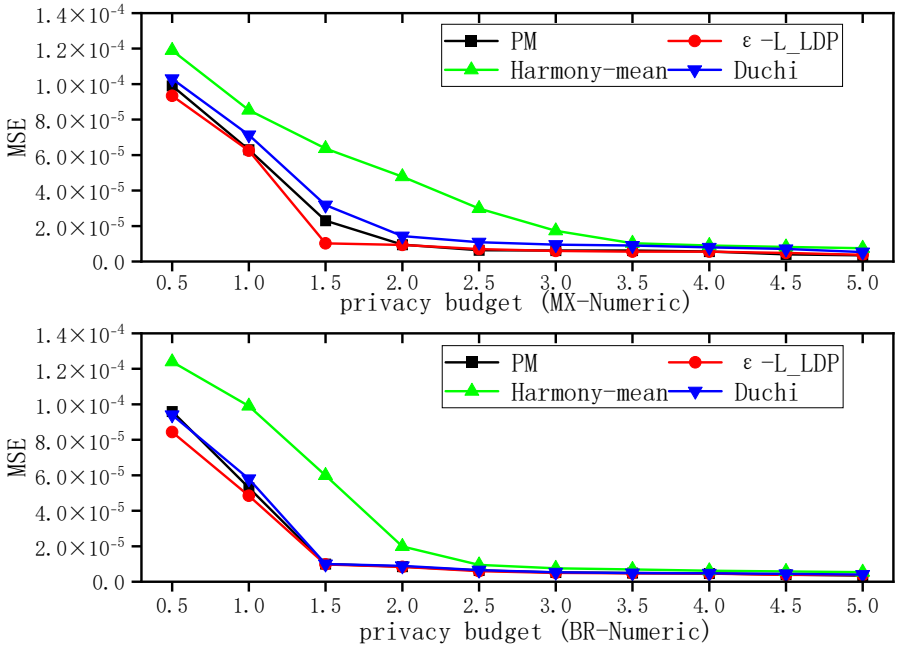

Figure 7. Experimental results for the MX dataset and the BR dataset.

In summary, the ϵ-L_LDP algorithm proposed in this paper can meet the carrier's multidimensional numerical sensitive dataset privacy protection, in which it can achieve the carrier's on-demand individually set level of privacy protection ϵ_i. Moreover, the MSE of the data after perturbation protection by the ϵ-L_LDP algorithm is low, indicating that the numerical data received by the server also have good usability.

5.3. Experimental Analysis of ϵ-LT_LDP Algorithms

Two performance metrics are specified in this experiment to study the locations received at the data server side compared to the original dataset, namely the performance of the trajectory proportion estimation and the performance of the location proportion estimation. In order to visualize the estimation performance, this experiment uses the trajectory (location) proportion rather than the trajectory (location) frequency as the performance metric. The evaluation criteria used in the experiment are the mean absolute percentage error (MAPE) and root mean square error (RMSE), which are defined as follows (28) and (29).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{p_t - \widehat{p_t}}{p_t}\right| \tag{28}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{1}(p_t - \widehat{p_t})^2} \tag{29}$$

where p_t is the true proportion of the t trajectory (position) and $\widehat{p}_t$ is the estimated proportion of the t trajectory (position). The QLP algorithm and the QJLP algorithm from the [62] were chosen for the comparison experiments. The range of values for the privacy budget was set to [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]. Here, personalization is applied in the ϵ-LT_LDP algorithm designed in this paper to set their privacy budget ϵ_i, ϵ_i in the interval $[\epsilon/\sqrt{d}, \epsilon]$ of randomly selected values. In the experiments to estimate the performance, each method was repeated 100 times to take the mean value in order to eliminate the effect of randomly generated errors.

For the experiments, we used GPS data from over 14,000 taxis in Chengdu from 3 August to 30 August 2014. The original dataset includes more than 1.4 billion GPS records, and each record has a total of six attributes, namely vehicle ID, latitude, longitude, passenger load, date, and time. After removing the obviously invalid track records, we selected a rectangular geographical area with high coverage track records, a total of 49,851,265 track records, and set the time resolution to 300 s to obtain a spatio-temporal dataset of 9858 $\times$ 265 GPS points.

5.3.1. The Performance of the Trajectory Proportion Estimation

In experiments researching the performance of trajectory proportion estimation, the MAPE and RMSE of the ϵ-LT_LDP algorithm, the QLP algorithm, and the QJLP algorithm for trajectories of length 3 are shown in Figure 8a,b, and the MAPE and RMSE of the ϵ-LT_LDP algorithm, the QLP algorithm, and the QJLP algorithm for trajectories of length 2 are shown in Figure 8c,d. In Figure 8a, the MAPE of both the QLP algorithm and the QJLP algorithm are greater than that of the ϵ-LT_LDP algorithm for the same privacy budget, indicating that the availability of the ϵ-LT_LDP algorithm data are overall higher than that of the QLP algorithm and the QJLP algorithm. The MAPE of the ϵ-LT_LDP algorithm decreased significantly for privacy budgets of 0.5–3.5, and slowly decreased for privacy budgets greater than 3.5. In Figure 8b, the RMSE of the ϵ-LT_LDP algorithm, QLP algorithm, and QJLP algorithm decrease and eventually converge as the privacy budget increases. The privacy budget has a small effect on the RMSE of the ϵ-LT_LDP algorithm and the QJLP algorithm, which slowly decreases as the privacy budget increases. In Figure 8c, the MAPE of the ϵ-LT_LDP algorithm, QLP algorithm, and QJLP algorithm gradually decrease and converge as the privacy budget increases. With the same privacy budget, the MAPE of the ϵ-LT_LDP algorithm is significantly lower than that of the QLP algorithm and QJLP algorithm, indicating that their data availability is higher. In Figure 8d, the RMSE of the ϵ-LT_LDP algorithm and the QJLP algorithm are very close when the privacy budget is 0.5. With such a privacy budget, the RMSE of the ϵ-LT_LDP algorithm is significantly smaller than the RMSE of the QLP algorithm. As the privacy budget increases, the RMSE of the ϵ-LT_LDP algorithm, the QLP algorithm, and the QJLP algorithm all gradually decrease.

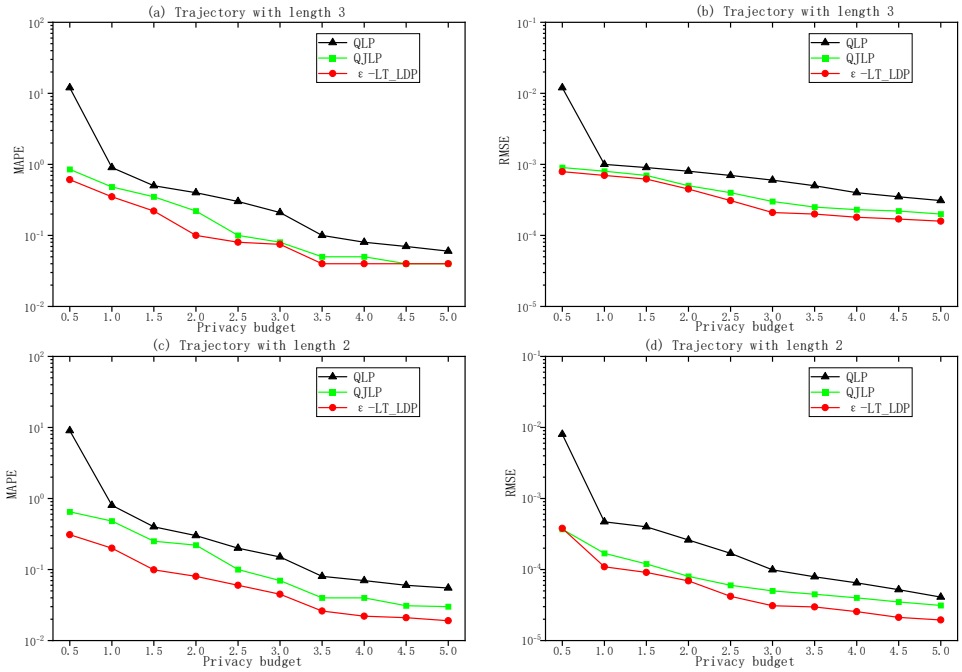

Figure 8. Experimental results of trajectory proportion estimation performance.

5.3.2. The Performance of the Location Proportion Estimation

Since the trajectory is a sequence of position points combined according to a temporal rule, experiments on the performance of position scaling estimation are conducted in this experiment. In the evaluation of the position proportion performance, the evaluation annotation was only chosen for the mean absolute percentage error MAPE. The experiments on the position proportion estimation performance of the ϵ-LT_LDP algorithm, the QLP algorithm, and the QJLP algorithm were conducted to compare the error between the perturbed position points and the true position points, and the experimental results are shown in Figure 9a,b. In Figure 9a,b, for the same privacy budget, the QLP algorithm has the lowest MAPE and the ϵ-LT_LDP algorithm has an MSPE that is again lower than the MAPE of the QJLP algorithm but higher than the MAPE of the QLP algorithm. The QLP algorithm has the highest availability of data for the same privacy budget and the ϵ-LT_LDP algorithm again has better availability than the QJLP algorithm. This is because both the ϵ-LT_LDP algorithm and the QJLP algorithm are privacy-protection algorithms for trajectory data, and privacy protection for individual location points results in lower data availability. However, the QLP algorithm is a privacy-protection algorithm for location point data. The MAPE of the ϵ-LT_LDP algorithm, the QLP algorithm, and the QJLP algorithm all gradually decrease as the privacy budget increases.

In summary, the ϵ-LT_LDP algorithm proposed in this paper can satisfy the privacy protection of the carrier's location track data, where ϵ_i can achieve the degree of privacy protection set by the carrier on demand, and the MAPE and RMSE of the location track data after perturbation protection by the ϵ-LT_LDP algorithm are both low, indicating that the location track data received by the server also have good usability.

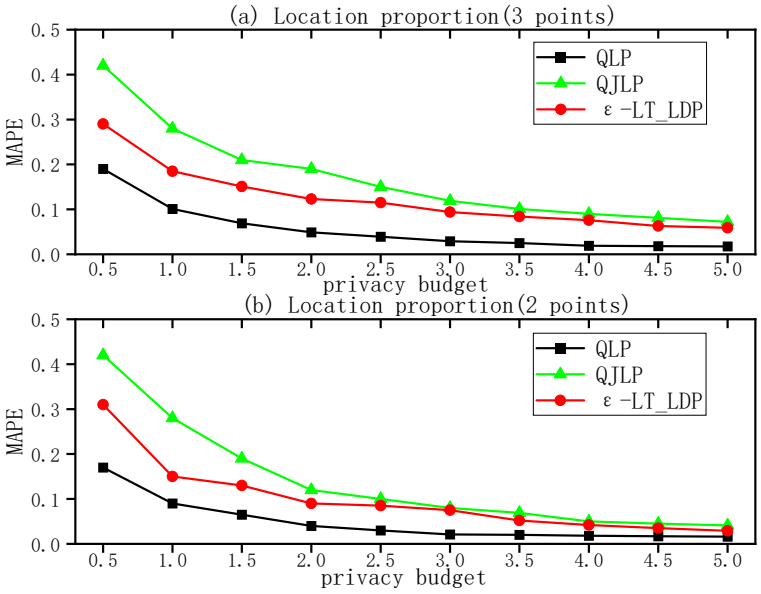

Figure 9. Experimental results of the location proportion estimation performance.

6. Conclusions

With the continuous improvement in people's living standards, new requirements for the logistics industry in terms of intelligence, information technology, and low cost have been put forward, promoting the creation and development of an intelligent logistics platform. The intelligent logistics platforms use smart collection devices to collect logistics process status data, and to provide users with better and more stable intelligent logistics services. At the same time, with the progress of a humanistic society, personal privacy awareness has begun to awaken, and the carrier in the intelligent logistics platform is one of the important components of the people object entity, so the carrier's privacy information needs to be effectively protected. The smart collection device collection of the logistics process state data contains the carrier's sensitive sense data, which if not for its privacy protection, will not only expose the carrier's private information, and even directly or indirectly affect the efficiency of the intelligent logistics platform. Of course, as an intelligent logistics system is a production system, when privacy protection is applied to its carrier data, a balance between the privacy protection and the visibility of production operations should be considered.

This paper proposes a privacy protection scheme for the sensitive data of carriers in an intelligent logistics system, which ensures that the logistics process data do not expose the carrier's private information under the condition of ensuring the availability of logistics process data. In this way, the privacy of the data in the intelligent logistics platform is improved while protecting the privacy of the carrier, which is conducive to the efficient and stable operation of the intelligent logistics platform. In this paper, we design the local differential privacy protection algorithm ϵ-L_LDP (ϵ-logistic local differential privacy, ϵ-L_LDP) for the carrier's multidimensional numerical data, and the local differential privacy protection algorithm ϵ-LT_LDP (ϵ-logistic trajectory local differential privacy, ϵ-LT_LDP) for the carrier's location data. Both the ϵ-L_LDP algorithm and the ϵ-LT_LDP algorithm allow carrier users to set personalized privacy budgets according to their privacy needs, and rigorously prove the privacy of both algorithms in terms of privacy theory. In the experiments, real and simulated datasets were used as experimental data, mean square error (MSE), mean absolute percentage error (MAPE), and root mean square error (RMSE) were used as evaluation criteria. Both the ϵ-L_LDP algorithm and the ϵ-LT_LDP algorithm were not only proved to have good usability and can be used for the privacy protection of the sensitive data of carriers in the intelligent logistics system to prevent the sensitive

data of carriers from being leaked; however, it also maintains the balance between privacy protection and the visibility of production operations.

Author Contributions: Conceptualization, L.T. and Z.Y. (Zhengyi Yao); methodology, Z.Y. (Zhengyi Yao); software, Z.Y. (Zhengyi Yao), J.Y. and L.F.; validation, Z.Y. (Zhengyi Yao), Z.Z., J.Y. and X.T.; formal analysis, J.X.; investigation, L.F. and K.S.; data curation, Z.Y. (Zhengyi Yao) and Z.Z.; supervision, L.T.; writing—original draft preparation, Z.Y. (Zhengyi Yao), P.Y., W.W., D.Y., J.Y. and Z.Y. (Ziyuan Yu); project administration, L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 61373126; and Sichuan Provincial Science and Technology Department Project under Grant No. 2022YFG0161, No.2023YFG0295.

Data Availability Statement: The data used in the experiment is detailed in the Experimental Results and Analysis section.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Humayun, M.; Jhanjhi, N.Z.; Hamid, B.; Ahmed, G. Emerging smart logistics and transportation using IoT and blockchain. *IEEE Internet Things Mag.* **2020**, *3*, 58–62. [CrossRef]
2. Ding, Y.; Jin, M.; Li, S.; Feng, D. Smart logistics based on the internet of things technology: An overview. *Int. J. Logist. Res. Appl.* **2021**, *24*, 323–345. [CrossRef]
3. Song, Y.; Yu, F.R.; Zhou, L.; Yang, X.; He, Z. Applications of the Internet of Things (IoT) in smart logistics: A comprehensive survey. *IEEE Internet Things J.* **2020**, *8*, 4250–4274. [CrossRef]
4. Speranza, M.G. Trends in transportation and logistics. *Eur. J. Oper. Res.* **2018**, *264*, 830–836. [CrossRef]
5. Belli, L.; Cilfone, A.; Davoli, L.; Ferrari, G.; Adorni, P.; Di Nocera, F.; Dall'Olio, A.; Pellegrini, C.; Mordacci, M.; Bertolotti, E. IoT-enabled smart sustainable cities: Challenges and approaches. *Smart Cities* **2020**, *3*, 1039–1071. [CrossRef]
6. Liu, W.; Long, S.; Liang, Y.; Wang, J.; Wei, S. The influence of leadership and smart level on the strategy choice of the smart logistics platform: A perspective of collaborative innovation participation. *Ann. Oper. Res.* **2021**, *324*, 893–935. [CrossRef]
7. Yang, M.; Mahmood, M.; Zhou, X.; Shafaq, S.; Zahid, L. Design and implementation of cloud platform for intelligent logistics in the trend of intellectualization. *China Commun.* **2017**, *14*, 180–191. [CrossRef]
8. Mathew, E. Swarm intelligence for intelligent transport systems: Opportunities and challenges. In *Swarm Intelligence for Resource Management in Internet of Things*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 131–145.
9. Liu, S.; Wang, J. A security-enhanced express delivery system based on NFC. In Proceedings of the 2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT), Hangzhou, China, 25–28 October 2016; pp. 1534–1536.
10. Lin, X.; Jing, P.; Yu, C.; Feng, X. TPLI: A traceable privacy-preserving logistics information scheme via blockchain. In Proceedings of the 2021 International Conference on Networking and Network Applications (NaNA), Lijiang, China, 29 October–1 November 2021; pp. 345–350.
11. Waters, D. *Supply Chain Risk Management: Vulnerability and Resilience in Logistics*; Kogan Page Publishers: London, UK, 2011.
12. Sativell, T.; Sabar, R. The threat of high value cargo faced by logistics companies. In Proceedings of the Simposium Pengurusan Teknologi, Operasi & Logistik (SIPTIK III), Changlun, Malaysia, 11–12 December 2012.
13. NBC. UPS Driver Kidnapped, Packages Stolen in Brazen Atlanta Heist. Available online: https://www.nbcnews.com (accessed on 29 December 2021).
14. Feng, C.; Tan, L.; Xiao, H.; Yu, K.; Qi, X.; Wen, Z.; Jiang, Y. PDKSAP: Perfected double-key stealth address protocol without temporary key leakage in blockchain. In Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Xiamen, China, 28–30 July 2020; pp. 151–155.
15. Yang, C.; Tan, L.; Shi, N.; Xu, B.; Cao, Y.; Yu, K. AuthPrivacyChain: A blockchain-based access control framework with privacy protection in cloud. *IEEE Access* **2020**, *8*, 70604–70615. [CrossRef]
16. Tan, L.; Shi, N.; Yang, C.; Yu, K. A blockchain-based access control framework for cyber-physical-social system big data. *IEEE Access* **2020**, *8*, 77215–77226. [CrossRef]
17. Tan, L.; Shi, N.; Yu, K.; Aloqaily, M.; Jararweh, Y. A blockchain-empowered access control framework for smart devices in green internet of things. *ACM Trans. Internet Technol. (TOIT)* **2021**, *21*, 1–20. [CrossRef]
18. Liu, X.; Hu, B.; Zhou, Q.; Liu, J. A Logistics Privacy Protection System Based on Cloud Computing. In *Frontier Computing: Theory, Technologies and Applications FC 2016 5*; Springer: Singapore, 2018; pp. 455–461.
19. Léauté, T.; Faltings, B. Coordinating logistics operations with privacy guarantees. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11), Barcelona, Spain, 16–22 July 2011; pp. 2482–2487.

20. Xu, F.J.; Tong, F.C.; Tan, C.J. Auto-ID enabled tracking and tracing data sharing over dynamic B2B and B2G relationships. In Proceedings of the 2011 IEEE International Conference on RFID-Technologies and Applications, Sitges, Spain, 15–16 September 2011; pp. 394–401.
21. Gao, Q.; Zhang, J.; Ma, J.; Yang, C.; Guo, J.; Miao, Y. LIP-PA: A logistics information privacy protection scheme with position and attribute-based access control on mobile devices. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 9436120. [CrossRef]
22. Qian, W.E.I.; Xing-Yi, L.I. Express information protection application based on K-anonymity. *Appl. Res. Comput./Jisuanji Yingyong Yanjiu* **2014**, *31*, 555–567.
23. Qi, H.; Chenjie, D.; Yingbiao, Y.; Lei, L. A new express management system based on encrypted QR code. In Proceedings of the 2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA), Nanchang, China, 14–15 June 2015; pp. 53–56.
24. Zhang, X.; Li, H.; Yang, Y.; Sun, G.; Chen, G. LIPPS: Logistics information privacy protection system based on encrypted QR code. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August 2016; pp. 996–1000.
25. Yan, W.; Yao, Y.; Zhang, W.M. Privacy-preserving scheme for logistics systems based on 2D code and information hiding. *Chin. J. Netw. Inf. Secur.* **2017**, *3*, 22–28.
26. Pournader, M.; Shi, Y.; Seuring, S.; Koh, S.L. Blockchain applications in supply chains, transport and logistics: A systematic review of the literature. *Int. J. Prod. Res.* **2020**, *58*, 2063–2081. [CrossRef]
27. Tijan, E.; Aksentijević, S.; Ivanić, K.; Jardas, M. Blockchain technology implementation in logistics. *Sustainability* **2019**, *11*, 1185. [CrossRef]
28. Rožman, N.; Corn, M.; Požrl, T.; Diaci, J. Distributed logistics platform based on Blockchain and IoT. *Procedia CIRP* **2019**, *81*, 826–831. [CrossRef]
29. Yi, H. A secure logistics model based on blockchain. *Enterp. Inf. Syst.* **2021**, *15*, 1002–1018. [CrossRef]
30. Duan, H.; Yang, J.; Yang, H. A Blockchain-Based Privacy Protection Application for Logistics Big Data. *J. Cases Inf. Technol. (JCIT)* **2022**, *24*, 1–12. [CrossRef]
31. Li, H.; Han, D.; Tang, M. A privacy-preserving storage scheme for logistics data with assistance of blockchain. *IEEE Internet Things J.* **2021**, *9*, 4704–4720. [CrossRef]
32. Zhou, S.; Li, F.; Tao, Y.-F.; Xiao, X.-K. Privacy preservation in database applications: A survey. *Chin. J. Comput.* **2009**, *32*, 847–861. [CrossRef]
33. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2002**, *10*, 557–570. [CrossRef]
34. Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M. l-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data (TKDD)* **2007**, *1*, 3-es. [CrossRef]
35. Li, N.; Li, T.; Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In Proceedings of the IEEE 23rd International Conference on Data Engineering, Istanbul, Turkey, 17–20 April 2007; pp. 106–115.
36. Wong, R.C.W.; Li, J.; Fu, A.W.C.; Wang, K. (α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 754–759.
37. Ganta, S.R.; Kasiviswanathan, S.P.; Smith, A. Composition attacks and auxiliary information in data privacy. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 265–273.
38. Wong, R.C.W.; Fu, A.W.C.; Wang, K.; Yu, P.S.; Pei, J. Can the utility of anonymized data be used for privacy breaches? *ACM Trans. Knowl. Discov. Data (TKDD)* **2011**, 5, 1–24. [CrossRef]
39. Dwork, C. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–29 April 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
40. Dwork, C.; Lei, J. Differential privacy and robust statistics. In Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, Bethesda, MD, USA, 31 May–2 June 2009; pp. 371–380.
41. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006*; Proceedings 3; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
42. McSherry, F.; Talwar, K. Mechanism design via differential privacy. In Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), Providence, RI, USA, 20–23 October 2007; pp. 94–103.
43. Dwork, C.; Naor, M.; Pitassi, T.; Rothblum, G.N.; Yekhanin, S. Pan-Private Streaming Algorithms. In Proceedings of the Innovations in Computer Science—ICS 2010, Beijing, China, 5–7 January 2010; pp. 66–80.
44. Duchi, J.C.; Jordan, M.I.; Wainwright, M.J. Local privacy and statistical minimax rates. In Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, Berkeley, CA, USA, 26–29 October 2013; pp. 429–438.
45. Ye, Q.; Meng, X.; Zhu, M.; Huo, Z. Survey on local differential privacy. *J. Softw.* **2018**, *29*, 1981–2005.
46. Chen, R.; Li, H.; Qin, A.K.; Kasiviswanathan, S.P.; Jin, H. Private spatial data aggregation in the local setting. In Proceedings of the 2016 IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, Finland, 16–20 May 2016; pp. 289–300.
47. McSherry, F.D. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, Providence, RI, USA, 29 June–2 July 2009; pp. 19–30.

48. Warner, S.L. Randomized response: A survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* **1965**, *60*, 63–69. [CrossRef] [PubMed]
49. Fanti, G.; Pihur, V.; Erlingsson, Ú. Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries. *arXiv* **2015**, arXiv:1503.01214.
50. Erlingsson, Ú.; Pihur, V.; Korolova, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp. 1054–1067.
51. Bassily, R.; Smith, A. Local, private, efficient protocols for succinct histogram. In Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing, Portland, OR, USA, 14–17 June 2015; pp. 127–135.
52. Kairouz, P.; Oh, S.; Viswanath, P. Extremal mechanisms for local differential privacy. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
53. Duchi, J.C.; Jordan, M.I.; Wainwright, M.J. Local privacy, data processing inequalities, and statistical minimax rates. *arXiv* **2013**, arXiv:1302.3203.
54. Duchi, J.C.; Jordan, M.I.; Wainwright, M.J. Privacy aware learning. *J. ACM (JACM)* **2014**, *61*, 1–57. [CrossRef]
55. Nguyên, T.T.; Xiao, X.; Yang, Y.; Hui, S.C.; Shin, H.; Shin, J. Collecting and analyzing data from smart device users with local differential privacy. *arXiv* **2016**, arXiv:1606.05053.
56. Andrés, M.E.; Bordenabe, N.E.; Chatzikokolakis, K.; Palamidessi, C. Geo-indistinguishability: Differential privacy for location-based systems. In Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, Berlin, Germany, 4–8 November 201 3; pp. 901–914.
57. Duchi, J.C.; Jordan, M.I.; Wainwright, M.J. Minimax optimal procedures for locally private estimation. *J. Am. Stat. Assoc.* **2018**, *113*, 182–201. [CrossRef]
58. Wang, N.; Xiao, X.; Yang, Y.; Zhao, J.; Hui, S.C.; Shin, H.; Shin, J.; Yu, G. Collecting and analyzing multidimensional data with local differential privacy. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macao, 8–11 April 2019; pp. 638–649.
59. Qardaji, W.; Yang, W.; Li, N. Differentially private grids for geospatial data. In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, QLD, Australia, 8–12 April 2013; pp. 757–768.
60. Bentley, J.L. Multidimensional binary search trees used for associative searching. *Commun. ACM* **1975**, *18*, 509–517. [CrossRef]
61. Samet, H. The quadtree and related hierarchical data structure. *ACM Comput. Surv. (CSUR)* **1984**, *16*, 187–260. [CrossRef]
62. Yang, Z.; Wang, R.; Wu, D.; Wang, H.; Song, H.; Ma, X. Local trajectory privacy protection in 5G enabled industrial intelligent logistics. *IEEE Trans. Ind. Inform.* **2021**, *18*, 2868–2876. [CrossRef]
63. Integrated Public Use Micro-Samples, "IPUMS". UPS Driver Kidnapped, Packages Stolen in Brazen Atlanta Heist. Available online: https://www.ipums.org (accessed on 10 October 2021).