

Research Article

Information Processing Methods and LOT Cloud Platform for Fusion of 5G Urban Component Sensors

Jing He 

College of Artificial Intelligence and Big Data, Chongqing Industry Polytechnic College, China

Correspondence should be addressed to Jing He; hejing@cqipc.edu.cn

Received 20 January 2022; Accepted 1 March 2022; Published 7 April 2022

Academic Editor: Hye-jin Kim

Copyright © 2022 Jing He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of intelligent applications, the development of the Internet of Things changes with each passing day. The Internet of Things, as an essential component of the new information technology, has already become another revolutionary information industry after computers and modern communications. Although the Internet of Things has developed and grown, at the application level, the applications of Internet of Things devices of different architectures are relatively independent of their data information. This paper is based on the research of 5G urban component sensor information processing method and LOT cloud platform, and proposes a data fusion algorithm based on the data sequence of the same time. It analyzes the system process and architecture diagram of the LOT cloud platform, examines the performance of the Hadoop platform on different scale data sets, and uses the stand-alone system as a reference. The experimental results show that the time increase of the cluster decreases with the increase of the amount of data, which shows that the system exerts the advantages of parallel computing under the massive amount of data, and as the data source file increases, the accuracy rate decreases. When the data source size is 2000 MB, the Hadoop cluster system with 3 and 4 cluster nodes has the highest accuracy, 79.21% and 76.28%, respectively.

1. Introduction

As LOT technology evolves, users can deploy their own needs on the Internet of Things cloud platform, remotely manage applications, perform real-time data analysis, and integrate traditional enterprises including the Internet of Things. Research on the cost of industrial products and development methods to achieve the successful transformation of traditional industries and the promotion of manufacturing capabilities. But in the Internet of Things, there are many types of Internet of Things device applications. Access system protocols are also diverse, such as WiFi, LoRa, ZigBee, GPRS, NB-IOT, and so on. Numerical value collection tools, image capture applications, local location monitoring applications and other applications use various protocols to connect to the LOT cloud platform before the applications can be configured and controlled.

The research trend of the Internet of Things technology has been carried out in many countries, and the Internet of Things has great research significance and practical significance from both academic and practical perspectives. The Internet of Things has a wide range of applications in many

industries such as manufacturing, transportation, chemical engineering, and medicine, and has promoted huge application prospects. The rapid development of many platforms on the Internet announces that the widespread application of the Internet of Things platform has become an irreversible trend. With the further development of the Internet of Things technology, it is very important for platform companies to capture massive amounts of data and achieve powerful object analysis. However, the development problems encountered by certain LOT platforms with relatively single processing objects are not simple application problems, and the most important thing is to keep pace with the times.

Recently, research on cloud integrated of Things (LOT), which integrates the Internet of Things (LOT) and the cloud environment, is actively being carried out. Kim H W found that LOT should provide highly reliable services corresponding to various kinds of political factors in daily life. To ensure that LOT services are highly available with high degree of stability, optimal modeling, simulation and resource management techniques that integrate physical and computational elements are required. For these reasons, many systems are

under development, in which autonomous computing technology is applied, and part of human work is replaced by existing computer technology. So that the computer system can be self-tuned, self-configured, self-protected, and self-repairing, and the efficiency of the computer system can be improved by technical management techniques to reduce management costs, and it can sense any internal errors or changes in the external environment that occur during the operation of the system, and the system can self-adjust or evolve. Autonomic computing requires high levels of efficiency in an LOT environment comprising large-scale nodes, but it has not yet been realized in real life [1]. Currently, the 21st century is in a period of information explosion. But for the fast-developing information age, the research progress of sensor-based information processing devices is relatively slow. Y Motohashi provides an apparatus for processing messages that prioritizes inference results deduced from contextual information. The message processing apparatus comprises a reasoning cell that gets reasoning outcomes derived by the application of the reasoning reg's to pieces of contextual messages; a calculation cell for index values of reasoning outcomes that calculates an index value based on the reading user's level of knowledge of each reasoning rule used in the reasoning process, giving an overall view of the reading user's depth of knowledge of the outcomes. The inference result display unit, which displays the inference results based on the index value; and the knowledge depth update unit, which updates the reading user's knowledge level for each inference rule used in the inference process based on the evaluation information obtained. The unit modules are set up in detail, but the exact effect is not yet clear[2]. Lu J investigates the state-estimation problem for the best convergence of multistage asynchronous multi-scale transducers having both uncertain measurements and correlated noise. Noise from different agents is correlated and combined with a priori flow and corresponding system noise at the same time step. The problem is characterized by the maximum possible summary rate, where different sensors observe a target independently at more than one rate. He proposes an optimal state estimation method based on iterative estimation with a white noise estimator, the special feature of this algorithm is that it provides a simple, efficient and reliable uniformity measurement method, which can avoid including blocks with structural information into the measurement process, and improves the accuracy and robustness of the system state estimation. Numerical examples show the effectiveness of the algorithm, but its practicality is not strong[3]. Xing Z proposed three nonlinear centralized scale unscented Kalman filter (SUKF) algorithms for multi-sensor data fusion, which are enhanced measurement, measurement weighting, and sequential filter fusion. First, he studied in detail the accuracy analysis of Extended Kalman Filter (EKF) and SUKF. Secondly, by comparing the error covariance trajectory of the centralized SUKF of the multi-sensor data fusion algorithm and the centralized EKF of the multi-sensor data fusion algorithm and the absolute average estimation error in the X and Y directions. It can be seen that among the six algorithms, the centralized augmented measurement SUKF multi-sensor data fusion algorithm has the best performance. In other words, the algorithm (Iu) performs best in terms of accuracy. Finally, the run-

ning time of the six algorithms is combined and comprehensively analyzed, which shows that among the six algorithms, the algorithm (Iu) is the best in terms of synthesis, but it is a bit cumbersome in terms of experimental operation [4]. Among the greatest remaining operational surprises for LOTplatforms is the calibration of sensors in uncontrolled environments, which measures cross-sensitivity to compensate for interfering contaminants and environmental conditions. Ferrer-Cid P studied how data fusion collected by sensor arrays can improve the calibration process. In particular, he compared sensor array calibration, weighted average multi-sensor data fusion calibration and machine learning model multi-sensor data fusion calibration. He evaluated the calibration by combining data from various sensors with linear and non-linear regression models, but there were too few subjects [5]. The basic principle of multi-sensor data fusion technology is like the process of comprehensive processing of information by the human brain. Malakar B proposed an adaptive multi-sensor data fusion technology based on bilinear recursive least squares, which is used to accurately locate railway vehicles and detect accidental separation of trains, and is used in the Indian Railway Train Collision Avoidance System (TCAS). Robust resolution of one of the task is to augment GPS with an airborne multi-sensor system. Dual Linear recursive least squares adaptive filters are used to estimate and offset the location errors of the on-board multi-sensor system. The capability of the method is compared with that of the observation error-based method, the bounded offset-based method, and the technique of pseudo-measurement state bounding. Simulation results show that the superiority of the proposed method in terms of accuracy of positioning and detected unexpected train separation at least separation distance has not yet been applied in practice[6].

The innovations of this article: (1) The number of search cluster nodes is increased on the Hadoop platform, which provides flexibility for maintenance and control under complex control situations that require multiple devices, and applications are connected to the platform through networks and systems. (2) Connecting the device to the LOTplatform, and the application devices can communicate directly. It uses Internet + thinking to build a cloud platform for the Internet of Things, which is more convenient and flexible than the detection and display of the upper computer in the stand-alone mode, and the processing speed is faster without affecting the detection accuracy of the data source file.

2. Information Processing Method and Cloud Platform Research Method

2.1. The LOT Cloud Platform of 5G Urban Component Sensors. In the Internet of Things, data is a fundamental part of the Internet of Things. Cloud platform data mainly includes collected data transmitted by application programs and data generated by users working on the cloud platform. Cloud platform functions are analyzed, designed, and developed around data types. The data on the LOT cloud platform mainly has the following characteristics: There are many types of data structures:

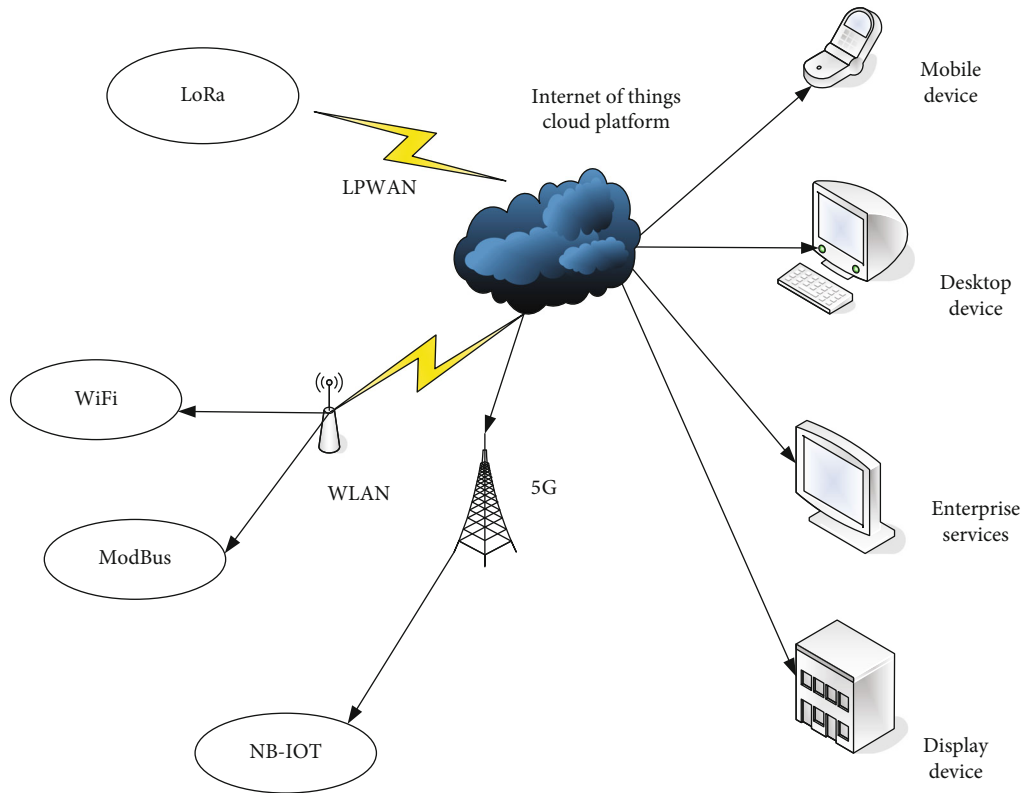


FIGURE 1: System flow chart of Internet of things cloud platform.

- (1) In real industrial manufacturing, the format and meaning of data collected by different devices in the Internet of Things, and the amount of data collected by sensors are also different. For example, a monitoring application captures the temperature, humidity, and real-time location of mobile devices in the workshop, and collects four-dimensional data. Magnetic card readers can capture five-dimensional or more multi-dimensional data in the card information, and output the data, and the data structure uploaded by the device is not the same [7].
- (2) Data compatibility: The data contained in the enterprise Internet of Things database is very important. For example, user data is associated with user device data including device configuration data or command data. In the cloud platform, the data that is not uploaded is not independent, but is related to each other or between multiple and multiple [8].

Data is temporal and spatial. The data collected by sensors on devices in the private Internet of Things represents the state of the device at a certain time or the state of the environment around the device.

For enterprises, in daily production, a piece of equipment cannot support normal production and operation requirements. Companies need different 5G urban component sensor components, such as LoRa sensors, NB-IOT-sensors, WiFi sensors, ModBus central management application equipment, and communication networks used

by different platforms. LoRa equipment uses LPWAN, NB-IOT uses the 5G network provided by the operator, and WiFi and ModBus industrial equipment uses the local area network to transmit data[9]. The system process of the LOT cloud platform is shown in Figure 1:

The cloud platform only needs to provide users with the services they need. The internal implementation of the cloud platform and the collection of underlying data are transparent to users, but it is always confidential to other terminals. In order to make full use of the control of cloud platform system equipment and resources, and provide higher performance access and data processing capabilities, the system integrates SOA ideas, as a service source, provides data processing and data services between modules [10]. The structure of the cloud platform system is shown in Figure 2:

As shown in Figure 2, the cloud platform is composed of platform application services, management systems, storage systems, and communication systems. Each system can exchange information and update data. The device is connected to the cloud platform through the device access interface provided by the cloud platform, and the use of wireless network transmission can solve the problem of Internet access in complex areas. Its main core components are wireless routers, LoRa base stations, routers and three-layer switches [11–12]. The network detection diagram of the system is shown in Figure 3.

The wireless ad hoc network base station is an important tool for building back-end IP networks. Through the wireless MESH protocol network, a personal network system with multiple wireless links can be created, which can be easily connected to the network to support links. There

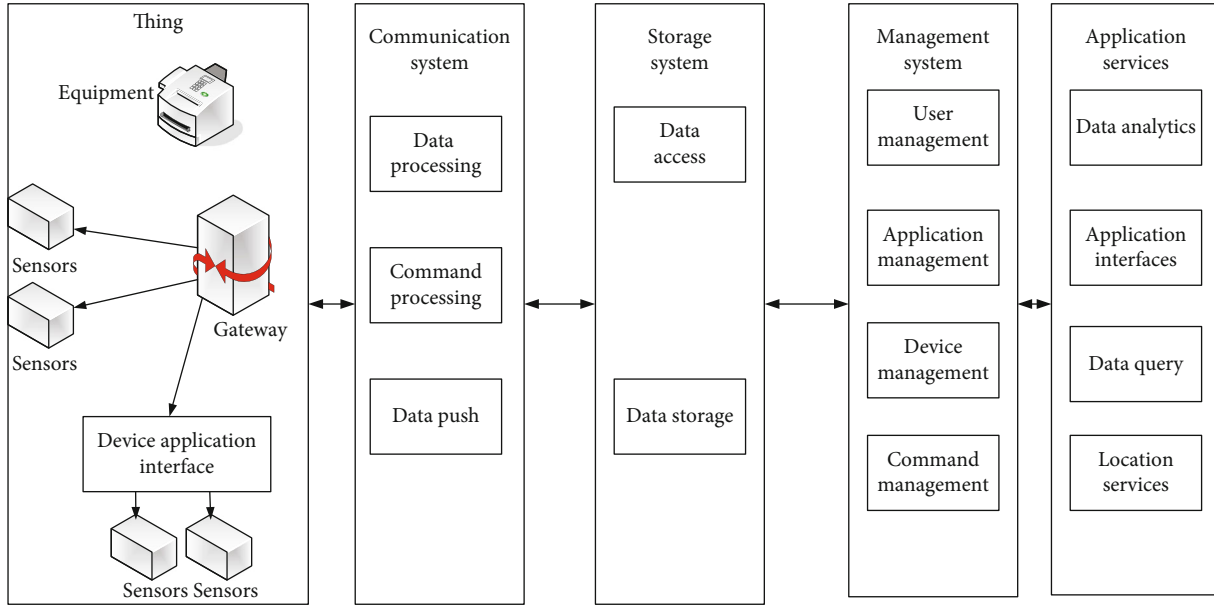


FIGURE 2: Cloud platform system architecture diagram.

are many types of routing structures, including the advantages of high bandwidth and strong stability [13]. Because the equipment is easy to install, the construction cost of the optical fiber switcher is low, and it can be used all the time. The wireless installation is simple, and when the site does not require a wireless connection, the application can be switched to other places where a wireless network is needed to continue using it. The LoRa base station is a protocol converter, which is responsible for converting the radio frequency signal transmitted by the node into an IP network signal and sending it to the LoRa server for transmission. It is a bridge between the temperature storage terminal and the server [14], with a built-in LoRaWAN protocol, and supports access to a network of 5000 nodes at most. The wireless transmission distance is up to 20 kilometers, and the 470MhzISM unlicensed frequency band is used for communication.

2.2. Mathematical Model of Data Fusion Algorithm. Data fusion includes many aspects, and the models are also diverse. Figure 4 is a general model of data fusion, which is divided into four levels. There is no time sequence characteristic between levels, and these sub-processes are processed in parallel [15].

The first level: Including the registration, association, recognition, etc. between data and images. The so-called registration is to preprocess the information sent back from the sensors of various 5G urban components, so that they are all distributed on the same coordinate or platform. Association mainly refers to the association of data, and identification mainly refers to identity recognition and identity fusion [16].

The second level: Situation assessment, which mainly includes situation extraction, situation analysis and situation prediction.

The third level: Practical evaluation, is about the applicability analysis of the data fusion model.

The fourth level: Optimizing fusion processing, including optimizing the use of resources, optimizing 5G urban component sensor management, and optimizing weapon control [17]. The purpose is to enable decision makers to make correct decisions.

Using a single 5G urban component sensor cannot accurately track the target's trajectory, nor can it accurately determine the target's identity information. Using data fusion in a multi-sensor information system can bring many benefits, but first of all, it can strengthen the robustness of the system. A single sensor is greatly affected by the external environment, and the results obtained are often error-prone. Using multi-sensor fusion can reduce the dependence on the environment and make it easier to adapt to changes in the environment. Secondly, the space and time coverage of the system is enhanced. For example, the visible light sensor can only work during the day. If the visible light sensor is used alone, we cannot get the information at night [18], and the use of a multi-sensor information system can cover the range that a single sensor cannot reach.

At present, there are many classification methods for data fusion. Among them, the most commonly used classification method is based on the level and substance of the fusion, which is divided into pixel level, feature level and decision level [19]. As shown in Table 1 is a comparison of the advantages and disadvantages of the three fusions, with 1 representing the highest level; 2 representing the medium and 3 representing the lowest [20].

The 5G urban component sensor sensing node is the most basic element in the wireless sensor network system. It usually has the function of sensing and detecting certain spatial parameters, temporary storage and calculation of data, and wireless communication, and it is powered by the battery of each node [21]. Data fusion is based on the sampled data from multiple sensors, and the data is integrated and calculated to obtain the final result [22].

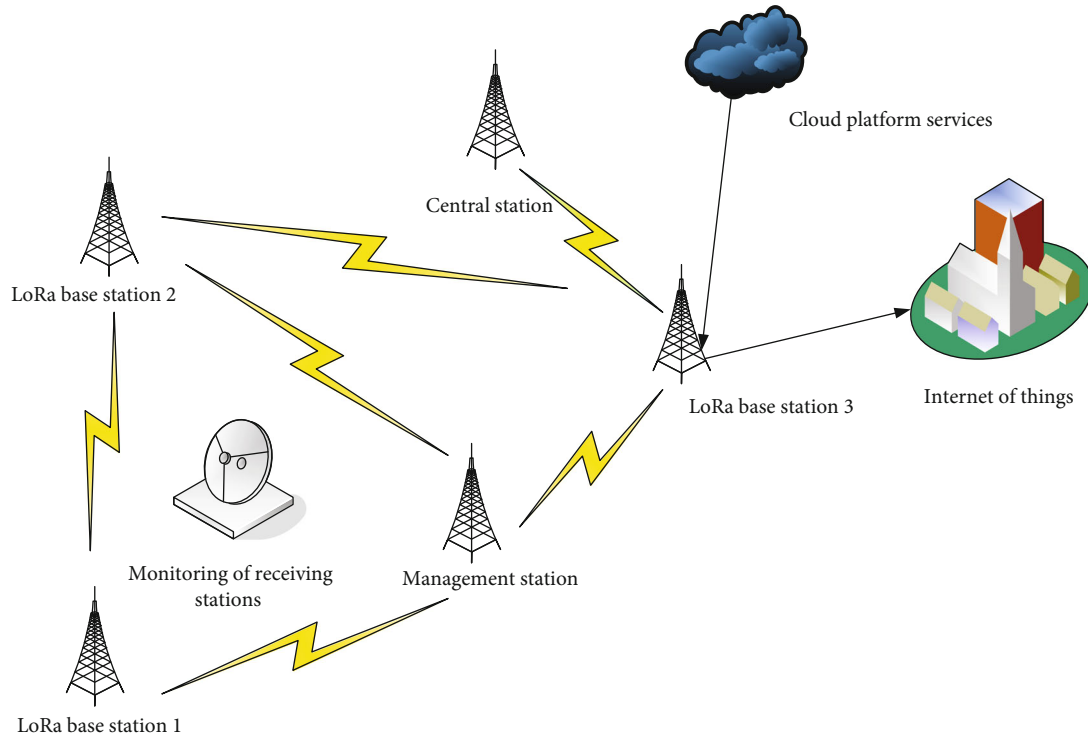


FIGURE 3: Schematic diagram of system network detection.

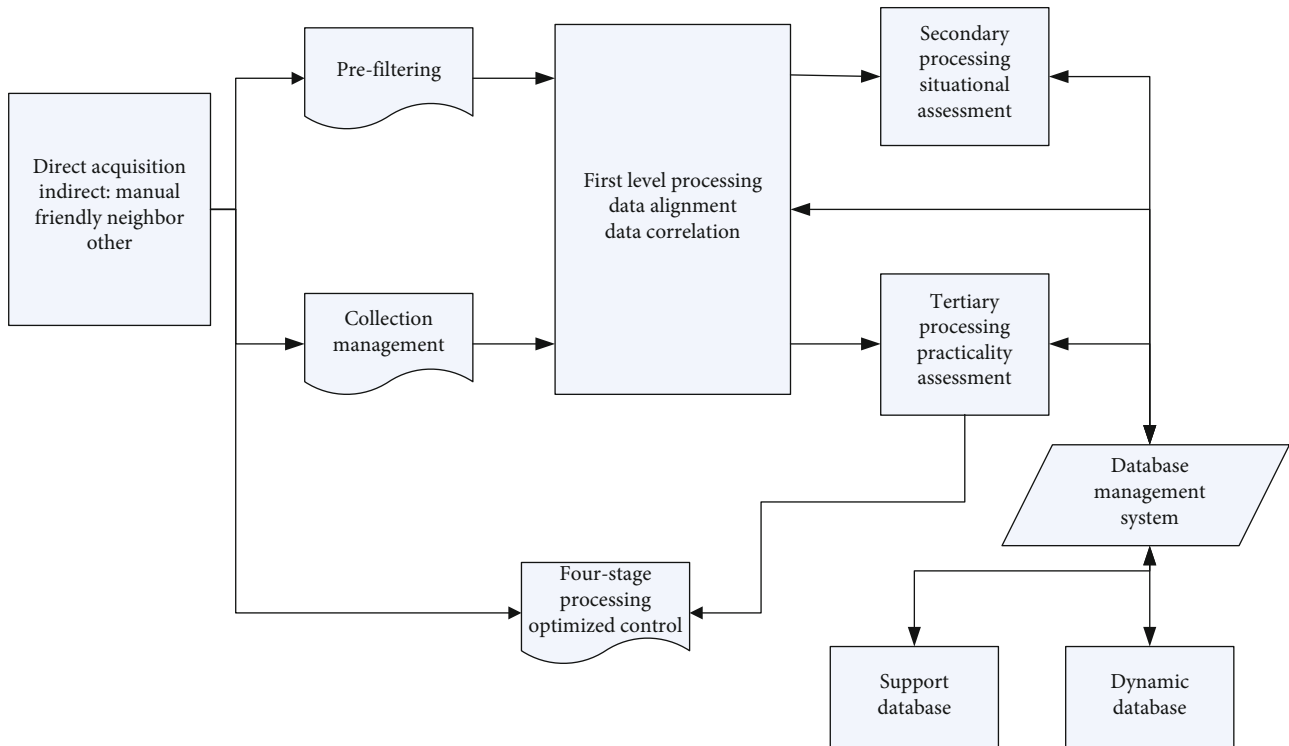


FIGURE 4: General model of data fusion.

Sensor sequences and sensors may be represented as follows.

$$A(t) = \{A_1(t), A_2(t), \dots, A_p(t)\} \quad (1)$$

Then there are:

$$a_i(t) = K + \varepsilon_i(t), i = 1, 2, \dots, p \quad (2)$$

TABLE 1: Comparison of advantages and disadvantages of three fusion levels.

	Pixel level fusion	Feature level fusion	Decision level fusion
Amount of information processed	1	2	3
Loss of information	3	2	1
Anti interference performance	3	2	1
Fault tolerance performance	3	2	1
Fusion performance	1	2	3
Dependence on sensors	1	2	3

$A_i(t)$ is the measured value of the sensor at time t , K is the true value [23].

It records the sampled data from t_1 to t_m , respectively, and the matrix of summarized values is shown below.

$$R = \begin{pmatrix} a_1(t_1) & a_1(t_2) & \cdots & a_1(t_q) \\ a_2(t_1) & a_2(t_2) & \cdots & a_2(t_q) \\ \cdots & \cdots & \cdots & \cdots \\ a_p(t_1) & a_p(t_2) & \cdots & a_p(t_q) \end{pmatrix} \quad (3)$$

In each element of the matrix R , the elements represent the recorded value of the sample sensor i at time t . The columns in the matrix represent the measured values of the sensor sequence at the same time. The rows in the matrix represent the sequence of sampled values of the same sensor over time.

The data fusion algorithm studied in this paper is based on the data sequence at the same time and does not consider time correlation. Therefore, the data sampling model used by this type of data fusion algorithm is relatively simple, expressed as follows:

$$R = \{a_1, a_2, \dots, a_n\} \quad (4)$$

The sequence R represents the sampling data sequence of the sensor sequence to the parameter X at a certain moment. This sequence expression method omits the time parameter because the algorithm does not involve time correlation [24]. This expression facilitates the selection of values and is more favorable to the processing and computation of data by this type of algorithm [25–26].

(1) Fusion function

For the purposes of measuring values of differences between data, “differences” can be quantified in terms of absolute distances. This distance is given by:

$$D_{ij} = |a_i - a_j| \quad (5)$$

Although absolute distance is defined as a measured discrepancy between 2 pieces of data, the concept of convergence function is implemented as a measure of the role of a piece of data in the integration process.

The confluence functional can be represented as below.

$$g_{ij} = \exp \{U \bullet D_{ij}\} \quad (6)$$

Equation (6) is considered to be the best explanation and description of the convergence degree factor. However, the convergence function is still a broad concept and is not limited to a particular formula. The convergence degree function is used to describe the level of relative convergence between data, and the expression based on absolute distance is only one of them. In addition, in some algorithms, there are expressions based on the quotient of the two, and so on. Therefore, the fusion function should be more understood as a way of expression, and its form is not fixed [27–28].

(2) Fusion matrix

An integration matrix has been produced to facilitate integrated management and analysis[27]. This convergence matrix has the following representation:

$$G = \begin{pmatrix} 1 & g_{12} & \cdots & g_{1q} \\ g_{21} & 1 & \cdots & g_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ g_{q2} & g_{q2} & \cdots & 1 \end{pmatrix} \quad (7)$$

Element A of the matrices represents the integration of data a_i with respect to the data a_j in series D . The convergence matrix has its diagonal elements all 1. That’s due to the fact that the integration of the data with respect to itself is meaningless and can be considered to be 1 with respect to the integration of itself[30].

(3) Fusion result expression

A weight factor can change from line to line in different algorithms. However the weighting factor has important characteristics and rules [31]. Then:

$$m_1 + m_2 + \cdots + m_p = 1 \quad (8)$$

a

For extreme data, the value of m_i is not excluded. But in the case of the sum of the gravity factor, the coefficient of unity is not negative and is less than or equivalent to 1 [32].

The power factor can be considered in some sense as the product of the outcome of the merging method. The final expression of the fusion result is as follows.

$$Z = m_1 \times a_1 + m_2 \times a_2 + \cdots + m_p \times a_p \quad (9)$$

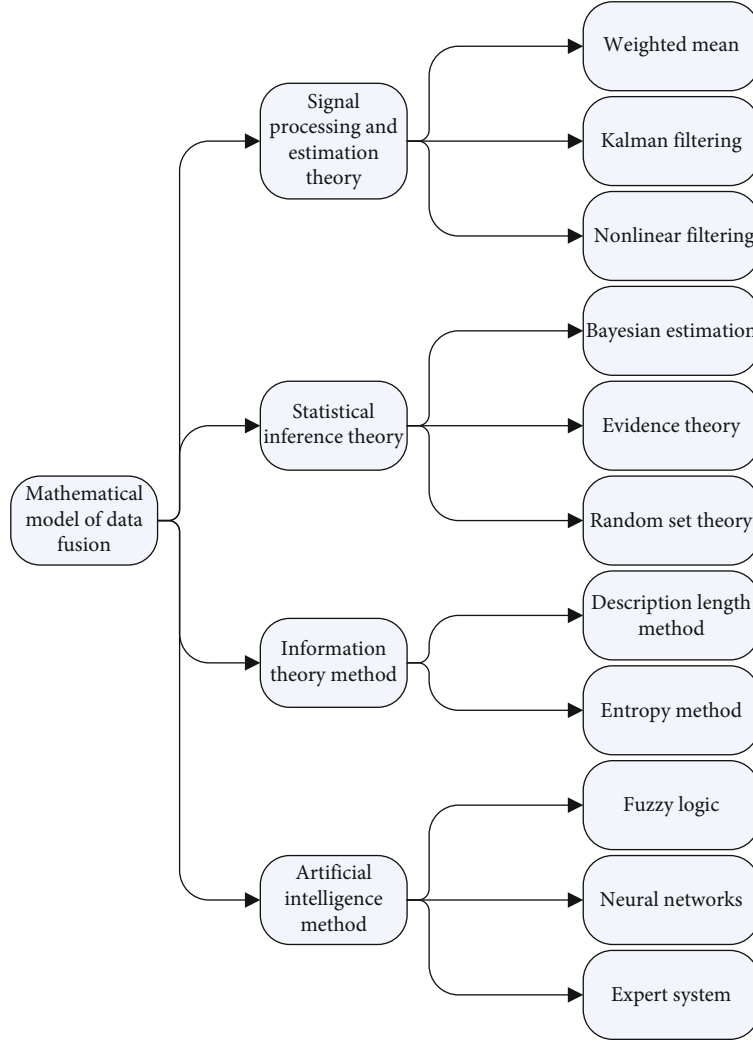


FIGURE 5: Typical information fusion algorithm

2.3. Overview of Data Fusion Algorithms in the Data Layer.

Fusion algorithm is the basic content of data fusion system. At present, most computing algorithms are researches based on the fusion of similar information. However, the data fusion of multiple types of sensors lacks effective algorithms and integrated models. The study found that with the help of modern statistical technology, the complex problems in the data fusion of different information can be solved. Multi-sensor fusion algorithms from different sources strengthened the thinking innovation in method. The data fusion algorithm is shown in Figure 5. Different application settings choose different fusion algorithms. From experience, the weakness of the algorithm makes the development of another algorithm always quick, usually, the deficiencies of this kind of algorithm are often room for improvement of another algorithm, so the complementarity between algorithms makes it a trend to combine two or more algorithms and then data fusion. So it can be combined and used according to the compatibility between the algorithms, and then the data is merged.

(1) Sampling data model

In wireless sensor networks, multiple sensors of the same type usually estimate and measure the same parameter. Due to the large differences in the space and geographic location of each sensor, there will be different situations of noise affecting each sensor. Therefore, the measured value is often in a certain area of the true value.

Due to the difference and suddenness of noise, the deviation of each measured value from the true value is also different. Normally, the component sensor has a set and known sampling frequency. According to the distribution of sampling time points, the sampling sequence of sensor A_i from time t_1 to time t_p is $A_i = \{a_{i1}, a_{i2}, \dots, a_{ip}\}$. Supposing that in a wireless sensor network, there are q nodes that measure parameters in the same context. The sampled data can be represented by the following matrix:

$$H = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{q1} & a_{q2} & \dots & a_{pq} \end{Bmatrix} \quad (10)$$

The element a_{ij} in parameter H denotes the value of the sensor A_i measured at time t_j . Because the application scenarios of sensor networks are different, in some applications, the measured value of the sensor node at the current moment has nothing to do with the measured value at the previous moment. Therefore, the sequential sampling data model is used in some fusion algorithms that only fuse the sampled data at the current moment. That is, at time t, the sampling data of the sensor sequence $A = \{A_1, A_2, \dots, A_p\}$ can be represented by the following sequence:

$$T_t = \{a_1, a_2, \dots, a_p\} \quad (11)$$

The element a_i in the sequence T_t represents the sampling data of the sensor A_i at the time t.

Let K be the true value of the sampled data at time t, then:

$$a_i = K + v_i \quad (12)$$

Where v_i represents the measurement error at time t for transducer A_i . Prior knowledge $E(v_i)$ and $D(v_i)$ are both unknown.

(2) Fuzzy mathematics and closeness

Fuzzy mathematics is a mathematical theory and method for studying and dealing with fuzzy phenomena. It uses precise mathematical methods to describe and model a large number of fuzzy concepts and fuzzy phenomena in the real world in order to achieve the purpose of properly processing them. Supposing there is a universe S, then a mapping from S to the unit interval [0,1]: $f_A : S \rightarrow [0, 1]$ is called a fuzzy set on S, which can also be called a fuzzy subset, denoted as L. The mapping $f_A : S \rightarrow [0, 1]$ is called the membership function of the fuzzy set A.

The gap or space between fuzzy sets can be expressed through a custom measurement process. For this reason, a metric needs to be established on the fuzzy power set. In addition, this metric can also be standardized, that is, mapped to the interval [0,1]. In the fuzzy pattern recognition method, the degree of closeness is used to identify the pattern category of the fuzzy subset to be discriminated. In order to measure the category of the subset to be identified, it is necessary to determine the relative closeness between each stage and the benchmark fuzzy set. In a sense, the closeness is the distance, and the closeness is proportional to the distance. Therefore, it should be the opposite of the value of distance. In addition, there are other definitions of closeness closely related to fuzzy sets, such as the maximum and minimum closeness:

$$\alpha(L, H) = \frac{\sum_{i=1}^p (L(f_i) \wedge H(f_i))}{\sum_{i=1}^p (L(f_i) \vee H(f_i))} \quad (13)$$

Arithmetic mean minimum closeness:

$$\alpha(L, H) = \frac{\sum_{i=1}^p (L(f_i) \wedge H(f_i))}{0.5 \sum_{i=1}^p (L(f_i) + H(f_i))} \quad (14)$$

Geometric mean minimum closeness:

$$\alpha(L, H) = \frac{\sum_{i=1}^p (L(f_i) \wedge H(f_i))}{\sum_{i=1}^p \sqrt{L(f_i) H(f_i)}} \quad (15)$$

Index closeness:

$$\alpha(L, H) = \frac{1}{e^{|L-H|}} \quad (16)$$

(3) Credibility matrix

Due to the difference in the location of the sensor itself and the quality of its parts, as well as some external random factors, such as electromagnetic radiation, strong weather, human factors, etc., the measurement parameters of the sensor node must have some deviations. These deviations are large or small, creating the problem of how to verify the measurement parameters. Credibility is used to measure the degree of closeness between the measured value of the current node and the measured value of other nodes, so as to judge the validity of the data. The establishment of the credibility function is based on the concept of maximum and minimum closeness in fuzzy mathematics.

The credibility of a_i and a_j at time t can be expressed as:

$$a_{ij}(t) = \min \{a_i, a_j\} / \max \{a_i, a_j\} (1 \leq i, j \leq n) \quad (17)$$

The confidence level matrix can then be denoted as follows:

$$S_n(t) = \begin{pmatrix} s_{11}(t) & s_{12}(t) & \dots & s_{1p}(t) \\ s_{21}(t) & s_{22}(t) & \dots & s_{2p}(t) \\ \dots & \dots & \dots & \dots \\ s_{p1}(t) & s_{p1}(t) & \dots & s_{pp}(t) \end{pmatrix} \quad (18)$$

(4) Mean and variance of credibility

This is expressed as follows so that the effect of self and self proximity may reduce the mean credibility of the i row:

$$w_i(t) = \sum_{j=1}^n a_{ij} - a_{ii} / (n-1) \quad (19)$$

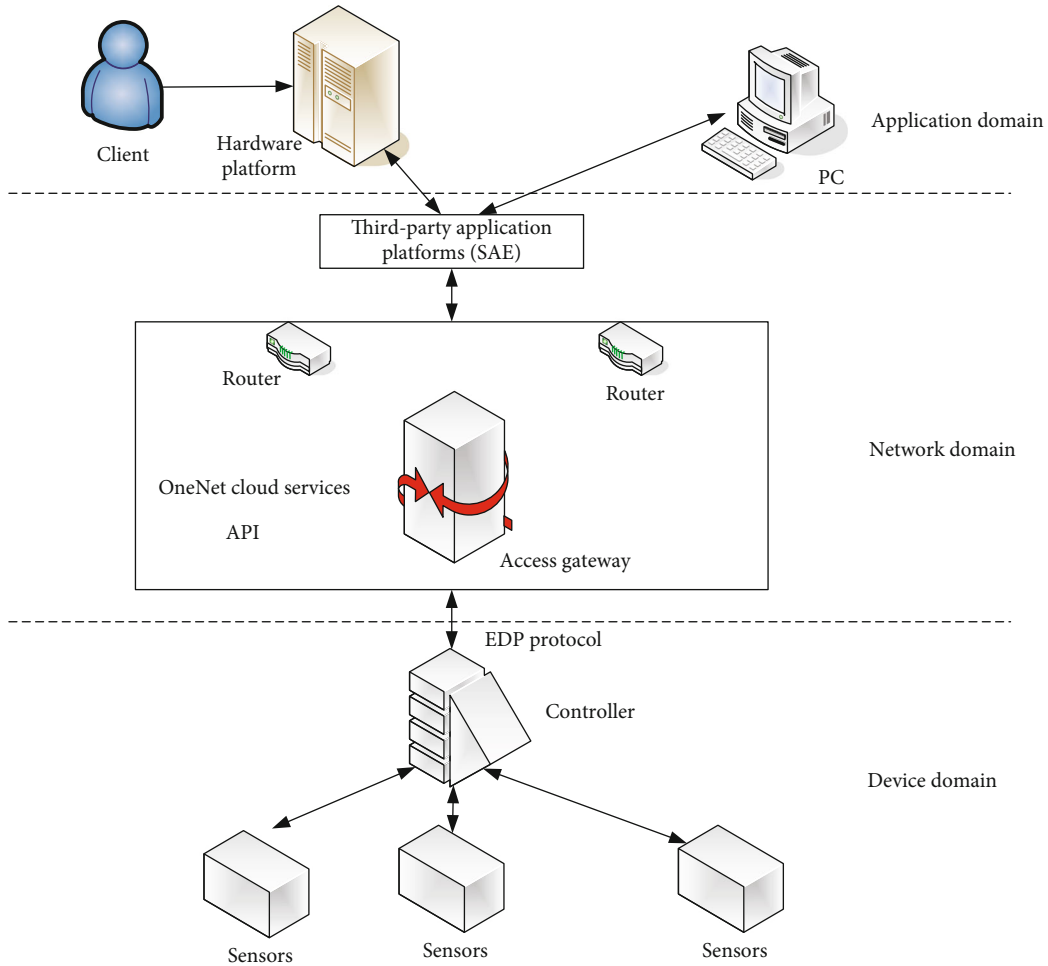


FIGURE 6: Schematic diagram of the system structure.

TABLE 2: Server specifications.

The server	Parameter	The server	Parameter
Processor	2	Memory	64GB
Processor dominant frequency	2GHz	Outline specification	4 U rack mounted
Cache capacity	18 MB	Storage device	Dell Fibre Channel disk

The expression for the variance of parentage in row i is as follows:

$$o_i^2(t) = \sum_{j=1}^n \frac{(\sum_{j=1}^n a_{ij}/n - a_{ij})^2}{n} \quad (20)$$

3. Experiment of LOT Cloud Platform

The LOT cloud platform system studied in this paper combines embedded technology and Internet technology. The system consists of 3 parts, and its structure is shown in Figure 6. The network domain uses Onenet as the device access platform, Sina Cloud (SAE) as the third-party application platform, and the application domain is the mobile client and PC terminal. Hadoop is a distributed system infrastructure

developed by the Apache Foundation. Users can develop distributed programs without understanding the underlying details of the distribution, and make full use of the power of the cluster for high-speed computing and storage.

The domain uses the Onenet Internet of Things platform, which provides Restful API interface, Socket interface, and EDP and Modbus device access protocols. The mobile application client checks the bound device and enters the device control panel. The client sends a command request to the SAE server through the application bar, and the SAE server sends the client request to the equipment installed in the machine area, and the machine returns a corresponding message. Finally, the SAE server returns the message generated after running to the application library, and the platform sends the message to the client, thereby completing the data and communication interaction between the

TABLE 3: Hadoop cluster planning.

Address	10.10.7.200	10.10.7.201	10.10.7.202	10.10.7.203
Host name	Node1	Node2	Node3	Node4
Function	Namenode,secondary namenode,HMaster	Datanode,tasktracker,HReginServer	Datanode,tasktracker,HReginServer,zookeeper	Datanode,tasktracker,HReginServer,zookeeper

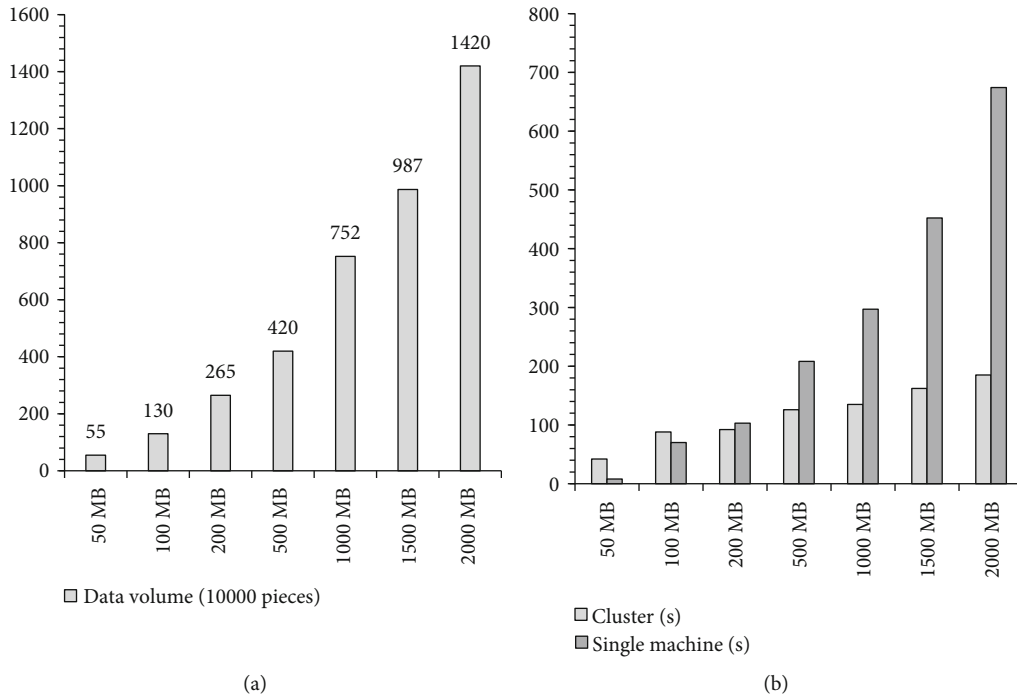


FIGURE 7: Comparison of the performance of different size data sets.

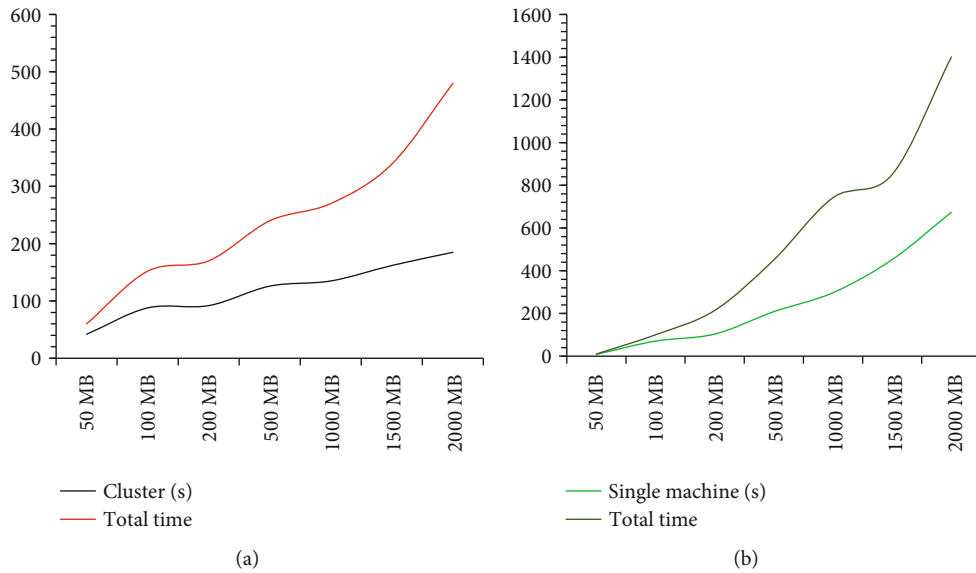


FIGURE 8: Comparison of processing time and total time in both modes.

platform and the application device. The PC side obtains the SAE server resource service through HTTP request, which indirectly realizes the user’s management and monitoring of the hardware equipment.

The sensor information processing system of the service platform is realized by using the existing equipment in the laboratory to deploy a distributed parallel computing cluster. The laboratory has an existing Dell PowerEdge R910 server and 20 TB storage. The specifications are shown in Table 2. By using server virtualization solutions, create multiple virtual machines on the server, check memory access permis-

sions and network connectivity. The virtual server running on it has better performance and stability and can be connected remotely. Remote management is carried out through the client, and 4 virtual machines form a Hadoop cluster. The addresses and division of labor are shown in Table 3.

4. Hadoop Cluster Data Processing Performance Analysis

Investigate the performance of the Hadoop platform on data sets of different scales, and take the stand-alone version of

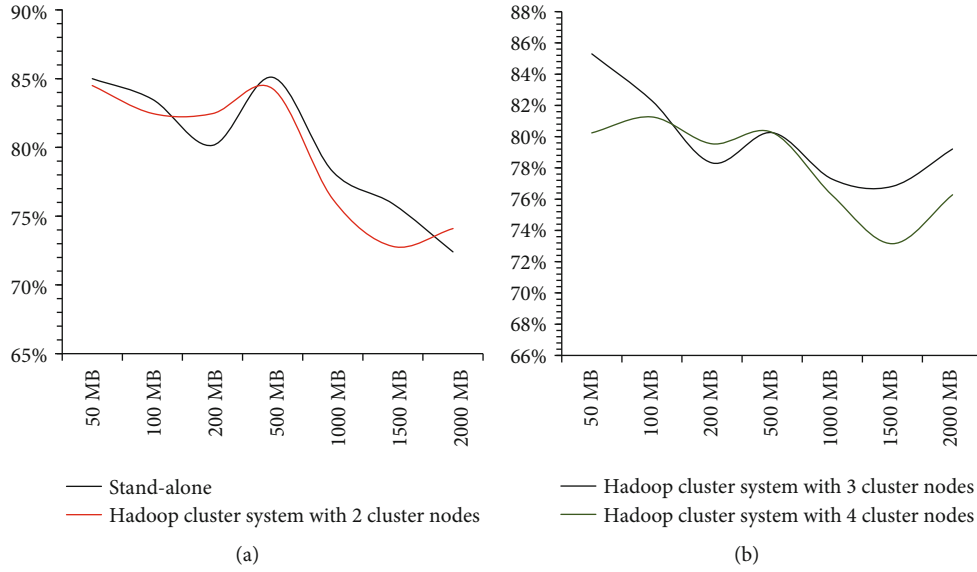


FIGURE 9: Accuracy of Hadoop clustered system and standalone mode.

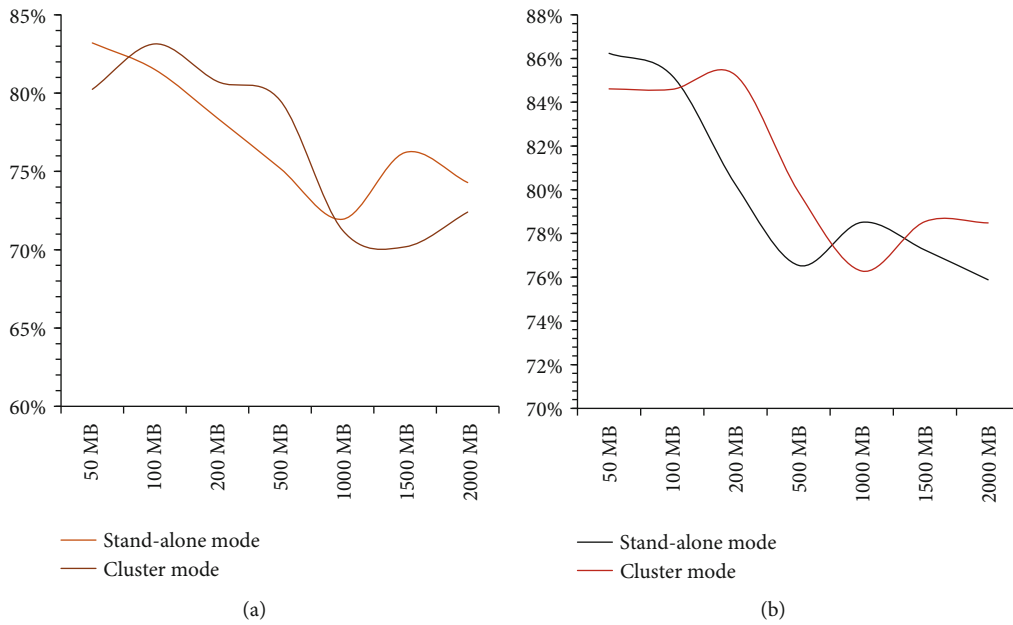


FIGURE 10: Performance comparison of two cases with 2 and 5 sensors.

the system as a reference. In stand-alone mode, Hadoop runs completely locally, without using HDFS or loading any Hadoop daemons. The data source file size is 50 MB, 100 MB, 200 MB, 500 MB, 1000 MB, 1500 MB, 2000 MB, respectively. By default, MapReduce uses the TextInputFormat interface to read and process files in HDFS in line units. One line is an independent record, so the number of data items can represent the workload of the system. The data volume corresponding to the file is 550,000, 1.3 million, 2.65 million, 4.2 million, 7.52 million, 9.87 million, and 14.2 million, respectively. And it verifies two cases with 2 and 5 sensors. The number of cluster nodes is 4. In order to reduce the influence of other factors on the experimental effect, compare the two sides using the same hardware envi-

ronment, the experimental result is the average of three experiments. The time comparison is shown in Figure 7, and the comparison between the processing time and the total time in the two modes is shown in Figure 8.

It can be seen from Figure 7 that when the amount of data is small, the processing efficiency of the Hadoop cluster system is not as good as that of a single machine. When dealing with large data sets, the running time of a single machine increases almost linearly. Because the stand-alone version executes all input data sequentially and there is no parallel processing mechanism. However, the time increase of the cluster decreases with the increase of the amount of data, which shows that the system exerts the advantages of parallel computing under the massive amount of data.

It can be seen from Figure 8 that when processing small-scale data, the processing time of the Hadoop cluster system exceeds half of the total time, while the processing time of the stand-alone mode exceeds 2/3 of the total time.

Selecting the data source file with a size of 50 MB, 100 MB, 200 MB, 500 MB, 1000 MB, 1500 MB, 2000 MB. Figure 9 shows the accuracy of the detection file and the Hadoop cluster system and stand-alone mode with 2, 3, and 4 nodes in the cluster. Figure 10 shows the accuracy of the Hadoop cluster system and single-machine mode under different processing sensor numbers.

It can be seen from Figure 9 that the size of the data source file has a great impact on the accuracy of the Hadoop cluster system and single-machine mode detection. As the data source file increases, the accuracy rate decreases. When the data source size is 2000 MB, the Hadoop cluster system with 3 and 4 cluster nodes has the highest accuracy, 79.21% and 76.28%, respectively.

It can be seen from Figure 10 that with the increase in the number of processing sensors, although the size of the text has become larger, the detection accuracy of massive data has increased significantly in the processing of files of the corresponding size.

5. Conclusion

With the advent of the 5G information era, sensor networks are an important medium for information collection. With the reduction in cost and the smaller and smaller electronic components, a large number of wireless sensors have begun to be used. As the most important terminal node of the Internet of Things, the deployment scale of sensors will inevitably become larger and larger. Correspondingly, the amount of data and requests generated will become larger and larger. So far, managing and processing the massive data generated by a large number of sensor nodes has become a thorny issue. Due to the huge amount of raw data collection, wireless sensors have an inevitable demand for data fusion. However, the computing power of the microprocessor is weak, the energy is limited, and the data fusion algorithm with high computational complexity is not suitable for wireless sensor networks. This article aims at the information processing method of 5G urban component sensors, combined with the Internet and cloud service technology, to build an Internet of Things cloud platform. By using 4 virtual machines to form a Hadoop cluster, it analyzes its performance on different scale data sets. But in the research process, this article still has many shortcomings, for example, the detection accuracy of massive data and the number of sensors are the most appropriate. These issues still need to be studied.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research study is sponsored by Chongqing Yubei District Science and technology planning project. The name of the project is Big data development of service PAAS automatic delivery and management platform. Thank the project for supporting this article!

References

- [1] H. W. Kim, J. H. Park, and Y. S. Jeong, "Efficient resource management scheme for storage processing in cloud infrastructure with internet of things," *Wireless Personal Communications*, vol. 91, no. 4, pp. 1635–1651, 2016.
- [2] Y. Motohashi, H. Sakagami, and S. Kamei, "Information processing device, information processing system, information processing method and computer-readable medium," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 4, pp. 333–347, 2018.
- [3] L. Jiang, L. Yan, Y. Xia, Q. Guo, M. Fu, and K. Lu, "Asynchronous multirate multisensor data fusion over unreliable measurements with correlated noise," *IEEE Transactions on Aerospace & Electronic Systems*, vol. 53, no. 5, pp. 2427–2437, 2017.
- [4] Z. Xing and Y. Xia, "Comparison of centralised scaled unscented Kalman filter and extended Kalman filter for multisensor data fusion architectures," *IET Signal Processing*, vol. 10, no. 4, pp. 359–365, 2016.
- [5] P. Ferrer-Cid, J. M. Barcelo-Ordinas, J. Garcia-Vidal, A. Ripoll, and M. Viana, "Multisensor data fusion calibration in IoT Air pollution Platforms," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3124–3132, 2020.
- [6] B. Malakar and B. K. Roy, "Adaptive multisensor data fusion technique for train localisation and detection of accidental train parting," *Radar, Sonar & Navigation*, vol. 12, no. 8, pp. 853–861, 2018.
- [7] T. Narita, "Information processing apparatus and information processing method and computer program," *Plant Cell*, vol. 2574, no. 12, pp. 356–362, 2017.
- [8] D. K. Fisher, R. S. Fletcher, S. S. Anapalli, and H. C. Pringle III, "Development of an open-source cloud-connected sensor-monitoring platform," *Advances in Internet of Things*, vol. 8, no. 1, pp. 1–11, 2018.
- [9] Y. Shu, Y. Qiu, and S. Bo, "The Key Technology Study on Cloud Computing Platform for ECG Monitoring Based on Regional Internet of Things," *Zhongguo yi liao qi xie za zhi = Chinese journal of medical instrumentation*, vol. 40, no. 5, pp. 341–343, 2016.
- [10] Y. Wang, Y. Lin, R. Y. Zhong, and X. Xu, "IoT-enabled cloud-based additive manufacturing platform to support rapid product development," *International Journal of Production Research*, vol. 57, no. 12, pp. 3975–3991, 2019.
- [11] P. Shan, H. Lv, L. Yu, H. Ge, Y. Li, and L. Gu, "A multisensor data fusion method for ball screw fault diagnosis based on convolutional neural network with selected Channels," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 7896–7905, 2020.
- [12] Y. Zhou, X. Tao, Z. Yu, and H. Fujita, "Train-movement situation recognition for safety justification using moving-horizon TBM-based multisensor data fusion," *Knowledge Based Systems*, vol. 177, pp. 117–126, 2019.

- [13] Y. Li, Z. J. Yang, C. Chen, Y. X. Song, J. J. Zhang, and D. W. Du, "An integral algorithm for instantaneous uncut chip thickness measuring in the milling process," *Advances in Production Engineering & Management*, vol. 13, no. 3, pp. 297–306, 2018.
- [14] S. Corporation, "Input device and input method, information processing device and information processing method, information processing system and program," *Journal of the Robotics Society of Japan*, vol. 17, no. 7, pp. 916–919, 2016.
- [15] D. Guo, M. Zhong, and D. Zhou, "Multisensor data-fusion-based approach to airspeed measurement fault detection for unmanned aerial Vehicles," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, pp. 317–327, 2018.
- [16] A. Kanno, "Information processing system and information processing method," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 4, pp. 333–347, 2017.
- [17] H. Tamura, N. Koganezawa, F. Tamiya, K. Nakao, R. Oiwa, and Y. Fujino, "Tamiya F Storage medium storing information processing program, information processing apparatus and information processing method," *Group Processes & Inter-group Relations*, vol. 15, no. 5, pp. 637–651, 2017.
- [18] M. Elgharabawy, A. Schwarzhaupt, M. Frey, and F. Gauterin, "A real-time multisensor fusion verification framework for advanced driver assistance systems," *Transportation Research Part F Traffic Psychology and Behaviour*, vol. 61, pp. 259–267, 2017.
- [19] D. Gong, S. Liu, M. Tang, L. Ren, J. Liu, and X. Liu, "Revenue sharing or profit sharing? An internet production perspective," *Advances in Production Engineering & Management*, vol. 13, no. 1, pp. 81–92, 2018.
- [20] X. H. Wu and S. M. Song, "Covariance intersection-based fusion algorithm for asynchronous multirate multisensor system with cross-correlation," *Iet Science Measurement & Technology*, vol. 11, no. 7, pp. 878–885, 2017.
- [21] G. Muniandi and E. Deenadayalan, "Train distance and speed estimation using multi sensor data fusion," *Radar, Sonar & Navigation, IET*, vol. 13, no. 4, pp. 664–671, 2019.
- [22] P. Iervolino, R. Guida, D. Riccio, and R. Rea, "A novel multi-spectral, panchromatic and SAR data fusion for land Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 10, pp. 3966–3979, 2019.
- [23] N. Mezned, N. Mechrgui, and S. Abdeljaouad, "Mine wastes environmental impact mapping using Landsat ETM+ and SPOT 5 data fusion in the north of Tunisia," *Journal of the Indian Society of Remote Sensing*, vol. 44, no. 3, pp. 451–455, 2016.
- [24] M. Ramesh, S. Jana, and M. S. Bhat, "Modified Cascade Kalman Filter for Sensor Data Fusion in Micro Aerial Vehicle," *IFAC PapersOnLine*, vol. 49, no. 1, pp. 355–360, 2016.
- [25] L. I. Al-Shrouf, N. Szczepanski, and D. Söffker, "Online feature-based multisensor object detection system for bucket-wheel excavators," *International Journal of Advanced Manufacturing Technology*, vol. 82, no. 5-8, pp. 1213–1226, 2016.
- [26] C. M. de Farias, L. Pirmez, G. Fortino, and A. Guerrieri, "A multi-sensor data fusion technique using data correlations among multiple applications," *Future Generation Computer Systems*, vol. 92, pp. 109–118, 2019.
- [27] A. Verger, F. Baret, and M. Weiss, "A multisensor fusion approach to improve LAI time series," *Remote Sensing of Environment*, vol. 115, no. 10, pp. 2460–2470, 2011.
- [28] A. Fujii and I. Hamada, "Information processing apparatus, information processing method, and providing medium," *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 35, no. 8, pp. 2817–2822, 2018.
- [29] G. Aiello, I. Giovino, M. Vallone, P. Catania, and A. Argento, "A decision support system based on multisensor data fusion for sustainable greenhouse management," *Journal of Cleaner Production*, vol. 172, pp. 4057–4065, 2018.
- [30] E. Aïmeur, U. Ruhi, and M. Weiss, [Lecture Notes in Business Information Processing] *E-Technologies: Embracing the Internet of Things Volume 289 || Integrating Access Control Obligations in the Session Initiation Protocol for Pervasive Computing Environments*, Meur, Esma , U. Ruhi , and M. Weiss, 2017.
- [31] M. Villari, M. Fazio, S. Dustdar, O. Rana, D. N. Jha, and R. Ranjan, "Osmosis: the osmotic computing platform for microelements in the cloud, edge, and internet of Things," *Computer*, vol. 52, no. 8, pp. 14–26, 2019.
- [32] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, "A scalable multi-cloud storage architecture for cloud-supported medical internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1641–1654, 2020.