

## Security fusion method of physical fitness training data based on the Internet of Things

Bin ZHOU\*

Harbin Sport University, Harbin, China

Received: 25.08.2023

Accepted/Published Online: 28.11.2023

Final Version: ..202

**Abstract:** Physical fitness training, an important way to improve physical fitness, is the basic guarantee for forming combat effectiveness. At present, the evaluation types of physical fitness training are mostly conducted manually. It has problems such as low efficiency, high consumption of human and material resources, and subjective factors affecting the evaluation results. "Internet+" has greatly expanded the traditional network from the perspective of technological convergence and network coverage objects. It has expedited and promoted the rapid development of Internet of Things (IoT) technology and its applications. The IoT with many sensor nodes shows the characteristics of acquisition information redundancy, node energy sensitivity, network distribution openness, data demand reliability, etc. Thus, the research on data security fusion method of the IoT has important theoretical significance and application prospects. In order to ensure the authenticity and reliability of the fusion results of physical fitness training data, the security characteristics and performance of the IoT are analyzed, and the basic requirements for the security fusion of IoT sensory data are identified. An improved cluster-based data fusion model is proposed to address the shortcomings of the cluster-based data fusion model, and a security fusion method of physical fitness training data is studied. Finally, this article conducts a large number of simulation experiments. The experimental results show that the improved cluster-based data fusion model has better performance, further improving the security of physical fitness training data fusion based on the IoT. Finally, the article provides a security performance analysis.

**Key words:** Data security fusion, physical fitness training, Internet of Things, improved cluster-based data fusion model

### 1. Introduction

Physical fitness training is an important means to strengthen the physique of soldiers, enhance their combat effectiveness, and complete various battlefield operations [1]. Soldier physical fitness training, a crucial component of military training, is of great significance in achieving the strategic goal of building a strong military. Whether engaging in emergency rescue missions or offensive operations, adequate physical fitness is indispensable. Therefore, for soldiers, physical fitness training directly translates to visible combat effectiveness.

Physical fitness training, as the cornerstone of physical exercise, is an essential component of military training and preparation, constitutes the primary focus of such training. It carries important decisive factors in warfare at the micro level [2]. Given the increasingly complex and ever-changing battlefield conditions, maintaining a high level of physical fitness is crucial for readiness in any war scenario. The physical robustness of soldiers directly affect whether they can maintain the best physical condition when performing tasks in battle.

\*Correspondence: [zhoubin@hrbipe.edu.cn](mailto:zhoubin@hrbipe.edu.cn)

Superior physical fitness empowers soldiers to achieve twice the result with half the effort on the battlefield and better serve the operational needs.

Physical training data can intuitively reflect the training results of soldiers and is an effective way for officers to master the training results of soldiers [3]. Scientific physical fitness training activities can effectively improve the physical and psychological qualities of officers and soldiers. Currently, manual evaluation methods are often used to evaluate general physical fitness training projects for soldiers, and there are some problems with manual methods. One is that a large amount of manpower and material resources need to be invested in the evaluation of soldier training effectiveness, which results in poor economic benefits. Secondly, there are factors such as fatigue period in manual evaluation, which cannot guarantee long-term continuous evaluation. The judgment of the standard level for soldier training actions is influenced by subjective factors of the evaluator, which can easily lead to misjudgment and other situations.

With the widespread research on technologies such as the IoT, the computing and storage capabilities of mobile terminals will become stronger. IoT devices and even the entire IoT system will gradually become intelligent [4]. Physical fitness training systems are gradually moving towards automation, efficiency, and intelligence. Compared to manual evaluation methods, intelligent physical training systems have better real-time feedback. Hardware devices meet practical needs such as continuous high-intensity work, significantly improving the efficiency of physical training. The intelligent IoT mainly refers to the IoT that is connected by intelligent IoT nodes, which have strong computing and storage capabilities. It can support higher security protocols and algorithms [5].

By collecting useful information from a large number of nodes in the IoT environment, we can use this information for various analyses and processing. Due to energy limitations and data-centric characteristics, data fusion technology is widely used in the IoT [6]. As a key technology in the perception layer of the IoT, data fusion plays a crucial role in saving energy for the entire network, improving the accuracy of data collection, and improving the efficiency of data collection. However, in most practical application scenarios, nodes are deployed in remote, open, sensitive, and uncontrollable areas. Consequently, data privacy security protection has become a hot issue in IoT data fusion [7]. For example, in the node network deployed on the battlefield to collect enemy military information, attackers can not only eavesdrop on the information transmitted by nodes, but also capture nodes publishing forged information. The obtained data may be invalid or harmful.

Data fusion is a very important technology in the perception layer of the IoT and is currently a research hotspot [8]. By using this technology, a large number of raw data collected by sensor nodes can be processed through some specific algorithms to remove unnecessary information. Only a small amount of meaningful data results can be transmitted to the fusion node. The use of data fusion technology can greatly reduce the data traffic transmitted in wireless sensor networks, network burden, and network congestion, thereby extending the lifespan of the network. However, due to the frequent fusion and exposure of data at fusion nodes during the fusion process, the security of data fusion faces challenges [9].

A large number of sensor nodes in the perception layer of the IoT are generally distributed in unsupervised harsh environments or security sensitive areas. Wireless communication is used between nodes, making data fusion in the network face various information security risks, such as data eavesdropping, forgery, tampering, and replay. The security mechanisms used in traditional networks require high storage space and computational complexity for nodes [10]. In addition to its inherent resource and energy limitations, data fusion in the IoT also faces unprecedented security challenges [11]. Data fusion in the IoT is extremely necessary. The fused data can reduce data traffic in the network, thereby avoiding network resource waste. At the same time, the data

obtained from the fusion operation is more intuitive, which can bring great convenience to users' observation and analysis of physical training. However, in order to resist attacks such as node capture and data theft during the fusion process, previous works that only considered energy-saving data fusion would have significant practical limitations. The application of these security technologies in the IoT, which is severely lacking in communication capacity and storage space, is greatly limited. Therefore, it is necessary to propose a physical training data fusion scheme that adapts to the characteristics of the IoT network. The main contributions of this article are as follows.

(1) In order to ensure the authenticity and reliability of the fusion results of physical fitness training data, the security characteristics and performance of the IoT are analyzed in this paper. Additionally, the basic requirements for the security fusion of IoT sensory data are identified. An improved cluster-based data fusion model is proposed to address the shortcomings of the cluster-based data fusion model. And a security fusion method of physical fitness training data is studied.

(2) Through simulation experiment and evaluation analysis, the improved cluster-based data fusion model is compared with other models from four aspects: data fusion value, expected value of data fusion result evaluation, node reputation value, and malicious node detection rate. It is proved that the improved cluster-based data fusion model achieves better results in physical training data fusion. Finally, the safety performance is analyzed.

The remaining part of this article consists of four parts. Section 2 presents the related literature. Section 3 provides a detailed introduction to the secure fusion method of physical fitness training data based on the IoT. Section 4 analyzes the proposed method and its effectiveness through systematic experiments and indicators. Finally, in Section 5, the main research content and conclusions of this article are summarized.

## 2. Related work

Physical stamina comprises body shape, body function, and physical quality, serving as the fundamental cornerstone of training, and it is indispensable. Research on physical stamina has been extensively undertaken in various sectors including the military, schools, communities, and among the elderly, reflecting an increasingly wide range [12]. Therefore, it is necessary to conduct in-depth research on physical fitness training, deepen the emphasis on physical fitness, and demonstrate professionalism. The IoT is an information-based and intelligent network that utilized various communication technologies such as short distance wireless networks, local area networks, and private networks. The Internet connects things and people through various means, forming a connection between things and people, achieving remote management and control. The IoT extends the end of the Internet, encompassing not only the Internet and all its inherent resources, but also its existing applications, making all its elements personalized and private. At present, the IoT technology is actively being developed, and the application of the IoT has gradually developed towards large-scale promotion. However, security has become a core issue that restricts its further development. The process of information collection and data fusion in the IoT face various information security threats [13]. Therefore, constructing data security fusion in the IoT has practical significance[14].

He and Nguyen et al. [15] proposed two schemes, CPDA and SMART, based on perturbation and segmentation recombination methods to protect data fusion privacy and confidentiality. In CPDA, sensor nodes hid real data values by adding random seeds and private random numbers to the original data for perturbation processing. Cluster head nodes used the algebraic properties of polynomials to solve accurate summation and fusion results. SMART used segmentation and recombination technology to achieve privacy protection data

fusion. The basic idea of SMART was that the sensor node randomly divided the original data into multiple data slices, used hop by hop encryption mechanism to exchange data slices with randomly selected neighbor peer exchange. It performed summation operation on all received data slices, and uploaded the summation results to the base station. The base station summed all received data to obtain accurate summation and fusion results. Girao et al. [16] proposed a scheme that uses doomingo ferrer homomorphic encryption mode to achieve sensor Internet privacy data fusion. This scheme had the advantages of low transmission overhead and simple computation, but it came with lower security. In the scheme, since each node shared the same key with the base station and encrypted its own collected data, any node could easily obtain data from adjacent nodes. If an attacker captured a node, the entire network data was in control. Casteliuccia et al. [17] first proposed the scheme using the addition homomorphic encryption method, which was an ideal scheme with relatively simple calculation and communication in the fusion process. It could ensure good security. However, the scheme still had many drawbacks, and the significant communication overhead caused by the rapid expansion of transmission was the most obvious problem of this scheme. Secondly, data loss also resulted in the inability of this scheme to be applied in practice, as the network could not know the ID of a node after it dropped due to a sudden situation. Another scenario was that the unresponsive node ID might be lost during transmission, ultimately resulting in the base station not receiving the correct decryption result. Papadopoulos et al. [18] used summation homomorphic encryption functions and shared secret data and other technologies to complete data fusion on ciphertext, providing a certain degree of integrity verification. However, the scheme directly ignored ID inflation and data loss. In the aspect of full homomorphic encryption, Gentry et al. [19] proposed a lattice-based bounded homomorphic encryption scheme. The literature pointed out that the scheme could meet the addition homomorphism and could save the vector space structure of messages. However, the public key was quite large, but the encryption and decryption operations were quite fast. The process of encryption or decryption, or the key generation process was quite complex, which was difficult to apply in the IoT with limited resources.

The scheme introduced in reference [20] considered node weights as node reputation values to complete data fusion operations. Subsequently, the actual values were compared with the fusion results. If the values were inconsistent, it indicated that the corresponding node was suspicious. A penalty factor was then used to reduce the reputation value of the node. Through continuous iteration of the reputation value and fusion value, the presence of malicious nodes was determined through screening. The algorithm proposed in reference [21] had been improved to a certain extent based on this algorithm. It removed nodes with reputation values lower than the average reputation value and only allowed high reputation value nodes to perform subsequent fusion steps. This scheme could reduce the difference between the fusion value and the actual value. The authors in [22] adopted the principle of divide and conquer. The tree topology network of nodes was first dynamically divided into many logical groups of similar size. The next data fusion operation would be carried out step by step in each logical group, and the final fusion result would be transmitted to the base station. The main task of a base station was to analyze the fusion result set of these logical groups, then identify and eliminate suspicious logical groups. The authors in [23] proposed an efficient and secure data fusion protocol based on pattern codes. The secure data fusion operation carried out by this protocol used pattern codes without any physical significance. During data transmission, intermediate nodes did not care about the specific content of information, so there was no need to decrypt and encrypt the ciphertext. This ensured the confidentiality of the data and avoided the problem of message eavesdropping at intermediate nodes. The periodic broadcast of keys by the base station also helped to ensure the freshness of the data.

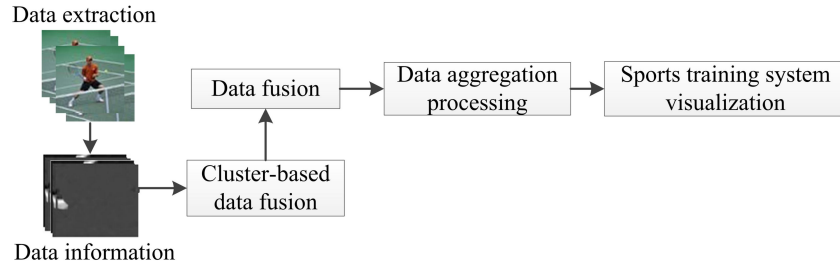
In summary, the first consideration in the IoT is how to reduce energy consumption and extend the life of network activities. Therefore, combining data fusion research has become an important branch of the IoT. Studying algorithms and methods for secure data fusion suitable in physical fitness training networks has important theoretical and practical significance. In order to ensure the authenticity and reliability of the data fusion results of the IoT, this article will focus on researching the secure data fusion technology of the IoT network, aiming to propose a safe and efficient IoT data fusion method, which is of great significance for the practical application and promotion of the IoT in various fields.

### 3. Security fusion method of physical fitness training data

Data fusion is a very important technology in the perception layer of the IoT, and it is also a research hotspot in current physical fitness training. By using this technology, a large number of physical training raw data collected by sensor nodes can be processed through some specific algorithms to remove redundant information, and only a small amount of meaningful physical training data results can be transmitted to the fusion node.

#### 3.1. Physical training data preprocessing

The processing process of physical fitness training involves first extracting optical flow from the collected physical fitness training videos to obtain optical flow information. The optical flow information and physical fitness training action videos are used as inputs to the fusion network to obtain key point information and fused feature information of the human skeleton in the video. The fused feature information is then input into the detection model and fused with key point information of the human skeleton to locate and classify the video action sequence. Finally, the key frames of the action are extracted, comparative analysis of the actions are performed, the results are obtained, and they are displayed on the system interface. The preprocessing process is displayed in Figure 1.



**Figure 1.** Physical training data preprocessing.

The reputation of the assumed malicious nodes in the cluster are calculated at the end of the experiment. According to the provisions of the cluster-based data fusion model [24], as these nodes' reputation is not higher than the average reputation, they do not belong to the high reputation group. In addition, the object of physical training data fusion operation is high reputation group data. Therefore, even if some nodes are captured, these captured nodes do not participate in the specific physical training data fusion operation and cannot affect the final fusion result. The security data fusion model can still guarantee the authenticity and reliability of the physical training data fusion result, that is, it has good fault tolerance. The simulation results of the fusion results conducted in the cluster-based data fusion model compared with real physical training data also prove the effectiveness of the model.

Optical flow is generated by the motion or relative motion of objects themselves and is the instantaneous velocity of pixels on the observation plane [25]. The essence of optical flow is a two-dimensional vector field, where each vector represents the displacement of a point in the current image from the previous frame to the next frame, reflecting the trend of grayscale changes of pixels in the image. Total variation-L1 (TV-L1) optical flow [26] is used to extract optical flow information. This method is improved on the basis of Horn-Schunck optical flow [27] to improve the robustness of the algorithm. TV-L1 uses an energy function different from Horn-Schunck, which includes a data item using L1 norm and a regularization using population variation. The energy function of TV-L1 is shown in Formula 1, where  $u$  and  $v$  are two-dimensional vector fields,  $\lambda$  is a weight constant, and  $\theta$  is an auxiliary variable.

$$E_{\theta}(u, v) = \int_{\Omega} \nabla u_1 + \nabla u_2 + \frac{1}{2\theta} |u - v|^2 + \lambda |\rho(v)| \quad (1)$$

Because wireless sensor networks are data-centric, utilizing data processing techniques within the network can reduce excessive energy consumption. This involves using data fusion technology to address these issues. The authors in [28] conducted a detailed discussion and research on the function of data fusion technology from two aspects: theoretical analysis and simulation testing. The research results indicate that when using data fusion technology, the ratio of network energy cost is displayed in Equation 2.

$$\lim_{d \rightarrow +\infty} \frac{N_D}{N_A} = \frac{1}{k} \quad (2)$$

Among them,  $N_D$  is the amount of data transmission times in the network that uses data fusion, and  $N_A$  is the amount of data transmission times in the network that does not use data fusion.  $D$  is the distance from the sensor node to the fusion node, and  $k$  represents the amount of data collection source nodes.

### 3.2. Improved cluster-based data fusion model

Although the cluster-based data fusion model has good fault tolerance and effectiveness, it also has some obvious problems. For example, the fusion node needs to calculate the reputation values of all ordinary member nodes, which is a waste of already scarce network resources. In addition, the method of calculating node reputation values in this model may conceal the current malicious behavior of the member nodes' historical accumulated reputation [29]. In response to these issues in the cluster-based data fusion model, this article proposes an improved cluster-based data fusion model, as shown in Figure 2.

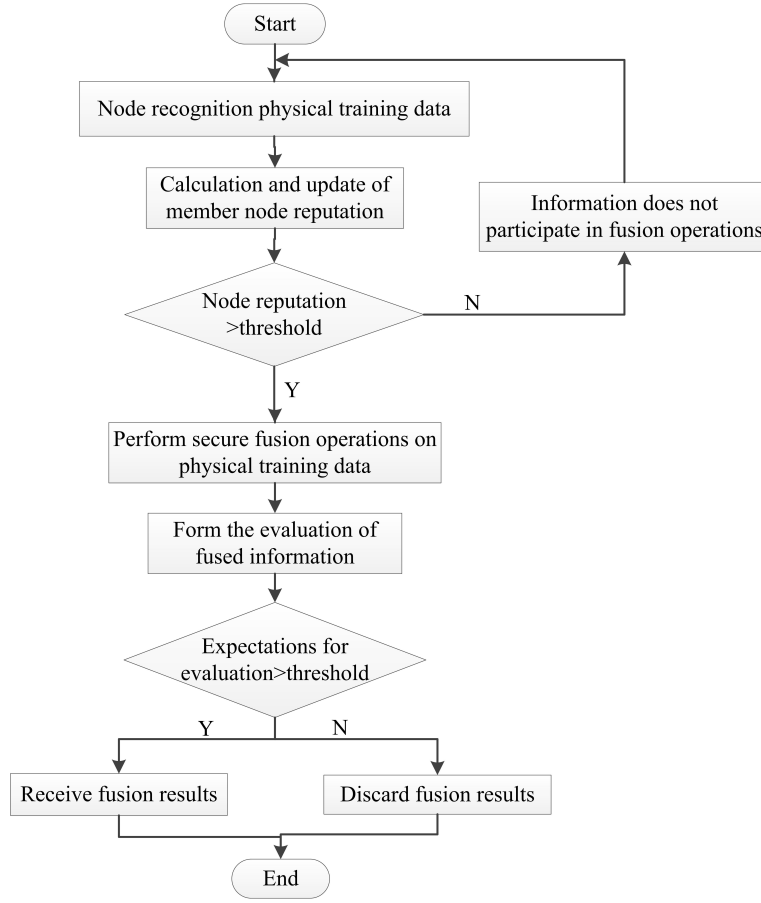
Based on the above process, the specific steps of the improved security data fusion model can be described as follows.

Step 1: After receiving the perception information sent by the member nodes, the fusion node first identifies data that significantly deviates from the perception results of nearby nodes through error theory, and removes it as malicious or erroneous data.

Step 2: Next, the reputation values of the remaining member nodes are calculated and updated to determine the trust node.

Step 3: The node reputation threshold is compared, and the fusion node only allows the perception data of nodes with high reputation to participate in the fusion step.

Step 4: The fusion result is evaluated, and then both the fusion result and the evaluation are sent to the base station together.



**Figure 2.** Security data fusion model flowchart.

Step 5: The credibility of the node determines whether to use the collected data for data fusion operations.

The fusion results and evaluation results are sent together to the base station for final decision-making and data utilization. Due to the dense distribution of sensing nodes in the IoT sensing layer, there must be redundancy in the data perceived by adjacent nodes in the ground. Based on statistics and information theory, the original data sent by the detection node is used to estimate the reputation of each node, which is used to measure the trustworthiness of each node. Based on the level of node reputation, it is determined whether to use the collected data for data fusion operations. Therefore, uncertainty in fusion results can be measured through evaluation based on node reputation. The main improvement of this method is that before calculating the reputation value of member nodes, the cluster head is required to first identify and eliminate data from nodes with malicious tendencies. The data based on sensor networks has redundancy and the data collected by such nodes deviates significantly from the data of nearby nodes, and their reputation value is no longer calculated. The data sent by them does not participate in fusion operations. This can not only save energy resources in the network but also avoid concealing the current malicious behavior of member nodes due to their historical accumulated reputation. It can also obtain a higher evaluation of the fusion results, meaning higher degree of trustworthiness of the fusion results.

For the distribution of the measured value  $u_1, u_2, \dots, u_n$  from small to large order statistic  $u(i)$ , when  $u_i$  follows normal distribution, the statistics of the maximum value  $u(n)$  is obtained as follows.

$$r_{10} = \frac{u(n) - u(n-1)}{u(n) - u(1)} \quad (3)$$

$$r_{11} = \frac{u(n) - u(n-1)}{u(n) - u(2)} \quad (4)$$

$$r_{21} = \frac{u(n) - u(n-2)}{u(n) - u(2)} \quad (5)$$

$$r_{22} = \frac{u(n) - u(n-2)}{u(n) - u(3)} \quad (6)$$

Similarly, the statistic of the minimum value  $u(1)$  is represented by the following equation.

$$r_{10} = \frac{u(1) - u(2)}{u(1) - u(n)} \quad (7)$$

$$r_{11} = \frac{u(1) - u(2)}{u(1) - u(n-1)} \quad (8)$$

$$r_{21} = \frac{u(1) - u(3)}{u(1) - u(n-1)} \quad (9)$$

$$r_{22} = \frac{u(1) - u(3)}{u(1) - u(n-2)} \quad (10)$$

In order to eliminate gross errors, the dixon criterion  $n \leq 7$  suggests that selecting  $r_{10}$  is effective. When  $8 \leq n \leq 10$  is selected,  $r_{11}$  has a good effect. When  $11 \leq n \leq 13$  is selected,  $r_{21}$  has a good effect. When  $n \geq 14$  is selected,  $r_{22}$  has a good effect. Here,  $n$  represents the number of data participating in the fusion. For example, when fusing in clusters,  $n$  is the number of nodes within the cluster. Significance  $\alpha$  is selected, and the values of 0.01 or 0.05 are obtained. And it is combined with the number of data  $n$  participating in the fusion. According to the Dickson criterion, the critical value  $r_0(n, \alpha)$  of the corresponding statistic can be obtained by looking up the table. If the measured statistical value  $r_{ij}$  is higher than the critical value, it is considered that  $u(n)$  contains coarse errors.

### 3.3. An improved physical fitness training data security fusion scheme based on the IoT

The perception layer of the IoT contains thousands of sensor nodes that independently sense the external environment within their respective distribution areas. In general, the data sensed by these sensor nodes follows the normal distribution rule. The data sent by the captured normal nodes or malicious nodes will significantly deviate from the normal distribution; otherwise, the purpose of destroying the system will not be achieved. Therefore, refer to the Josang trust model [30] and the normal distribution rule is used to calculate the reputation value of the node. Ideally, the probability that the value of a normal random variable is located



in the central value  $[-\sigma, +\sigma]$  area is 0.68, that is the Bernoulli distribution. When malicious nodes continue to send forged data, the actual probability distribution will be inconsistent with this probability. Using the ideal node probability as a benchmark, the difference between the true and ideal values of the node probability distribution can be represented by distance, which can represent the reputation value of the node. There is a negative proportional relationship between the two.

If the probability of the physical fitness training data output by a node falling within one time and the standard deviation from the center value is  $p_i$ , then the probability of being outside this area is  $1 - p_i$ . The degree of deviation of the node can be expressed as shown in Equation 11.

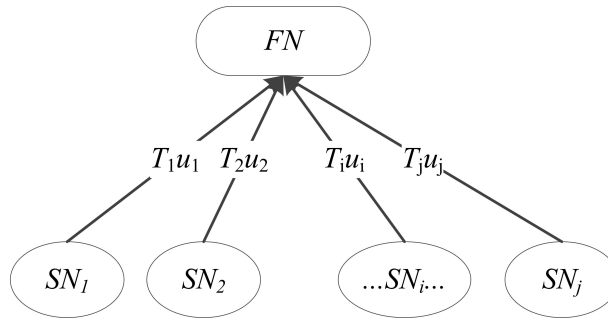
$$D_i = \left| (1 - p_i) \log_2\left(\frac{1 - p_i}{0.32}\right) - p_i \log_2\left(\frac{p_i}{0.68}\right) \right| \quad (11)$$

The definition of the reputation value of the corresponding node is shown in Equation 12.

$$T_i = e^{-\sqrt{D_i}} - kD_i \quad (12)$$

In the equation, the first half is an exponential operation that can reflect the reputation value of the node through the distance between the actual probability and the ideal probability of the node, and can also reflect the historical cumulative behavior of the node. The latter part is a penalty measure, where  $k$  is the penalty factor. The introduction of the penalty factor is another improvement of the improved cluster-based data fusion model. This factor can avoid the historical accumulated reputation of member nodes from masking their current malicious behavior, which helps to achieve a slow increase and fast decrease effect when calculating and updating node reputation values, making it easy to quickly discover and identify malicious nodes. This is very beneficial for the application of the IoT perception layer. As the number of iterations increases, the reputation value of nodes continues to accumulate and update. If the current reputation value of a node is lower than the system preset threshold  $T_0$ , the system automatically determines it as a malicious node. The data is no longer used.

The improved cluster-based data fusion model utilizes the output values of high reputation nodes for weighted fitness training data fusion. The weighted fitness training fusion method is shown in Figure 3 [31]. In Figure 3,  $FN$  represents the fusion node, such as the cluster head, and  $SN$  represents the ordinary node, such as the ordinary node within the cluster.  $T$  and  $u$  represent the reputation value and perception data of the node, respectively.



**Figure 3.** Schematic diagram of weighted physical fitness training data fusion.

The fusion node calculates the weighted fusion result according to equation 13 based on the reputation value and perception data of each ordinary node.

$$U = \frac{\sum_{i=0}^n T_i * u_i}{\sum_{i=0}^n T_i} (T_i > Th) \quad (13)$$

Only node data with a reputation value higher than the threshold is allowed to participate in the fusion here, which not only isolates the negative impact of malicious data but also reduces the computational complexity of physical training data fusion. In response to the uncertainty factors of the fusion results of physical fitness training data and the transmission of information, this article adopts the Josang trust model. This model measures the trustworthiness of fusion results through trust defined as evaluation.  $W = (b, d, u, a)$  is evaluated, where the vectors  $a$ ,  $b$ ,  $d$ , and  $u$  represent the relative coefficient, trust, distrust, and uncertainty of the fusion result  $U$ , respectively. The expected probability of evaluation can be expressed as equation 14.

$$E(W) = b + u \times a \quad (14)$$

The expected probability depends on the comprehensive result of trust and uncertainty, and the role of  $a$  is to reflect the effect of uncertainty on the expected probability of evaluation. The expectation of evaluating the fusion data results of physical fitness training is closely related to the current and cumulative reputation values of the node. Through equation 14, it can be known that the expected probability of evaluation depends on the trust and uncertainty of the node. That is the size of the node's reputation value, which includes the node's cumulative reputation value and the current reputation value. The improved cluster-based data fusion model incorporates a penalty measure when calculating the current reputation value of nodes, which helps to achieve a slow increase and fast decrease in the calculation and update of node reputation values. It can avoid the historical accumulated reputation of member nodes masking their current malicious behavior, and thus can achieve high expectations for evaluating the fusion data results of physical training.

## 4. Experiment and result analysis

### 4.1. Basic framework

The basic framework of a physical fitness training data security fusion system based on the IoT includes three core parts: data collection end, server end, and query client.

The data collection end is composed of physical training data collection nodes, which adopt a tree topology structure. These nodes are divided into leaf nodes and fusion nodes. The data fusion node at the top is connected to the server end through an internal network. The physical training data collection network simultaneously deploys a secure and efficient perception layer data fusion protocol. The physical training data collection end consists of an initialization module and a physical training data collection module. The initialization module is responsible for generating the topology structure, initializing the physical training data security fusion protocol, preparing the data collection module, and establishing a secure connection with the server. The physical training data acquisition module enters the waiting mode after initialization, waiting for the data update command from the server for physical training data acquisition and safety fusion.

The server is composed of a personal computer (PC) with strong computing and storage capabilities, as well as high security. It is connected to the top data fusion node of the physical training data collection end through an internal network. On the other hand, it provides data query and download functions for the query client through access control. The server side is composed of initialization module, data update module, user

registration module, query module, and backend database. The initialization module starts the main process, starts the data update module, regularly sends update requests for fused physical training data to the physical training data collection end, and responds to the registration and data query requirements of multiple terminals by establishing a multithreaded web server. In the experiment, the server side was handled by a PC, the database used was MySQL database. The main program was written in Java.

The query client is a user terminal that can be connected to the server through the internet. After user verification, the data stored on the server by the physical training data collection end can be queried and downloaded.

#### 4.2. Experimental results and analysis

Theoretical analysis is conducted on the security performance, energy consumption, algorithm complexity, and hardware consumption of the improved cluster-based data fusion model proposed in this article to verify its security, efficiency advantages, and adaptability [32]. Joint Directors of Laboratories (JDL) model is a data processing method in the data fusion system [33]. In order to test the performance of improved cluster-based data fusion model, we will compare the experiment with JDL model. Simulation testing is conducted to compare and evaluate the performance of the improved cluster-based data fusion model and other models from four aspects: data fusion value, expected value of data fusion result evaluation, node reputation value, and malicious node detection rate [34].

Suppose that the sensor network of the sensing layer of the IoT has used some algorithm to cluster, each cluster has a cluster head node and several member nodes, and the sensing data of each node follows the normal distribution law [35]. The simulation parameters involved in this article's simulation experiments mainly include the number of experimental iteration rounds  $R$ , the malicious node ratio  $P$  in the network, the reputation penalty factor  $k$ , the data fusion trust threshold  $Th$ , and the amount of nodes in the cluster  $n$ . Considering the actual network structure and the relevant parameter settings of the original model, in the simulation experiment of this article, their values are 20, 0.03, 0.04, 0.3, and 20, respectively. The experimental results recorded using MATLAB are displayed in Figures 4–7, respectively.

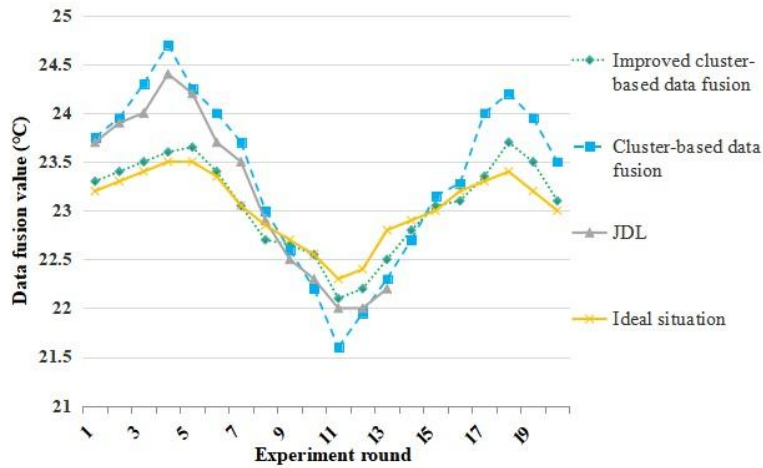
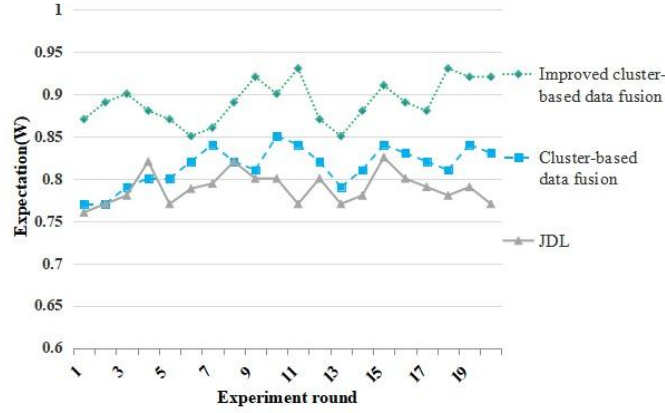
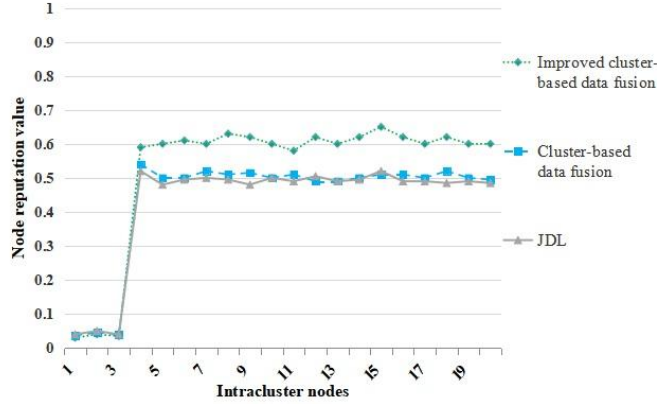


Figure 4. Comparison of data fusion values.



**Figure 5.** Comparison of expected values for evaluation of data fusion results.



**Figure 6.** Comparison of node reputation values.

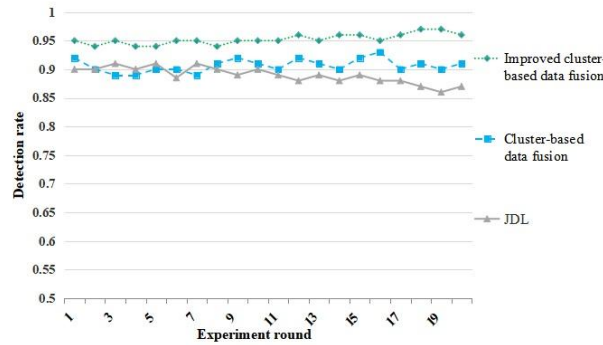
From the comparison in Figure 4, it can be seen that the fusion values of the physical training data for the improved cluster-based data fusion model are relatively close to the true values, while other models deviate more from the true values. This is because the improved cluster-based data fusion model strictly screens the physical training data participating in the fusion, increasing the proportion of real perception data participating in the fusion of physical training data, thereby improving the accuracy of the fusion results of physical training data.

The comparison of the expected values for evaluating the fusion results of physical training data between the improved cluster-based data fusion model and other models is shown in Figure 5. The horizontal axis in the figure represents the number of rounds conducted in the simulation experiment, and the vertical axis represents the expected values for evaluating the fusion results of physical training data. From the figure, it can be seen that the expected value of the improved cluster-based data fusion model is significantly higher than that of other models. This is because the improved cluster-based data fusion model introduces a penalty measure when calculating the node reputation value, which can avoid the historical accumulated reputation of member nodes from concealing their current malicious behavior and ensure that the data participating in the physical training

fusion operation is truly trustworthy, thereby increasing the expectation of evaluating the results of physical training data fusion. The fusion results of physical fitness training data from the improved cluster-based data fusion model are more reliable.

Node reputation value is an important indicator of secure data fusion models and a major factor affecting the reliability, credibility, and fusion efficiency of physical fitness training data fusion results. The comparison of node reputation values between the improved cluster-based data fusion model and other models is shown in Figure 6, where the horizontal axis represents the node number within the cluster and the vertical axis represents the node reputation value. As shown in the figure, the reputation values of the first three nodes in models are significantly lower than those of the other nodes, as the simulation experiments assume that the first three nodes are malicious nodes. In addition, compared to other models, the improved cluster-based data fusion model has a lower reputation value for malicious nodes, while the reputation value for normal nodes is significantly higher.

The detection rate of malicious nodes is an important criterion for determining the security of physical training data participating in fusion operations, and is an important factor reflecting the authenticity and reliability of the fusion results of physical training data. The comparison of malicious node detection rates between the improved cluster-based data fusion model and other models is shown in Figure 7. The horizontal axis in the figure represents the number of rounds conducted in the simulation experiment, and the vertical axis represents the malicious node detection rate. As shown in the figure, the detection rate of the improved cluster-based data fusion model is generally higher than that of other models. This is because the improved cluster-based data fusion model preprocesses the physical training data for detecting malicious nodes and compares the node reputation value with the high reputation threshold, which can avoid missed detection of malicious nodes and improve the detection rate of malicious nodes. In addition, as the number of simulation experiments continues to increase, the detection rate of the improved cluster-based data fusion model has increased to a certain extent, while the detection rate of the cluster-based data fusion model shows a significant downward trend. This is because the introduction of penalty factors in the improved cluster-based data fusion model allows for a slow increase and fast decrease in node reputation value calculation and update, which can quickly detect malicious nodes and improve the detection rate of malicious nodes.



**Figure 7.** Comparison of malicious node detection rates.

In summary, simulation experiments and performance evaluations are conducted to compare the improved cluster-based data fusion model and other models from four aspects, proving that the improved model in this article has obvious advantages.

### 4.3. Security performance analysis and algorithm complexity

Compared to other models, our model can enhance the security of data fusion models in different ways. Data processing is the initial judgment and screening of the sensor node's perceived data by the fusion node before conducting secure data fusion. This can achieve the identification and removal of abnormal data that deviates significantly from normal data due to network failures or hostile attacks so that abnormal data no longer participates in subsequent fusion operations, thereby eliminating the impact of abnormal data on data fusion results and ensuring the reliability of data fusion results. The introduction of penalty factors helps achieve a slow increase and fast decrease in the calculation and update of node reputation values, which can avoid the historical accumulated reputation of member nodes masking their current malicious behavior, facilitate the rapid detection of malicious nodes, and thus improve the security of data fusion.

Compared to other models, our model introduces node reputation values for evaluation. This increases the complexity of the algorithm. However, a data fusion model that achieves high security performance at the cost of increasing model complexity is desirable.

## 5. Conclusions

Data fusion is an information processing technology that has developed in recent years and has been widely applied in both military and nonmilitary fields. Utilizing data fusion technology in the IoT can decrease redundancy in perception node data and reduce data traffic within the network. Consequently, it extends the network's service life. However, the security of the IoT is relatively low, which leads to many security challenges for data fusion and results in huge losses for security-sensitive fields. Therefore, secure data fusion is necessary. This article mainly studies the security fusion method of physical training data in the IoT, aiming to propose a secure and efficient physical training data fusion scheme suitable for the IoT, which is conducive to improving the digitalization, informatization, and intelligence level of physical training. Firstly, the data preprocessing of physical fitness training is analyzed. Based on a thorough analysis of the performance for the cluster-based data fusion model, an improved cluster-based data fusion model is proposed. Finally, the experimental simulation system architecture is introduced. Through simulation experiments, improved cluster-based data fusion model and other models are compared from four aspects. The performance advantages of the improved model are verified, providing strong support for physical fitness training. However, the proposed model in this article is based on an ideal state, which assumes that the base station and fusion node are completely trustworthy. Therefore, in the future, we should improve the security data fusion model assuming that the base station or cluster head is attacked.

## Acknowledgment

This work was supported by Heilongjiang Province Higher Education Teaching Reform Project (Grant no. SJGY20200404).

## Conflict of interest

The author declares that there are no conflicts of interest regarding this paper.

## Author contributions

Bin Zhou contributed to the entirety of the paper.

## References

- [1] Nindl BC, Billing DC, Drain JR, Beckner ME, Greeves J et al. Perspectives on resilience for military readiness and preparedness: report of an international military physiology roundtable. *Journal of science and medicine in sport*, 2018, 21(11): 1116-1124.
- [2] Pangrazi RP, Beighle A. *Dynamic physical education for elementary school children*. Human Kinetics Publishers, 2019.
- [3] Smith R. The long history of gaming in military training. *Simulation & Gaming* 2010; 41 (1): 6-19.
- [4] Zhang C, Lu Y. Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration* 2021; 23: 100224.
- [5] Miao J, Wang Z, Xue X, Wang M, Lv J et al. Lightweight and secure D2D group communication for wireless IoT. *Frontiers in Physics*, 2023; 11: 1210777.
- [6] Jan M A, Zakarya M, Khan M, Mastorakis S, Menon V G et al. An AI-enabled lightweight data fusion and load optimization approach for Internet of Things. *Future Generation Computer Systems* 2021; 122: 40-51.
- [7] Miao J, Wang Z, Ning X, Xiao N, Cai W et al. Practical and secure multifactor authentication protocol for autonomous vehicles in 5G. *Software: Practice and Experience*, 2022.
- [8] Lv Z, Song H. Mobile internet of things under data physical fusion technology. *IEEE Internet of Things Journal* 2019; 7 (5): 4616-4624.
- [9] Ding W, Jing X, Yan Z, Yang L. A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Information Fusion* 2019; 51: 129-144.
- [10] Zhang W, Sheng Q, Mahmood A, Tran D, Zaib M et al. The 10 research topics in the Internet of Things 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC). *IEEE*, 2020: 34-43.
- [11] Miao J, Wang Z, Wang M, Feng X, Xiao N et al. Security Authentication Protocol for Massive Machine Type Communication in 5G Networks. *Wireless Communications and Mobile Computing* 2023.
- [12] Thomas JR, Martin P, Etnier JL, Silverman SJ. *Research methods in physical activity*. Human kinetics, 2022.
- [13] Qiu T, Chen N, Li K, Atiquzzaman M, Zhao W et al. How can heterogeneous internet of things build our future: A survey. *IEEE Communications Surveys & Tutorials* 2018; 20 (3): 2011-2027.
- [14] Miao J, Wang Z, Miao X, Xing L. A secure and efficient lightweight vehicle group Authentication protocol in 5G networks. *Wireless Communications and Mobile Computing* 2021: 1-12.
- [15] He W, Liu X, Nguyen H, Nahrstedt K, Abdelzaher T. Pda: Privacy-preserving data aggregation in wireless sensor networks. *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*. *IEEE*, 2007: 2045-2053.
- [16] Girao J, Westhoff D, Schneider M. CDA: Concealed data aggregation for reverse multicast traffic in wireless sensor networks. *IEEE International Conference on Communications*, 2005. *ICC 2005*. 2005. *IEEE*, 2005, 5: 3044-3049.
- [17] Castelluccia C, Mykletun E, Tsudik G. Efficient aggregation of encrypted data in wireless sensor networks. *The second annual international conference on mobile and ubiquitous systems: networking and services*. *IEEE*, 2005: 109-117.
- [18] Papadopoulos S, Kiayias A, Papadias D. Secure and efficient in-network processing of exact SUM queries. *2011 IEEE 27th International Conference on Data Engineering*. *IEEE*, 2011: 517-528.
- [19] Gentry C, Halevi S. Implementing gentrys fully-homomorphic encryption scheme. *Advances in Cryptology CEURO-CRYPT 2011: 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Tallinn, Estonia, May 15-19, 2011. *Proceedings 30*. Springer Berlin Heidelberg, 2011: 129-148.
- [20] Atakli IM, Hu H, Chen Y, Ku W, Su Z. Malicious node detection in wireless sensor networks using weighted trust evaluation. *Proceedings of the 2008 Spring simulation multiconference*. 2008: 836-843.

- [21] Xiangdong H, Pengqin Y, Qinfang W. Securing sensor networks based on optimization of weighted confidence. *China Communications*, 2012, 9(8): 122-128.
- [22] Yang Y, Wang X, Zhu S, Cao G. SDAP: A secure hop-by-hop data aggregation protocol for sensor networks. *ACM Transactions on Information and System Security (TISSEC)* 2008; 11 (4): 1-43.
- [23] Cam H, Ozdemir S, Nair P, Muthuavinashiappan D. ESPDA: energy-efficient and secure pattern-based data aggregation for wireless sensor networks. *SENSORS*, 2003 IEEE. IEEE 2003; 2: 732-736.
- [24] Dhanaraj RK, Lalitha K, Anitha S, Khaitan S, Gupta P et al. Hybrid and dynamic clustering based data aggregation and routing for wireless sensor networks. *Journal of Intelligent & Fuzzy Systems* 2021; 40 (6): 10751-10765.
- [25] He H, Li Y, Tan J. Relative motion estimation using visual Cinertial optical flow. *Autonomous Robots* 2018; 42: 615-629.
- [26] Liu Y, Li Y, Yi X, Hu Z, Zhang H et al. Micro-expression recognition model based on TV-L1 optical flow method and improved ShuffleNet. *Scientific Reports* 2022; 12 (1): 17522.
- [27] Blachut K, Kryjak T. Real-time efficient FPGA implementation of the multi-scale Lucas-Kanade and Horn-Schunck optical flow algorithms for a 4K Video Stream. *Sensors* 2022; 22 (13): 5017.
- [28] Krishnamachari B, Estrin D, Wicker S. Modelling data-centric routing in wireless sensor networks. *IEEE infocom*. 2002; 2: 39-44.
- [29] Xiao F. CEQD: A complex mass function to predict interference effects. *IEEE Transactions on Cybernetics* 2021; 52 (8): 7402-7414.
- [30] Josang A. PKI trust models. *Theory and Practice of Cryptography Solutions for Secure Information Systems*. IGI Global, 2013: 279-301.
- [31] Xiao X, Huang H, Wang W. Underwater wireless sensor networks: An energy-efficient clustering routing protocol based on data fusion and genetic algorithms. *Applied Sciences*, 2020; 11 (1): 312.
- [32] Meng T, Jing X, Yan Z, Pedrycz W. A survey on machine learning for data fusion. *Information Fusion* 2020, 57: 115-129.
- [33] Swart I, Irwin B, Grobler MM. Adaptation of the JDL Model for Multi-Sensor National Cyber Security Data Fusion. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, 2016; 6 (3).
- [34] Wang T, Wang P, Cai S, Zheng X, Ma Y et al. Mobile edge-enabled trust evaluation for the Internet of Things. *Information Fusion* 2021; 75: 90-100.
- [35] Erhan L, Ndubuaku M, Di Mauro M, Song W, Chen M et al. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion* 2021; 67: 64-79.