Hindawi
Journal of Applied Mathematics
Volume 2023, Article ID 1046050, 10 pages
https://doi.org/10.1155/2023/1046050

Research Article

Data Fusion and Processing Technology of Wireless Sensor Network for Privacy Protection

Lusheng Shi, Kai Li, and Huibo Zhu

School of Information Engineering, Suqian University, Suqian 223800, China

Correspondence should be addressed to Lusheng Shi; shilusheng@mjc-edu.cn

Received 20 September 2022; Revised 26 April 2023; Accepted 15 May 2023; Published 29 May 2023

Academic Editor: Theodore E. Simos

Data fusion and privacy protection technologies are both the research focuses in the field of wireless sensor networks. When the sensor network is in a harsh environment, the sensor nodes will face the danger of malicious entity attack in the data fusion progress. The efficiency and privacy protection of sensor network data fusion are very important. The traditional data fusion privacy protection algorithm has the problems of low data fusion efficiency and low privacy protection level. These problems are to be solved in this study. An improved cluster-based privacy data aggregation (I-CPDA) is proposed, which combines data slicing and false interference data technology. The experimental results of the algorithm show that the data fusion accuracy of the I-CPDA algorithm increases faster than the traditional algorithm with the time interval increasing, and the highest value reaches 90.7%. The fusion accuracy of the traditional CPDA algorithm under the same environment is 68.7%. In the actual test, the interception success rate of the I-CPDA algorithm for data attacks reached 90.74%, while the traditional CPDA was only 76.66. In addition, when the number of nodes in the cluster is 15, the data traffic of the I-CPDA is 56, while the data traffic of the traditional CPDA algorithm in the same environment exceeds 200. Compared with the currently widely used traditional algorithms, the I-CPDA algorithm has obvious advantages in terms of fusion effect, privacy, and efficiency and can be put into practical application.

1. Introduction

1.1. Background. Wireless sensor network refers to a self-organizing network that is formed by multiple wireless nodes, and each wireless sensor node has computing power and perception ability and can operate without human labor and supervision [1]. This technology has been used in the detection of ecological environment, machinery manufacturing, natural disasters, engineering construction, and other fields [2]. Wireless sensor nodes are easily attacked by malicious entities in data fusion and transmission process. Malicious people can extract the privacy information of network nodes or inject malicious codes to control the nodes [3]. The traditional data fusion privacy protection algorithm has the problems of low data fusion efficiency and low privacy protection level. These problems are to be solved in this study. In the field of data fusion privacy protection, cluster-based privacy data aggregation (CPDA) is a widely used traditional algorithm [4]. This research corrects and improves the defects of the traditional CPDA algorithm and integrates the idea of data slicing and false data interference into the privacy protection module. The improved I-CPDA algorithm theoretically has lower communication overhead and higher privacy protection level. It is hoped that the research of I-CPDA algorithm will make a practical contribution to the field of wireless sensor networks.

1.2. Planning. This article has 5 sections. The second is a literature review of related work, in which previous researches are discussed. The third is the description and construction of the proposed algorithm. The forth is the result of the experiment. The fifth is the conclusion of the research.

2. Literature Review

In terms of privacy information protection in the data and network environment, many researchers have explored and practiced. Kumar et al. proposed a blockchain data privacy protection technology for automatic driving, which ensures the security of data transmission in automatic driving. However, the application scope of this technology is small, and it is generally applied to automatic driving and related fields [5]. Mahdavisharif et al. proposed a big data-based identification system for hackers to crack the state of computer networks. The system is based on deep learning long- and short-term storage, and its detection accuracy is 20% higher than the current intrusion detection system. However, the application range of the system is limited, and its capability in wireless sensor networks is reduced [6]. Feng and Liu proposed a low-power 3D wireless sensor privacy protection algorithm [7]. The communication energy consumption of the algorithm is very low, and the accuracy of data fusion is improved to some extent. However, the privacy protection ability of the algorithm still has a lot of room for improvement [7]. Table 1 shows the pros and cons of the literature of those privacy information protection methods.

In the research and application of wireless sensor networks, Pramod et al. proposed a monitoring method based on wireless sensor networks. When the target is present, sensors observe an (unknown) deterministic signal with attenuation depending on the unknown distance between the sensor and the target. Simulation results confirm the promising performance of the proposed approaches [8]. Nasurulla and Kaniezhil studied the long-range transmission problem of wireless sensor network and established an optimized system for wireless sensor network for fuzzy subordinate support system. The accuracy level of the system is discussed in [9]. Lakshmi and Deepthi proposed a solution with homomorphic encryption [10]. The data fusion encryption scheme in traditional wireless sensor networks cannot effectively support data aggregation and the solution that fixed this problem. It has good resistance to various attack methods, which is significantly better than the existing schemes [10]. Zhang et al. led his team members to address the information security problem of wireless sensors in the power grid and proposed a data sharing model combined with blockchain. The analysis tells that the data sharing model safely and effectively protects, stores, and shares confidential data [11]. Jiang et al. made an improved scheme for the high energy consumption and security defects of wireless sensor networks in applications such as habitat and military monitoring [12]. The scheme deploys the sensor distribution strategically. Compared with traditional deployment schemes, energy consumption and information delay are optimized, and privacy is greatly improved [12]. Alam et al. proposed a data clustering technology for data compression in wireless sensor networks. It utilizes the characteristics of adaptive recursion and smooth data compression [13]. Experiments show that this technology can efficiently work for data compression with minimal space-time complexity [13]. Giri et al. proposed an optimized wireless sensor network system, which can effectively monitor landslide problems using wireless inertial measurement units, and the test shows that the system can predict the failure intensity of landslides correctly [14]. Combined with the research results in related fields, there is a lot of constructive research in the development and optimization of wireless sensors, but there are still some research gaps in the optimization of data privacy issues based on traditional algorithms, so the traditional CPDA algorithm is revised and optimized, in order to bring practical potential research results to this field.

3. Data Fusion Technology of Wireless Sensor Network Applied to Privacy Protection

3.1. Design of Wireless Sensor Data Fusion Algorithm and Selection of Clustering Protocol. Currently widely used privacy protection algorithms in the field of wireless sensors include CPDA algorithm and slice-mix-aggregate (SMART) method [15]. The principle of CPDA method is to add noise to the data to cause data disturbance, thus achieving the effect of privacy data protection. It uses polynomial data perturbation to protect data, which will not disclose other node information when calculating the fusion result. Although this algorithm can ensure the accuracy of fusion results while protecting privacy, its communication consumption is very high, the process is complex, and the level of privacy protection is also low [16]. The principle of SMART algorithm is partition and reorganization. The original sensing data is processed in pieces and sent to other nodes. Attackers cannot obtain complete information unless they obtain data from all nodes. The SMART method has stronger communication consumption, cost, and privacy protection ability than the CPDA algorithm, but its data fusion time is longer, so it cannot protect the integrity of data privacy [17]. According to the definition, the mathematical expression of CPDA to calculate the disturbance data is shown in

$$v = a + rx + rx^2. \tag{1}$$

In Equation (1), v represents the disturbance data, r represents the random number generated by the node, and x is the seed generated by the node. In addition, the probability of data eavesdropping under SMART is shown in

$$PS = \frac{h}{H}. \tag{2}$$

In Equation (2), h represents the number of keys, and H is the total number of keys in the key pool. Aiming at the problems existing in the data fusion algorithm of traditional wireless sensor privacy protection, an improved cluster-based privacy data aggregation (I-CPDA) is proposed. The improved algorithm adopts dynamic election for cluster head nodes. The slicing mechanism is integrated to separate the privacy data that needs to be protected, which increases the difficulty of eavesdropping on data. In addition, the algorithm will create false information to further interfere with data eavesdropping, so as to increase the security level of private information. The operation flow chart of the I-CPDA algorithm is shown in Figure 1. In the step of node data

Table 1: The pros and cons of the literature.

Methods	Pros	Cons
Literature [5] (blockchain)	Excellent protection capability and transmission speed	The application scope is small and can only be applied to specific fields
Literature [6] (big data and deep learning)	Fast recognition speed and high accuracy	Capability in wireless sensor networks is reduced
Literature [7] (CPDA)	Low communication energy consumption	Privacy protection ability is not enough

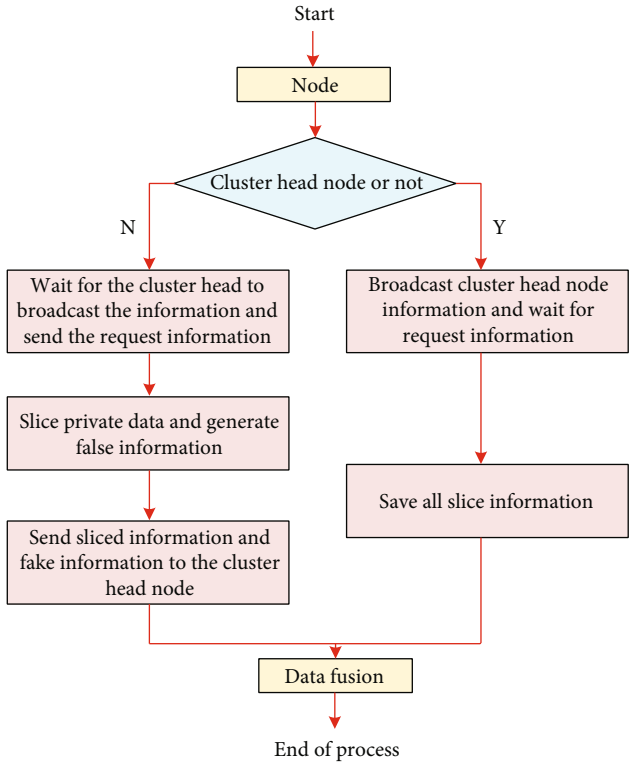

Figure 1: Flow chart of I-CPDA algorithm operation.

processing, the algorithm slices the private data, adds disturbing false data, and finally performs data fusion.

The I-CPDA algorithm adopts the Hive architecture, which is the basic framework of the Hadoop database and is friendly to the Structured Query Language. Users can perform MapReduce operations through Structured Query Language [18]. Figure 2 presents the structure of the Hive architecture. In the structure, the server is saved to allow users to access Hive. After the user submits the command, the grammar will convert the user input into a MapReduce task and then submit it to the cluster.

What I-CPDA can use in various clustering protocols are low energy adaptive clustering hierarchy (LEACH) and secure exchange protocol (SEP). The role of LEACH protocol in I-CPDA is to dynamically select specific cluster head nodes and form different clusters. The cluster is the product of the sensor node organization, and the cluster head node needs numerous energy, which will cause the cluster head node in the model to die quickly. The LEACH protocol will determine a new cluster head node in each round of operation. In this mode, the node energy will not be quickly exhausted, so the overall energy consumption will be decreased, and the life will be extended. The LEACH protocol selects cluster head nodes in line with the preset percentage of cluster heads and the number of times each node has become a cluster head. The mathematical expression of the cluster head node selected by the protocol is shown in

$$F(n) = \begin{cases} \dfrac{p}{1 - pr * \bmod (1/p)}, & n \in G, \\ 0, & n \notin G. \end{cases} \tag{3}$$

In Equation (3), n represents the nth node, and G represents the set of nodes that have not been cluster heads among all nodes. p represents the cluster head node probability, and r represents the current number of rounds. When selecting a cluster head node, the judged node generates a random number between 0 and 1, which is used to compare with the preset threshold value. When the random number is less than the threshold value, the corresponding node will become the cluster head node. The SEP protocol treats the probability of a node being selected as a cluster head as a variable, which makes the energy distribution in the whole network more uniform and stable. In the SEP protocol, the creation and transmission of clusters are periodic and have a random proportion of advanced nodes, which have higher energy than ordinary nodes. Assuming that the probability of an advanced node becoming the cluster head node in the SEP protocol is pa and the probability of an ordinary node selected as the cluster head node is pn, then the probability of the two types of nodes being selected as the cluster head is shown in

$$\begin{cases} \mathrm{pa} = \dfrac{p(1+\beta)}{\alpha\beta + 1}, \\ \mathrm{pn} = \dfrac{p}{\alpha\beta + 1}. \end{cases} \tag{4}$$

In Equation (4), p is the proportion of cluster head nodes in the entire cluster, and α is the proportion of advanced nodes, and β means that the energy of advanced nodes is β times larger than ordinary node energy. After defining the probability of the two kinds of nodes becoming cluster head nodes, their thresholds can be described. The threshold of the advanced node is shown in

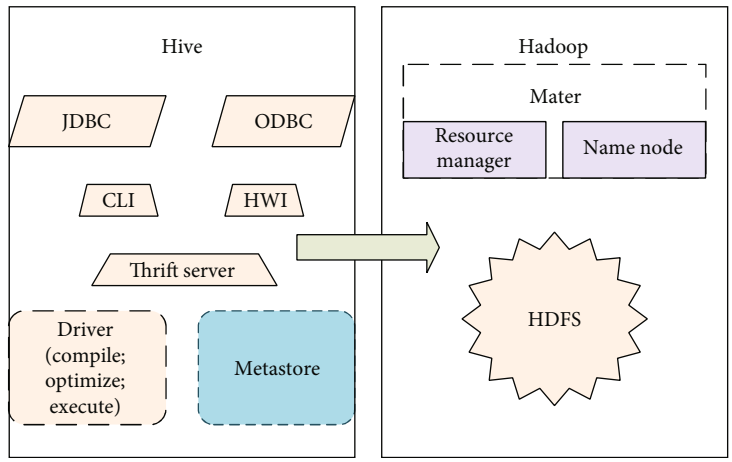

Figure 2: Schematic diagram of Hive architecture.

$$F(\mathrm{na}) = \begin{cases} \dfrac{\mathrm{pa}}{1 - \mathrm{pa}[r * \mathrm{mod}\ (1/\mathrm{pa})]}, & n \in \mathrm{Ga}, \\ 0, & n \notin \mathrm{Ga}. \end{cases} \tag{5}$$

In Equation (5), $F(\mathrm{na})$ is the threshold of the advanced node, and na is the nth advanced node. r is the current round, and Ga is the combination of advanced nodes that have not become cluster heads. See Equation (6) for the different node thresholds in the SEP protocol.

$$F(\mathrm{nn}) = \begin{cases} \dfrac{\mathrm{pn}}{1 - \mathrm{pn}[r * \mathrm{mod}\ (1/\mathrm{pa})]}, & n \in \mathrm{Gn}, \\ 0, & n \notin \mathrm{Gn}. \end{cases} \tag{6}$$

In Equation (6), $F(\mathrm{nn})$ is the threshold of the advanced node, and nn is the first n ordinary node, and Gn is the combination of ordinary nodes that have not become cluster heads.

3.2. Construction of Wireless Sensor Data Fusion Algorithm. In the I-CPDA algorithm, the nodes of wireless sensors mainly refer to the components composed of small computing units with low power consumption, wireless antennas and sensors, and some nodes are composed of base stations. The clustering process of the node is shown in Figure 3. When the base station sends a data request, the node will send its address to the base station. Then other noncluster head nodes will use the signal strength as the criterion to decide which cluster to join.

In the clustering process of data fusion, several rules need to be followed to ensure the efficiency and correctness of the model. First of all, there should be no less than three nodes in a cluster, because when there is only one node in the cluster, the fusion operation cannot be performed. When there are only two nodes, although the fusion can be performed, the data privacy is extremely low and it is easy to be eavesdropped. Secondly, the probability of a node becoming a cluster head in the model is proportional to the distance between the node and its base station. This is because the node needs to bear more data volume, communication consumption, and energy consumption when it is closer to its base station. This characteristic determines that the node closer to the base station needs more cluster heads to disperse the pressure. Finally, the more remaining energy and the more surrounding neighbor nodes, the easier it is to become the cluster head node. In the fusion process, I-CPDA divides the private information into different parts and generates different false information for each part to protect the private information. In addition, each node will send the processed information to other nodes to increase the level of privacy protection. Here, there are three nodes in a cluster, X, Y, and Z, respectively; then the mathematical expression of encrypted transmission of information is shown in

$$\begin{cases} \mathrm{EN}(x_2, K_{XY}), \mathrm{EN}\left(x_2{}', K_{XY}\right), \\ \mathrm{EN}(x_3, K_{XZ}), \mathrm{EN}\left(x_3{}', K_{XZ}\right). \end{cases} \tag{7}$$

Equation (7) represents the process of a node sending the segmented private information and corresponding false information to other nodes. Among them is x, the private data in the x node, which will be X divided into x_1, x_2, and x_3 and $x_1{}'$, $x_2{}'$, and $x_3{}'$, the false data corresponding to the private data. K_{XY} represents the shared secret key between the two nodes. When the node X sends data, the node Y will also send data to node Z. Y and Z are the mathematical expressions of the data sent to the node (see the following equation).

$$\begin{cases} \mathrm{EN}(y_2, K_{YX}), \mathrm{EN}\left(y_2{}', K_{YX}\right), \\ \mathrm{EN}(y_3, K_{YZ}), \mathrm{EN}\left(y_3{}', K_{YZ}\right). \end{cases} \tag{8}$$

In Equation (8), y is the private data in the node Y. After segmenting y_1, y_2, and y_3, $y_1{}'$, $y_2{}'$, and $y_3{}'$ are the false data corresponding to the private data. The process of nodes Z sending data to other nodes can also be expressed in similar mathematics. After a node receives data sent by other nodes,

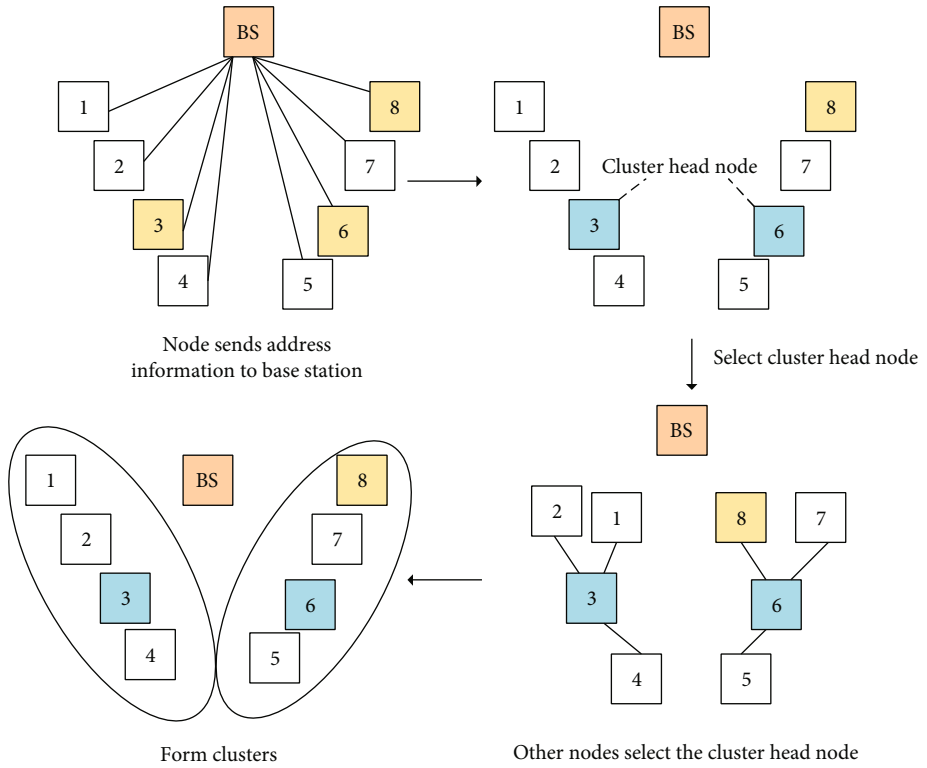

Figure 3: Schematic diagram of the clustering process.

it needs to decrypt the data through the shared secret key. The mathematical expression of data decryption is shown in

$$\begin{cases} \mathrm{DE}(y_2, K_{YX}), \mathrm{DE}\left(y_2{}', K_{YX}\right), \\ \mathrm{DE}(z_2, K_{ZX}), \mathrm{DE}\left(z_2{}', K_{ZX}\right), \\ F_X = y_2 + y_2{}' + z_2 + z_2{}'. \end{cases} \tag{9}$$

Equation (9) describes the X process of decrypting the received private data and false data by the node. DE( ) represents the process of the node receiving slice information and false slice information, in which F_X is the data value collected by the fusion node X. The mathematical expression of the decryption of the shared secret key of node Y can also be obtained:

$$\begin{cases} \mathrm{DE}(x_2, K_{XY}), \mathrm{DE}\left(x_2{}', K_{XY}\right), \\ \mathrm{DE}(z_2, K_{ZY}), \mathrm{DE}\left(z_2{}', K_{ZY}\right), \\ F_Y = x_2 + x_2{}' + z_2 + z_2{}'. \end{cases} \tag{10}$$

In Equation (10), F_y represents the data value collected by the fusion node Y. After the data is decrypted, the data fusion operation can be performed, and the data fusion process is shown in Figure 4. Each node will receive the data values of the other two nodes, besides all the fake data and the first part of the private data received by the cluster head

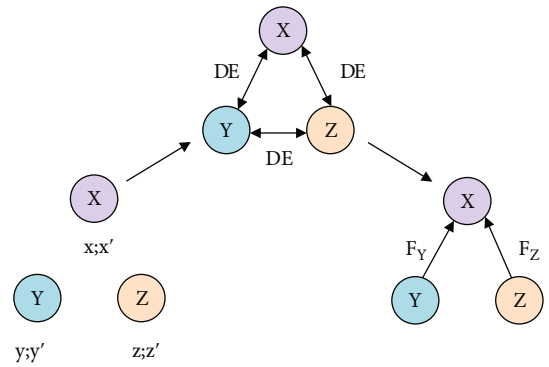

Figure 4: Schematic diagram of data fusion.

note. This mode allows each node to obtain the fusion value of all private data.

3.3. Analysis Strategy of Data Fusion Privacy Protection Algorithm. When analyzing the network data fusion algorithm and privacy data protection capability of wireless sensors, its data traffic and privacy characteristics need special attention [19]. Data traffic generally refers to communication overhead, that is, the data used and consumed when encrypting and transmitting the same private information. In the I-CPDA algorithm, each node will randomly match with the other two nodes and send their own private data slice and corresponding false information to them, and finally, all nodes will transfer the fusion value to the cluster

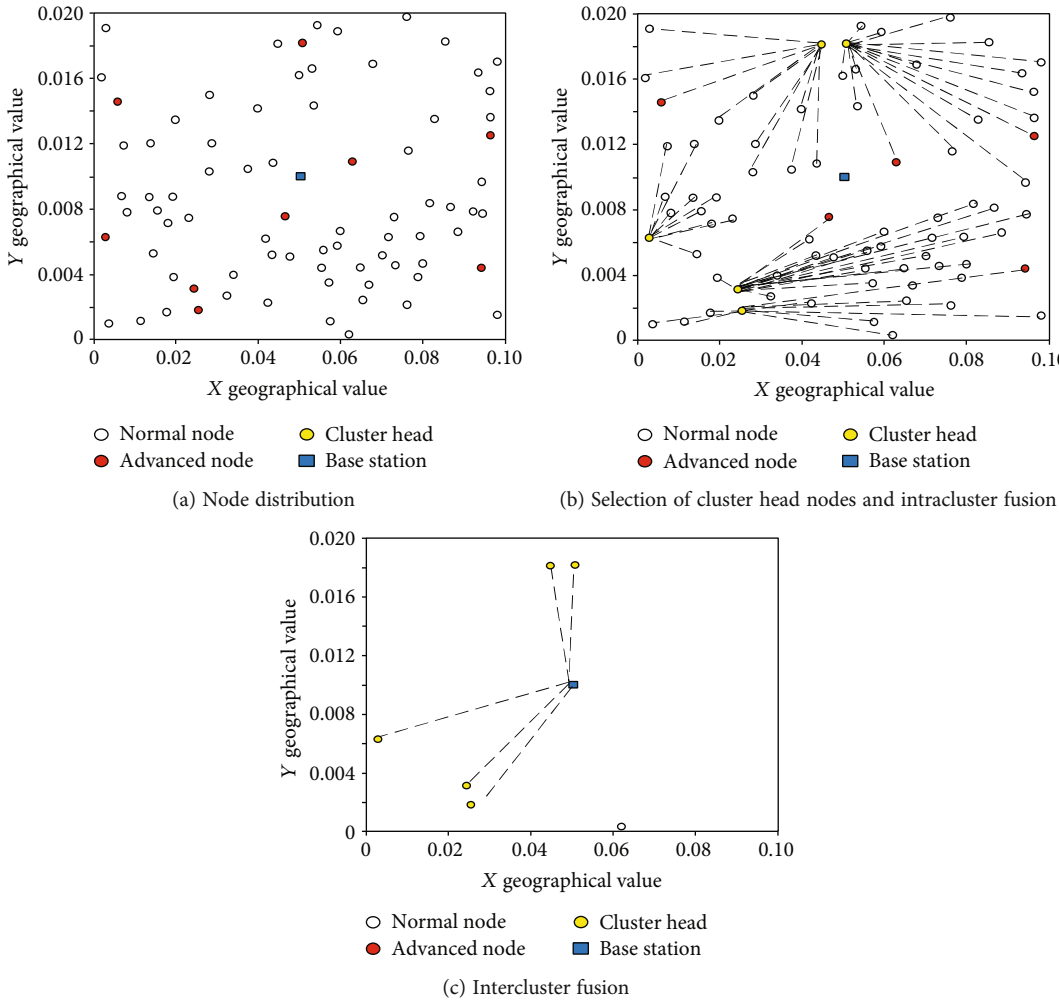

Figure 5: Distance matrix fusion process and results of sensors.

head node. In this transmission mode, the communication overhead of the I-CPDA algorithm is shown in

$$C_{\text{I-CPDA}} = 4m + m - 1. \tag{11}$$

In Equation (11), $C_{\text{I-CPDA}}$ represents the communication overhead of the I-CPDA algorithm, and m is the number of sensor nodes. After determining the communication cost of the I-CPDA algorithm, the communication cost of the SMART and CPDA algorithms can be compared, which are traditional algorithms in the field of network data fusion and privacy protection of infinite sensors. The communication overhead of CPDA is shown in

$$C_{\text{CPDA}} = m + m(m-1) + m - 1. \tag{12}$$

In Equation (12), m represents the number of sensor nodes contained in a cluster, and the broadcast seeds of sensor nodes are also m. In the CPDA algorithm, each sensor node sends encrypted interference data to its neighbor nodes, and finally, each node sends the fusion data to the cluster head node. The communication cost expression of SMART algorithm is shown in

$$C_{\text{SMART}} = NM. \tag{13}$$

In Equation (13), N represents the total number of data packets generated in the data fusion stage, which means that the original privacy data is cut into M pieces. After the calculation method of the data overhead is confirmed, the privacy calculation method of the privacy protection algorithm needs to be checked. In I-CPDA, each node sends two encrypted data, but the number of encrypted data received is not certain, so stealing information needs to crack the data sent and received by the node. To mathematically express the privacy of the I-CPDA algorithm, first make the definition as shown in

$$\begin{cases} Q_1 = C_2^{\,1} Q C_2^{\,1} Q, \\ Q_2 = \sum_{K=0}^{n-1} P(\text{in} = k) C_2^{\,1} Q^k. \end{cases} \tag{14}$$

In Equation (14), Q represents the possibility of data eavesdropping on the node link, Q_1 is the probability of information being stolen, and Q_2 is the probability that the information received by the node is eavesdropped. In addition, $P(\text{in} = k)$ is the probability that a node receives k information from other nodes, and its mathematical expression is shown in

$$P(\text{in} = k) = C^k_{n-1}\left(\frac{1}{n-1}\right)\left(\frac{n-2}{n-1}\right)^{n-1-k}. \quad (15)$$

After defining the relevant probability, the privacy measurement equation of the I-CPDA algorithm can be obtained, in which the privacy is expressed by the average probability of the node data being cracked:

$$P_{\text{I-CPDA}} = Q_1 Q_2 = C_2{}^1 Q C_2{}^1 Q \sum_{K=0}^{n-1} P(\text{in} = k) C_2{}^1 Q^k. \quad (16)$$

In addition, the privacy of the traditional CPDA algorithm can be used to compare with the I-CPDA algorithm. In the traditional CPDA algorithm, the difficulty for an eavesdropper to crack a node is influenced by the size of the cluster. When the size of the cluster is n, the eavesdropper must obtain the private $nn-1$ key sent by the node to obtain complete information. Therefore, the average node data cracked by the traditional CPDA algorithm the probability is shown in

$$P_{\text{CPDA}} = \sum_{K=n}^{d\ \max} P(n = k)\left(1 - \left(1 - Q^{k-1}\right)k\right). \quad (17)$$

According to the construction process of the algorithm, the algorithm effectively reduces the communication overhead by ensuring the privacy protection ability. In theory, the encryption complexity of this algorithm is lower than that of ordinary CPDA, but it can bring better protection effect.

4. Performance Test of Data Fusion Algorithm for Wireless Sensor Network Applied to Privacy Protection

To confirm the reliability of the I-CPDA algorithm, it is necessary to test and compare the performance of the algorithm. This test uses MATLAB as the simulator of the sensor fusion privacy protection algorithm and randomly sets nodes in a 100×100 area for fusion and evaluates and compares the performance of the algorithm during the fusion process and after the fusion is completed. First, the distance matrix fusion process and results of the sensor are observed and recorded, and the results are shown in Figure 5.

Figure 5(a) describes the node distribution of sensors in the simulator and the positions of the base station and cluster head nodes. It can be seen that in this simulation, ordinary nodes account for the vast majority, and advanced nodes only account for about one-tenth of the total number of nodes. Figure 5(b) describes the selection of cluster head nodes and the process of fusion within the cluster. There are 5 cluster head nodes in the simulation, 4 of which are advanced nodes and 1 is an ordinary node. The number gap is in line with the design principles of the algorithm. In addition, by observing the distribution of fusion within the cluster, the behavior of nodes joining the cluster follows the distance principle of the cluster. Figure 5(c) shows the process of intercluster fusion. After the data processing in the cluster is completed, the private data between different clusters will continue to be processed through the base station. The simulated sensor under the I-CPDA algorithm also successfully completes the process. After confirming the fusion process and results of the distance matrix of the sensor, the privacy, communication overhead, computational load, and accuracy of the I-CPDA algorithm are tested and analyzed. The first is the privacy of the algorithm. This test can measure the algorithm's ability to protect and encrypt private information. The test results are shown in Figure 6.

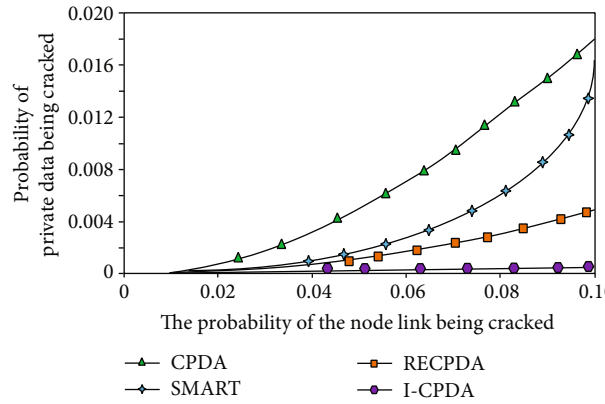

Figure 6: Privacy test results of the I-CPDA algorithm.

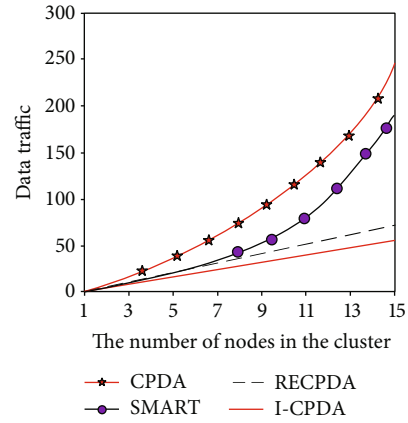

Figure 7: Communication overhead test results of I-CPDA algorithm.

Figure 6 evaluates the privacy of the I-CPDA algorithm with the probability of private data being stolen under the link cracking probability of different nodes as the index and also uses the SMART algorithm, the traditional CPDA algorithm, and another improved CPDA algorithm as the

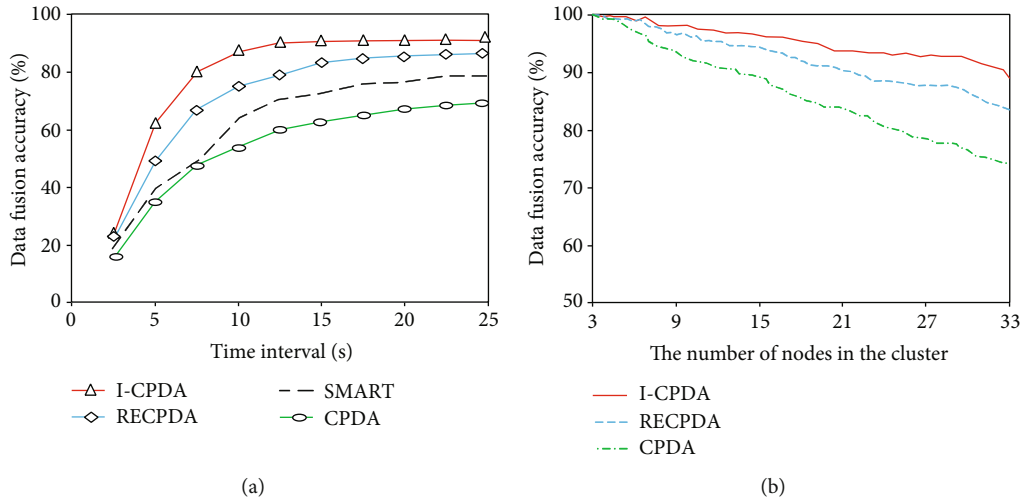

Figure 8: Data fusion accuracy test results of I-CPDA algorithm.

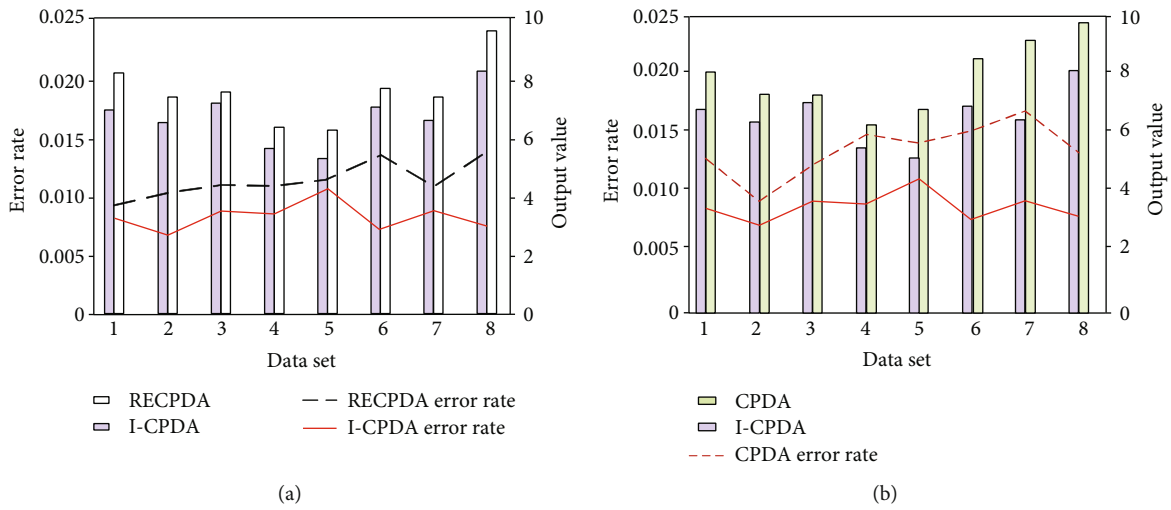

Figure 9: Data error evaluation results of I-CPDA algorithm.

comparison algorithm. In addition to the I-CPDA algorithm, another improved algorithm becomes the LECPDA algorithm. It can be seen that under the probability of the same node link being cracked, the private data theft probability of the traditional CPDA algorithm and the SMART is higher than that of the two improved algorithms. Among them, the traditional CPDA has the worst performance. When the probability of the node link being cracked is 0.1, the probability of its private data being cracked reaches 0.018. At this time, the probability of the private data being cracked by the I-CPDA algorithm is less than 0.001. From the image point of view, the slope of the I-CPDA algorithm curve is also the lowest. After analyzing the data privacy of the algorithm, analyze its communication overhead. The test results of communication overhead are shown in Figure 7.

From Figure 7, the test uses the number of nodes in the cluster as a variable to measure and compare the data traffic of different algorithms under different numbers of nodes. The comparison results show that in the whole data network, the data overhead curve of the I-CPDA is always lower than other algorithms' curve, and the gap between the algorithms increases when the number of nodes in the cluster increases. When the number of nodes in the cluster is 15, the data traffic of the I-CPDA algorithm is 56, while the data traffic of the traditional CPDA algorithm has reached 244. By comparing the results, it can be concluded that the I-CPDA algorithm has lower communication overhead than the mainstream traditional algorithms, and when there are more nodes in the data network cluster, the I-CPDA algorithm has more obvious advantages in terms of communication overhead. After comparing the communication overhead indicators, the next step is to test the data fusion accuracy of the I-CPDA algorithm. The result is shown in Figure 8.

Figure 8(a) describes the variation of the data fusion accuracy of the algorithm at different time intervals. It can

Table 2: The actual use test results of the I-CPDA algorithm.

Experiment/algorithm		I-CPDA	RE-CPDA	CPDA
EXP 1	Time (s)	116	127	133
	Interception success rate (%)	96.6	94.3	79.4
EXP 2	Time (s)	36	39	47
	Interception success rate (%)	93.5	86.6	79.1
EXP 3	Time (s)	64	71	87
	Interception success rate (%)	84.4	80.6	72.3
EXP 4	Time (s)	13	19	Twenty-three
	Interception success rate (%)	96.7	87.4	80.3
EXP 5	Time (s)	Twenty-four	31	39
	Interception success rate (%)	82.5	77.4	72.2
A level value	Time (s)	50.6	57.4	65.8
	Interception success rate (%)	90.74	85.26	76.66

be seen that the data fusion accuracy of several algorithms increases rapidly when the time interval gets larger. Among them, the data fusion of the I-CPDA algorithm accuracy rose the fastest, peaking at 90.7%. Figure 8(b) describes the variation of data fusion accuracy under different numbers of nodes in the cluster. The data fusion accuracy of several algorithms decreases with the increase of the number of nodes in the cluster. The reason for this change trend is that when the number of nodes in the cluster increases, the data that each node needs to send will increase rapidly. At this time, the probability of node collision in the cluster increases, and the accuracy of the data fusion result will decrease accordingly. After evaluating the accuracy of data fusion, the next step is to evaluate from the perspective of data error, and the evaluation results are shown in Figure 9.

The error evaluation of the I-CPDA algorithm is judged by the data error rate of the algorithm in different datasets. The bar graph in Figure 9 is the output value of different algorithms after data fusion, and the line graph is the error between the output value and the correct value. Figure 9(a) is the comparison result between the I-CPDA algorithm and the LECPDA algorithm, and Figure 9(b) is the comparison result between the I-CPDA algorithm and the traditional CPDA algorithm. Observing the data error, the error of the I-CPDA algorithm is always the lowest among the three algorithms, while the error of the LECPDA algorithm is closer to the I-CPDA than the traditional CPDA algorithm. In dataset 2, the I-CPDA algorithm exhibits the lowest error rate, which is only 0.0065. Finally, the performance of the I-CPDA algorithm in actual use is tested. Here, the running time of the algorithm and the success rate of interception and eavesdropping when dealing with private data and attacks are used as comparison indicators. The result is shown in Table 2.

The time and interception success rate of the I-CPDA algorithm are higher than those of the traditional CPDA and LECPDA algorithms used for comparison in each experiment. From the average of multiple trials, the average time of the I-CPDA algorithm is 50.6 s, while that in the traditional CPDA algorithm is 65.8 s. The average interception success rate of the I-CPDA algorithm is 90.74 (%), and that in the traditional CPDA is only 76.66 (%). The results show that the improvement of the CPDA algorithm is practical.

5. Conclusion

The problems of data fusion and privacy data protection in data processing and transmission of wireless sensor networks have always been the difficulties that need to be overcome in the field of wireless sensors. The traditional data fusion privacy protection algorithm has the problems of high communication overhead, insufficient security level, and slow speed. Aiming at these problems, an improved I-CPDA algorithm is designed on the basis of traditional CPDA. The algorithm uses data slicing technology and generates false interference data to improve the security level and is optimized for the problem of excessive data traffic in traditional algorithms. The performance of the I-CPDA data fusion privacy protection algorithm is tested in an all-round way, mainly using the traditional CPDA and SMART to compare with the I-CPDA. The test results show that when the probability of the node link being cracked is 0.1, the probability of cracking the private data of the I-CPDA algorithm is less than 0.001, while the probability of cracking the private data of the traditional CPDA algorithm reaches 0.018. The data traffic of the I-CPDA algorithm is 56 under the condition of 15 nodes, while the data traffic of the traditional CPDA algorithm has reached 244, and the data traffic of the SMART algorithm is 191. The research on the I-CPDA algorithm has achieved relatively successful results. The algorithm is better than the traditional one in terms of communication overhead, privacy, and fusion accuracy. In conclusion, the contribution of this research is to provide an effective privacy data protection algorithm for wireless sensor networks, which can protect data more effectively than traditional algorithms. There is still room for further exploration in this study. In the actual application scenario of wireless sensor networks, there are many potential influencing factors, which may affect the stability of the method. The experiment of I-CPDA algorithm in this study is mainly simulated by MATLAB software and only for the

performance of the algorithm itself, and it is not applied to the scene where wireless sensors are needed. This experimental method leads to the lack of practical application data in research to a certain extent, so the performance of I-CPDA algorithm in real scenes is the next research direction.

Acronyms

CPDA: Cluster-based privacy data aggregation
I-CPDA: Improved cluster-based privacy data aggregation
SMART: Slice-mix-aggregate
SEP: Secure exchange protocol
LEACH: Low energy adaptive clustering hierarchy
LE-CPDA: Low energy cluster-based privacy data aggregation.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the authors that this article is free of conflict of interest.

References

[1] Y. Jiang and N. Kabaoglu, "SEEHPIP: secure energy efficient homomorphism based privacy and integrity preserving data aggregation for wireless sensor networks," *International Journal of Computational Intelligence Research*, vol. 14, no. 11, pp. 893–910, 2018.

[2] S. Hu, L. Liu, L. Fang, F. Zhou, and R. Ye, "A novel energy-efficient and privacy-preserving data aggregation for WSNs," *IEEE Access*, vol. 8, pp. 802–813, 2020.

[3] X. Qi, X. Liu, J. Yu, and Q. Zhang, "A privacy data aggregation scheme for wireless sensor networks," *Procedia Computer Science*, vol. 174, pp. 578–583, 2020.

[4] V. B. Christopher and J. Jasper, "Jellyfish dynamic routing protocol with mobile sink for location privacy and congestion avoidance in wireless sensor networks," *Journal of Systems Architecture*, vol. 112, no. 3, article 101840, 2020.

[5] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, N. Kumar, and M. M. Hassan, "A privacy-preserving-based secure framework using blockchain-enabled deep-learning in cooperative intelligent transport system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16492–16503, 2022.

[6] M. Mahdavisharif, S. Jamali, and R. Fotohi, "Big data-aware intrusion detection system in communication networks: a deep learning approach," *Journal of Grid Computing*, vol. 19, no. 4, pp. 1–28, 2021.

[7] L. Feng and B. Liu, "Low-energy data fusion privacy protection algorithm for three-dimensional wireless sensor network," *Mobile Information Systems*, vol. 2022, Article ID 3580607, 10 pages, 2022.

[8] K. Pramod, S. Pierluigi, and C. Domenico, "Distributed detection in wireless sensor networks under multiplicative fading via generalized score tests," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9059–9071, 2021.

[9] I. Nasurulla and R. Kaniezhil, "Integration of fault-tolerant feature to OMIEEPB routing protocol in wireless sensor network," *International Journal of Intelligent Computing and Cybernetics*, vol. 15, no. 3, pp. 414–424, 2022.

[10] V. S. Lakshmi and P. P. Deepthi, "A secure channel code-based scheme for privacy preserving data aggregation in wireless sensor networks," *International Journal of Communication Systems*, vol. 32, no. 1, article e3832, 2019.

[11] X. Zhang, L. Zhao, X. Gao, and X. Zhang, "A data-sharing model based on blockchain for power grid big data," *Journal of Physics: Conference Series*, vol. 1792, no. 1, article 012051, 2021.

[12] S. Jiang, M. Li, and Z. Tang, "A new scheme for source-location privacy in wireless sensor networks," *International Journal of Network Security*, vol. 20, no. 5, pp. 879–888, 2018.

[13] M. K. Alam, A. A. Aziz, S. A. Latif, and A. A. Aziz, "Error-control truncated SVD technique for in-network data compression in wireless sensor networks," *IEEE Access*, vol. 9, pp. 13829–13844, 2021.

[14] P. Giri, K. Ng, and W. Phillips, "Wireless sensor network system for landslide monitoring and warning," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 1210–1220, 2019.

[15] K. Roopakumar, "Novel energy-efficient secure routing protocol for wireless sensor networks with Mobile sink," *Peer-to-Peer Networking and Applications*, vol. 12, no. 4, pp. 881–892, 2019.

[16] S. M. Elsherif, M. Elshrkawey, and M. E. Wahed, "An efficient secure scheme for data aggregation in wireless sensor networks using the additive property of complex numbers," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 9, pp. 2649–2664, 2018.

[17] C. Gu, M. Bradbury, J. Kirton, and A. Jhumka, "A decision theoretic framework for selecting source location privacy aware routing protocols in wireless sensor networks," *Future Generation Computer Systems*, vol. 87, pp. 514–526, 2018.

[18] S. Ji, Z. Gui, T. Zhou, H. Yan, and J. Shen, "An efficient and certificateless conditional privacy-preserving authentication scheme for wireless body area networks big data services," *IEEE Access*, vol. 6, pp. 69603–69611, 2018.

[19] K. Renuka, S. Kumar, S. Kumari, and C. M. Chen, "Cryptanalysis and improvement of a privacy-preserving three-factor authentication protocol for wireless sensor networks," *Sensors*, vol. 19, no. 21, p. 4625, 2019.