Security Technology of Wireless Sensor Internet of Things Based on Data Fusion

https://doi.org/10.3991/ijoe.v13i11.7748

Jie Zhang
Henan University Puyang Institute of Engineering, Henan, China
zhangjie@pyvtc.edu.cn

Abstract—In order to prove the effect of data fusion technology in the Internet of things, a wireless sensor Internet of things security technology based on data fusion is designed, and the impact of data fusion in the field of communication technology is studied. Therefore, two security fusion algorithms are designed on the basis of analyzing and comparing the advantages and disadvantages of various security fusion algorithms, namely, data security fusion algorithm EDCSDA and approximate fusion algorithm PADSA. By analyzing the probability distribution model of the data collected by the nodes, the disturbance data is superimposed on the original data to hide the effect of the original data. A test bed system for perception layer of the Internet of things is designed and implemented. The test results prove the feasibility of the two algorithms. Meanwhile, it shows that the two algorithms can reduce the transmission overhead of the network while guaranteeing the security. Based on the above finding, it is concluded that data fusion technology is very effective for improving network efficiency and prolonging the network life cycle as one of the key technologies in the perception layer of Internet of things.

Keywords—data security fusion, Internet of things, test bed system

1 Introduction

The Internet of things (IOT) refers to the acquisition of physical world information or control of objects in the physical world by deploying a variety of devices that have certain capabilities such as perception, computation, execution, and communication. Through the transmission, coordination and processing of network information, the network of human-object communication and object- object communication is realized. According to the universal hierarchical model of Internet of things, the Internet of things is logically divided into perception extension layer, network / service layer and application layer [1]. The perceived extension network of the Internet of things perception layer has the characteristics of large number of nodes, large scale and data multi hop transmission. In actual application, huge transmission and computation overhead will be generated, resulting in network congestion and excessive consumption of resources [2]. Therefore, in the routing protocol, the transmission overhead can be reduced by adding the data fusion to the nodes.

Data fusion is a data process by using computer technology to analyze and integrate some perceptual data acquired by timing under certain criteria, so as to complete the required decision-making and evaluation tasks. Data fusion brings new security problems while reducing network overhead. Therefore, a certain security mechanism in routing protocols is needed to improve the security of the fusion algorithm. Relevant research shows that the algorithm security is positively related to the consumption of resources. The higher the security, the greater the consumption of resources. If resource consumption is too large, the advantage of reducing overhead due to convergence will cease to exist. Therefore, the main task at this stage is to find high efficiency, high energy saving and high stability data security fusion algorithm to improve network performance and ensure the communication security of the Internet of things perception layer.

2 The data security fusion scheme

Although the CMT algorithm solves the problem of the original data and the confidentiality of the fusion data in the process of data fusion, the algorithm requires all nodes involved in the fusion to be a transmission ID number. There are many information of forwarding ID numbers, and the transmission overhead of nodes near the base station is very high, resulting in unequal network overhead and redundancy. Aiming at the shortage of ID number in CMT algorithm, a data fusion algorithm (EDCDSA) based on error correction packet compression coding is proposed in this paper. Based on the CMT algorithm, the algorithm solves the redundancy problem of ID number, and improves the robustness of the system by introducing the automatic error correction mechanism [3].

2.1 Network model and security hypothesis

Assuming that the Internet of things perception layer is larger, and the number of nodes is more. After the node is deployed, it will be fixed and will not move again. The network topology is a hybrid structure of cluster tree, and its model is shown in figure 1.

The central node of the whole network is called base station, and the base station collects all the network information [4]. Each small area in the network constitutes a cluster, and the nodes in the cluster responsible for data aggregation and aggregation are called cluster head. A tree structure is formed between the cluster heads, and the fusion results are uploaded to the base station in a multi hop manner. The following assumptions about node performance, network topology and security is assumed. Firstly, the base station cannot be attacked, and it has higher computing and storage capability. Second, ordinary nodes and cluster heads have certain computing and storage capabilities, and have certain probabilities to be attacked. Finally, when data is not needed to transmit and receive data, the node can switch to low-power or sleep mode to reduce energy loss.

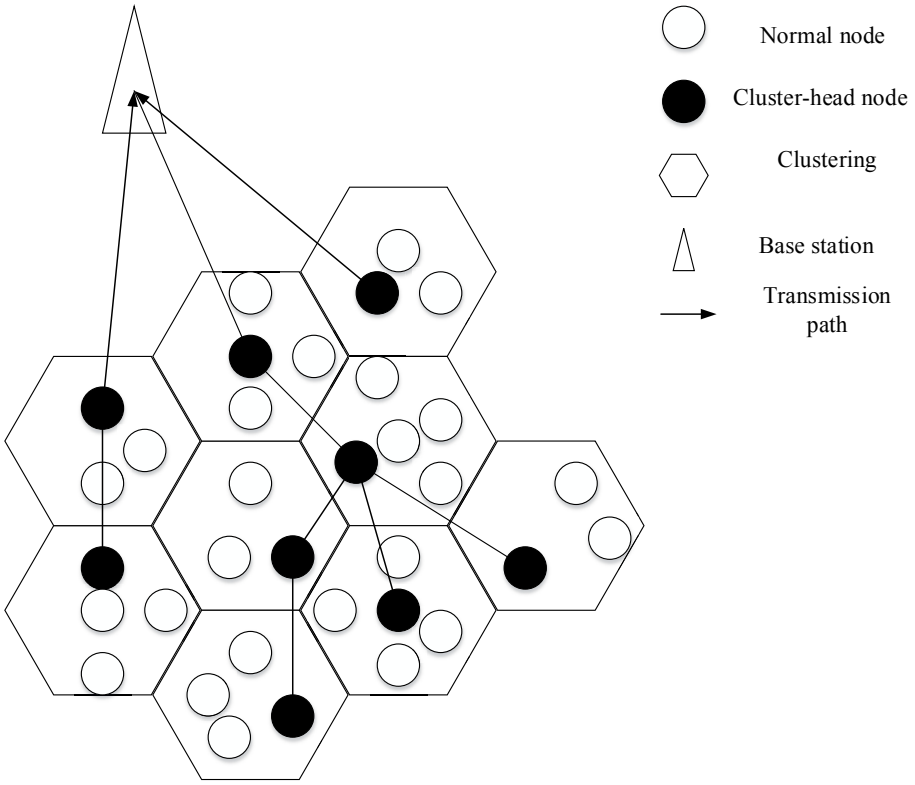

Fig. 1. Network topology diagram

The goal of the system design is to resist network eavesdropping attack, and to ensure the secret integrity of the message during the transmission and fusion. At the same time, it is necessary to ensure the integrity of the fusion data, and prevent the attacker from tampering with the identity of the nodes involved in the fusion, resulting in errors in the results of the system fusion [5].

2.2 Document content

The data security fusion algorithm designed in this paper consists of two parts: data encryption fusion and ID compression fusion. Table 1 shows the meanings represented by symbols in the system.

The algorithm flow consists of two parts, namely data encryption fusion and ID compression. Among them, the data fusion part uses the same steps as the CMT algorithm [6]. In the second chapter, it has given a detailed description, which is no longer detailed here. The focus of this chapter is the compression of the ID number.

Table 1. System parameters and meaning description

Parameter symbols	Meaning
S_i	A node whose node number is i, referred to as node i
$Seed_i$	The secret random number seed shared by node i and base station
$plain_i$	Data collected by node i
f(x)	It generates the function of the session key, and the parameter x is the seed of random numbers
key_i	Session key for node i
$cipher_i$	The data collected by node i is encrypted to ciphertext.
Q	It stands for the large prime number Q and is used for modulo
$Table_i$	List of sub- nodes of cluster head i
Count	Number of nodes within a cluster
length	Packet length for ID compression coding
num	Packet number for ID compression coding

CMT algorithm uses homomorphic encryption mechanism to ensure the confidentiality of the original data and the fusion data. In the algorithm, the key is dynamically generated by the ID number and timestamp of the node, which can effectively resist the replay attack. In terms of computational overhead, the CMT algorithm uses only addition and modulo operations, and the computational overhead is very low. But the algorithm takes a lot of resources because of the need to forward all the node ID numbers that participate in the data acquisition. In order to solve this problem, this chapter proposes a ID number fusion algorithm, which is suitable for the terminal nodes of the Internet of things perception layer. It can not only reduce the transmission overhead, but also ensure the reliability of the system [7].

The operation of fusing ID numbers is done by cluster heads. Before fusion, the base station stores all the ID of the cluster head and the ID table of the sub- nodes of all cluster heads. The cluster head fuses the data by homomorphic encryption, and fuses the ID according to the method in this article. After the fusion, the cluster is uploaded to the base station in the required packet format. After receiving data packets, the base station decodes the ID number first. The ID number of the cluster head is checked to obtain the ID number of the members in the cluster, so as to decrypt the data.

3 Example verification and performance analysis

3.1 Example verifications

Model hypothesis: Cluster head ID is 0 x 8000, and there are 16 common nodes in cluster. Each node has a unique 16-bit ID number, 0 x 1001~0 x 100f. Among them, node 0 x 1001, 0 x 1007, 0 x 100b, 0 x 100e and 0 x 1010 are all involved in data fusion. In accordance with the CMT algorithm, the cluster head attaches the ID number directly to the back of the fusion data. The ID number of each node takes up 32-

bit of memory, and the ID number of the 5 nodes will bring the transmission overhead of 32*5=160 bit per hop [8]. According to the method proposed in this paper, the ID number is coded in two steps: the first step is compression coding, and the second step is error correction coding.

According to the method proposed in this paper, the ID number is coded in two steps: the first step is compression coding, and the second step is error correction coding. The first step: In accordance with the member's ID number, the cluster head will re number the node from 1 to 16 through the order from small to large. After generating a 16- bit binary sequence, each member corresponds to a node number. If the number of nodes is involved in the fusion, the bit is set to 1, and the other is set to 0, generating a sequence of 1000001000100101. According to the following formula, the coding sequence number of the sequence is calculated $C_{12}^4 - C_{10}^4 + C_9^3 - C_6^3 + C_5^2 - C_3^2 = 357$. Its binary form is 101100101, and the number of '1' in the original sequence is 5, and the binary form is 101. It is combined with the encoding sequence into a binary sequence 101101100101 as the input signal of the error correction code.

The encoder is considered as a linear network, and its response is analyzed by the shock response of the linear system. u represents the input sequence. The output sequences V_1 and V_2 are the convolution of u and the corresponding impulse response, respectively [9]. The impact response is expressed in g_1 and g_2. For this encoder, there is g_1 = (111) and g_2 = (101), and the coding equation is:

$$v_1 = u * g_1, v_2 = u * g_2 \tag{1}$$

The input signal u and the shock response g_1 and g_2 are brought into the equation. When calculating, the modulus 2 addition operation is used, and the coding output is as follows:

$$\begin{aligned} v_1 &= 101101100101 * 111 = 11000000111011 \\ v_2 &= 101101100101 * 101 = 10011011110001 \end{aligned} \tag{2}$$

After encoding V_1 and V_2, coding results can be obtained, which are v = (11,10,00,01,01,00,01,01,11,11,10,00,10,11). The decoding process is also divided into two steps: the first step is to decode the error correcting code, and the second step is to decode the compressed code to find the ID number of the nodes involved in the fusion.

In order to verify the correction ability of error correcting codes, it is assumed that coding v has error code and it becomes v' during the transmission process. The equation is v'= (11,10,00,01,01,01,01,01,10,11,10,00,10,11). There are two errors compared to v' and v. The maximum likelihood decoding sequence is 102101100101 by stack memory decoding, and the error correction is successfully completed.

3.2 Performance analysis

Performance analysis is divided into two parts: security performance and transmission overhead.

Safety performance: The EDCDSA algorithm is equivalent to the CMT method for data fusion [10]. Security is the confidentiality of pseudo - random seeds shared by nodes and base stations. The CMT algorithm takes the form of plaintext transmission for the transmission of ID numbers because the attacker cannot know the pseudo-random seed shared by the node and the base station through the ID number. However, the attacker can tamper with the ID number to destroy the fusion result, so that the base station cannot decrypt the correct data. The EDCDSA algorithm proposed in this paper has fused the ID number and added the error correction mechanism. The error correction mechanism can not only prevent the error in the process of message transmission, but also detect the malicious tampering of attackers, and improve the security of the system.

Transmission overhead: EDCDSA algorithm has greatly improved the transmission overhead of the system. The transmission overhead of CMT algorithm, EDCDSA algorithm and simple ID encoding algorithm is compared and analyzed. It is assumed that the number of nodes in a cluster is Count, and the ID number of the node takes up r bits of memory space. In each cycle, the probability of nodes participating in the fusion is n, and the participating probability is equal. The simple ID encoding algorithm uses a Count bit binary number to represent the nodes involved in the fusion, and each bit of the binary number corresponds to a node in the cluster. If the node is involved in the fusion, the corresponding bit is set to '1', whereas the other is set to '0'. Since the error correction coding is independent of ID fusion, and the CMT algorithm does not provide error correction mechanism, the error correction coding is not considered when comparing the transmission overhead. Under the changing situation of the number of nodes, the occupied space of node ID, and the node's participation in the fusion probability, the transmission overhead of CMT algorithm, EDCDSA algorithm and simple old coding algorithm are simulated by MATLAB respectively. The simulation results are shown in figure 2, 3 and 4.

The transmission overhead of CMT algorithm is greater than that of simple ID encoding algorithm and ECDSDA algorithm under the premise of changing the number of nodes, the length of ID and the probability of node participation in fusion. When the node participates in the fusion probability, the number of nodes and the length of ID have little influence on the simple ID encoding algorithm and the ECDSDA algorithm. There are only two kinds of nodes in a cluster: participation, fusion and non-participation. When the node of the job exceeds Count/2, we only need to process the ID number of the node that does not participate in the fusion. Therefore, for the probability of node participation in fusion, the effect of η and $1-\eta$ on transmission overhead is the same. Although we only consider one possibility that $0 \leq \eta \leq 0.5$, it covers all the possibilities. If the number of nodes and ID length are provided, and the probability of node fusion is of $n \leq 0.35$ or $n \geq 0.65$, The ECDSDA algorithm has advantages over simple ID coding methods. What kind of encoding is adopted depends

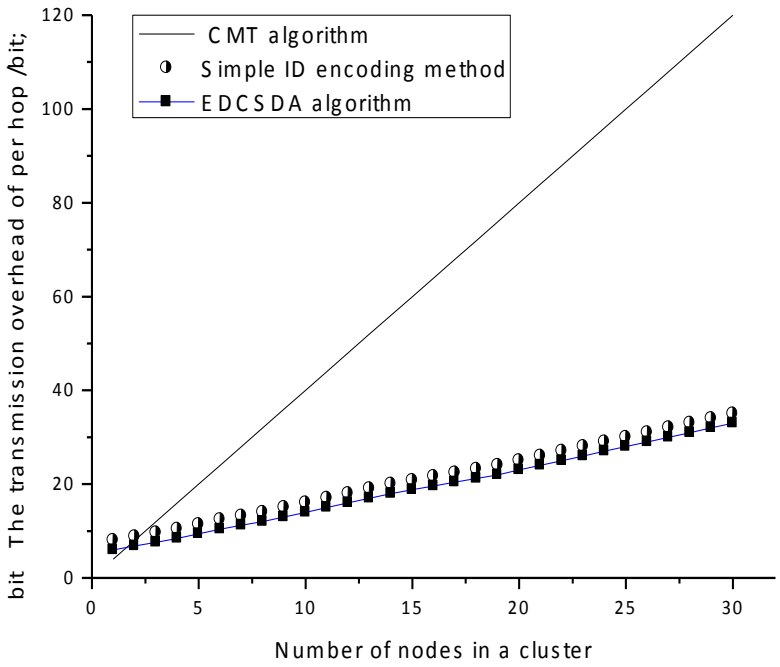

Fig. 2. Relation between the number of nodes and the transmission overhead (r=8 bits, η =0.5)

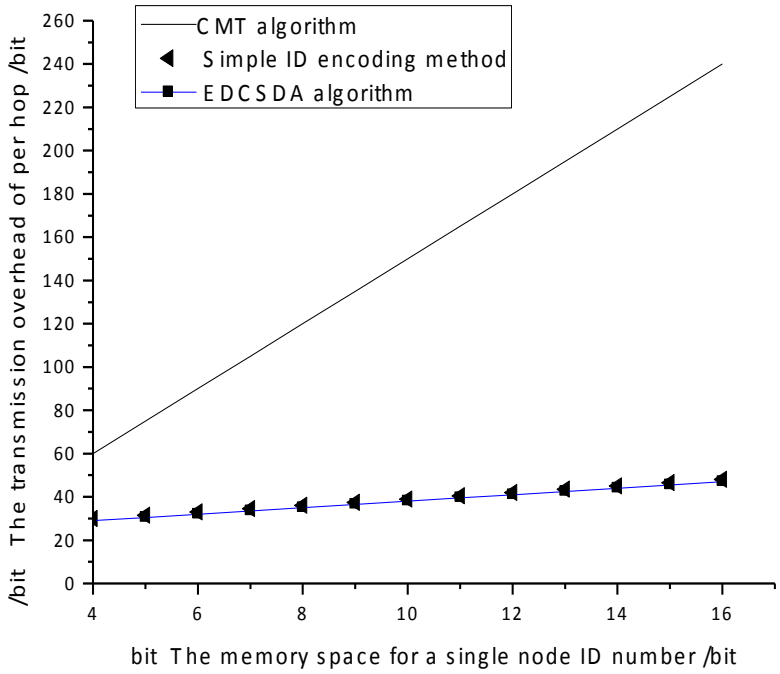

Fig. 3. Relation between ID length and transmission overhead (Count=30, η =0.5)

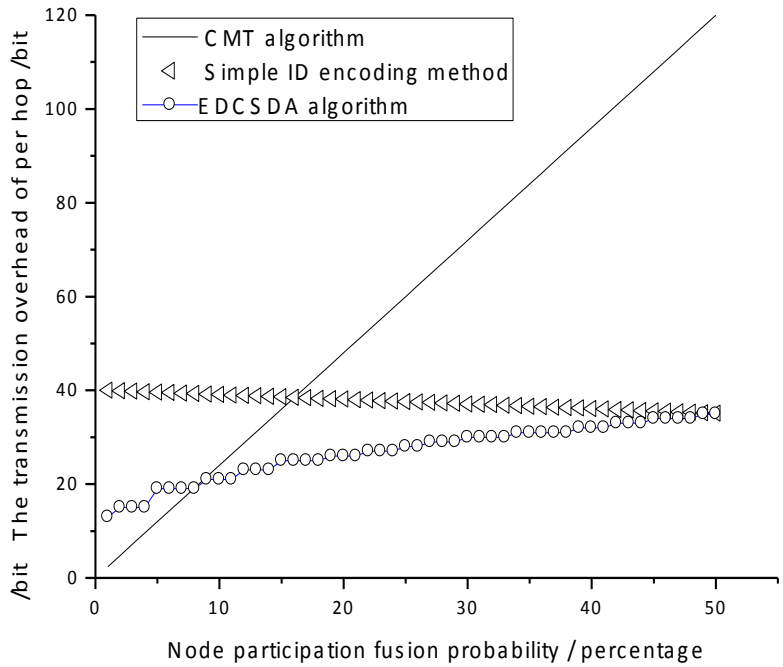

Fig. 4. Relation between the probability of node participation in fusion and the transmission overhead (Count=30, r=8 bits)

on the actual participation of network nodes. The method presented in this paper has obvious advantages for the network with larger or smaller proportion of fusion nodes.

A ECDSDA algorithm based on packet compression and error correction coding is proposed. The specific flow of the algorithm is described. The feasibility, security and energy saving of the algorithm are proved by theoretical analysis, example verification and MATLAB simulation. The results show that the ECDSDA method has obvious advantages over the CMT method in terms of transmission overhead, and also improves the security.

4 Approximate fusion method based on probability distribution

Aiming at the shortage of ID redundancy in CMT algorithm, the ECDSDA algorithm proposed in the third chapter provides a method of ID fusion. The error correction mechanism is added to improve the robustness of the system. However, there is still a security problem that corresponds to the ID number and the key, and the attacker can destroy the fusion result by tampering with the ID number of the data packet sent by any source node. Moreover, the node management and update mechanism is complex and will consume a lot of resources. For this reason, an approximate fusion algorithm based on probability distribution (PADSA) is proposed, which reduces the transmission overhead at the expense of partial accuracy. First, under certain meas-

urement conditions, the absolute value of accidental error will not exceed the upper limit. Second, the probability of occurrence of error with small absolute value is greater than large one. Third, the probabilities of positive and negative errors of absolute equality are equal. Finally, under the same measurement conditions, the arithmetic average of accidental error caused by repeated measurements of infinity tends to 0.

System errors are determined by component manufacturing process. It can be reduced or eliminated by calibration, which usually has little effect on the measured results. The influence of the actual measurement result is mainly caused by accidental error. According to the characteristics of accidental error, it can be approximately considered that the accidental error conforms to the normal distribution. Since the range of normal distribution is $(-\infty, +\infty)$ in the real field, the measured data cannot be completely consistent, and can only be approximately estimated. Because the nodes in the same cluster are similar in location and the distribution is similar, in order to facilitate the analysis, it is assumed that the nodes of the same clustering data are independent of each other. At the same time, the variables satisfy the same normal distribution, and the mean is the actual value to be measured. The variance of the data reflects the fluctuation of the accidental error of the measured data.

4.1 Acquisition data distribution model

In a certain area, the node data of perception layer of the Internet of things usually have strong correlation, such as temperature and light. Therefore, it can be approximately considered that the data collected by the nodes in the same cluster are approximately equal, while the data collected by the similar cluster nodes are similar in numerical value. But in actual measurement, because of the different manufacturing process of sensor nodes and the noise interference of the environment, it will lead to measurement errors. There are mainly two kinds of measurement errors: system error and accidental error.

System error: The same physical quantity is repeated n times under the same observation condition and the same operation. If the size and symbols of the error are the same or they change according to a certain law, then this error is called the system error. System error is generally caused by the lack of accuracy of the measurement tool. For example, A ruler with an actual length of 1.001 meters is used to measure length. The measure of each meter will result in an error of ++0.001 meters. The longer the length to be measured is, the greater the error is.

Accidental error: The same physical quantity is measured n times under the same observation condition and the same operation. If the size and the symbol of the error are uncertain, the error is called accidental error, also known as random error. Accidental errors are usually caused by environmental noise and man-made readings. For single measurements, accidental error cannot predict its size and symbol. However, under the same observation condition, after repeated measurements of a physical quantity, the accidental error will show a certain statistical law. With the increase of observation times, its regularity is more and more obvious. The regularity of accidental error is reflected in the following aspects:

Model hypothesis: The network consists of N_group clusters, and the number of sub- nodes C_Num of each cluster head is equal, and the number of them is 100. In the same cluster, the variables collected by different nodes are independent of each other and obey the same normal distribution. For a specific cluster $\mathrm{Agg_i}$, the mean u_i of normal distribution, which is obtained by the data collected from intra cluster nodes, obeys the uniform distribution U (20,30). The variance a_i of the normal distribution obeys the uniform distribution u_i, and the uniform distribution U (λ, λ+1). The system error threshold is β, and the perturbation data added to each node in the network is X~U (0, R). The normal distribution image selects 0.025 upper sub loci.

Goal achievement: According to the parameters, the value of the R is found. For the single cluster, the cluster head cannot make the estimated value accord with the error threshold according to the statistical method. For the base station, the approximation error threshold can be estimated by statistical method.

As shown in figure 5 and 6. by controlling variables, the influence of different parameters on the value of R is studied, and the simulation image is designed by matlab. The section between the two curves in the graph satisfies the required range of the R. On the basis of meeting the requirements, the size of the R value is set by the user according to the actual situation. As the R value increases, the security of the system increases, whereas vice versa. As shown in figure 5, the maximum and minimum values of R increase as the error threshold increases. As shown in figure 6, when the number of nodes in the cluster is fixed, the R maxima increase with the number of clusters, while the minimum values are independent of the number of clusters.

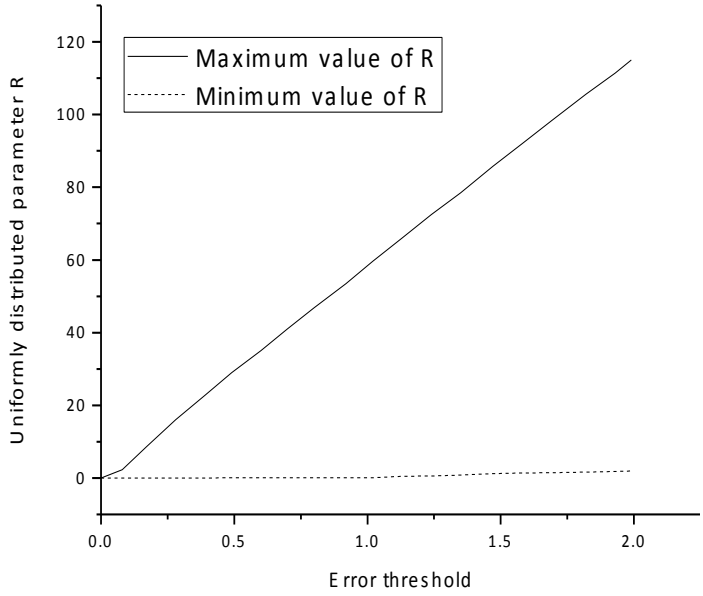

Fig. 5. Rrelation of error threshold and R

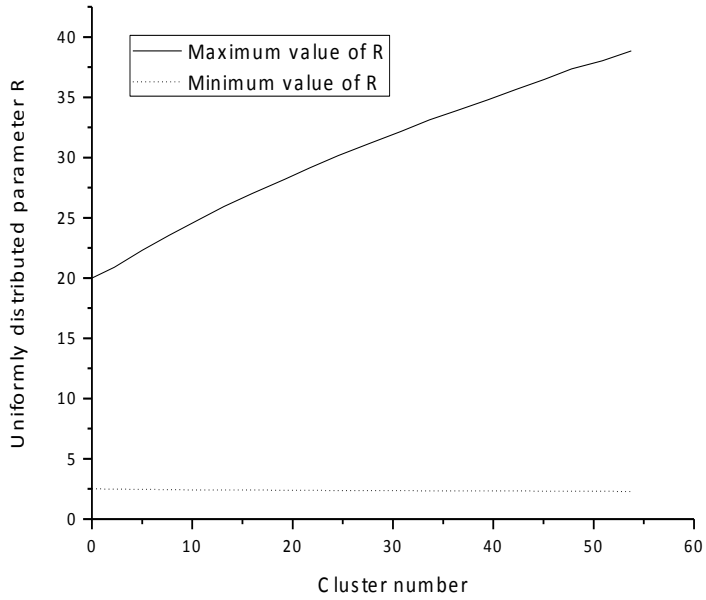

Fig. 6. Relation between the number of clusters and the R

This chapter proposes a PADSA algorithm based on probability distribution, and describes the specific process of the algorithm. The feasibility, energy saving and security of the system are proved by theoretical analysis, case study and MATLAB simulation. The results show that the PADSA method has obvious advantages over the EDCDSA algorithm proposed in the previous chapter and the CMT algorithm in terms of transmission overhead, but the fusion result is approximate. The accuracy of the fusion results is related to the network size. The higher the number of nodes in the network is, the higher the accuracy of the fusion results is.

5 Conclusions

Two algorithms for data security fusion of the perception of Internet of things are designed. EDCDSA algorithm improves the classical CMT algorithm, and solves the problem that ID transmission takes too much transmission overhead. The PADSA algorithm is suitable for large-scale networks and achieves approximate fusion of data. At the same time, the test bed system is designed and implemented. First, the key technologies of the security integration of the Internet of things are discussed in detail, while the advantages and disadvantages of homomorphic encryption, data fragmentation and approximate fusion are summarized. Second, based on the analysis of the advantages and disadvantages of the existing schemes, the compression coding and error-correction coding techniques are studied. The classical algorithm CMT is improved, and a block compression coding algorithm (EDCDSA) with error correction capability is designed. This solves the redundancy problem of CMT algorithm ID

number transmission. Matlab simulation shows that the algorithm can reduce the transmission overhead of the network. Third, an approximate security fusion algorithm (PADSA) for large-scale networks is designed. The algorithm eliminates the overhead of key management, and does not need to transmit the node ID number, thus further reducing the transmission overhead of the network. The security is proved, while the effectiveness and energy saving of the algorithm are verified by MATLAB simulation.

6 References

[1] Ranjan, R., Wang, M., Perera, C., Jayaraman, P. P., Zhang, M., & Strazdins, P., et al. (2015). City data fusion: sensor data fusion in the internet of things. International Journal of Distributed Systems & Technologies, 7(1), 15-36.

[2] Zou, P., & Liu, Y. (2015). An efficient data fusion approach for event detection in heterogeneous wireless sensor networks. Applied Mathematics & Information Sciences, 9(1), 517-526. https://doi.org/10.12785/amis/090160

[3] Luo, X., & Chang, X. (2015). A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks. International Journal of Control Automation & Systems, 13(3), 539-546. https://doi.org/10.1007/s12555-014-0309-8

[4] Fei, X., & Xiaofang, L. I. (2016). Wireless sensor network data fusion algorithm based on compressed sensing theory. Journal of Jilin University, 54(3), 575-579.

[5] Liu, L., Luo, G., Qin, K., & Zhang, X. (2017). An algorithm based on logistic regression with data fusion in wireless sensor networks. Eurasip Journal on Wireless Communications & Networking, 2017(1), 10. https://doi.org/10.1186/s13638-016-0793-z

[6] Xiao, L., & Jian, Y. (2016). Wireless sensor network data fusion model based on compressed sensing theory. Journal of Computational & Theoretical Nanoscience(12), 9515-9520. https://doi.org/10.1166/jctn.2016.5875

[7] Liu, Y., Zhao, Y., Zhao, Y., & Wang, L. (2016). The reliability analysis of wireless sensor networks based on the energy restrictions. International Journal of Wireless & Mobile Computing, 10(4), 399-406. https://doi.org/10.1504/IJWMC.2016.078220

[8] Dai, Z., & Yuanxiang, L. I. (2015). Research on wireless sensor decision network of multilayer agent data fusion and its multiplicity. Computer Engineering, 41(3), 198-203,217.

[9] Chen, S., Gao, H., Liu, Y., Liang, D., & Wu, D. (2016). In network data fusion for agricultural information on wireless sensor nodes based on jn5139. Journal of Agricultural Mechanization Research, 91(16), 7648-7652.

[10] Tian, L., & Jing, Z. (2015). A data fusion algorithm based on neural network research in building environment of wireless sensor network. International Journal of Future Generation Communication & Networking, 8(78), 2082-7.

7 Author

Jie Zhang is with Henan University Puyang Institute of Engineering, Department of mathematics and Information Engineering, Henan, China (zhangjie@pyvtc.edu.cn).

Article submitted 24 September 2017. Published as resubmitted by the author 28 October 2017.