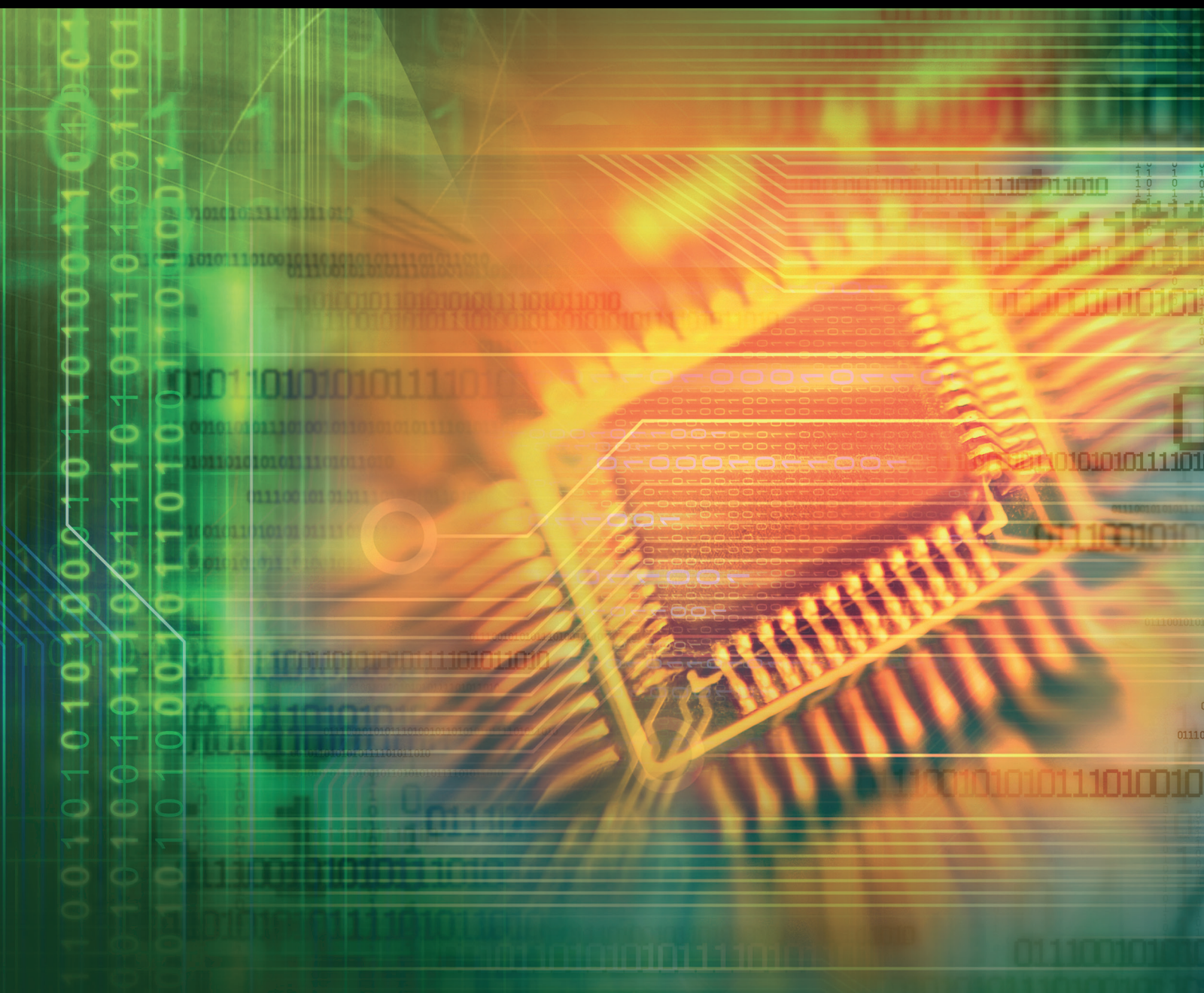
Journal of Electrical and Computer Engineering

Security and Privacy in Internet of Things with Crowd-Sensing

Lead Guest Editor: Liangmin Wang
Guest Editors: Hongjian Sun, Zhuo Lu, Yantian Hou, and Mengxing Huang



Security and Privacy in Internet of Things with Crowd-Sensing

Journal of Electrical and Computer Engineering

Security and Privacy in Internet of Things with Crowd-Sensing

Lead Guest Editor: Liangmin Wang
Guest Editors: Hongjian Sun, Zhuo Lu, Yantian Hou, and Mengxing Huang

Editorial Board

Contents

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 2057965, 2 pages
https://doi.org/10.1155/2017/2057965

Editorial
Security and Privacy in Internet of Things with Crowd-Sensing

Liangmin Wang,[1] Zhuo Lu,[2] Hongjian Sun,[3] Yantian Hou,[4] and Mengxing Huang[5]

[1]*Department of Cyber Security, Jiangsu University, Zhenjiang 212013, China*
[2]*Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA*
[3]*School of Engineering and Computing Sciences, University of Durham, Durham DH1 3HP, UK*
[4]*Department of Computer Science, Boise State University, Boise, ID 83706, USA*
[5]*College of Information Science and Technology, Hainan University, Haikou 570100, China*

Correspondence should be addressed to Liangmin Wang; wanglm.ujs@gmail.com

Received 28 November 2017; Accepted 28 November 2017; Published 21 December 2017

The rapid proliferation of mobile sensing devices, such as smartphones, wearable devices, and mobile vehicles, has promoted the emergence of a novel sensing paradigm for Internet of Things (IoTs). Crowd-sensing, known as a promising data collection method, is becoming increasingly popular in IoTs due to its low deployment cost and large-scale spatial coverage. Crowd-sensing-based IoTs interconnects various physical objects, including human, sensors, and smart devices, to collect data through the enhanced communication technology and process and share information with the assistance of cloud servers. Currently, a wide range of crowd-sensing applications are fostered in various domains such as environmental monitoring, assistive healthcare, social network, business, and intelligent transportation.

Numerous research challenges arise in the crowd-sensing-based IoTs, among which security and privacy are two critical issues that hinder the ubiquitous deployment of relevant applications. For data sensing in the front-end IoTs, sensed data may contain some sensitive information of mobile users. Moreover, data can be falsified and tampered with by external attackers or even be polluted by internal attackers (i.e., malicious users). For data storage and processing in the back-end IoTs, cloud servers may be curious to infer private information of data owners and query users or even maliciously modify some query results. In these circumstances, many factors, including user privacy, data confidentiality, integrity, reliability, and access control, should be taken into consideration in a holistic perspective.

This special issue provides the opportunity for researchers, practitioners, and application developers to discuss the recent technical advances and future challenges in security and privacy protection for crowd-sensing-based IoTs. The topic of accepted papers pertains to security and privacy issues from different aspects, ranging from secure network architecture, secure and privacy-preserving data sensing, and data transmission to data processing in IoTs with crowd-sensing.

Over the numerous submissions, six papers have been accepted after the rigorous review process. We now list and give a brief summary of these papers as below.

The paper "A Student Information Management System Based on Fingerprint Identification and Data Security Transmission" by P. Yang et al. studies the secure and reliable data transmission in the student information management system. Based on an improved AES algorithm, the authors propose a novel data encryption method and design a new S-box, which significantly reduces the encryption time. Experimental results also validate the efficiency of their algorithm.

The paper "Vulnerability Analysis of Interdependent Scale-Free Networks with Complex Coupling" by C. Cao et al. mainly analyzes the vulnerability of interdependent scale-free networks with complex coupling based on the BA model. The results indicate that these networks have the same vulnerability against the maximum node attack, the load of the maximum node attack, and the random node attack, indicating that the coupling relationship between network nodes is an important factor in network design.

The paper "The Anonymization Protection Algorithm Based on Fuzzy Clustering for the Ego of Data in the Internet of Things" by M. Xie et al. introduces the concept of ego of data and implements two steps of data clustering for the IoTs, which obscures the specific location information and achieves the anonymization protection. Experimental results show that their proposed algorithm can protect the data more efficiently, without sacrificing the anonymization quality.

The paper "A Variable Weight Privacy-Preserving Algorithm for the Mobile Crowd Sensing Network" by J. Zhong et al. addresses the problem of user privacy leakage in mobile crowd-sensing scenarios. The authors propose a variable weight privacy-preserving algorithm of secure multiparty computation. Its effectiveness and feasibility are demonstrated through experiments.

The paper "Abnormal Event Detection in Wireless Sensor Networks Based on Multiattribute Correlation" by M. Wang et al. focuses on the issue of abnormal event detection in wireless sensor networks. A novel approach is proposed to improve the quality of detection results, which considers both spatiotemporal and attributes correlations. Experiments prove the effectiveness and high detection accuracy of their scheme.

The paper "Health Monitoring System for Nursing Homes with Lightweight Security and Privacy Protection" by Y. Jiang et al. mainly designs a secure and effective monitoring system for nursing homes. A mobile authentication protocol based on hash function is proposed to realize secure access and privacy protection. Compared with the traditional protocols, lower computation and communication cost is induced, which satisfies the high real time and stronger security requirements.

Acknowledgments

We would like to thank all the authors for submitting their manuscripts to this issue and all the reviewers for every effort in the professional review process.

Liangmin Wang
Zhuo Lu
Hongjian Sun
Yantian Hou
Mengxing Huang

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 9598581, 6 pages
https://doi.org/10.1155/2017/9598581

Research Article

A Student Information Management System Based on Fingerprint Identification and Data Security Transmission

Pengtao Yang, Guiling Sun, Jingfei He, Peiyao Zhou, and Jiangjiang Liu

College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China

Correspondence should be addressed to Guiling Sun; sungl@nankai.edu.cn

Received 18 February 2017; Revised 17 July 2017; Accepted 17 August 2017; Published 19 September 2017

Academic Editor: Liangmin Wang

A new type of student information management system is designed to implement student information identification and management based on fingerprint identification. In order to ensure the security of data transmission, this paper proposes a data encryption method based on an improved AES algorithm. A new *S*-box is cleverly designed, which can significantly reduce the encryption time by improving ByteSub, ShiftRow, and MixColumn in the round transformation of the traditional AES algorithm with the process of look-up table. Experimental results show that the proposed algorithm can significantly improve the encryption time compared with the traditional AES algorithm.

1. Introduction

At present, there are a large number of college students, so the identification and verification of student identity information occur at all times in the campus, as well as the corresponding services given by the students' identification. Therefore, safe and efficient student information management, convenient identification to obtain the required service, and safe and reliable information transmission have become an important task for the student information management [1–3]. Three main features of the proposed system are the following:

(1) This system uses the fingerprint identification terminal to collect the fingerprint information. By means of replacing the campus card with the physiological characteristics of lifelong invariance, uniqueness, and convenience, it has become the basis of student identity authentication. The maturity of the fingerprint identification technology ensures the safety and speed of the process and also eliminates the disadvantages of the campus card which is easy to be stolen and forged and easily lost.

(2) In order to ensure the safety of the students' information, the fingerprint characteristic value is encrypted and transmitted, using the improved AES encryption algorithm [4], which has the same security guarantee with traditional AES algorithm but reduces the required time for encryption. Therefore, this student management system not only is convenient for students in the college, but also protects the privacy of students.

(3) After the system has been built, because it is easy to maintain and popularize, the modular system design is easier to improve, and it can be widely used in other fields.

2. Description of the Student Information Management System

The system is mainly composed of two parts: terminal and host computer. The terminal is composed of fingerprint identification module and micro controller. The host computer can use personal computers or large servers according to the number of users, and the management of student information database uses SQL Server. The terminal fingerprint sensor uses optical fingerprint recognition module, while the microcontroller uses STM32F4, with 192 KB of SRAM [5]. Each terminal processes and encrypts the collected fingerprint data and then transmits it to the host computer. To ensure the safety of data, the fingerprint data is only stored in the host

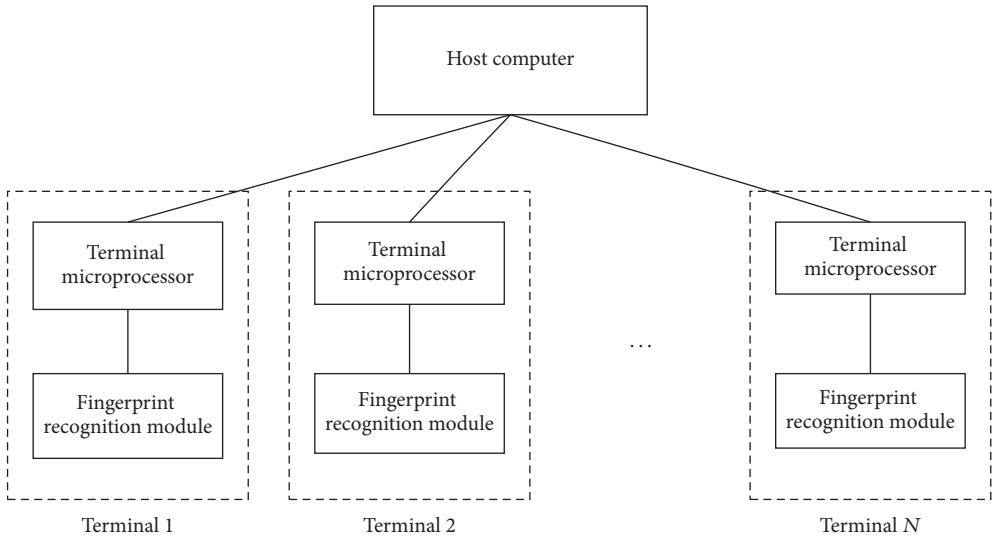


Figure 1: System structure diagram.

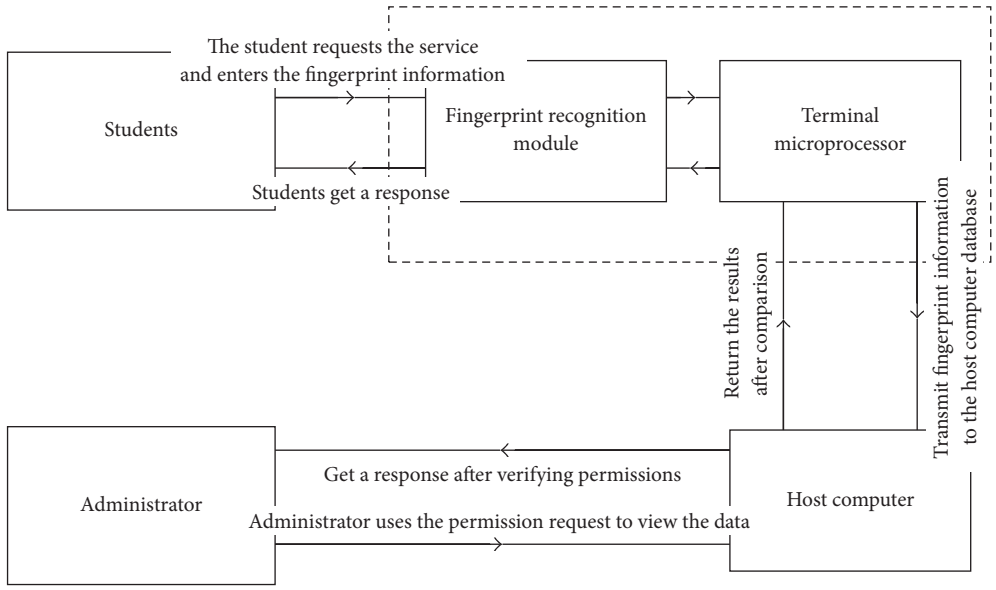


Figure 2: System flow diagram.

computer database, and the terminals are only responsible for collection and processing. The system structure is shown in Figure 1.

3. Implementation of the Student Information Management System

The system collects fingerprints through the terminal fingerprint identification sensor. And the microprocessor processes and encrypts the fingerprint information and then transmits it to the server. On the server side, it compares the fingerprint information transmitted from terminal with the fingerprint information stored in the server database. If the identity is consistent, the user is allowed to operate by verification. The overall process is shown in Figure 2.

4. Data Transmission Encryption Method

In order to achieve the campus student consumption, identity registration, and other functions, the student information identification management system based on fingerprint identification and data security transmission needs to transmit student fingerprint information, identity information, and bank card information among the terminal. There is a risk of being intercepted during data transmission. Students' private information has a high commercial value; once intercepted by criminals, the consequences could be disastrous. When using plaintext transmission, security is very low; therefore, the entire data transmission using ciphertext transmission, to achieve a plaintext view and ciphertext transmission effect, greatly improve the security, so that criminals cannot take the opportunity. In order to ensure the security of encrypted

transmission and user-friendliness, the encryption process uses the optimized AES algorithm.

AES algorithm is a variable data block length and variable key length iterative block cipher algorithm, and the length of the data block and the key length can be 128, 192, or 256 bits [6]. The most important operation in the AES algorithm is the round transformation operation, where the various operations applied to the process give a high encryption strength. The round transformation operation consists of four steps: ByteSub, ShiftRow, MixColumn, and AddRoundKey, and these steps will be mathematically transformed to eventually construct a new S-box [7, 8].

4.1. Matrix Representation of AES Algorithm Round Transformation. AES algorithm mainly consists of three modules: encryption module, decryption module, and key expansion module. Each round transformation of the encryption module consists of ByteSub, ShiftRow, MixColumn, and AddRoundKey four operations [9]. The decryption module is also composed of four similar operations; the difference is that ByteSub, ShiftRow, and MixColumn are the inverse operation of the encryption module. And the extension key used in AddRoundKey is generated by the key expansion module. The encryption module and the decryption module are the core of the AES algorithm, which are the repetition process of the round transformation, so the simplified round function can improve the operation speed of the AES algorithm [10, 11].

For convenience of description, 128-bit (16 bytes) data is used here and the key is 128 bits.

In the ByteSub transformation, it is assumed that the input is A, $A = [a_{i,j}]$, $(0 \le i, j \le 3)$; output is B, $B = [b_{i,j}]$, $(0 \le i, j \le 3)$. ByteSub transformation can be expressed as

$$B = (A). \tag{1}$$

And it can also written as

$$b_{i,j} = B\left(a_{i,j}\right). \tag{2}$$

In practice, this transformation can be converted to look-up table operation. The table is the AES algorithm byte conversion table, also known as S box.

In the ShiftRow transformation, the schematic diagram is shown in Figure 3. It is assumed that the output is C, $C = [c_{i,j}]$, $(0 \le i, j \le 3)$.

Then C can be expressed as a matrix:

$$\begin{bmatrix} c_{0,j} \\ c_{1,j} \\ c_{2,j} \\ c_{3,j} \end{bmatrix} = \begin{bmatrix} b_{0,(j+0)\%4} \\ b_{1,(j+1)\%4} \\ b_{2,(j+2)\%4} \\ b_{3,(j+3)\%4} \end{bmatrix}. \tag{3}$$

In the MixColumn transformation, each column of the state array obtained in ShiftRow is treated as a polynomial on GF(2^8) and modulo $x^4 + 1$ multiplication with a fixed polynomial $03x^3 + 01x^2 + 01x + 02$.

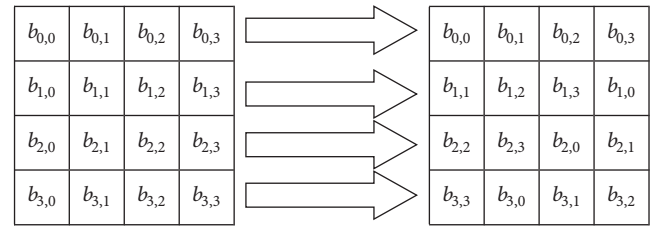


Figure 3: ShiftRow transformation schematic diagram.

It is assumed that the output is D, $D = [d_{i,j}]$, $(0 \le i, j \le 3)$; then MixColumn can also be written as matrix multiplication [12–14]:

$$\begin{bmatrix} d_{0,j} \\ d_{1,j} \\ d_{2,j} \\ d_{3,j} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} c_{0,j} \\ c_{1,j} \\ c_{2,j} \\ c_{3,j} \end{bmatrix}. \tag{4}$$

In the AddRoundKey transformation, the expansion round key generated by the key expansion module begins to function. Set the round key to K, $K = [k_{i,j}]$, $(0 \le i, j \le 3)$. Set the output to E, $E = [e_{i,j}]$, $(0 \le i, j \le 3)$. Then AddRoundKey can be expressed as a matrix:

$$\begin{bmatrix} e_{0,j} \\ e_{1,j} \\ e_{2,j} \\ e_{3,j} \end{bmatrix} = \begin{bmatrix} d_{0,j} \\ d_{1,j} \\ d_{2,j} \\ d_{3,j} \end{bmatrix} \oplus \begin{bmatrix} k_{0,j} \\ k_{1,j} \\ k_{2,j} \\ k_{3,j} \end{bmatrix}. \tag{5}$$

Equations (2), (3), and (4) into (5) can get

$$\begin{aligned} \begin{bmatrix} e_{0,j} \\ e_{1,j} \\ e_{2,j} \\ e_{3,j} \end{bmatrix} = {} & \begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} S\left[a_{0,(j+0)\%4}\right] \oplus \begin{bmatrix} 03 \\ 02 \\ 01 \\ 01 \end{bmatrix} S\left[a_{1,(j+1)\%4}\right] \\ & \oplus \begin{bmatrix} 01 \\ 03 \\ 02 \\ 01 \end{bmatrix} S\left[a_{2,(j+2)\%4}\right] \oplus \begin{bmatrix} 01 \\ 01 \\ 03 \\ 02 \end{bmatrix} S\left[a_{3,(j+3)\%4}\right] \\ & \oplus \begin{bmatrix} k_{0,j} \\ k_{1,j} \\ k_{2,j} \\ k_{3,j} \end{bmatrix}. \end{aligned} \tag{6}$$

Above we have come to a matrix representation between input A and output E of each round transformation of AES algorithm [15–17].

4.2. Optimized AES Algorithm. In (6), to calculate $\begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} S[a_{0,(j+0)\%4}]$ requires one xtime [4] operation and one exclusive-OR operation. Thus, getting each column vector of a round transformation result E requires four xtime operations and eight exclusive-OR operations (regardless of round key generation). According to the observation we can see in the column vector multiplied by $S[a_{0,(j+0)\%4}]$, $S[a_{1,(j+1)\%4}]$, $S[a_{2,(j+2)\%4}]$, and $S[a_{3,(j+3)\%4}]$, only the three elements: 01, 02, and 03. So we can create a new S box to get directly each element in the $\begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} S[a_{0,(j+0)\%4}]$, $\begin{bmatrix} 03 \\ 02 \\ 01 \\ 01 \end{bmatrix} S[a_{1,(j+1)\%4}]$, $\begin{bmatrix} 01 \\ 03 \\ 02 \\ 01 \end{bmatrix} S[a_{2,(j+2)\%4}]$, and $\begin{bmatrix} 01 \\ 01 \\ 03 \\ 02 \end{bmatrix} S[a_{3,(j+3)\%4}]$ four column vectors by look-up table method, so that we can save four xtime operations and four exclusive-OR operations and get each column vector of a round transformation result E which requires only four exclusive-OR operations (regardless of round key generation). Let data in the original S box operate, respectively, with 01, 02, 03, and we get a new byte conversion table, as shown in Table 1.

In the use of C language to implement, the table will be set to a two-dimensional array $S_{\text{new}}[256][3]$, so that we can get each element of $\begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} S[a_{0,(j+0)\%4}]$, $\begin{bmatrix} 03 \\ 02 \\ 01 \\ 01 \end{bmatrix} S[a_{1,(j+1)\%4}]$, $\begin{bmatrix} 01 \\ 03 \\ 02 \\ 01 \end{bmatrix} S[a_{2,(j+2)\%4}]$, and $\begin{bmatrix} 01 \\ 01 \\ 03 \\ 02 \end{bmatrix} S[a_{3,(j+3)\%4}]$ four column vectors by look-up table method. For example, in $\begin{bmatrix} 02 \\ 01 \\ 01 \\ 03 \end{bmatrix} S[a_{0,(j+0)\%4}]$, the lower four bits and higher four bits of $a_{0,(j+0)\%4}$ correspond separately to the abscissas and ordinates of the table, so that we get the row coordinates of the two-dimensional array, which is equivalent to determining which grid is in Table 1. The 2, 1, 1, 3 of the column vector correspond separately to the 1, 0, 0, 2 in two-dimensional array column coordinates, which is equivalent to determining which element of the grid is in Table 1. The optimized AES encryption algorithm flow chart is shown in Figure 4.

Likewise, a similar new byte conversion table can be created at the time of decryption to achieve decryption optimization.

4.3. Experimental Results and Analysis. In order to test the encryption speed between classical AES algorithm and optimized AES algorithm in this paper, we use C ++ language to implement the two algorithm encryption processes, respectively, the encryption process in Windows 7 operating system, Core i5-3230M 2.60 GHz CPU, and 8 G memory environment. In each experiment we take 100,000 times the encryption time, and we get a total of 10 sets of data in five experiments. The data obtained in the experiments are shown in Table 2.

Through the test results in Table 2 we can see that the encryption speed of optimized AES algorithm has a great improvement compared to the classic AES algorithm. In terms of memory footprint, this optimized AES encryption algorithm requires $256 \times 3 \times 2 = 1536$ B = 1.5 KB to store two new byte conversion tables (encryption and decryption). The traditional AES algorithm requires $256 \times 2 = 512$ B = 0.5 KB to store two bytes conversion tables, so the optimized AES algorithm does not significantly increase the memory resource occupancy.

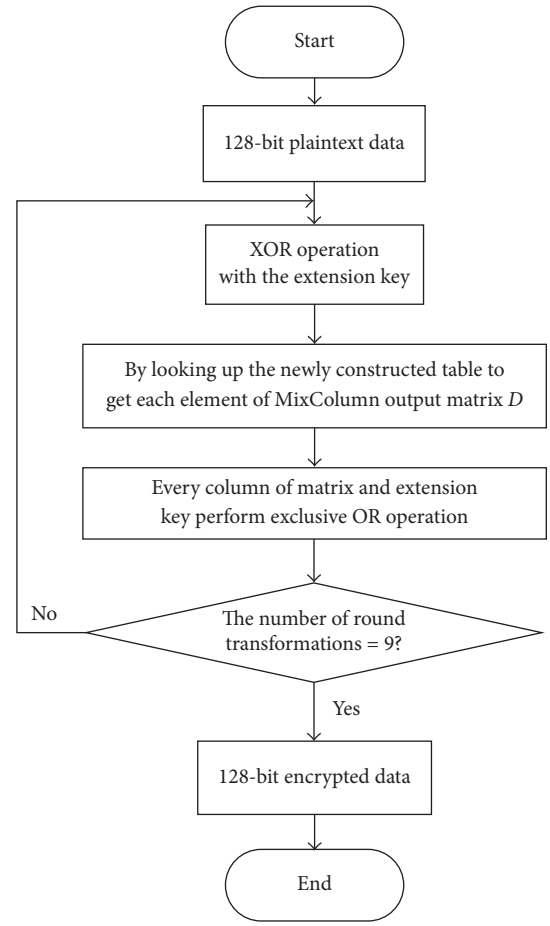


Figure 4: Optimized AES encryption algorithm flow chart.

5. Conclusion

The system implements the verification of the student identity through the fingerprint, which can make the campus life more convenient. The fingerprint data is only stored in the host computer database after encryption transmission, which makes the convenience greatly improved on the basis of ensuring security. Each terminal connected with the host computer constitutes an integral system to achieve the information sharing among each terminal, and the host computer stores the terminal data and manages the students' information efficiently with less time. The encryption method based on the improved AES optimizes the implementation method of algorithm in the process of simplifying the operation step, and the mathematical structure of the original algorithm is not changed, so that the encryption speed increases rapidly under the condition that the security is not reduced, while the memory occupation does not increase significantly, so it is easy to be achieved in the embedded system. Taking an example of AES with 128-bit plaintext length and key

Table 1: The byte conversion table of optimized AES algorithm.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
	63	7c	77	7b	f2	6b	6f	5c	30	01	67	2b	fe	d7	ab	76
0	c6	f8	ee	f6	ff	d6	de	91	60	02	ce	56	e7	b5	4d	ec
	a5	84	99	8d	0d	bd	b1	54	90	03	a9	7d	19	62	e6	9a
	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
1	8f	1f	89	fa	ef	b2	8e	fb	41	b3	5f	45	23	53	e4	9b
	45	9d	40	87	15	eb	c9	0b	ec	67	fd	ea	bf	f7	96	5b
	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
2	75	e1	3d	4c	6c	7e	f5	83	68	51	d1	f9	e2	ab	62	2a
	c2	1c	ae	6a	5a	41	02	4f	5c	f4	34	08	93	73	53	3f
	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
3	08	95	46	9d	30	37	0a	2f	0e	24	1b	df	cd	4e	7f	ea
	0c	52	65	5e	28	a1	0f	b5	09	36	9b	3d	26	69	cd	9f
	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
4	12	1d	58	34	36	dc	b4	5b	a4	76	b7	7d	52	dd	5e	13
	1b	9e	74	2e	2d	b2	ee	fb	f6	d4	61	ce	7b	3e	71	97
	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
5	a6	b9	02	c1	40	e3	79	b6	d4	8d	67	72	94	98	b0	85
	f5	68	02	2c	60	1f	c8	ed	be	46	d9	4b	de	d4	e8	4a
	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
6	bb	c5	4f	ed	86	9a	66	11	8a	e9	04	fe	a0	78	25	4b
	6b	2a	e5	16	c5	d7	55	94	cf	10	06	81	f0	44	ba	e3
	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
7	a2	5d	80	05	3f	21	70	f1	63	77	af	24	20	e5	fd	bf
	f3	fe	c0	8a	ad	bc	48	04	df	c1	75	63	30	1a	0e	6d
	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
8	81	18	26	c3	be	35	88	2e	93	55	fc	7a	c8	ba	32	e6
	4c	14	35	2f	e1	a2	cc	39	57	f2	82	47	ac	e7	2b	95
	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
9	c0	19	9e	a3	44	54	3b	0b	8c	c7	6b	28	a7	bc	16	ad
	a0	98	d1	7f	66	7e	ab	83	ca	29	d3	3c	79	e2	1d	76
	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
a	db	64	74	14	92	0c	48	b8	9f	bd	43	c4	39	31	d3	f2
	3b	56	4e	1e	db	0a	6c	e4	5d	6e	ef	a6	a8	a4	37	8b
	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
b	d5	8b	6e	da	01	b1	9c	49	d8	ac	f3	cf	ca	f4	47	10
	32	43	59	b7	8c	64	d2	e0	b4	fa	07	25	af	8e	e9	18
	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
c	6f	f0	4a	5c	38	57	73	97	cb	a1	e8	3e	96	61	0d	0f
	d5	88	6f	72	24	f1	c7	51	23	7c	9c	21	dd	dc	86	85
	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
d	e0	7c	71	cc	90	06	f7	1c	c2	6a	ae	69	17	99	3a	27
	90	42	c4	aa	d8	05	01	12	a3	5f	f9	d0	91	58	27	b9
	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
e	d9	eb	2b	22	d2	a9	07	33	2d	3c	15	c9	87	aa	50	a5
	38	13	b3	33	bb	70	89	a7	b6	22	92	20	49	ff	78	7a
	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16
f	03	59	09	1a	65	d7	84	d0	42	29	5a	1e	7b	a8	6d	2c
	8f	f8	80	17	da	31	c6	b8	c3	b0	77	11	cb	fc	d6	3a

Table 2: Experimental test results.

The algorithm used	Experiment number					The average time of 100,000 times encryption (s)
	1	2	3	4	5	
Traditional AES algorithm	8.472	8.443	8.382	8.427	8.430	8.43
Optimized AES algorithm	1.550	1.471	1.471	1.469	1.533	1.50

length, this paper proposes an optimization scheme based on actual requirement. The scheme can also be extended to the AES with other data lengths, which is suitable for various situations of data encryption, so it has a wide range of applications and strong practicability.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Special thanks are due to National University Student Innovation Program and Nankai University for the assistance provided to this project.

References

[1] Z. Kai, "Design and implementation of college students' entrepreneurship management system based on B/S structure," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2016, no. 17, pp. 102–113, 2016.

[2] S. R. Bharamagoudar, R. B. Geeta, and S. G. Totad, "Web based student information management system," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 6, 2013.

[3] R. Ahmad and W. Ismail, "Performance comparison of advanced encryption standard-128 algorithms for wimax application with improved power-throughput," *Journal of Engineering Science and Technology*, vol. 11, no. 12, pp. 1678–1694, 2016.

[4] J. Daor, J. Daemen, and V. Rijmen, "Aes proposal: rijndael. Vazirani, efficient and secure pseudo-random number generation," in *Proceedings of the 25th IEEE FOCS*, 1999.

[5] STMicroelectronics, *STM32 Reference Manual*, 10th edition, 2009.

[6] US Department of Commerce and NIST, "Advanced Encryption Standard," in *Proceedings of the National Computer Conference*, pp. 83–87, 2006.

[7] R. Ahmad and W. Ismail, "A survey of high performance cryptography algorithms for WiMAX applications using SDR," *Self-Organization and Green Applications in Cognitive Radio Networks*, pp. 231–246, 2013.

[8] C. Monteiro, Y. Takahashi, and T. Sekine, "Low-power secure S-box circuit using charge-sharing symmetric adiabatic logic for advanced encryption standard hardware design," *IET Circuits, Devices and Systems*, vol. 9, no. 5, pp. 362–369, 2015.

[9] A. M. Youssef and S. E. Tavares, "Affine equivalence in the AES round function," *Discrete Applied Mathematics*, vol. 148, no. 2, pp. 161–170, 2005.

[10] G. Bertoni, L. Breveglieri, I. Koren, P. Maistri, and V. Piuri, "Error analysis and detection procedures for a hardware implementation of the advanced encryption standard," *IEEE Transactions on Computers*, vol. 52, no. 4, pp. 492–505, 2003.

[11] J. Blömer and J. P. Seifert, "Fault Based Cryptanalysis of the Advanced Encryption Standard (AES)," in *Proceedings of the Financial Cryptography, International Conference, FC 2003*, vol. 2742, pp. 162–181, DBLP, Guadeloupe, French West Indies, France, 2003.

[12] J. Daemen and V. Rijmen, *The Design of Rijndael: AES-The Advanced Encryption Standard*, Springer, Berlin, Germany, 2002.

[13] B. Schneier, *Applied Cryptography: Protocols, Algorithms and Source Code in C*, Wiley Publishing, Indianapolis, IN, USA, 2015.

[14] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 1999.

[15] M. McLoone and J. V. McCanny, "Rijndael FPGA implementation utilizing look-up tables," in *Proceedings of the IEEE Workshop on Signal Processing Systems-Design and Implementation-(SIPS) 2001*, pp. 349–360, October 2001.

[16] J. Gong, W. Liu, and H. Zhang, "Multiple lookup table-based aes encryption algorithm implementation," *Physics Procedia*, vol. 25, pp. 842–847, 2012.

[17] J.-F. Wang, S.-W. Chang, and P.-C. Lin, "A novel round function architecture for AES encryption/decryption utilizing look-up table," in *Proceedings of the 37th Annual 2003 International Carnahan Conference on Security Technology*, pp. 132–136, October 2003.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 9080252, 5 pages
https://doi.org/10.1155/2017/9080252

Research Article

Vulnerability Analysis of Interdependent Scale-Free Networks with Complex Coupling

Chunjie Cao,[1,2] Zhiqiang Zhang,[2] Jingzhang Sun,[2] Xianpeng Wang,[1,2] and Mengxing Huang[1,2]

[1]*State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China*
[2]*College of Information Science & Technology, Hainan University, Haikou, China*

Correspondence should be addressed to Zhiqiang Zhang; zhiqiang_zhang@hainu.edu.cn

Received 10 February 2017; Accepted 10 July 2017; Published 14 August 2017

Academic Editor: Jit S. Mandeep

Recent studies have shown that random nodes are vulnerable in interdependent networks with simple coupling. However, relationships in actual networks are interrelated and complex coupling. This paper analyzes the vulnerability of interdependent scale-free networks with complex coupling based on the BA model. The results indicate that these networks have the same vulnerability against the maximum node attack, the load of the maximum node attack, and the random node attack, which explain that the coupling relationship between network nodes is an important factor in network design.

1. Introduction

The small-world network model [1] (Watts and Strogatz, 1998) and the scale-free network model [2] (Barabasi and Albert, 1999) inspired research on complex networks. At present, research of complex networks penetrates into almost all fields such as computer networks, social networks, and biological networks. Based on the small-world network model and scale-free network model, researchers have improved the moment conditions. In the field of computer networks, researchers presented an algorithm for building a new small-world network model by combining the two archetypal small-world networks [3], which can better describe the real instant messaging chat network. In the research of the social networks, researchers established a navigable small-world network based on game-theory model [4] and used this to solve the problem of reciprocity and navigability in social networks. In terms of biological networks, researchers found that the FRC network topology is highly robust and revealed the critical role of the network integrity in the activation of adaptive immune responses [5]. Researchers continue to improve the complex network model and laid the foundation for the research of complex network theory.

Currently, network vulnerability analysis is a very important topic in the theoretical research of complex networks, which mainly studies the problem of vulnerability of complex networks under various attacks (such as deliberate attacks and random failures), especially the vulnerability of an interdependent network system. Recently, domestic and foreign scholars have carried out a series of researches on the impact of the complex network vulnerability and triggered cascading failures. In order to analyze constantly changing impact of component failure propagation on the system, Cheng et al. proposed an optimizing model based on game theory [6]. They found that the failure of components has a great negative impact on the physical network system and used a linear programming method to solve this problem. As for how to effectively model critical infrastructures with interdependent relationships, Rinaldi established the interdependent network system using dynamic simulation and identified the range of cascading security operation [7]. According to the complex relationship between the power system and the rapid processing data network, Cai et al. established an interdependent network model to analyze the complex effects caused by cascading failures [8]. The modern society relies on the complex network of critical infrastructure

systems (such as power networks, telecommunication networks, transportation networks, and water supply systems). Although urban communities rely on each independent infrastructure, recent disasters, such as hurricanes, large-scale power outages, and terrorist attacks, have shown that the most dangerous vulnerabilities were hidden between different interdependencies infrastructures. In a system, the failure of a subsystem is likely to result in the failure of the other, and then the failure of cascading is triggered. On September 28, 2003, a power plant in Italy's grids failed, which led to disabling of the corresponding Internet node. This also resulted in other power stations going offline. Eventually, this effect caused a blackout in the southern peninsula [9]. Regarding this event, Buldyrev et al. published the results of a study (April 2010, Nature) [10] showing that the coupled complex network system facing random faults is also fragile.

Buldyrev et al.'s conclusion reveals the effect of network dependence on network vulnerability. However, as Vespignani said, Buldyrev et al.'s model is simply coupled and ideal. It cannot reflect the complex dependency relationship of actual network systems [11]. This BA model based paper studies the network vulnerability of scale-free networks with complex dependency. Results show that the scale-free networks with complex dependencies are vulnerable against the node of the maximum degree, the node of the maximum load, and the random node attack, and when a single node is attacked, most nodes in the network are compromised. This presents a serious challenge to the complex network systems' design and optimization.

2. Network Model

In recent years, researchers have discovered many complex networks, including the Internet, WWW, and metabolic networks; all these networks' connection degree distribution functions have power laws [12–15]. There is no obvious feature length in these networks' nodes connectivity. These networks are called scale-free networks. Currently, researchers propose many optimized methods for the scale-free networks model by using sparse matrix vector multiplication to construct scale-free networks [16], using the internal weighted average method to calculate the configuration parameters of scale-free networks [17], and using boosting regression algorithm and Bayesian algorithm to construct prior information and establish the scale-free networks based on prior information [18]. Random scale-free networks are modeled with chain fault [19]. We add the strategy of multiscale networks to generate artificial neural networks when constructing scale-free networks, which is used to simulate the network performance of the multiscale [20].

2.1. BA Network Model. Barabasi and Albert proposed a scale-free network model (BA model) to describe two important features of actual networks.

(i) Growth characteristic: the network scale is expanding. As a large number of new articles are published every month, such as on the WWW, there are a lot of new web pages every day.

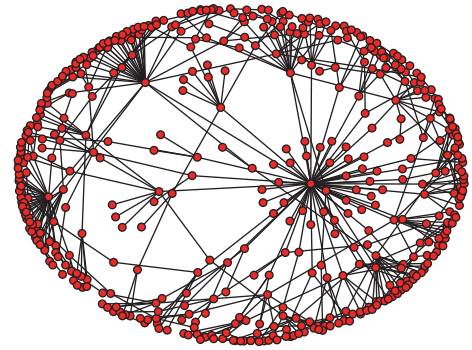


Figure 1: Scale-free network A.

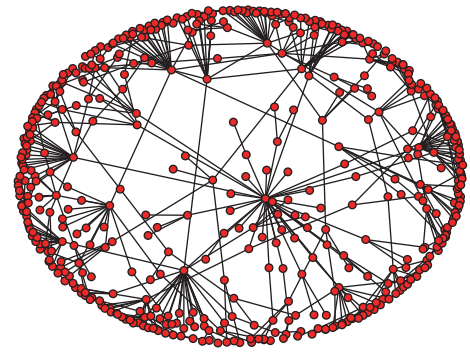


Figure 2: Scale-free network B.

(ii) Priority connection characteristic: the new node tends to connect with nodes with a higher degree. This phenomenon is called "the rich get richer effect" or "Matthew effect."

Based on the growth characteristic and priority connection characteristic, the algorithm for constructing the BA scale-free network model is as follows.

(iii) Growth: the network starts with m_0 nodes, with the addition of a new node and chain to the m existing nodes, where $m \le m_0$.

(iv) Priority connection: define P_i as the probability of a new node connecting with node i. The relationship between P_i, degree k_i of node i, and degree k_j of node j is as follows:

$$P_i = \frac{k_i}{\sum_j k_j}. \tag{1}$$

(v) After t steps, the BA model construction algorithm generates a network with $N = t + m$ nodes and mt edges.

2.2. Complex Coupling Interdependent Scale-Free Network. Based on the BA scale-free network model constructing algorithm, altogether, at least $T = 100$ samples are made, each time generating two scale-free networks (A and B) with 1000 nodes, as shown in Figures 1-2, and randomly generating coupling coefficient matrixes (M_{AB} and M_{BA}) existing in between networks A and B, as shown in Figures 3-4. If $M_{AB}(i, j) = 1$, then nodes A_i and B_j will have a coupling relationship; that is, if node A_i loses, node B_j fails. If $M_{AB}(i, j) = 0$, then nodes A_i and B_j will have no coupling relationship; that is,

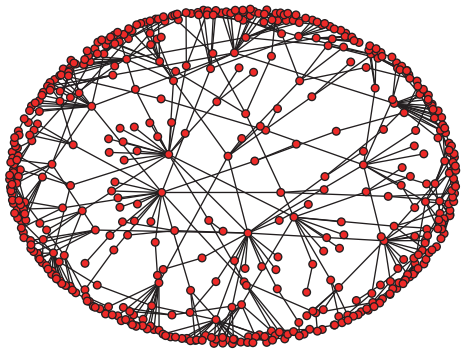


Figure 3: Coupling coefficient matrix M_{AB}.

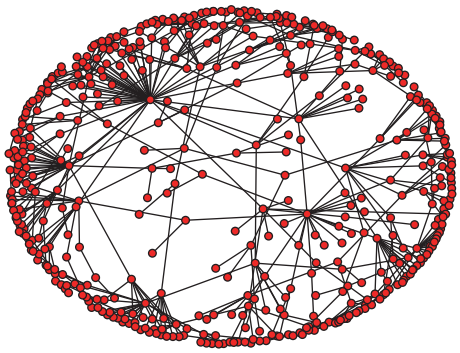


Figure 4: Coupling coefficient matrix M_{BA}.

node A_i loses, while node B_j is validated. Coupling coefficient matrix depicts that the relationship of interdependence network nodes does not have to be a two-way direct coupling relationship. And this method also reflects the reality of actual networks. In the network model proposed by Buldyrev et al., interdependence network nodes have a two-way direct coupling relation. That is, if node A_i loses, node B_j fails, and there is no influence on other nodes. And if node B_j loses, node A_i also fails. However, in the actual network, coupling relationships among different network nodes are complex; node A_i possibly has a coupling relationship with node B_{j1} or may have a coupling relationship with node B_{j2}. And node B_j loss does not necessarily result in node A_i failure (e.g., Internet, WWW, and electric power networks). This is our improvement on Buldyrev et al.'s network model.

3. Vulnerability Analysis

In many real networks, a single or a few failed nodes or failed edges (this fault may happen randomly or may be due to a deliberate attack) can cause other nodes to break down through the coupling relationships between nodes and then cause a chain reaction, eventually causing the breakdown of a lot of nodes, even the entire network. This phenomenon is called cascading failure, sometimes called "avalanche" phenomenon. This phenomenon is a widespread problem in traffic networks, Internet networks, power networks, social and economic networks, and other real complex networks.

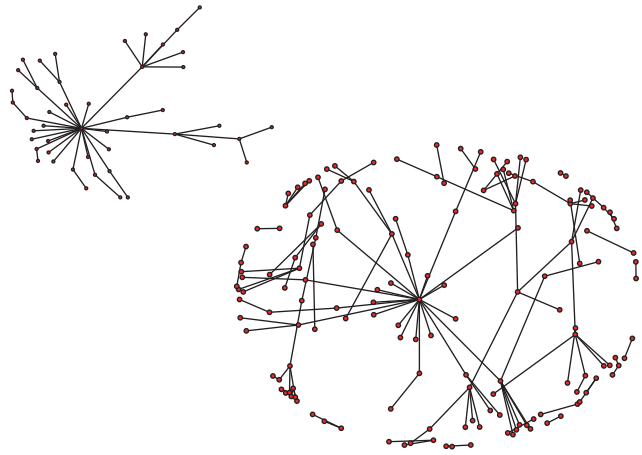


Figure 5: Maximum node attack.

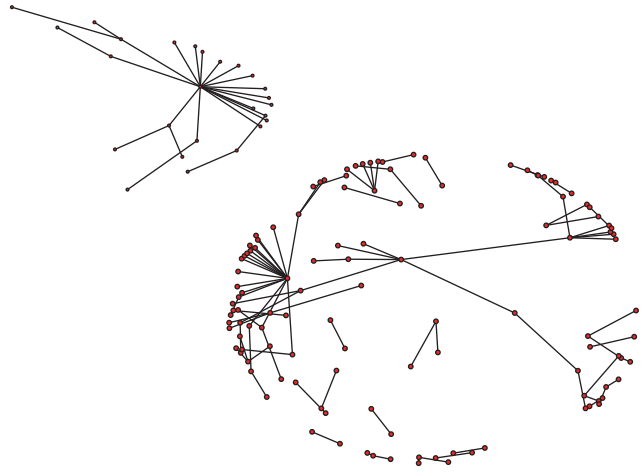


Figure 6: Node of maximum load attack.

We perform an analysis based on the above network model and attack the node of maximum degree, the node of maximum load, and the random node in network A by using the maximum degree node attack, the maximum load node attack, and the random node attack, negate the effects of the node, and then observe the effect of connectedness of network B. The connectivity is analyzed through the quantity of nodes contained in the maximal connected subset (subgraph) in network B; as shown in Figures 5–7, the nodes of maximum degree and random nodes are easy to select and calculate, and selection of the maximum load node uses the betweenness of the node as a selection standard. Betweenness adopts the Floyd algorithm to calculate the shortest path between nodes of network A.

From a large amount of data analysis under three attack modes, we can see that nonscale networks maintained a similar network topology, and the scale of the largest connected subgraph is approximately equal. As shown in Figure 8, the straight line in the graph is the average value of the maximal connected subgraph under three attack methods. In this figure's bottom left corner is the probability distribution maximal connected subgraph under three attack methods.

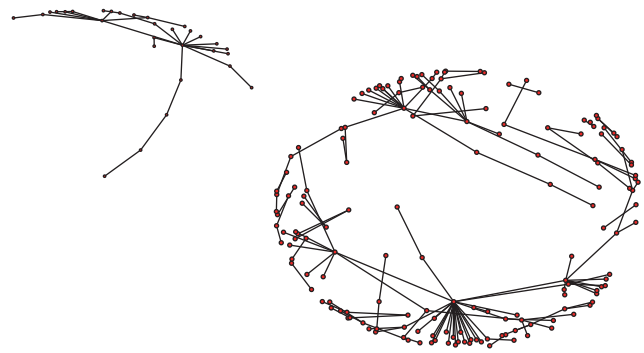


Figure 7: Random node attack.

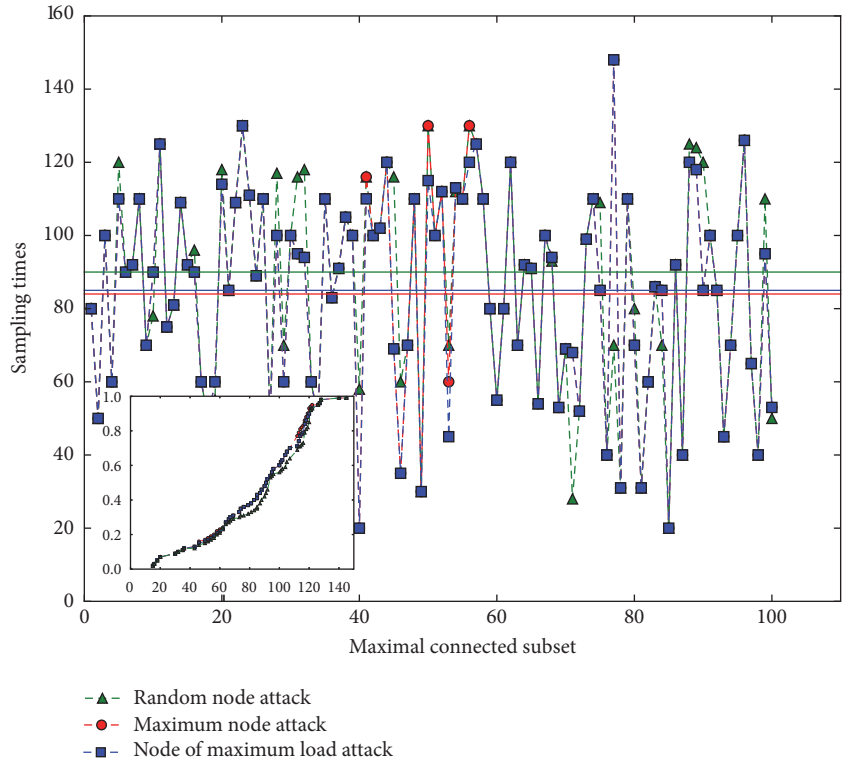


Figure 8: Connected subset size contrast.

Statistical results show that the three attack methods have the same influence on the scale of the largest connected subgraph node of scale-free networks with complex coupling relationships. From the perspective of connectivity, the attack caused the decline of the network performance by 90%.

4. Conclusion

The model proposed by Buldyrev et al. is simply coupled and ideal. It cannot fully present the complicated dependence of the actual network system. This paper studies scale-free networks with complex dependency based on the BA model. Results show that the scale-free networks with complex dependencies are vulnerable against the node of the maximum degree, the node of the maximum load, and the random node attack, and when a single node is attacked, most nodes in the network are compromised. This presents a serious challenge to the complex network systems' design and optimization. Therefore, the coupling between networks must be taken into account in the design of reliable systems.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (no. 61661019), the Major Science and Technology Project of Hainan Province

(no. ZDKJ2016015), the Natural Science Foundation of Hainan Province (no. 20156217), and the Higher Education Reform Key Project of Hainan Province (no. Hnjg2017ZD-1).

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[2] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *American Association for the Advancement of Science. Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[3] J. Guan, M. Tang, G. Huang, W. Zhu, S. Zhou, and G. Ji, "A new small-world network model for instant messaging chat network," in *Proceedings of the 11th Systems of Systems Engineering Conference (SoSE '16)*, June 2016.

[4] Z. Yang and W. Chen, "A game theoretic model for the formation of navigable small-world networks," in *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*, pp. 1329–1339, May 2015.

[5] M. Novkovic, L. Onder, J. Cupovic et al., "Topological Small-World Organization of the Fibroblastic Reticular Cell Network Determines Lymph Node Functionality," *PLoS Biology*, vol. 14, no. 7, Article ID e1002515, 2016.

[6] M. X. Cheng, M. Crow, and Q. Ye, "A game theory approach to vulnerability analysis: Integrating power flows with topological analysis," *International Journal of Electrical Power and Energy Systems*, vol. 82, pp. 29–36, 2016.

[7] S. M. Rinaldi, "Modeling and simulating critical infrastructures and their interdependencies," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, IEEE, January 2004.

[8] Y. Cai, Y. Cao, Y. Li, T. Huang, and B. Zhou, "Cascading failure analysis considering interaction between power grids and communication networks," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 530–538, 2016.

[9] L. A. Nunes Amaral, A. Scala, M. Barthelemy et al., "Classes of behavior of small-world networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 21, Article ID 0001458, pp. 11149–11152, 2000.

[10] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[11] A. Vespignani, "Complex networks: The fragility of interdependency," *Nature*, vol. 464, no. 7291, pp. 984-985, 2010.

[12] C. Petersen, J. G. Simonsen, and C. Lioma, "Power law distributions in information retrieval," *ACM Transactions on Information Systems*, vol. 34, no. 2, article 8, 2016.

[13] I. Dobson, J. Chen, J. S. Thorp, B. A. Carreras, and D. E. Newman, "Examining criticality of blackouts in power system models with cascading events," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS '02)*, January 2002.

[14] A. Medina, I. Matta, and J. Byers, "On the origin of power laws in internet topologies," *Computer Communication Review*, vol. 30, no. 2, pp. 18–28, 2000.

[15] G. J. Fakas, Z. Cai, and N. Mamoulis, "Diverse and proportional size-l object summaries for keyword search," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '15)*, pp. 363–375, June 2015.

[16] W. T. Tang, R. Zhao, M. Lu et al., "Optimizing and auto-tuning scale-free sparse matrix-vector multiplication on Intel Xeon Phi," in *Proceedings of the IEEE/ACM International Symposium on Code Generation and Optimization (CGO '15)*, pp. 136–145, February 2015.

[17] S. F. Muldoon, E. W. Bridgeford, and D. S. Bassett, "Small-world propensity and weighted brain networks," *Scientific Reports*, vol. 6, Article ID 22057, 2016.

[18] B. Yang, J. Xu, B. Liu, and Z. Wu, "Inferring gene regulatory networks with a scale-free property based informative prior," in *Proceedings of the 8th International Conference on BioMedical Engineering and Informatics (BMEI '15)*, pp. 542–547, October 2015.

[19] R.-R. Yin, B. Liu, H.-R. Liu, and Y.-Q. Li, "Research on invulnerability of the random scale-free network against cascading failure," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 458–465, 2015.

[20] A. Gutfraind, I. Safro, and L. A. Meyers, "Multiscale network generation," in *Proceedings of the 18th International Conference on Information Fusion (Fusion '15)*, pp. 158–165, July 2015.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 2970673, 10 pages
https://doi.org/10.1155/2017/2970673

Research Article

The Anonymization Protection Algorithm Based on Fuzzy Clustering for the Ego of Data in the Internet of Things

Mingshan Xie,[1,2,3] Mengxing Huang,[1,2] Yong Bai,[1,2] and Zhuhua Hu[1,2]

[1]*State Key Laboratory of Marine Resource Utilization in South China Sea, Haikou 570228, China*
[2]*College of Information Science & Technology, Hainan University, Haikou 570228, China*
[3]*College of Network, Haikou College of Economics, Haikou 571127, China*

Correspondence should be addressed to Mengxing Huang; huangmx09@163.com

Received 10 February 2017; Accepted 5 March 2017; Published 8 June 2017

Academic Editor: Jit S. Mandeep

In order to enhance the enthusiasm of the data provider in the process of data interaction and improve the adequacy of data interaction, we put forward the concept of the ego of data and then analyzed the characteristics of the ego of data in the Internet of Things (IOT) in this paper. We implement two steps of data clustering for the Internet of things; the first step is the spatial location of adjacent fuzzy clustering, and the second step is the sampling time fuzzy clustering. Equivalent classes can be obtained through the two steps. In this way we can make the data with layout characteristics to be classified into different equivalent classes, so that the specific location information of the data can be obscured, the layout characteristics of tags are eliminated, and ultimately anonymization protection would be achieved. The experimental results show that the proposed algorithm can greatly improve the efficiency of protection of the data in the interaction with others in the incompletely open manner, without reducing the quality of anonymization and enhancing the information loss. The anonymization data set generated by this method has better data availability, and this algorithm can effectively improve the security of data exchange.

1. Introduction

In the information age, many things such as software, websites, and Internet of things are providing data, and there are also various interactive processes of data, for example, collaborative analysis of data, classification of data, integration of heterogeneous data, big data analysis, and trading. But in these interactions of the data many units or agencies are reluctant to open or publish their data collection. In fact, the data can be published after the appropriate treatment and can also play a positive role in our data analysis and mining. Data on the one hand need to be open and, on the other hand, need to be conservative. How to adjust? We need to recognize the ego characteristics of the data. We need to pay attention to the ego characteristics of the data. If we do not pay attention to it, the data sovereignty is not clear. Many disputes will be triggered. The insufficient cooperation will arise in data collaboration.

The ego of the data in the IOT has its own characteristics. It has the sensitivity of time and space and contains a lot of privacy information. In the process of data interaction, in order to achieve the common goal, the data need to share some information, but some sensitive information cannot be exposed. We should fully understand and respect the ego of data.

This paper is structured as follows. In Section 2, we present related work. Section 3 introduces the definition of the ego of data. In Section 4, we introduce the characteristics of the ego of data in the IOT. Section 5 provides the concepts involved in the anonymization protection of incompletely open cooperation for the ego of data in the IOT. Section 6 provides the design and analysis of an anonymization protection algorithm in the data interaction process in an incompletely open manner for the ego of data in the IOT. In Section 7 we have the experiment to validate the algorithm. We conclude our work and lay out future research in Section 8.

2. Related Work

The protection of data, especially the protection of privacy information, has been studied by experts and scholars. The K-anonymous model is proposed by [1], which requires that each data in the data set has at least k-1 data that can not be distinguished from the quasi identifier. According to the literature [2], the k-anonymity model can not resist the attacks of homogeneous attacks and background knowledge. The literature [3] proposed a l-diversity model, which requires that the sensitive values in the cluster have l different values. The (α, k)-anonymization model has been proposed in [4] to realize the diversity of the sensitive value, so as to improve the security of the data. The k-anonymous method based on clustering is proposed in [5].

The method of anonymization is often used in data privacy protection. The literature [6] proposed an anonymous proof protocol based on property certificate. By the trusted ring signature scheme based on property certificate, the authors have achieved the anonymity of computing nodes and prevented the leakage of platform configuration information. Muftic et al. have combined this method with the characteristics of business data to build the business information exchange system with security, privacy, and anonymity which provides innovative features of privacy and full anonymity of users in the paper [7].

Data privacy information protection has accumulated a lot of scholars' work in [8–10]. However, with the explosion of data and the arrival of the era of big data, the interaction between the data is more and more frequent. The coordination between the openness and the protection of data has attracted more and more attention. For the Internet of things, this problem is particularly prominent. The data of the Internet of things has the characteristics of mass and spatiotemporal sensitivity. Literature [11] introduced the data characteristics of the Internet of things. A lot of people's privacy information exists in the Internet of things. Each of its records has the specific location and time attributes that represent the place at which the tag of the Internet of things is located at a certain moment. In addition, the layout of several objects adjacent to each other, which has certain stability, makes it easier for the IOT to leak privacy information. There are few studies on the protection of data for the Internet of things. Reference [12] studied the privacy information protection of the Internet of things and proposed the concept of the data set distribution sequence to optimize the generation of cluster seeds. The authors studied the problem of privacy information protection in the IOT and proposed the concept of data set distribution sequence to optimize the generation of cluster seeds. They clustered the data in parallel so that the equivalent class contains the data of multiple nodes; thus the specific location information of the data can be blurred, and the layout characteristics of tags could be eliminated. A privacy preserving k-anonymous algorithm has been designed for the Internet of things.

The special research on the data protection for Internet of things is not too much to adapt to the requirements of the times. In the paper [13] k-anonymity notion and a method based on bottom-up clustering have been adopted to be used in wireless sensor networks (WSN) as a security framework with two levels of privacy. Samani et al. have modeled the privacy concepts and concerns in the IOT and proposed a privacy protection management framework for CDS at the interaction level in [14]. The application of the framework has been demonstrated by extending Contract Net Protocol (CNP) to support privacy protection for CDS.

The literature [12] is only for the privacy protection of the IOT. However, the research on the issue of privacy protection extension to the coordination problem between data openness and protection of the Internet of things is very low. In the process of data interaction in the Internet of things, on the one hand, we need to protect their privacy; on the other hand, we need to open their cooperation with other data. This paper is aimed at this problem, puts forward the concept of the ego of data, and then analyzes the characteristics of the ego of data in the Internet of things. The two parameters of acceptable node set and acceptable sampling period are given. We will build an incompletely open anonymous protection model for the ego of data which is suitable for the Internet of things environment. In this paper, anonymity protection algorithm based on the fuzzy clustering for the data in the IOT eliminates the layout characteristics of the Internet of things label in order to improve the security of the data and ensure the openness of the data.

3. Models and Definitions

3.1. The Ego of Data Definition. The idea of the ego of data is derived from the observation of data aggregation in the Internet of things: when the data from a variety of Internet of things aggregated to the cloud platform, it was always difficult to deal with the problem of balance between data opening and data protection for some data providers. Some of the data, in fact, did not affect the data provider but was completely shielded. In another case, some of data should be completely removed, because it was useless to the cloud platform for the Internet of things but has been occupying the transmission channel and storage space.

The ego of data characterizes the fact that the data assesses its own value and importance, requires a certain degree of protection of its own and restrictions on opening, and pursues the best balance between data opening and privacy protection in the process of data interaction.

Definition 1 (the ego of data). It is the sense of the value and importance of data. It quantifies the value of data using the method of maximum approximation and evaluates the effect of using the data. Let $E(d)$ represent the numerical expression of the ego of data. There is the following formula:

$$E(d) = \max\left(\|M(D) - M(D \ominus d)\|_q\right), \quad (1)$$

where d represents the concerned data. D represents the data set that contained d. $D \ominus d$ means the data set in which d is removed. M is the quantitative impact calculated by the algorithm; $\|M(D) - M(D \ominus d)\|_q$ denotes the q order norm distance of $M(D)$ and $M(D \ominus d)$.

The connection and difference between the ego of data and several concepts are shown in Table 1.

Table 1: The connection and difference between the ego of data and several concepts.

Concepts	Points of focus
Data sovereignty	Researching on the ownership of data
The value of data	Researching on the usefulness of data
The ego of data	Focusing on the study of the degree of data coordination, in the process of data interaction, including the estimation of the value of data, which is more extensive than the concept of data privacy
Data privacy	Focusing on data confidentiality and precautions

3.2. Characteristics of the Ego of Data

Interactivity. The ego of data reflects the ability of data to absorb other data or open itself to accomplish a specific work.

Various Forms. There are text, pictures, and other forms of data. In addition, there are all kinds of data structures.

Complex Relation of Subject. Each data has its own actors. The value and importance of data are determined by the behavior subject of the data. The complex relationship of the behavior subject of the data makes the data have the ego. The three main behavior subjects are as follows. The first is the owner of the data, namely, data collection platform owners. The second is the data producers who are concerned with and studied by the owner of the data collection platform. The third is the data user who can tap the value of the data. There are times when there is three-behavior subject unity, sometimes two-behavior subject unity, and sometimes the separation of three-behavior subject unity. Now, medical data can be an example to illustrate their relationship. The producers of the data are the patients, the users of the data are doctors, and the owner of the data is the data collection platform builder who is always a medical institution. In addition, farm data also can be an instance, the producer of the data are the crop, the user of the data and the owner of the data are farm managers. It is a case of two-behavior subject unity.

Nonintuitive Value. The value of data is not intuitive, which can be reflected by the data mining and data-processing technology.

Value Variability. The value of data is difficult to measure and can only be approximated to the real value. The number that can reflect the value of data is always varied with time and scene.

Different Domain. The value of data is varied with the users from different domains. The same data is very important for some of the actors, while it is not useful for any other actors.

Various Sensitive Information. Due to the complexity of the relationship of the three-behavior subject, the sensitivity of the data is not the same and the sensitivity of the record is not the same.

Uniqueness of Data Set Distribution. Different data sets have different distribution patterns. There are two forms of attribute distribution and record distribution.

3.3. Data Interactive Mode Based on the Ego of Data.

The research content of the ego of data is mainly focused on the interaction mechanism of the data and the adjustment of the algorithm complexity. The research methods of the ego of data are mainly a variety of quantitative analysis methods based on artificial intelligence technology and computational theory.

Different modes of cooperation are adopted by the ego of data, according to the evaluation of its own value, in cooperation with other data. There are two types of data cooperation mode.

The first type is to publish its own data. The ego of data provides its own data to other data users to complete a task. The second type is to absorb other data. Some data, because they are not complete, interact with other data in order to supplement and improve and modify its own data. Each type has the following data open modes.

Totally Enclosed. The data is highly encrypted, so that it is unable to extract useful information in the process of the data mining. The most extreme case of this mode was that the data was not provided at all; that is, the data was deleted completely.

Incompletely Open. Only part of the information is published, in order to protect the sensitive information of the data. At the same time, the degree of influence of the protection method on the cooperation result is controlled in a certain range. The sensitive information can not correspond to the specific behavior subject.

Completely Open. Useful data information is readily available. The most extreme case of completely open mode is to provide raw data.

Definition 2 (the degree of opening). Set function $f : d \to I$. The data d is inputted; the useful information I can be output. Then

$$\mathrm{OP}(d) = \frac{f(d)}{d}. \tag{2}$$

$\mathrm{OP}(d)$ is known as the degree of openness of data d.

Both the totally enclosed mode and the completely open mode are the special cases of the incompletely open mode. When $\mathrm{OP}(d) = 1$, the data d is completely open. $\mathrm{OP}(d) \simeq 0$ indicates that the data d is close to the totally enclosed mode. $0 < \mathrm{OP}(d) < 1$ indicates that the data d take the incompletely open mode.

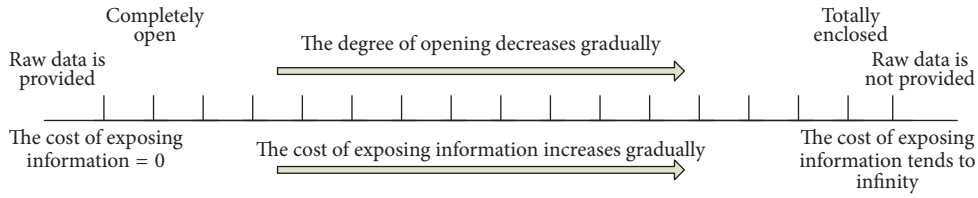


Figure 1: The relationship between OP(d) and COST(d).

The data that is incompletely open is evaluated. The lower the opening degree OP(d) is, the higher the requirement of the protection function and the cost would be. The relationship between OP(d) and COST(d) is shown in Figure 1.

Definition 3 (sensitive information). Some attributes or records that cannot be opened are always important. The range and sensitivity of sensitive information are determined by the behavior subject of the ego of data. Different sensitivity levels are set according to the degree of openness of sensitive information.

When the sensitive information is included in the data set, the interaction mode of the data set can be the mode of incompletely open.

The cost of data protection is paid in order to protect the sensitive information. The cost of exposing the sensitive information also has to be paid. "Unable to decrypt" indicates that the cost of decryption tends to infinity.

Definition 4 (the cost of exposing information). The cost was made up by the computing time and computing complexity of the data mining algorithm used to obtain useful information from the processed data. That is,

$$\text{COST}(d) = \begin{cases} 0, & \text{completely open} \\ \max \text{cost}(d) \times (1 - \text{OP}(d)), & \text{incompletely open} \\ \max \text{cost}(d), & \text{totally enclosed.} \end{cases} \quad (3)$$

COST(d) becomes the cost of exposing information exposure costs for the ego of data, where $\max \text{cost}(d)$ represents the maximum cost of exposing information. There is a decreasing function relationship between the cost of exposing information and the degree of opening. The relationship between OP(d) and COST(d) is shown in Figure 1.

4. The Characteristics of the Ego of Data in the Internet of Things

The ego of data needs to be more concerned with the Internet of things. For example, the producer of the data, which is the condition monitoring value like the patient's pulse, blood pressure, blood sugar, and so on, is the patients for the ego of data in the wireless body area network. Many patients and their families do not want these monitoring data to be leaked out. In the process of data interaction, the interaction of sensitive information is always needed. The incompletely open data cooperation mode is taken by the ego of data in the Internet of things in the process of data exchange.

The space attribute and time attribute of the Internet of things are taken into account, when the ego of data in the Internet of things is studied. Different IOT service system and application system have different requirements on the space attribute and time attribute.

The data collection of the Internet of things is completed by the acquisition nodes deployed in space. In order to study and use these nodes conveniently, they are usually numbered. The spatial information of nodes usually corresponds to the number of nodes. The data of the Internet of things are the data sampled at a certain time, so the data of the Internet of things corresponds to a node and a sampling time. These data collected by the adjacent nodes deployed in the same spatial range have intimate relationships with respect to spatial attributes. The sampling time of each node of the Internet of things also has correlation. Data sampled in the same time period have similar characteristics.

5. The Concepts Involved in the Anonymization Protection of Incompletely Open Cooperation for the Ego of Data in the Internet of Things

In this paper, anonymization protection algorithm has been adopted to complete the incompletely open cooperation for the ego of data. In some data interaction processes, the data processed by anonymization is needed. The data, which is replaced by the equivalence class in the anonymization protection algorithm, is more likely to be found in its own laws, is usually used to reflect the development of the object, and is conducive to data mining for partners. At the same time, it can also realize the sensitive information protection.

The anonymization model based on the fuzzy clustering method is adopted in the incompletely open cooperation for the ego of data in this paper. Considering the characteristics of the ego of data in the Internet of things, the two parameters of tolerable nodes set and tolerable sampling period are introduced, in order to ensure that anonymization results are available. Data in the Internet of things can be anonymous effectively, so as to achieve the incompletely open cooperation protection for the ego of data.

In the Internet of things, the set of data sampling nodes can be divided into many subsets according to the spatial deployment, and the sampling time of the node also can be included in a series of short continuous time fragments. The availability of data will not be affected by the case that the space position and sampling time attributes in each record from the nodes could be replaced separately by the abstract spatial attributes of the set and intermediate value of sampling period, as long as the placement of the sampling point set is covered to a sufficiently small space and the difference of sampling time is small enough.

Definition 5 (NODESP (x, y, z, λ)). It means the spatial attributes of equivalent node. Consider an equivalence class T $(d_1, \ldots, d_n)$, where d_i $(1 \leqslant i \leqslant n)$ means the ith data.

s_i describes the spatial attributes of nodes. NODESP (x, y, z, λ) describes the subset of nodes formed by spatial adjacency clustering in the Internet of things. The subset whose threshold of adjacency is λ is the minimal subset of nodes which meets $\forall s_i \asymp \text{NODESP}(x, y, z, \lambda)$, where $x = \sum_{i=1}^{i=n} x_i/n$, $y = \sum_{i=1}^{i=n} y_i/n$, and $z = \sum_{i=1}^{i=n} z_i/n$. x_i, y_i, and z_i represent, respectively, the abscissa, ordinate, and height of s_i in the three-dimensional space.

Definition 6 (SAMPTIME(t)). It denotes the equivalent sampling time of the record. Let an equivalence class $T(d_1, \ldots, d_n)$, where d_i $(1 \leqslant i \leqslant n)$ is the ith data. t_i denotes the sampling time attribute of d_i. SAMPTIME(t) is the time which meets $\forall t_i \asymp \text{SAMPTIME}(t)$, where $t = \text{mid}(t_1, \ldots, t_n)$.

Definition 7 ($C(\lambda_s)$). It means the subset of nodes determined by clustering threshold λ_s.

Definition 8 ($T(t_s)$). It represents the sampling period in which the time interval is t_s.

Definition 9 ($\text{TNC}(\lambda_{\text{th}})$). It is the acceptable subset of nodes in which clustering threshold is λ_{th}.

It will not affect the availability of data where the spatial attributes of $T(d_1, \ldots, d_n)$ have been replaced by the spatial attributes of NODESP (x, y, z, λ).

Definition 10 ($\text{TST}(t_{\text{th}})$). It means acceptable sampling period. It means the period in which the duration is t_{th}. It will not affect the availability of data where the time attributes of $T(d_1, \ldots, d_n)$ have been replaced by SAMPTIME(t).

Definition 11 (a clustering anonymization algorithm for realizing the ego of data protection in the data interaction process in an incompletely open manner in the Internet of things). Given the group of the ego of data in the Internet of things $D(d_1, \ldots, d_n)$, $\text{TNC}(\lambda_{\text{th}})$, and $\text{TST}(t_{\text{th}})$, any data d_i $(d_i \in D)$ has at least another k-1 records $d_1, \ldots, d_j$ $(j \geqslant k - 1)$ with the same identifier as d_i whose equivalent node has these attributes that the equivalent space attribute is $\text{NODESP}(x, y, z, \lambda)$ and the equivalent sampling time is SAMPTIME (t). If $\lambda_s \leqslant \lambda_{\text{th}}$, $t_s \leqslant t_{\text{th}}$, the set $D(d_1, \ldots, d_n)$ meets the Internet of things' ego of data anonymization protection.

Table 2: The spatial location information of the sampling node.

Node index	x-axis	y-axis	z-axis
p_1	x_1	y_1	z_1
p_2	x_2	y_2	z_2
⋮	⋮	⋮	⋮
p_{Nd}	x_{Nd}	y_{Nd}	z_{Nd}

6. The Design and Analysis of an Anonymization Protection Algorithm in the Data Interaction Process in an Incompletely Open Manner for the Ego of Data in the Internet of Things

Because the ego of data has temporal and spatial characteristics in the Internet of things, the data which has the different sampling time and different node indexes is assigned to the same equivalence class by means of fuzzy clustering of nodes according to the layout of nodes and fuzzy clustering of sampling time. It is required that each equivalence class has to contain 2 or more nodes in the process of fuzzy clustering of nodes and sampling time. The data of multiple nodes is contained in an equivalence class. After anonymization generalization, the data in the equivalence class has the same attributes of node position sampling time, so as to hide the correspondence between the data records and the nodes. The corresponding relationship between the data record and the sampling time is hidden. The location information and time information of the data are blurred.

6.1. The Analysis of Node Fuzzy Clustering. Let Nd represent the number of nodes in the networking environment; p represents the node. Those nodes that are close to each other are divided into a cluster.

The First Step (node coordinates standardization). The spatial location information of the sampling node is denoted by Spactab. It is shown in Table 2.

The Euclidean distance formula is used to calculate the distance between nodes:

$$\text{dist}_{p_i,p_j} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2 + \left(z_i - z_j\right)^2}. \quad (4)$$

Mutual distance matrix is

$$\begin{bmatrix} \text{dist}_{p_1,p_1} & \text{dist}_{p_1,p_2} & \cdots & \text{dist}_{p_1,p_{\text{Nd}}} \\ \text{dist}_{p_2,p_2} & \text{dist}_{p_2,p_2} & \cdots & \text{dist}_{p_2,p_{\text{Nd}}} \\ \cdots & \cdots & \cdots & \cdots \\ \text{dist}_{p_{\text{Nd}},p_1} & \text{dist}_{p_{\text{Nd}},p_2} & \cdots & \text{dist}_{p_{\text{Nd}},p_{\text{Nd}}} \end{bmatrix}. \quad (5)$$

The Second Step (calibration (the establishment of fuzzy adjacent matrix)). The distance matrix is standardized and the maximum range formula is used:

$$\gamma_{p_i,p_j} = \frac{\text{dist}_{p_i,p_j} - \text{dist}_{\min}}{\text{dist}_{\max} - \text{dist}_{\min}}. \quad (6)$$

Input: Spactab and R;
Output: Subndgrp; % clustered node set
BEGIN
(1) Subndgrp = ∅
(2) FOR each r in R
(3) $\lambda_s = r$; % Value from R is assigned to variable r
(4) IF $\lambda_s \le \lambda_{\text{th}}$ then
(5) Subndgrp = Subndgrp $\bigcup_{m\le \text{NUMGRP}(\lambda_s)} \{[\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m \mid m \in N\}$
% Subndgrp is a node set which spatial adjacent degree is λ_s. $\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i$ represents the subset of nodes, where, NUMNd(λ_s, m) denotes the node number of the subset of spatial adjacent nodes determined by λ_s. The number of nodes in each cluster subset is equal or unequal. $[\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m$ is shown as the mth subset determined by the intercept λ_s. $\bigcup_{m\le \text{NUMGRP}(\lambda_s)} [\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m$ denotes the union of all sub sets. NUMGRP(λ_s) represents the subset number determined by λ_s. %
(6) ENDIF
(7) ENDFOR
END

Algorithm 1: Node clustering.

Table 3: The original sampling data set.

DATA index	Sampling time	Node index	Sensitive information
1	t_1	p_i	data_1
2	t_2	p_j	data_2
⋮	⋮	⋮	⋮
Ndata	$t_{N\text{data}}$	p_k	$\text{data}_{N\text{data}}$

Fuzzy adjacent matrix R':

$$\begin{bmatrix} \gamma_{p_1,p_1} & \gamma_{p_1,p_2} & \cdots & \gamma_{p_1,p_{\text{Nd}}} \\ \gamma_{p_2,p_1} & \gamma_{p_2,p_2} & \cdots & \gamma_{p_2,p_{\text{Nd}}} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{p_{\text{Nd}},p_1} & \gamma_{p_{\text{Nd}},p_2} & \cdots & \gamma_{p_{\text{Nd}},p_{\text{Nd}}} \end{bmatrix}. \tag{7}$$

The Third Step. See Algorithm 1.

6.2. The Analysis of Fuzzy Clustering of Sampling Time. The original sampling data set is denoted by Sampdata. We can assume that the number of records in the set is denoted by Ndata. It is shown in Table 3.

Let the ego of data of the mth nodes subset $[\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m$ group be D_m. We can assume that there are n records of data in D_m.

$$\begin{pmatrix} \text{The index of } D_m & d_1 & d_2 & \cdots & d_n \\ \text{Sampling time} & t_i & t_j & \cdots & t_k \\ \text{Node index} & P_u & P_v & \cdots & P_w \end{pmatrix}, \tag{8}$$

where $\{P_u, P_v, \ldots, P_w\} = [\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m \subseteq$ Subndgrp, $D_m \subseteq$ Sampdata.

They are clustered according to the sampling time difference of the nodes. The nodes with small sampling time difference are divided into a subset, that is, to form an equivalence class.

The First Step. Using the following formula to calculate the sampling time difference between each other:

$$\begin{aligned} &\text{epoch}_{d_i,d_j} \\ &\quad = \left| d_i \cdot \text{Sampling time} - d_j \cdot \text{Sampling time} \right|, \end{aligned} \tag{9}$$

where $d_i \cdot$ Sampling time denotes the sampling time of the d_i record.

Mutual time difference matrix R' is

$$\begin{bmatrix} \text{epoch}_{d_1,d_1} & \text{epoch}_{d_1,d_2} & \cdots & \text{epoch}_{d_1,d_n} \\ \text{epoch}_{d_2,d_1} & \text{epoch}_{d_2,d_2} & \cdots & \text{epoch}_{d_2,d_n} \\ \cdots & \cdots & \cdots & \cdots \\ \text{epoch}_{d_n,d_1} & \text{epoch}_{d_n,d_2} & \cdots & \text{epoch}_{d_n,d_n} \end{bmatrix}. \tag{10}$$

The Second Step. See Algorithm 2.

6.3. The Analysis of the Ego of Data Anonymization Protection Algorithm in the Internet of Things. See Algorithm 3.

In the Internet of things, the physical location of the nodes in the subset whose intercept is λ_s is fixed, and the number of nodes in each cluster node is equal or unequal. Arithmetic average of the number is s. If the data is evenly distributed, the location information leakage (the probability

Input: R', TST(t_{th}) and $[\bigcup_{i\le \text{NUMNd}(\lambda_s,m)} P_i]_m$;
Output: Substgrp; % the set of equivalent class
BEGIN
(1) Substgrp = ∅;
(2) FOR each r in R'
(3) $t_s = r$;
(4) if $t_s < t_{\text{th}}$ then
(5) Substgrp = $\bigcup_{n\le \text{NUMGRP}(t_s) \text{ and } \max(\text{NUMd}(t_s))}\{[\bigcup_{j\le \text{NUMd}(t_s,n) \text{ and } \text{NUMNd}(n)\ge 2} d_j]_n \mid n \in N\}$
% The set of equivalence classes with sampling interval of t_s is obtained. $\bigcup_{j\le \text{NUMd}(t_s,n) \text{ and } \text{NUMNd}(n)\ge 2} d_j$ denotes the data subset whose intercept is t_s, namely, an equivalent class. The function of NUMd(t_s, n) is that the number of the nth equivalent class would be counted out. $[\bigcup_{j\le \text{NUMd}(t_s,n) \text{ and } \text{NUMNd}(n)\ge 2} d_j]_n$ denotes the nth subset, namely, the nth equivalent class. It is asked that the data for each equivalence class contains at least 2 nodes. The function of NUMNd(n) is that the number of nth equivalent class. In this paper, NUMNd(n) ≥ 2. $\bigcup_{n\le \text{NUMGRP}(t_s) \text{ and } \max(\text{NUMd}(t_s))}[\bigcup_{j\le \text{NUMd}(t_s,n) \text{ and } \text{NUMNd}(n)\ge 2} d_j]_n$ represents that the union of all subsets, namely the union of all equivalence classes. The function NUMGRP(t_s) is that total number of equivalence classes whose intercept is t_s is calculated. The maximum number of records would be covered in an equivalence class. We should choose the set of equivalence classes which can cover the maximum number of records.
(6) ENDIF
(7) ENDFOR
END

Algorithm 2: Sampling time clustering algorithm.

of the specific location of the data being guessed) is $1/s$. Similarly, if the sampling time of the data in the equivalence class is distributed in the period, the probability of causing the leakage of time information is $1/t_s$. The larger the t_s, the smaller the probability of leakage of the time information, but also the longer the sampling time of the equivalence class and the greater the loss of information caused by the anonymization result. However, when t_s is over, it will also affect the security of the data. People need to find a balance between security and information loss.

In this paper, the Internet of things' ego of data anonymization model sets acceptable subset of nodes and acceptable sampling period. Since the number of nodes in the fuzzy clustering is required to be at least 2, each equivalence class contains multiple nodes. The time fuzzy clustering algorithm ensures that the sampling time span of the equivalence class is in the appropriate range, so as to achieve the effect of sensitive information protection under the premise of ensuring the availability of data. The anonymization protection algorithm based on fuzzy clustering for the ego of data in the Internet of things in this paper ensures that each equivalence class contains the data of multiple computing nodes by means of fuzzy clustering method. The records in the equivalence class can not correspond to the sampling nodes, thus reducing the risk of leakage of location information.

7. Experiments

The data set used in this paper is the measured data provided by Intel Berkeley Laboratory [12]. The data is collected from 54 sensors deployed in the Intel Berkeley Research lab between February 28th and April 5th, 2004. Mica2Dot sensors with weather boards collected timestamped topology information, along with humidity, temperature, light, and voltage values once every 31 seconds. Data was collected using the Tiny DB in-network query processing system, built on the TinyOS platform.

15 nodes are randomly selected from the 54 nodes to perform fuzzy clustering, and 499 records of the 15 nodes are randomly selected. Due to the data of the Internet of things being cyclical, focusing on the transformation of the day, we can delete the date column in order to facilitate research. Because in the Internet of things time sampling is always in a continuous period and sampling interval is very short, the sampling time difference of this experiment is accurate to hours.

Set a fixed threshold of the acceptable subset of nodes for clustering. In this paper we randomly set $\lambda_{\text{th}} = 0.3$. The set of clusters of nodes is Subndgrp = {{1, 3, 6}, {8, 9, 12}, {15, 17}, {18, 21}, {37, 39, 44}, {47, 53}}. Fuzzy clustering is used to realize the clustering of all nodes. Figure 2 depicts the respective total number of records processed by anonymization at different acceptable sampling periods t_{th}. It is shown in Figure 2 that t_{th} is very small, resulting in the fact that a lot of data do not meet the conditions of clustering. With the increase of t_{th}, more and more records satisfy the clustering conditions. When $t_{\text{th}} = 18$, almost all the records meet the conditions of clustering.

The value of the spatial attributes of the node in an equivalent class is the average value of the spatial attribute value of the nodes in the acceptable subset, and the sampling time of nodes in an equivalent class is within the sampling period of the subset. The spatial and temporal attributes of the data after anonymous generalization are not greatly affected, and this algorithm does not affect the availability of the data. In the data exchange process, data

Input: λ_{th}, t_{th}, Spactab, Sampdata;
Output: processed data set;
Begin
(1) Do Algorithm 1; % All nodes are implemented by fuzzy clustering according to the spatial location and get the node clustering set Subndgrp.
(2) FOR each $[\bigcup_{i\leq \mathrm{NUMd}(\lambda_s)} P_i]_m$ in Subndgrp % Get a subset of nodes from Subndgrp orderly.
(3) Do Algorithm 2; % The sampling time fuzzy clustering is performed on the data records of the nodes in the subset, and the equivalence class set is obtained. Anonymization processing for each equivalence class.
(4) FOR each $[\bigcup_{j\leq \mathrm{NUMd}(t_s,n) \text{ and } \mathrm{NUMNd}(n)\geq 2} d_j]_n$ in Substgrp % Get a subset of nodes from Substgrp orderly.
(5) Psum = 0; % Psum denotes the total number of nodes in an equivalence class.
(6) FOR each p_j in $[\bigcup_{i\leq \mathrm{NUMd}(\lambda_s)} P_i]_m$
(7) $x$sum = $x$sum + $p_j \cdot x$; % $p_j \cdot x$ represents the x-axis of p_j node in the Spactab.
(8) $y$sum = $y$sum + $p_j \cdot y$; % $p_j \cdot y$ represents the y-axis of p_j node in the Spactab.
(9) $z$sum = $x$sum + $p_j \cdot z$; % $p_j \cdot z$ represents the z-axis of p_j node in the Spactab.
(10) ENDFOR
(11) $x$temp = $x$sum/$p$sum; $y$temp = $y$sum/$p$sum; $z$temp = $z$sum/$p$sum;
(12) FOR each p_j in Spactab % Replace the spatial location information of the node number in the equivalent class data record with the spatial attribute NODESP (x, y, Z, lambda) of the equivalent node of the equivalence class.
(13) IF $p_j \in [\bigcup_{i\leq \mathrm{NUMd}(\lambda_s)} P_i]_m$ then
(14) $p_j \cdot x = x$temp; $y = y$temp; $p_z = z$temp;
(15) ENDIF
(16) ENDFOR
(17) t = mid (d_1·Sampling time,. . ., d_n·Sampling time) % Equivalent sampling calculated with the intermediate value of the sampling time of all data records of time is equivalence class.
(18) For each d_j in $[\bigcup_{j\leq \mathrm{NUMd}(t_s,n) \text{ and } \mathrm{NUMNd}(n)\geq 2} d_j]_n$
(19) $d_j \cdot$ Sampling time = t% Replace the sampling time attribute of the record with the sampling time in the equivalent class.
(20) ENEFOR
(21) ENDFOR
(22) ENDFOR
(23) Count the rest of the data to delete.
% A small number of records cannot be clustered, because the special distance or the duration of sampling time between those records and the most number of records. If the few records are putted into the equivalent class, the nodes' sampling duration of the equivalent class is greater than t_{th}, or the spatial contiguity of nodes in the equivalent class goes beyond λ_{th}. %
END

Algorithm 3: Anonymization protection of the ego of data based on fuzzy clustering in the Internet of things.

after anonymous generalization can be open and can be used by partners. The information of the data after anonymous generalization is approximately equal to the useful information. The open degree OP(D) of the whole data set is approximated as the ratio of the data after anonymous generalization to the whole data, where D represents the data set used in this experiment. At the same time, OP(D) also reflects the anonymity efficiency of anonymous protection algorithms. Figure 3 describes the OP(D) of the proposed algorithm in this paper under different acceptable sampling periods.

It is shown in Figure 3 that OP(D) is increased with the increase of t_{th}. OP(D) value is high, that is, a higher rate of data anonymization.

The node loss rate of clustering K depends on the number of lost nodes in the process of clustering. K = LNd/ND, where LNd represents the number of lost nodes in the clustering; Nd is the total number of nodes in the IOT. All records of lost nodes will not be classified into equivalent classes. K hasa direct impact on the efficiency of anonymous protection and the openness of data and is an important indicator of the data interaction of Internet of things.

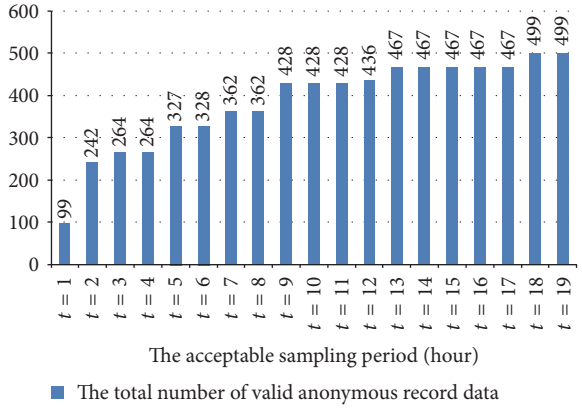


Figure 2: The respective total number of anonymous records at different acceptable sampling periods.

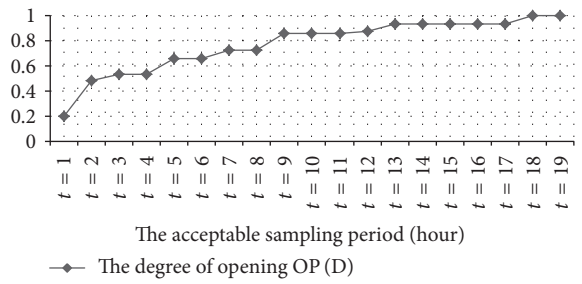


Figure 3: The OP(D) of the proposed algorithm under different acceptable sampling periods.

Figure 4 represents the two node loss rates of clustering K based on the two algorithms at different acceptable mutual distances between sampling nodes. The first is the anonymization protection algorithm based on fuzzy clustering proposed in this paper; the second is the anonymization protection algorithm based on seed clustering [12]. In this experiment, the value λ_{th} is converted to the acceptable mutual distance between nodes for convenient comparison. The second algorithm selects the node numbers for {6, 17, 37, and 47} as seed nodes, according to the law of the maximum number of nodes and the scattered position.

It is shown in Figure 4 that the two loss rates of node clustering reach 0, when the acceptable mutual distance between nodes increases to 20 meters. As the distance becomes larger, the loss rate decreases, down to 0. Figure 4 represents the fact that the node loss rate of clustering based on the fuzzy clustering algorithm is smaller than that based on the seed clustering algorithm. The loss rate is reduced to 0 when the distance is about 15 meters in our algorithm, while the distance is reduced to about 20 meters, which makes the loss rate reduce to about 0 in the seed clustering algorithm. The algorithm proposed in this paper makes the loss rate reach 0 faster than the seed clustering algorithm.

8. Conclusions

In the process of data exchange, data needs to be open on the one hand, and, on the other hand, it needs to be conserved. In order to solve this contradiction, this paper puts forward the concept of the ego of data. After analyzing the characteristics of the ego of data in IOT, we use anonymization protection model to make the data get a certain degree of security in the case of a certain degree of openness. In this paper, we introduce the acceptable subset of nodes and the acceptable sampling period, based on the analysis of the time attribute and spatial properties of the ego of data in the IOT. We can obtain the equivalent class by the fuzzy clustering method and guarantee that the equivalent class contains many nodes. Finally, an anonymization protection algorithm which is suitable for the data exchange in incompletely open manner for the ego of data in the IOT is designed in this paper. The anonymous data set generated by the algorithm can effectively protect the sensitive information of the Internet of things under the premise of ensuring the availability of the data. As a future work, we will continue to extend the concept extension of the ego of data. We are planning to solve the problem of data protection in the IOT by integrating differential privacy protection.

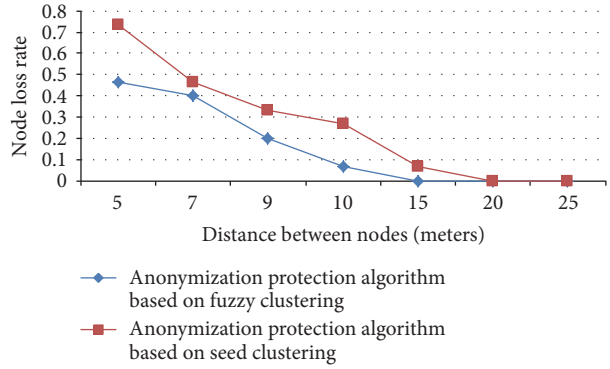


Figure 4: The two loss rates of node clustering.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the Project of Natural Science Foundation of Hainan Province in China (Grant nos. 20166232 and 617033), the National Natural Science Foundation of China (Grant nos. 61462022 and 61561017), and Open Project of State Key Laboratory of Marine Resource Utilization in South China Sea (Grant no. 2016013B).

References

[1] L. Sweeney, "K-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[2] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," in

Proceedings of the 22nd International Conference on Data Engineering (ICDE '06), pp. 24–36, Atlanta, Ga, USA, April 2006.

[3] A. Machanavajjhala, D. Kifer, J. Gehrke et al., "L-diversity: privacy beyond *k*-anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, article 3, 52 pages, 2007.

[4] R. Wong, J. Li, A. Fu, and K. Wang, "(α, k)-anonymous data publishing," *Journal of Intelligent Information Systems*, vol. 33, no. 2, pp. 209–234, 2009.

[5] G. Aggarwal, T. Feder, K. Kenthapadi et al., "Achieving anonymity via clustering," in *Proceeding of the 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '06)*, pp. 153–162, New York-NY-USA, June 2006.

[6] Ning Z. H., W. Jiang, J. Zhan et al., "Property-based anonymous attestation in trusted cloud computing," *Journal of Electrical and Computer Engineering*, vol. 2014, no. 17, pp. 1–7, 2014.

[7] S. Muftic, N. B. Abdullah, and I. Kounelis, "Business information exchange system with security, privacy, and anonymity," *Journal of Electrical and Computer Engineering*, vol. 2016, no. 1, pp. 1–10, 2016.

[8] S. Landau, "Control use of data to protect privacy," *Science*, vol. 347, no. 6221, pp. 504–506, 2015.

[9] J. C. Doshi and B. Trivedi, "Hybrid intelligent access control framework to protect data privacy and theft," in *Proceeding of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '15)*, pp. 1766–1770, 2015.

[10] Y.-S. Jeong and S.-S. Shin, "An efficient authentication scheme to protect user privacy in seamless big data services," *Wireless Personal Communications*, vol. 86, no. 1, pp. 7–19, 2016.

[11] Y. Ge and F. Li, "Data management in the internet of things," *Chinese society of computer communication*, vol. 6, no. 4, pp. 30–34, 2010.

[12] H. Wei and C. Zhong, "Information technology of Internet of things k-anonymous algorithm based on parallel clustering," *Information Technology*, vol. 12, pp. 6–10, 2013.

[13] H. Bah and A. Lev, "k-anonymity based framework for privacy preserving data collection in wireless sensor networks," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 18, no. 2, pp. 241–271, 2010.

[14] A. Samani, H. H. Ghenniwa, and A. Wahaishi, "Privacy in internet of things: a model and protection framework," *Procedia Computer Science*, vol. 52, no. 538, pp. 606–613, 2015.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 3053202, 7 pages
https://doi.org/10.1155/2017/3053202

Research Article

A Variable Weight Privacy-Preserving Algorithm for the Mobile Crowd Sensing Network

Jiezhuo Zhong,[1] Wei Wu,[1,2] Chunjie Cao,[1,3] and Wenlong Feng[1]

[1]*College of Information Science and Technology, Hainan University, Haikou, Hainan, China*
[2]*Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan, China*
[3]*State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, Haikou, Hainan, China*

Correspondence should be addressed to Chunjie Cao; chunjie_cao@126.com

Received 6 February 2017; Accepted 3 May 2017; Published 5 June 2017

Academic Editor: Zhuo Lu

Mobile crowd sensing (MCS) network collects scenario, environmental, and individual data within a specific range via the intelligent sensing equipment carried by the mobile users, thus providing social decision-making services. MCS is emerging as a most important sensing paradigm. However, the person-centered sensing itself carries the risk of divulging users' privacy. To address this problem, we proposed a variable weight privacy-preserving algorithm of secure multiparty computation. This algorithm is based on privacy-preserving utility and its effectiveness and feasibility are demonstrated through experiment.

1. Basic Theories

1.1. Architecture of Mobile Crowd Sensing Network. Mobile crowd sensing (MCS) network [1] takes the ordinary mobile terminals as the basis sensing units. The sensing task distribution and sensing data collection are achieved through collaboration via the mobile Internet. This represents a large-scale complex social sensing task. "Crowd" refers to the aspect of mobilizing the power and intelligence of the general public, and "sensing" is the process of acquiring the users' behavioral data under different scenarios using the sensors.

Figure 1 shows a typical MCS framework, which consists of the mobile users and the sensing platform. Mobile users are millions of mobile intelligent terminals, into which sensors are embedded (e.g., GPS, gravity sensor, temperature sensor, camera, microphone, and acceleration sensor). These sensors collect various sensing data, which are updated to the sensing platform via the mobile network or short-range wireless communication network. Upon receiving the data, the sensing platform will commence data analysis and processing. The processed data will be directly applied to a diversity of universal social sensing services. After the data analysis and processing are finished, each parcel of data will be evaluated. The mobile users participating in the sensing tasks will be awarded based on the specific incentive mechanism, so as to attract more users into the large-scale sensing task. Liu proposed schemes based on both the Monopoly and Oligopoly models enhancing the location privacy of MCS applications by reducing the bidding and assignment steps in the MCS cycle [2]. Jin proposed a differentially private incentive mechanism that preserves the privacy of each worker's bid against the other honest-but-curious workers [3]. Furthermore, many researchers focused on the detailed information extraction processing in MCS including Hybrid Deep Learning Architecture [4] and Fog Computing and Data Aggregation Scheme [5, 6].

1.2. Application of the MCS Network. The MCS network comprising the mobile intelligent terminals and the mobile sensors is capable of large-scale, complex, fine-grained, and thorough data sensing and collection. For example, the use of MCS network for the collection, analysis, and fusion of the urban traffic flow information can provide highly efficient and convenient path planning and driver assistance system for the mobile users. The MCS network can also provide the decision-making support for urban transport planning and for the formulation of a safe and highly efficient urban transportation network. The MCS network-based sensing

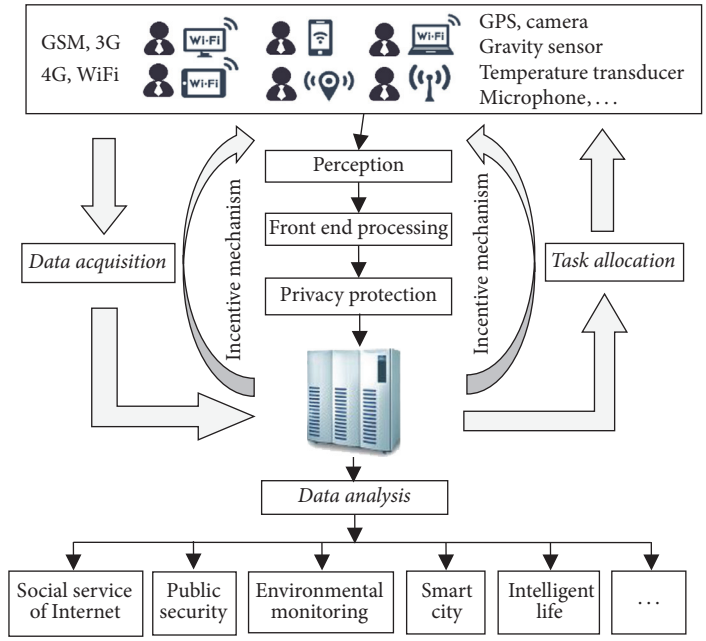


Figure 1: System structure diagram of mobile crowd sensing network.

and monitoring of urban domestic infrastructures offer convenient life services for the local residents. The wide prevalence of the mobile intelligent terminals is a solid guarantee for the high-efficiency and low-cost and large-scale monitoring of natural environment in the cities.

2. Privacy Protection Mechanism for MCS Network Users

2.1. Privacy Preserving in MCS Network. The sensing data collected by the MCS network are largely user privacy. Location data usually contain the sensitive information such as users' address, scope of activity, and transportation route. The mining of users' state of motion can obtain the sensitive information of users' living habits and health conditions. The biological data collected contains the information of users' voice, fingerprints, and basic physiological characteristics. The routine usage data of the mobile intelligent terminals are associated with the user privacy of a deeper level, including the users' hobbies and behavioral traits. Once the user privacy is divulged, there may be violation of privacy, harassment, fraud, or even direct economic loss. Therefore, designing the data security architecture for dynamic privacy protection under the MCS network is an urgent issue.

2.2. Related Technology of Privacy Preserving. The major privacy-preserving techniques used for MCS network are divided into the following types.

(1) Generalized Privacy-Preserving Algorithm. Anonymization is performed while sharing the sensing data, so that the sensitive information about the user's identity is removed without harming the meaningful deduction based on the anonymized sensing data. However, the currently used anonymization methods are usually greedy algorithms which have low execution efficiency.

(2) Perturbation-Based Privacy-Preserving Algorithm. The raw sensing data are perturbed by adding a random number, noises, and exchanges, so that the other party cannot mine the raw sensing data and privacy policies. The main difficulty with data perturbation is how to strike a balance between data correctness, privacy, and security.

(3) Secure Multiparty Computation (SMPC). This technique integrates data encryption and multiple parties are involved in the computation and mining. Because none of the parties have access to complete data, the users' privacy can be ensured. SMPC is now used for collaborative computing among a group of untrusted parties. Many researches have been carried out over the SMPC problem. In 2000, Lindell proposed the method of secure multiparty decision tree (ID3) to protect the data privacy of users [7]. Asharov proposed the threshold homomorphic encryption scheme to improve efficiency of the privacy protection algorithm [8]. In 2014, the threshold-based encryption of K-means outsourcing computing proposed by Liu is a more efficient privacy protection algorithm [9].

Proper application of information technology and algorithm design are the two major concerns in privacy protection. However, the users' attitudes towards privacy are generally neglected. A survey [10] indicates that 17% of the Internet users are still unwilling to provide their authentic information even under privacy protection; 56% of the Internet users are more willing to provide their authentic information in the presence of proper privacy protection; the remaining 27% of the Internet users do not particularly care about their privacy

and will provide the authentic information with or without privacy protection. It is obvious that the users' attitudes towards privacy affect their willingness to share the personal information. Users may react differently to the prospect of disclosing different personal information. But under some incentive mechanisms, the psychological response of the users to the disclosure of different sensitive information may vary.

This study constructed an MCS network-based privacy-preserving algorithm by reference to SMPC. The weight function of privacy preference was built by combining the analysis of the users' sensitivity to the disclosure of privacy and classification of the privacy level of the sensing data. This proposed algorithm can effectively prevent the divulging of privacy information while achieving a maximal acquisition and analysis of the sensing data.

3. Variable Weight Privacy-Preserving Algorithm

3.1. Measure of User Privacy Sensing

3.1.1. User Multiattribute Assumption. Suppose there exists Euclidean space, in which n dimensions represent n solutions to one problem; f_j denotes the attribute j, and G is a set of attributes, $G = \{f_1, f_2, \ldots, f_n\}$. x_i denotes one solution, and X is a set of one solution, $X = \{x_1, x_2, \ldots, x_m\}$. $x_{ij} = f_j(x_i)$ denotes the attribute value of the solution under attribute f_j. $D = (x_{ij})_{m\times n}$ denotes a decision-making matrix of solution X under the attribute G:

$$D = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}. \tag{1}$$

Considering the varying sensitivity to privacy, the users show different willingness to share their privacy in the MCS network. The influence factors of this willingness are divided into profit factors and risk factors, each of which is measured differently. Let $M = \{1, 2, \ldots, m\}$ be the set of the profit attributes, and $N = \{1, 2, \ldots, n\}$ be the set of risk attributes. The two sets are normalized by multiple attribute decision-making using the following formula:

$$y_{ij} = \frac{\left(x_{ij} - \min_i x_{ij}\right)}{\left(\max_i x_{ij} - \min_i x_{ij}\right)} \quad i \in M,\ j \in N. \tag{2}$$

After the transform, the synthetic matrix is $Y = (y_{ij})_{m\times n}$.

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{pmatrix}. \tag{3}$$

3.1.2. Weight Determination of Privacy Perception Attributes. As the users differ in privacy perception, each attribute will carry the information of different user preferences. Therefore, the given user preference can be expressed as the weight of the individual, and the weight of each attribute is expressed as

$$W = (w_1, w_2, \ldots, w_n)^T, \quad \sum_{j=1}^{n} w_j = 1. \tag{4}$$

The utility of each user is expressed as the sum of the weighted attributes. Hence, the user utility U_i is

$$U_i = \sum_{j=1}^{n} y_{ij} \cdot w_j, \quad (j = 1, 2, \ldots, n). \tag{5}$$

The utility analysis of users' privacy perception will provide not only the weight parameters for the SMPC, but also some suggestions for the collection modes of the sensing data under the MCS network. For example, the privacy information sensitive to most users will be prevented and a reasonable incentive mechanism can be designed on this basis. This is very important for increasing the confidence and participation level of users with a lower utility of privacy perception.

3.2. Variable Weight SMPC-Based Privacy-Preserving Algorithm

3.2.1. SMPC-Based Algorithm. SMPC can be conceptualized by the following mathematical model: n participants $P_1, P_2, \ldots, P_n$ of the protocol jointly implement the function $f(x_1, x_2, \ldots, x_m)$. $S_{\text{input}} = \{x_1, x_2, \ldots, x_m\}$ is the set of input variables. The set of input variables S_{P_i} provided by the participant P_i $(i \in \{1, 2, \ldots, n\})$ is a subset of S_{input}, which satisfies $\bigcup_{P_i} S_{P_i} = S_{\text{input}}$, $S_{P_i} \cap S_{P_j} = \phi$ $(i \neq j)$. It is required in the computing of the function that the input S_{P_i} from any participant P_i $(i \in \{1, 2, \ldots, n\})$ is not known to other participants P_j $(j \neq i)$.

The essence of SMPC is a data encryption algorithm using the encryption scheme so as to ensure data privacy. Rivest et al. [11] proposed the concept of fully homomorphic encryption in 1978, aiming to construct an encryption mechanism that supports ciphertext retrieval. Goldwasser [12] studied the strategies used by mobile attackers in the secure channel model. They generalized the threshold mechanism to the ordinary SMPC. The plaintext will be revealed only when at least t participants are involved in the collaborative decryption. This effectively restricts the access to the final SMPC output and the participants will not disclose the data.

3.2.2. Weighted Threshold Secret Sharing Scheme Based on Mignotte Sequence. The weighted threshold secret sharing scheme refers to that each participant assumes a different role, based on which different weights are assigned. The conventional weighted threshold secret sharing schemes achieve only works on the premise of assigning more secret shares

to those who are given special permission. However, this will increase the insecurity of key management and transmission. In this study, we adopted the weighted threshold secret sharing scheme based on *Mignotte* sequence. Regardless of the weight, each participant is only allowed one private key and there is no transmission of secret information between the participants and the dealer. Therefore, the cost of key transmission and storage is spared.

Mignotte sequence is defined as follows [13]:

Let $k, n \in Z,\ n \geqq 2,\ 2 \leqq t \leqq n$. If the integer sequence $m_1, m_2, \ldots, m_n$ satisfies

(1) $m_1 < m_2 < \cdots < m_n$;

(2) $(m_i, m_j) = 1$, where $1 \leqq i < j \leqq n$;

(3) $\prod_{i=0}^{t-2} m_{n-i} < \prod_{i=1}^{t} m_i$,

then sequence $m_1, m_2, \ldots, m_n$ is called a (t, n)-*Mignotte* sequence.

The weighted threshold secret sharing scheme based on the *Mignotte* sequence is designed.

(1) Parameter Configuration. In this scheme, the dealer assigns the weights to each participant using a digit with a length of large prime. The secret to be shared is determined and the relevant system parameters are configured. There are n participants and they constitute the set $U = \{u_1, u_2, \ldots, u_n\}$. The weight vectors of the participants are correspondingly $W = (w_1, w_2, \ldots, w_n)$. The threshold is t, and the secret to be shared is s.

(2) Construction and Expansion of Mignotte Sequence. The dealer needs the system parameters to construct an expanded Mignotte sequence fit for the weighted threshold secret sharing scheme. Meanwhile, the converted scheme should be equivalent to the original scheme. A (t, n)-*Mignotte* sequence is constructed as $m_1', m_2', \ldots, m_n'$, which is expanded into

$$\overbrace{m_1', \ldots, m_1'}^{w_1}, \overbrace{m_2', \ldots, m_2'}^{w_2}, \ldots, \overbrace{m_n', \ldots, m_n'}^{w_n}. \tag{6}$$

The above sequence is a sequence of n' primers, where $n' = \sum_{i=1}^{n} w_i$ which makes the sequence satisfy the following conditions:

(a) The product β of the last $t-1$ numbers is smaller than the product α of the first t numbers.

(b) $\beta < s < \alpha$.

Let $m_1 = (m_1')^{w_1}, m_2 = (m_2')^{w_2}, \ldots, m_n = (m_n')^{w_n}$.

From above, it can be known that sequence $m_1, m_2, \ldots, m_n$ has the following property: $(m_i, m_j) = 1$ $(1 \leqq i \leqq j \leqq n)$.

When $w_n + w_{n-1} + \cdots + w_j < t < w_1 + w_2 + \cdots + w_i$, $m_n m_{n-1} \cdots m_j < s < m_1 m_2 m_i$, $(1 \leqq i \leqq j \leqq n)$.

Thus sequence $m_1, m_2, \ldots, m_n$ is the expanded Mignotte sequence, denoted as (W, t, n)-*Mignotte* sequence. This sequence is revealed.

(3) Generation of Secret Shares. The dealer computes the secret shares of each participant according to the *Mignotte* sequence $m_1, m_2, \ldots, m_n$ and the shared secret:

$$\begin{aligned} s_1 &= s \bmod m_1, \\ s_2 &= s \bmod m_2, \\ &\vdots \\ s_n &= s \bmod m_n. \end{aligned} \tag{7}$$

This S_i is sent to the participant u_i via the secret channel.

(4) Secret Restoration. Suppose there are k participants who constitute the set A, $A = (u_1, u_2, \ldots, u_k)$, and restore the secret. The vector weights for each participant in A constitute the set $W = (w_1, w_2, \ldots, w_n)$.

When the sum of the weight vectors of each participant in A is above or equal to the threshold, that is, $\sum_{i=1}^{t} w_i \geq t$, the following congruence equations are constructed:

$$\begin{aligned} x &= s_1 \bmod m_1, \\ x &= s_2 \bmod m_2, \\ &\vdots \\ x &= s_k \bmod m_k. \end{aligned} \tag{8}$$

$x = \sum_{i=1}^{k} s_i M_i^{-1} M_i (\bmod m)$, where $M_i = m \mid m_i$, $M_i^{-1} M_i \equiv 1 (\bmod m_i)$, $(1 \leq i \leq k)$ and the solution x is the shared secret s.

4. Implementation and Deployment

MapReduce System was used for the high-efficiency parallel processing in the large-scale matrix multiplication in the weighted threshold secret sharing scheme. On the simple data center comprising 5 host machines, the Hadoop distributed storage and computing environment was deployed as a mimic of the sensing platform in the MCS network. One host machine was the Master node, which was deployed with the roles of NameNode and JobTracker for the management of distributed data and task decomposition; 4 host machines were the Slaver and were deployed with the roles of DataNode and TaskTracker for the distributed data storage and task execution. The implementation and deployment (Figure 2) are illustrated below.

(1) The initialization program at the data center would preset the system parameters. The threshold t was determined. The weight vectors of each participant were initialized. The key management system as the trusted third party generated n pairs of homomorphic public and private keys. The public keys hom_PK were the same, and the n private keys were distributed to different participant nodes.

(2) The block function MR_Splitter( ) in the MapReduce System was responsible for dividing the sensing data files submitted by the clients in the MCS network into blocks. Each block was 64 M. The data blocks were encrypted using the

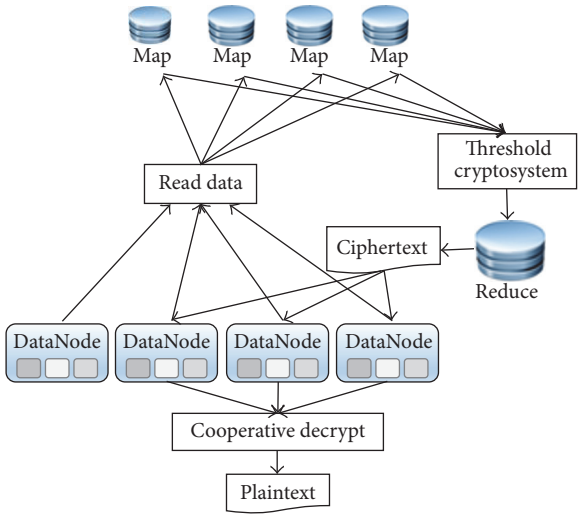


Figure 2: Flowchart of deployment.

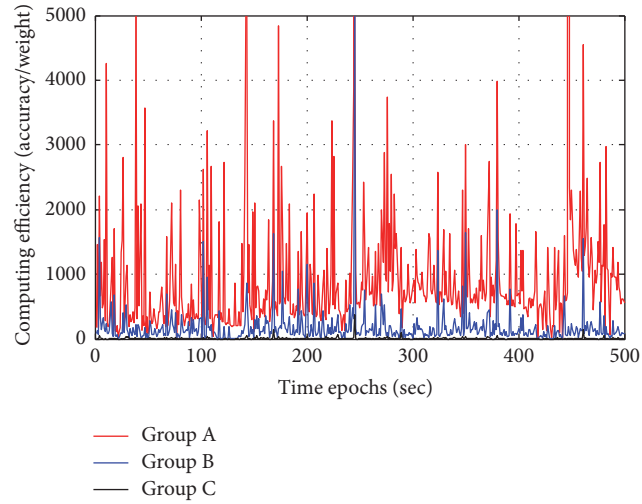


Figure 3: Privacy computing efficiency iteration results of all groups.

public key hom_PK. The encrypted data block file is stored in the distributed file system of the DataNode.

(3) An intermediate ⟨key, value⟩ pair was computed during the matrix multiplication in the privacy-preserving algorithm. The map nodes were allocated to each operation. Before the mapper output the ⟨key, value⟩ pair, the ciphertext for each participant was generated using formulae (7).

(4) Reducer replicated the intermediate output of the corresponding division from the mapper output terminal to the local file system.

(5) At least t participants were involved in the decryption of the ciphertext using the decryption algorithm in formulae (8). These participants would share the decrypted information with other participants. The information decrypted by the t participants was then combined with the information decrypted by the remaining participants to obtain the final result.

For users in network society, we divide them into three groups according to the weights aforementioned in Sections 2 and 3, that is, privacy careless person (group A), practical privacy person (group B), and the group who protect their privacy strictly (group C). In group A, they are not so sensitive with privacy and willing to share their true information. In group B, they may share personal files while policies and regulations are carefully learned. In group C, they are not interested in any sharing information activities at any circumstances.

In a certain survey, the percentage results of groups A, B, and C are obtained as 33.1%, 57.4%, and 9.5% from 352 users on the Internet, and we can initiate the weights of sharing by 0.9, 0.5, and 0.1. These parameters are easily adjusted during privacy protection mechanism proposed here.

As is shown in Figure 3, the three groups in privacy iteration results are given. Group A indicates that since they are not concerned about their information, those provided data are true and the efficiency is acceptable. Group B is matching data on the condition that they believe the privacy is protected, so that their efficiency is not stable and high. Group C are not willing to share their information, and their provided information is not all correct, which influences the computing fundamentally.

5. Algorithm Performance Analysis

5.1. Security Analysis. The private information of each participant is randomly divided into m fragments in a certain way. Each participant selects one fragment randomly and preserves it. The remaining fragments are randomly allocated to other participants. After the fragments are reallocated according to the protocol, each participant will own an equal amount of fragments. Each participant owns one fragment of his or her information plus one fragment transmitted from another participant. Therefore, even if participants P_{i-1} and P_{i+1} conspire, they can only infer the reallocated information of participant P_i and do not know other private information N_i. Any two conspiring participants can only infer the reallocated information of the third party. Then, combining with the information fragments owned by themselves, they can infer the private information of the third party. But when there are more than 3 participants, it will be very difficult to infer all information of the other participants by conspiracy. When there are more than 4 participants and when most participants are honest, the possibility of information leak will approach 0.

5.2. Complexity Analysis. Computational complexity: each round of computation consists of m operations (different from the m aforementioned), and m rounds involve m^2 operations. Thus the computational complexity $S(m)$ is expressed as $S(m) = m^2$, as shown in Figure 4.

Communication Complexity. Each participant needs to transmit $m - 1$ fragments to other participants. Therefore, in the fragment transmission stage, $m(m - 1)$ communications will occur. In the computing stage, each participant needs

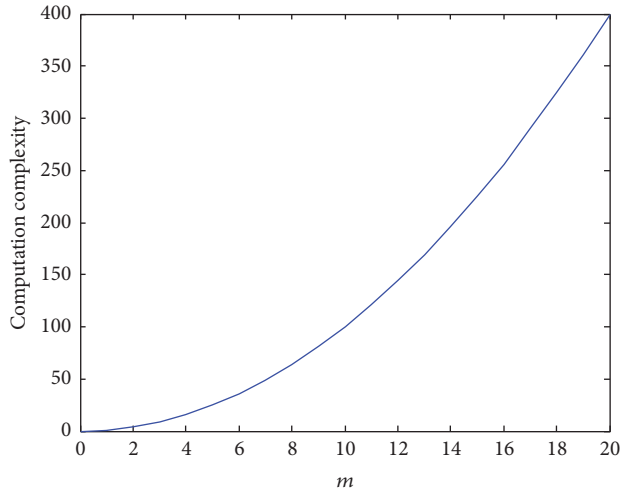


Figure 4: Computation complexity.

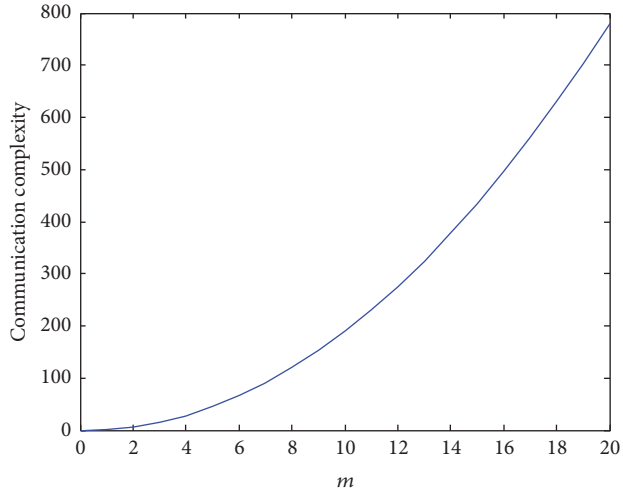


Figure 5: Communication complexity.

to transmit the summation of some fragments to other participants over the ring structure. Therefore, each round of computation consists of m communications, and m rounds involve m^2 communications. The overall communication complexity $C(m)$ of the algorithm is expressed as $C(m) = m(m-1) + m^2 = 2m^2 - m$, as shown in Figure 5.

6. Summary and Forecast

To protect against privacy violation in the MCS network, we proposed a variable weight SMPC-based privacy-preserving algorithm. The weighted threshold secret sharing scheme based on Mignotte sequence was applied for the encryption of the sensing data and private key management. Considering the different attitudes of users towards the disclosure of the private information, the privacy of the information was graded. Thus the weight parameters of the privacy-preserving algorithm were determined based on the utility analysis of the users' privacy perception. The proposed model was deployed in the Hadoop distributed environment to verify its effectiveness and validity. The implementation of the SMCP protocol requires several participants, among which communications are necessary. This will incur significant communication and computational costs. How to enhance the reliability of channel communication and to increase the efficiency of sensing data encryption are issues awaiting resolution.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper and confirm that the mentioned received funding in the Acknowledgments did not lead to any conflicts of interest regarding the publication of this manuscript.

Acknowledgments

This work was supported by the Natural Science Foundation of Hainan Province (no. 20166216 and no. 617033) and Education and Reaching Research Project of Hainan University (no. hdjy1325) investigated by Jiezhuo Zhong; National Natural Science Foundation of China (no. 61661019), the Major Science and Technology Project of Hainan Province (no. ZDKJ2016015), the Natural Science Foundation of Hainan Province (no. 20156217), and the Higher Education Reform Key Project of Hainan Province (no. Hnjg2017ZD-1) by Chunjie Cao; National Science and Technology Support Program (no. 2015 BAH55F01-5) and Natural Science Foundation of Hainan Province (no. 614232) investigated by Wenlong Feng.

References

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] B. Liu, W. Zhou, T. Zhu et al., "Invisible hand: a privacy preserving mobile crowd sensing framework based on economic models," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 1-1, 2017.

[3] H. Jin, L. Su, B. Ding et al., "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proceedings of the IEEE 36th International Conference on Distributed Computing Systems (ICDCS '16)*, pp. 344–353, Nara, Japan, June 2016.

[4] S. A. Ossia, A. S. Shamsabadi, A. Taheri et al., "A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics," https://arxiv.org/abs/1703.02952.

[5] S. Basudan, X. Lin, and K. Sankaranarayanan, "A privacy-preserving vehicular crowdsensing based road surface condition monitoring system using fog computing," *IEEE Internet of Things Journal*, no. 99, pp. 1-1, 2017.

[6] C. Xu, R. Lu, H. Wang, L. Zhu, and C. Huang, "PAVS: a new privacy-preserving data aggregation scheme for vehicle sensing systems," *Sensors*, vol. 17, no. 3, p. 500, 2017.

[7] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Advances in Cryptology—CRYPTO 2000, 20th Annual International Cryptology Conference, Santa Barbara, California, USA,*

August 20–24, 2000, vol. 1880 of *Lecture Notes in Computer Science*, pp. 36–54, 2000.

[8] G. Asharov, A. Jain, A. López-Alt, E. Tromer, V. Vaikuntanathan, and D. Wichs, "Multiparty computation with low communication, computation and interaction via threshold FHE," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7237, pp. 483–501, 2012.

[9] L. Liu and M. Tamer Özsu, *Encyclopedia of Database Systems*, Springer, New York, NY, USA, 2017.

[10] D.-H. Shin, "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption," *Interacting with Computers*, vol. 22, no. 5, pp. 428–438, 2010.

[11] R. L. Rivest, L. Adleman, and M. L. Dertouzos, *On Data Banks And Privacy Homomorphism Proc of Foundations of Secure Computation*, Academic Press, New York, NY, USA, 1978.

[12] S. Goldwasser, "Multi party computations: past and present," in *Proceedings of the sixteenth annual symposium on Principles of distributed computing (ACM '97)*, pp. 1–6, August 1997.

[13] M. Mignotte, "How to share a secret," *Lecture Notes in Computer Science*, vol. 149, no. 2, pp. 371–375, 1983.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 2587948, 8 pages
https://doi.org/10.1155/2017/2587948

Research Article

Abnormal Event Detection in Wireless Sensor Networks Based on Multiattribute Correlation

Mengdi Wang, Anrong Xue, and Huanhuan Xia

School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China

Correspondence should be addressed to Anrong Xue; xuear@ujs.edu.cn

Received 10 February 2017; Accepted 27 March 2017; Published 6 April 2017

Academic Editor: Mengxing Huang

Abnormal event detection is one of the vital tasks in wireless sensor networks. However, the faults of nodes and the poor deployment environment have brought great challenges to abnormal event detection. In a typical event detection technique, spatiotemporal correlations are collected to detect an event, which is susceptible to noises and errors. To improve the quality of detection results, we propose a novel approach for abnormal event detection in wireless sensor networks. This approach considers not only spatiotemporal correlations but also the correlations among observed attributes. A dependency model of observed attributes is constructed based on Bayesian network. In this model, the dependency structure of observed attributes is obtained by structure learning, and the conditional probability table of each node is calculated by parameter learning. We propose a new concept named attribute correlation confidence to evaluate the fitting degree between the sensor reading and the abnormal event pattern. On the basis of time correlation detection and space correlation detection, the abnormal events are identified. Experimental results show that the proposed algorithm can reduce the impact of interference factors and the rate of the false alarm effectively; it can also improve the accuracy of event detection.

1. Introduction

Abnormal event detection is one of the main problems in wireless sensor networks [1]. In wireless sensor networks, abnormal events are usually complex, because an event usually involves multiple observed attributes, and it is difficult to describe an abnormal event pattern [2]. Existing anomaly detection algorithms detect an abnormal event by comparing a single attribute threshold [3, 4] or by considering the spatiotemporal correlations of sensor readings [2, 5–8]. However, some important information may be hidden in the correlations among different attributes [9].

In [3], an adaptive distributed event detection method is proposed, which dynamically adjusts the decision threshold based on the trust value of the sensor nodes and uses the moving average filter to tolerate the transient faults of the sensor nodes. Although this method is fault-tolerant, it is still possible to misjudge the event nodes into faulty nodes. Particularly when the event range is large, the accuracy of detection will decrease significantly. Besides, this method computes a trust value for each sensor node, so it can only be applied to univariate applications. Paper [5] models the event region based on Dynamic Markov Random Field. This method can effectively capture the dynamic changes of local area; since the method needs to exchange information of space-time neighbor constantly, the detection efficiency is low. Besides, the detection of the events lacks a global perspective, which may lead to misjudgment of abnormal events. Paper [6] proposed an event detection scheme based on spatiotemporal correlations. In this method, the sensor nodes are divided into multiple working groups; the time correlation of the sensor data is used to eliminate low frequency errors. Different working groups cooperate to determine whether the anomalies represent an event. However, this method only constructs the model based on the single sensing attribute and does not consider the relations between the multisensory attribute and the abnormal event.

The attributes of the sensor readings usually contain time information, sensor topology information, and other attributes directly sensed by the sensor (e.g., temperature, humidity, and light intensity). When abnormal events occur in the network, events often show temporal correlation,

spatial correlation, and attributes correlation [9]. In most cases, event detection methods that take the spatiotemporal correlation of the data into account are susceptible to both sensor failures and external environmental noises. For observed attributes, a simple threshold comparison is insufficient to determine whether an abnormal event occurs. For instance, in an indoor fire monitoring application, the increase of the temperature and smoke concentration may be caused by cooking, rather than a fire accident.

In order to improve the accuracy of abnormal event detection in wireless sensor networks with multiple attributes and reduce the influence of environmental noises and sensor failures on the event detection results, this paper proposes a new method called Abnormal Event Detection based on Multiattribute Correlation (MACAED). First, considering that Bayesian network can effectively represent the dependencies among variables, a Bayesian network is used to establish the dependency model of observed attributes. In this model, the dependency structure of abnormal events is obtained by structure learning. Each node learns the parameters to get a conditional probability table. Then, the attribute correlation confidence is introduced to judge whether the attribute correlation mode of the point is an abnormal mode. Based on the sliding window model, the degree of temporal correlation was calculated; the spatial similarity was calculated by using the neighbor node information. Finally, the anomaly events were detected by three kinds of attribute correlation.

2. Attribute Dependency Model

In wireless sensor networks, abnormal events usually show the following three characteristics:

(1) For a single sensor node, the anomaly event will continue for a period of time once the event occurs; the adjacent time of the data shows a certain degree of similarity [7]. In addition, abnormal events will inevitably affect the physical environment of network monitoring, and the sensor data will change accordingly, showing a special mode.

(2) For a number of sensor nodes, sensor nodes in the event region will exhibit spatial similarity when abnormal events occur [10]; in other words, the readings of adjacent nodes exhibit similar patterns.

(3) When the abnormal events occur in the monitoring area, the sensed attributes of the sensor readings show a certain degree of relevance, and this correlation appears as probability relations [9].

According to the three kinds of characteristics of abnormal events in wireless sensor networks and the experience that Bayesian network can effectively represent the probability relationship among attributes, we construct the attribute dependency model. The attribute correlation confidence is proposed to measure the degree of similarity between the measured points and the anomalies in observed attribute probability model.

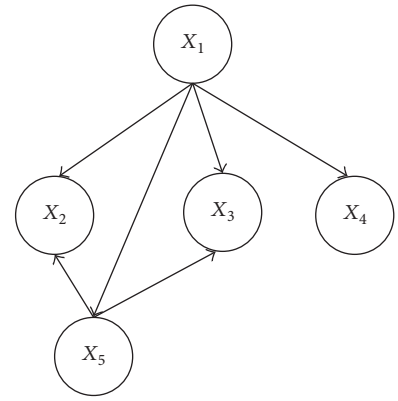


Figure 1: An example of attribute dependency model.

2.1. Bayesian Network. Bayesian network is a product of probability theory and graph theory. It is a directed acyclic graph with probabilistic annotations, which can represent the probability dependencies among random variables. It has a solid mathematical foundation [11]. On the one hand, the Bayesian network can reveal the structure of the problem intuitively by using graph theory. On the other hand, the Bayesian network can utilize the structure of the problem according to the principle of probability theory, which reduces the computational complexity of reasoning. In view of this, this paper establishes a dependency model of observed attributes based on the Bayesian network; each attribute is represented by a unique node, and the probabilistic dependencies are represented by arcs between nodes.

2.2. Formal Description. The attribute dependency model is represented by a triplet $B = (D, G, \theta)$, where D is the sample dataset that contains observed attributes, $D = \{d_1, d_2, \ldots, d_m\}$; G denotes a directed acyclic graph, which qualitatively describes the dependencies among attributes, $G = (X, U)$, where X is a set of nodes representing observed attributes, corresponding to the elements in D, and U is the directed edge set representing the dependencies among the attributes; θ is the set of conditional probability distributions for each node, which quantitatively describes the dependencies among attributes, $\theta = \{P(X_i \mid \pi(X_i))\}$, where X_i is the ith node in G and $\pi(X_i)$ is the set of parent nodes of node X_i. Figure 1 is an example of an attribute dependency model.

2.3. Structure Learning. For WSNs with large number of variables and implicit dependencies among variables, it is difficult to obtain a reasonable network structure relying on a priori information and expert knowledge, and the probability is subjective, so we learn the Bayesian network structure from training samples. This paper utilizes a strategy of scoring and searching. Specifically, we use a scoring function to evaluate the matching degree between a specific network structure and the training sample and select the appropriate search strategy to search the network structure with the highest scoring value.

Given a sample dataset $D(d_1, d_2, \ldots, d_m)$, let Bayesian network G take all the variables in the node set $X(X_1, X_2, \ldots, X_m)$ as nodes, and instantiate all the variables of X using the attribute value d_i in the sample dataset. The variable X_i has r_i possible values $(x_{i1}, x_{i2}, \ldots, x_{ir_i})$. Let the parent variable set of X_i be Π_i, w_{ij} denotes the jth instantiation value of the parent node Π_i with respect to D, and N_{ijk} denotes the number of instances in which the value X_{ik} of the variable X_i is taken and is instantiated into w_{ij} by Π_i, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. The Bayesian scoring criterion is used to compute the likelihood ratios of the two Bayesian network structures G_1 and G_2. Since $p(G_1 \mid D)/p(G_2 \mid D) = p(G_1, D)/p(G_2, D)$, we only need to compare the joint probability $p(G_1, D)$ and $p(G_2, D)$. This can be calculated by using the formula [12]

$$\begin{aligned} p(G, D) &= p(G)\, p(D \mid G) \\ &= p(G) \cdot \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{\left(N_{ij} + r_i - 1\right)!} \cdot \prod_{k=1}^{r_i} N_{ijk}!, \end{aligned} \tag{1}$$

where $p(G)$ is the prior probability and the arrangement order of Π_i is $(1, \ldots, q_i)$. Maximizing the joint probability $p(G, D)$ in (1)

$$\begin{aligned} &\max_G \{p(G, D)\} \\ &\quad = p(G) \prod_{i=1}^{n} \max_{\Pi_i} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{\left(N_{ij} + r_i - 1\right)!} \prod_{k=1}^{r_i} N_{ijk}! \right]. \end{aligned} \tag{2}$$

It can be seen that, for each variable X_i, it is only necessary to maximize

$$\max_{\Pi_i} \{g(i, \Pi_i)\} = \max_{\Pi_i} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{\left(N_{ij} + r_i - 1\right)!} \prod_{k=1}^{r_i} N_{ijk}! \right]. \tag{3}$$

In the initial stage of constructing the network structure, it is assumed that each node has no parent node. The nodes which meet the posterior probability maximization formula are recursively added to the parent set of nodes. When $p(G, D)$ is no longer increased, stop adding to the parent node set; then the network structure G' is obtained. For the current sample dataset D, G' is the optimal network structure under the Bayesian scoring standard.

2.4. Parameter Learning. According to the trained network structure, the parameter of each node in the network is learned to get the corresponding conditional probability table. The conditional probability table contains the probability relations among the variables. Using the maximum likelihood estimation method, suppose $(x_1, x_2, \ldots, x_n)$ is a set of possible values of random variable set $(X_1, X_2, \ldots, X_n)$, and the probability of $(X_1, X_2, \ldots, X_n)$ falling in the neighborhood of $(x_1, x_2, \ldots, x_n)$ (n-dimensional cubes with side length $dx_1, dx_2, \ldots, dx_n$, resp.) is approximated as $\prod_{i=1}^{n} f(x_i; \theta) dx_i$, where $\prod_{i=1}^{n} f(x_i; \theta)$ is the joint probability density of $(X_1, X_2, \ldots, X_n)$, θ is the structural parameters, and $\theta \in \Theta$. The maximum likelihood estimation value $\widehat{\theta}$ of θ is calculated through $\max_{\theta \in \Theta} L(x_1, \ldots, x_n; \theta)$. The conditional probability table for each node is obtained from the sample data and prior knowledge.

2.5. Attribute Correlation Confidence. Attribute correlation confidence is a concept we proposed to measure the fitting degree between the sensor reading and the abnormal event pattern. It is equal to the ratio of the joint probability distribution between the measured point and the abnormal point. Let $(y_1, y_2, \ldots, y_n)$ be the sensor reading at the current time. For an abnormal event E_i, the joint probability of all node variables $P(X_1, X_2, \ldots, X_n \mid E_i)$ is calculated according to the Bayesian network structure and the conditional probability table. Since in Bayesian network, not every node has an arc to the all the rest nodes, the conditional probability only depends on the direct parent node. In other words, given the values of parent variables, the probability of nondescendant node is conditionally independent of the parent node. So the calculation of joint probability $P(X_1, X_2, \ldots, X_n \mid E_i)$ can be simplified by using the chain rule [11],

$$p(x) = \prod_{i=1}^{n} p\left(x_i \mid x_{pa(i)}\right) \tag{4}$$

in which $x_{pa(i)}$ represents the parent node of x_i.

After calculating $P(X_1, X_2, \ldots, X_n \mid E_i)$, we can get the probability pattern of the reading in an event. According to the formula,

$$\alpha = \max_{i \in I} \frac{P\left(X_1 = y_1, \ldots, X_n = y_n\right)}{P\left(X_1 = x_1, \ldots, X_n = x_n \mid E_i\right)}, \tag{5}$$

the attribute correlation confidence α of the tested point is calculated. The higher the probability, the more the possibility for the anomaly to represent an abnormal event.

3. Abnormal Event Detection Algorithm Based on Multiattribute Correlation

In this paper, we propose a detection algorithm based on multiattribute correlation, which is divided into three phases: attribute correlation pattern decision, temporal similarity detection, and spatial similarity detection.

3.1. Description of Abnormal Event. For an abnormal event, define event information $Info = \{Tm, Loc, Attr, Parm, E_i\}$, where Tm is the time of occurrence of abnormal events, Loc is the location of abnormal events, and $Attr$ is the attribute set that an event involves. $Parm$ is the parameter set, which includes temporal similarity threshold ε, spatial similarity threshold δ, and attribute correlation confidence threshold φ. For different application environments, the values of each item in $Parm$ can be adjusted to achieve the best detection result adaptively. E_i represents the event type, $i = 0$ means no abnormal events occurred, $i > 0$ means that abnormal events occurred, and the higher the value i is, the more severity the abnormal event has.

3.2. Temporal Similarity Detection. The data sampling frequency of most wireless sensor networks is relatively high and data change range at the adjacent time is relatively small, so the sensor data is time-correlated. Combining with sliding window model and the attribute dependency model obtained, candidate anomalies that may represent abnormal events are detected.

Let s be the size of the sliding window, and for each data sequence t_i within the window, calculate the similarity between t_i and the current time series t

$$q(t_i, t) = \frac{1}{\left(1 + \sqrt{\sum_{k=1}^{m} \left(x_k^{t_i} - x_k^t\right)^2}\right)}. \tag{6}$$

Considering that the data sequence that is closest to the current time is most correlated, the average similarity between the current time data and the data in the window is calculated by the weighted summation method

$$\overline{q(t)} = \frac{\sum_{i=1}^{ww} w_i q(t_i, t)}{s}, \tag{7}$$

where the weight is $w_i = 1/(t - t_i)$. If the average similarity is smaller than the threshold ε and the confidence degree of the attribute correlation is greater than or equal to the threshold φ, it means that not only does the data sequence of the current time significantly deviate from the historical data, but also the relationship among the attributes is in accordance with the probability relation when the abnormal event occurs, which needs a further spatial correlation detection. In other cases, it will be filtered as a noise.

3.3. Spatial Similarity Detection. The similarity between the candidate anomaly and the neighbor node's data sequence is calculated. If the candidate anomaly and the neighbor node's data sequence satisfy certain similarity degree, it indicates that the abnormal event occurs in the region where the candidate anomaly is located and needs to be uploaded to the sink node.

The similarity between the candidate anomaly and the neighbor node sequence is calculated according to the following formula:

$$q(x_t, y_t) = \frac{1}{\left(1 + \sqrt{\sum_{k=1}^{m} \left(x_k^t - y_k^t\right)^2}\right)}. \tag{8}$$

If the spatial similarity $q(x_t, y_t)$ is greater than or equal to the threshold δ, it indicates that both nodes have detected an abnormal event at the same time and mark the candidate anomaly nodes and their neighbor nodes as abnormal event nodes. On the contrary, it indicates that no neighbor nodes detect abnormal information at this time, and the candidate anomaly belongs to noise data, which is also filtered out.

3.4. Description of MACAED Algorithm. Based on the calculation of attribute correlation confidence and the detection of temporal and spatial correlation of sensor data, an abnormal event detection algorithm based on multiattribute correlation is proposed. The pseudocode of the algorithm is shown in Algorithm 1.

In the pseudocode of Algorithm 1, rows (2)~(3) train the Bayesian network through the scoring-searching method and choose the network structure M with the highest score as the observed attribute dependency model, rows (4)~(26) detect abnormal events in real time, where rows (9)~(10) proceed parameter learning for each sensor in order to update the probability distribution in attribute dependency model, rows (10)~(14) compute the attribute correlation confidence of observed attributes, row (15) calculates the average similarity between the current time readings and the readings within the window, row (18) calculates the average similarity between the current node and the adjacent node readings, and rows (17)~(24) determine whether the current reading represents abnormal events readings.

3.5. Time Complexity Analysis. Let n be the number of observed attributes, which corresponds to the number of nodes in Bayesian network; m is the number of instances, that is, the number of readings; r is the number of possible values for each observed attribute; N is the number of nodes in WSN; s is the size of sliding window. For the structure learning part, the time complexity is $O(mn^4r)$ [12]. For abnormal event detection part, it contains two layers: outer layer loops $O(m - s - 1)$ times and inner loops $O(N)$ times. The parameter learning consists of a cycle of $O(n)$ times. The time correlation detection consists of a cycle of $O(s)$ times. The spatial correlation detection consists of a cycle of $O(N)$ times. The total time complexity of the algorithm is $O(mn^4r) + O(m-s-1)O(N)O(n+s+N)$. Since, for most wireless sensor networks, the value of n is small (less than 10) and sliding window s and the number of possible values of each attribute r are relatively small (in this experiment, $s = 10$; $r = 9$), the influence of these values on the total time complexity can be ignored, so the total time complexity can be simplified to $O(m) + O(mN^2) = O(mN^2)$.

4. Experimental Results and Analysis

4.1. Datasets. We test the performance of the MACAED algorithm by means of conducting simulation experiments on Matlab 2014a. The experiments are run on a PC with an Intel Core i3-2120 @3.30 GHZ Cpu, 4 GB memory, and Windows 7 operating system. For the instance of detecting fire event, the performance tests are based on the processed data of Intel Lab Data [13] from Intel Berkeley Lab. Except for the real data field, we insert the fire events and interference events data field into the dataset manually.

The experiment dataset contains the records of 54 sensors deployed in the IBRL lab during the time span from February 28th to April 5th in 2004. The MicaDot sensors collect temperature, humidity, light intensity, and voltage value every 31 seconds. Sensor node deployment is shown in Figure 2.

Input. WSN data set D
Output. Abnormal event Information *Info*
(1) standardize D into values between 0 and 1
(2) divide D into K subsets, choose the first set to learn Bayesian network
(3) choose the network with highest score as attribute
dependency model M
(4) for $t = s + 1$ to *epoch*//*epoch* is incremental tick
(5) if t%*period* = 0//*period* is parameter update period
(6) *flag* = true; //*flag* represents update parameter or not
(7) end
(8) for *id* = 1 to N//*id* is the id of WSN, N is the number of sensors
(9) learn parameter for each sensor node
(10) if dataPointer [*id*] < *group_length*
//prevent the *id* exceed the length of group
(11) if groupData_time [*id*] < t
//prevent a break caused by data loss
(12) compute α from M
(13) end
(14) end
(15) compute $q(t_i, t)$
(16) if $q(t_i, t) < \varepsilon$ && $\alpha \geq \varphi$
(17) compute $q(x_t, y_t)$;
(18) if $q(x_t, y_t) \geq \delta$
(19) report *Info* to sink node;
(20) else
(21) filter as noise;
(22) end
(23) end
(24) end
(25) *flag* = false;
(26) end

Algorithm 1: Abnormal event detection algorithm based on multiattribute correlation.

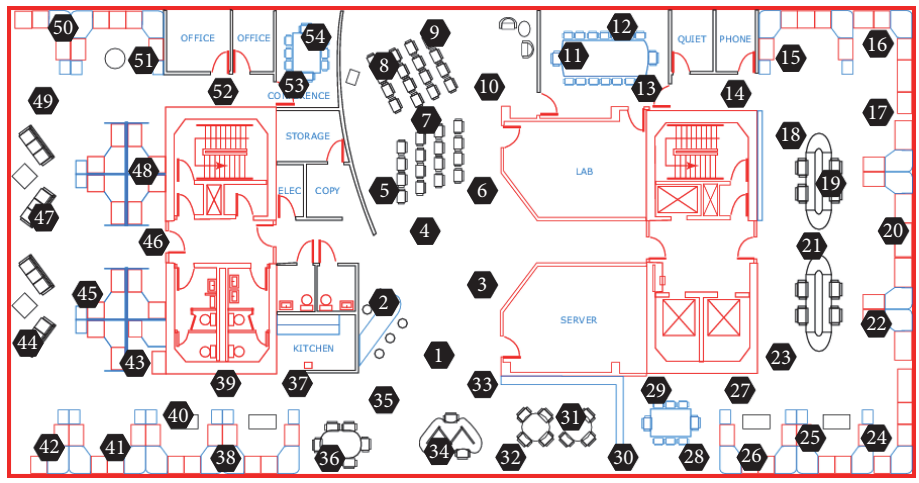


Figure 2: Location of sensor nodes deployed in IBRL lab.

4.2. Data Preprocessing. In our experiment, we choose the records within 24 hours in February 28th as our test data; we preprocess the raw data as follows:

(1) Since the unit of measurement attributes directly sensed by each sensor is different and the changing range of different attributes is wide, so the raw data needs to be standardized and mapped to [0, 1]; in this way, the relative distance can be calculated.

(2) Since the change of each attribute value is continuous and periodic, in order to facilitate the calculation, the experimental datasets are discretized, and the values of each attribute are divided into 10 intervals.

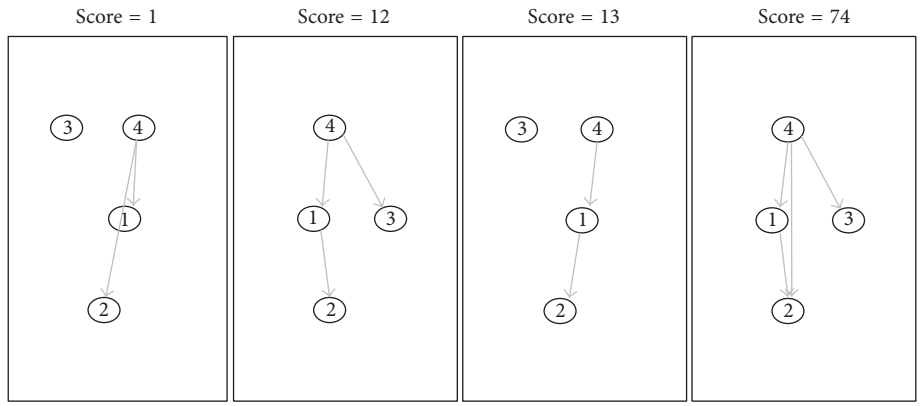


Figure 3: Network structure under Bayesian scoring criterion.

(3) For some parts of the raw IBRL datasets have missing values and the failure nodes (both node 5 and node 15 have no records; node 28 only has 3 attribute records), the NaN is used in this experiment to fill the missing values, and these values will be discussed in different situations, not for computation.

(4) In order to verify the performance of our algorithm on detecting abnormal events, abnormal readings that represent abnormal events are added in the dataset. In addition, the readings of the abnormal events with the interference are added (e.g., opening heater in the room will make the temperature rise).

4.3. Experimental Parameters. Temperature T, humidity H, light intensity L, and voltage V are numbered with 1, 2, 3, 4. In order to obtain relatively stable Bayesian network structure, we set the maximum number of parent nodes in structure learning $max\_fan\_in = 2$, learning step length $step = 10$, and the number of instances $ncases = 1000$. The optimal parameter learning cycle $period = 600$. Bayesian networks with four different scores are showed in Figure 3; the higher the score is, the more stable the network structure is. Thus, we choose the structure whose score = 74 as an attribute dependency model in this experiment.

In this method, the sliding window size has a direct impact on the detection results. The precision, the recall, and the $F1$-measure of anomaly detection under different sliding window sizes are experimented. The experimental results are shown in Figure 4.

From Figure 4, we can find that the recall decreases with the increase of the sliding window width; however, the overall change is not obvious. But the precision declines relatively faster, leading to the quick decrease of $F1$ value. This is because, with the increase of window width, the historical data increases, and the calculated average value declines ceaselessly, which means that the possibility of becoming candidate anomalies is higher. Considering that the sliding window width is small and the amount of uploaded data is small, so we set the sliding window size $s = 10$; in this way, we will make full use of historical data.

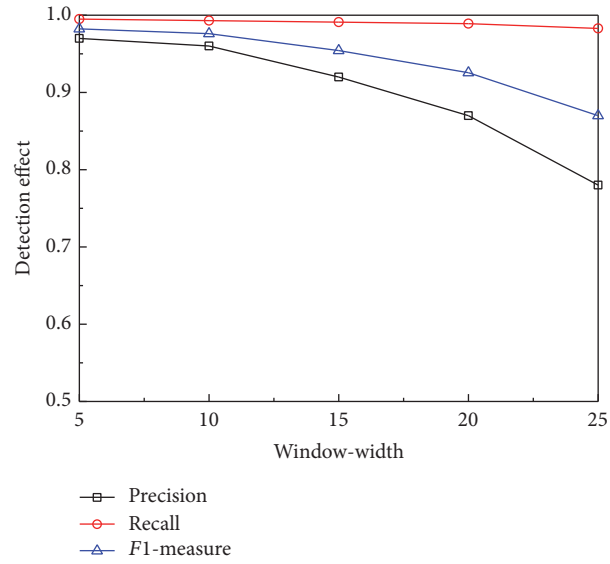


Figure 4: Influence of sliding window size on the test results.

There are different requirements for the threshold settings when the environment of wireless sensor networks differs. We change the value of three different thresholds and test the accuracy of the anomalies under the change of single threshold; the results are shown in Figure 5.

From Figure 5 it can be concluded that it gets the highest detection accuracy when temporal similarity threshold $\varepsilon = 0.1$, spatial similarity threshold $\delta = 0.2$, and attribute correlation confidence threshold $\varphi = 0.5$.

4.4. Contrast Experiment. In the contrast experiment, we still use the IBRL dataset, in which the number of sensor nodes is 54, and the deployment of nodes is shown in Figure 2. We use (T, H, L, V) to represent four different attributes: temperature, humidity, light intensity, and voltage. Since there are no interference factors in the dataset, we add some false abnormal events artificially, which are shown in Table 1.

The contrast algorithms include the Adaptive Fault-Tolerant Event Detection (AFTED) algorithm proposed in

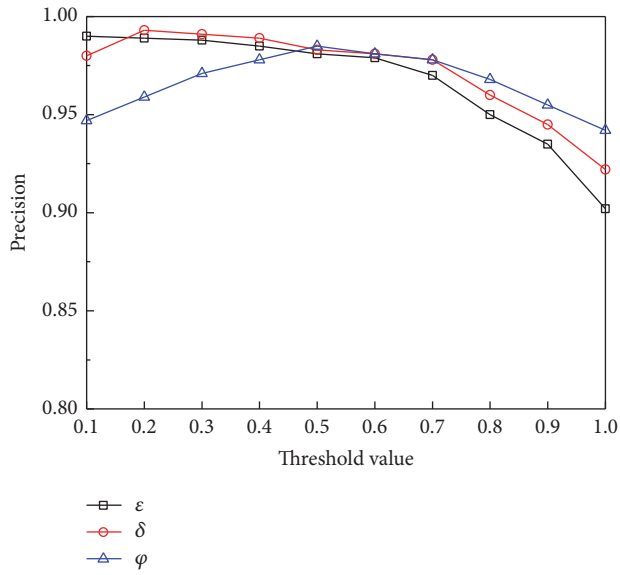


Figure 5: Influence of the three thresholds on the test results.

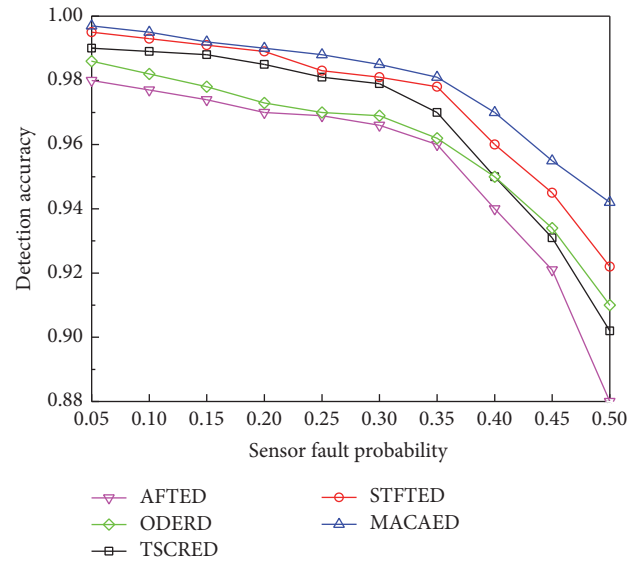


Figure 6: Detection accuracy of five algorithms.

Table 1: False abnormal events.

Number	Event name	Attributes	Id of nodes
1	Cooking	T, H, L, V	2, 37
2	Air-condition	T, H, V	5, 17, 24, 36, 44
3	Heater	T, V	11, 12, 13
4	Bath heater	T, L, V	53, 54
5	Humidifier	H	27, 28, 29, 30

[3], the Online Dynamic Event Region Detection (ODERD) algorithm proposed in [5], the Real-Time Event Detection Approach based on Temporal-Spatial Correlations (TSCRED) presented in [6], and the Spatiotemporal Correlation based Fault-Tolerant Event Detection (STFTED) scheme proposed in [8]. And we compare the detection accuracy, false alarm rate, and detection time of abnormal events.

In the proposed algorithm, we use the same parameter settings as the previous experiments. In AFTED algorithm, we set the window size for tolerating transient faults $M_{\text{AFTED}} = 4$, and the threshold for filtering transient faults $\delta_{\text{AFTED}} = 0.75$, which have been verified to be the most appropriate in their experiment. In ODERD algorithm, since we only focus on the static abnormal event detection, the parameters controlling the shift and deformation of event regions are set to 0 s. To compare these algorithms in an equivalent level, we set the sliding window size of TSCRED and STFTED to 10, which is the same as the proposed algorithm. Besides, all of the sensor nodes have the same communication range $R = 4$. And each event region is assumed to be a circle with radius $l = 2R$.

The results of the proposed algorithm compared with the other four algorithms in detection accuracy are shown in Figure 6. It can be seen from Figure 6 that when the node failure rate goes from 0.05 to 0.3, the detection accuracies of the five algorithms are similar, reaching 0.96 or more; this is because most of the noise is filtered out in the time correlation

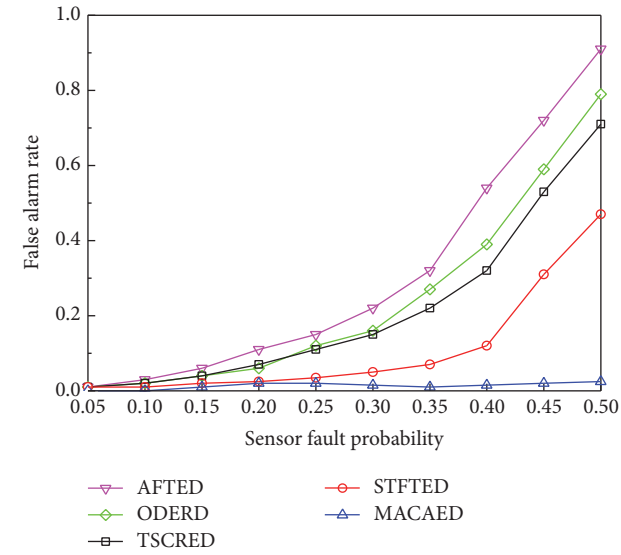


Figure 7: False alarm rate of five algorithms.

detection phase. When the node failure rate is greater than 0.3, the detection accuracies of the five algorithms decrease significantly, but the MACAED algorithm is significantly better than the other four algorithms. The reason is that all the five algorithms have the spatial correlation detection stage. With the increase of the failure rate, the faulty nodes are easily affected by the neighbor nodes which have not detected the abnormal events, and they are converted into the normal state, therefore misjudging that no abnormal events occurred.

As for the false alarm rate, these compared results are shown in Figure 7. It can be seen that MACAED has a significantly lower false alarm rate than the other four algorithms as the node failure rate increases. This is due to the fact that MACAED fully considers the impact of attribute correlations on abnormal event detection. By calculating the attribute

Table 2: Running time of five algorithms.

Algorithms	Time(s)
AFTED	8.381
ODERD	7.647
TSCRED	7.435
STFTED	10.917
MACAED	12.546

correlation confidence, the fitting degree between the data records and the abnormal event attribute dependency model can be determined, so the abnormal event and interference factor can be distinguished effectively.

The running time of the five algorithms is shown in Table 2.

It can be seen from Table 2 that the MACAED algorithm consumes the longest time. The reason is that the MACAED algorithm needs to train the network structure at the beginning. This process takes about 5 s on average. If the trained network structure is saved as the known result, the detection phase needs $12.546 - 5 = 7.546$ s, which is very close to TSCRED algorithm and ODRED algorithm.

5. Conclusion

In this paper, we present a new approach to detect abnormal events in wireless sensor networks. We construct a dependency model of observed attributes based on Bayesian network and propose a new method to measure the dependency of the attributes. Combining with the temporal correlation detection based on sliding window and the spatial correlation detection based on neighbor node information, the influence of noise and interference event factors on event detection results is effectively reduced. Experimental results show that the algorithm proposed in this paper can effectively eliminate the influence of interference events. It not only reduces the false alarm rate of abnormal events but also improves the accuracy of event detection compared with the other four algorithms.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is sponsored by Innovative Fund Project of Science and Technology Enterprises of Jiangsu Province in China no. BC2014212.

References

[1] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[2] A. De Paola, S. Gaglio, G. L. Re, F. Milazzo, and M. Ortolani, "Adaptive distributed outlier detection for WSNs," *IEEE Transactions on Cybernetics*, vol. 45, no. 5, pp. 888–899, 2015.

[3] S.-J. Yim and Y.-H. Choi, "An adaptive fault-tolerant event detection scheme for wireless sensor networks," *Sensors*, vol. 10, no. 3, pp. 2332–2347, 2010.

[4] Y. Wang, D. Wang, F. Chen, and W. Fang, "Efficient event detection using self-learning threshold for wireless sensor networks," *Wireless Networks*, vol. 21, no. 6, pp. 1783–1799, 2014.

[5] T. Wu and Q. Cheng, "Online dynamic event region detection using distributed sensor networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 393–405, 2014.

[6] F. Li and Z. Feng, "An efficient real-time event detection approach based on temporal-spatial correlations in wireless sensor networks," in *Proceedings of the International Conference on Computer Science and Network Technology (ICCSNT '11)*, pp. 1245–1249, December 2011.

[7] J. Yin, D. H. Hu, and Q. Yang, "Spatio-temporal event detection using dynamic conditional random fields," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI '09)*, pp. 1321–1326, IEEE, Pasadena, Calif, USA, July 2009.

[8] K. Liu, Y. Zhuang, Z. Wang, and J. Ma, "Spatiotemporal correlation based fault-tolerant event detection in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 643570, 2015.

[9] C. Luo, J.-G. Lou, Q. Lin et al., "Correlating events with time series for incident diagnosis," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 1583–1592, August 2014.

[10] K.-Z. Liu, Y. Zhuang, S.-L. Zhou, and S.-J. Liu, "Event detection method based on belief model for wireless sensor networks," *Journal of Beijing University of Posts and Telecommunications*, vol. 38, no. 1, pp. 61–66, 2015.

[11] D. Heckerman, "Bayesian networks for data mining," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 79–119, 1997.

[12] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.

[13] IBRL, "Intel Lab Data[EB/OL]," 2004 http://db.lcs.mit.edu/labdata/labdata.html.

Hindawi
Journal of Electrical and Computer Engineering
Volume 2017, Article ID 1360289, 11 pages
https://doi.org/10.1155/2017/1360289

Research Article

Health Monitoring System for Nursing Homes with Lightweight Security and Privacy Protection

Yu'e Jiang[1,2] and Jiaxiang Liu[2]

[1]*School of Computer and Information and the University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246011, China*
[2]*School of Computer and Information, Anqing Normal University, Anqing 246011, China*

Correspondence should be addressed to Yu'e Jiang; wsxiaoe@163.com

Received 9 November 2016; Revised 21 December 2016; Accepted 23 January 2017; Published 7 March 2017

Academic Editor: Liangmin Wang

With the rapid growth of aged population in China, it is urgent to design a safe and effective monitoring system for the nursing homes. An optimized scheme and high performance security and privacy protection for monitoring system have already become the focus studied especially. So this paper proposed a health monitoring system with lightweight security and privacy protection for nursing homes. Dual-band RFID, virtual routing location algorithm, and diet and exercise data collection based on RFID were adopted to obtain the location and health information. And that fused a mobile authentication protocol based on Hash function to realize security access and privacy protection, which can improve security and reduce the complexity of calculation and the implementation cost compared with the typical authentication protocols. The experiment results show that the ratio of relative network delay is below 35%. The system has strong real-time, high security, more comprehensive data, and lower cost of computation and communication. It can satisfy the requirements of health monitoring for nursing homes.

1. Introduction

It is estimated that China will enter the aging society. How to provide more comprehensive pension services becomes more and more urgent. As families and community can only provide limited elderly services, a tendency to meet the rocketing demands is to promote healthy monitoring system for nursing homes. As environment in nursing homes is complex, managers cannot focus on every one of the elderly people. So the challenge is to provide location service and health service in the event of dangerous conditions [1–5]. With the rapid development of monitoring technology, RFID get more attention due to its advantage, such as unique identification, moveable identification, multitargets identification, and good environmental adaptability. In a word, a safe and effective monitoring system based on RFID for the nursing homes will be required urgently.

Existing monitoring systems based on RFID usually lack an optimized scheme of data collection and privacy protection. For example, in the aspect of data collection, some are operating in a single frequency band, transmitting collection-data via line, even short of effective location algorithm and health data collection. As we known, there are obvious disadvantages in wireline monitoring system, such as routing restriction, the lack of flexibility, and high-cost. In terms of operation of frequency, RFID systems can be divided into four categories: low frequency (LF), high frequency (HF), ultra-frequency (UHF), and microwave, which have different properties. Due to low power, strong penetrating in RFID LF system, moveable identification, and multitargets identification in other RFID systems, how to realize the combination of the two advantages is a worthy topic [6–10]. Findings indicated that healthy diet and sufficient and regular exercises not only contribute to reduce and resist chronic diseases, but also promote physical and psychological health of elderly people. So to establish a system that can gather diet and exercise information and do some

statistical analysis for residents in nursing homes is of great signification. Manual entry and a survey are applied for data collection in traditional method, which still have some problems and drawbacks: low-efficiency, poor data quality, and less data quantity [11–14]. In another aspect of privacy protection, security and privacy threats exist in the monitoring system for nursing homes based on RFID, in which personal information, recording information in daily life, and personal property information are prone to attacks such as replay attacks, counterfeit attacks, and tracking attacks. It is urgent to establish policies and standards for avoiding security issues, especially the encryption algorithm based on Hash function. Presently, there are Hash-Lock protocol, Randomized Hash-Lock protocol, and Hash-Chain protocol. However these have some flaws and fail to solve the security problems [15–20].

This paper adopts a RFID system with dual-band, in which low frequency is used to activate the passive tag and the HF is used for communicating. A simple algorithm named virtual routing location algorithm is introduced to realize area location. It effectively uses RFID and ZigBee technology to establish an expanded RFID wireless network for collecting location data. A method of diet and exercise data acquirement based on RFID technology is also proposed later. At the same time, the system effectively integrates a lightweight RFID authentication protocol based on Hash function to achieve secure access and privacy protection. This protocol is effective against counterfeiting, tracking, and replay attacks and realizes the interaction between tag, reader, and back-end server. These works show that the new system overcomes the security leaks and has the merits of an optimized scheme of data collection, low communication and computation complexity, and so on.

The rest of this paper can be outlined as follows. Section 2 describes the architecture of the system, along with main modules, which is divided into vertical and horizontal aspects: function module and security and privacy protection module. Section 3 discusses the key technology in the function module, such as area location and data collection based on RFID. Section 4 designs a kind of lightweight RFID authentication protocol based on Hash function from the vertical aspect. Section 5 sets up experimentation environment and the network latency is tested, along with related work. Finally, in Section 6 we summarize our discussion.

2. System Description

The section mainly focuses on two aspects as follows: on the one hand, the architecture of this system is proposed, and the frequency selectivity of the RFID system, workflow, and communication mode are described in detail. On the other hand, the more detailed function requirements are analyzed from vertical and horizontal aspects. In the vertical aspect, the system is divided into monitoring subsystem and service subsystem. Security and privacy are discussed from the horizontal aspect. Security mode is embedded into every subsystem, such that a unified secure design makes sure that the monitoring system runs well.

2.1. Architecture. This system conforms to the IOT concept, which contains three layers, the perceptual layer, the network layer, and the application layer [3, 7]. The perceptual layer is at the most front-end of information collection and includes location nodes, other data collection nodes, sticking tags, and RFID wristbands. Then the information collected is uploaded to the back-end serve by ZigBee network with tree topology in the network layer. Finally, the system uses .NET+MYSQL technologies to realize the development of the whole system, including monitoring subsystem, service subsystem, and web service. Its design architecture is shown in Figure 1.

According to different requirements, we choose two kinds of nodes for data collection. Location node is designed to collect the location information, and the collection node is for another data collection. Each node communicates with CC2530 module by RS485 bus. Then ZigBee network formed by CC2530 modules will transmit the whole data to the back-end server.

Every location node and RFID wristbands work together in two phases: activation phase and communication phase. 125 KHz band is selected to activate passive part in a RFID wristband, and 433 MHz band is for communication between one node and the active part of a wristband. Compared with a single frequency band, the Dual-band RFID system combines the advantage of low frequency and high frequency effectively: low power and strong penetrating in LF system, moveable identification, and multitargets identification in HF system. According to the structure of antenna in a location node, it can also be divided into single-channel mode and dual-channel mode, which suit for different areas to be located. Every collection node mainly communicates with sticking tags, which follows the EPC standards. Desktop RFID reader with 433 MHz frequency band is selected for collection node to collect diet data, part of exercise data, and other service data.

In the system, tags or wristbands, any one of RFID reader, and back-end service are connected with each other by wireless network: RFID or ZigBee. This quite suits the monitoring system, but it also brings a problem. The wireless channel is prone to attacks such as replay attack, counterfeit attack, and tracking attack. So security and privacy should be paid adequate attention.

2.2. Function Module. The monitoring system mainly contains monitoring subsystem and service subsystem according to the actual function from the vertical aspect. Each subsystem combines with RFID wristbands to realize all-around automatically monitoring management. Security and privacy are designed to resist attacks from the horizontal aspect. The detailed functions are shown in Figure 2.

Monitoring subsystem offers a safe and secure environment for the nursing homes residents. All residents in the nursing homes wear RFID wristbands that can help the staff monitor their locations. So, when dangers occur, the staff can quickly locate and provide appropriate help [5–7]. The specific function is analyzed as follows.

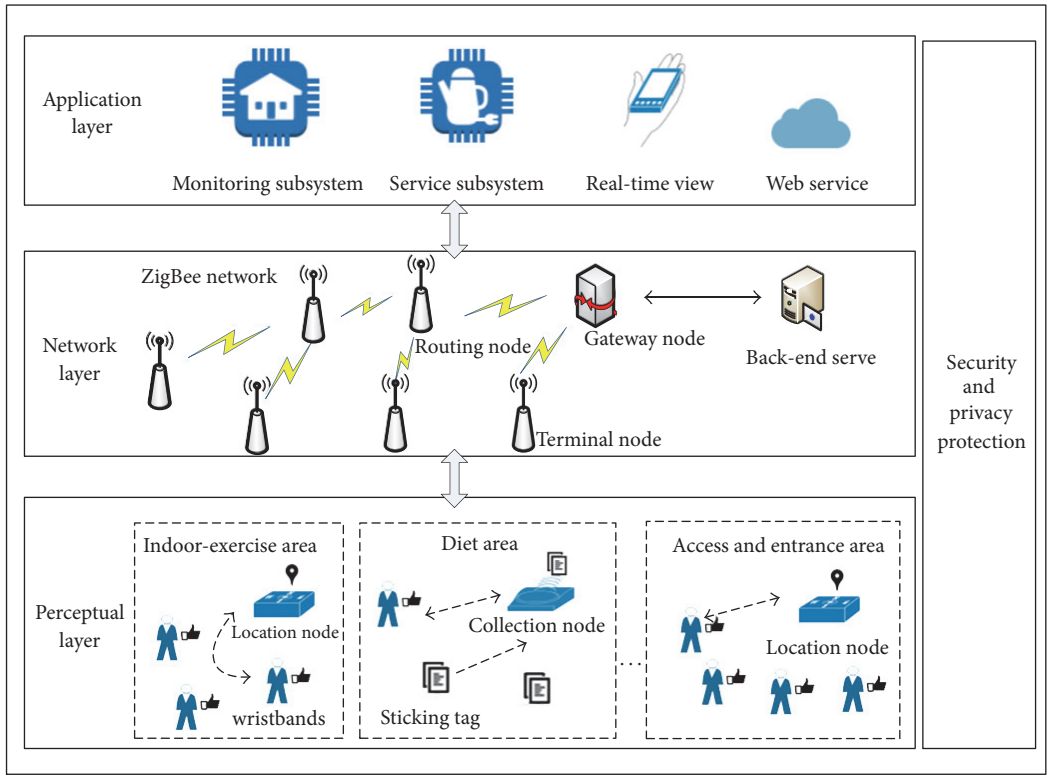


Figure 1: The overall architecture.

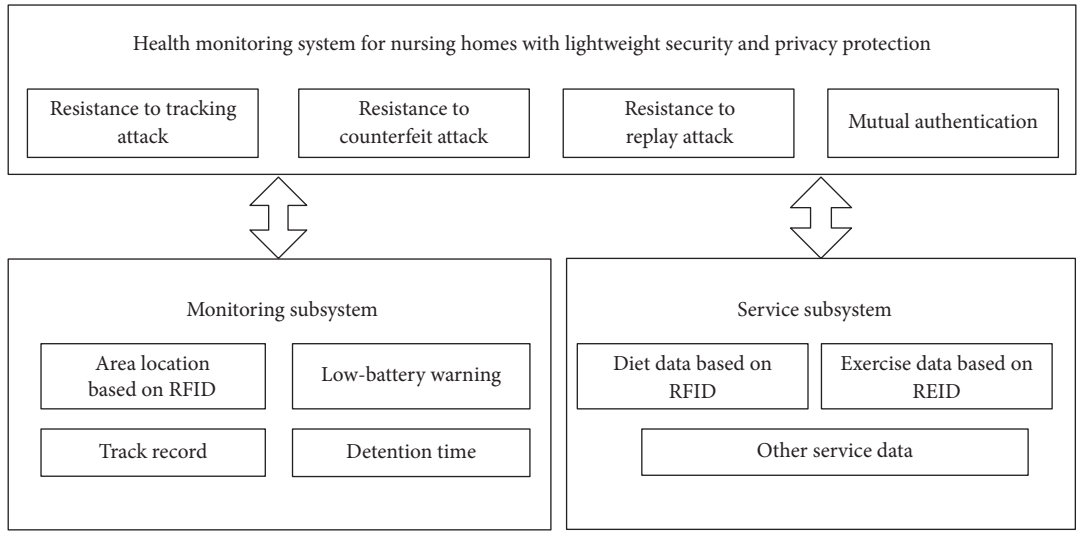


Figure 2: Function diagram.

(1) Area location: this module is mainly used for location Query for the residents wearing RFID wristbands. Or use the module to achieve a certain period of time statistics.

(2) Low-battery warning: the module can monitor the consuming situation of battery for the RFID wristbands. A warning signal is transmitted to the staff when the battery capacity is below the threshold.

(3) Detection time: calculate the detection time in some dangerous and privacy areas, such as bathrooms and toilets. For example, a resident wearing a RFID wristband stays in these areas longer than normal; it is very likely that an emergency has happened.

(4) Track record: record the daily activities of the residents wearing the RFID wristbands, and then provide the information for health assessment or other requirements.

Service subsystem is designed to offer health service and others and collect diet and exercise data which partly reflects the health of residents in the nursing homes. Concrete analysis is as follows.

(1) Diet data based on RFID: the diet data for every resident in the nursing homes are collected by the collection node, tricking tags, and RFID wristband. The back-end serve will record the data and compare with the standard.

(2) Exercise data based on RFID: the exercise data comes from three kinds of activities: the usage of athletic facilities, the usage of indoor recreational area, and outdoor-activities. All the exercise data and diet data are supported to help professionals to access health status for every resident in nursing homes.

(3) Other service data: the module is for other data services: querying relational information for every resident, sending blessing on holiday, reminding the resident to take medicine, and so on.

Security and privacy are designed to defend all the subsystems from threats. It can be detailed as follows.

(1) Resistance to Counterfeit Attacks. It prevents attackers from counterfeit RFID tags or readers, illegal access to personal information of the resident, guardian information and sensitive data, and so on.

(2) Resistance to Tracking Attacks. It prevents attackers from gaining traces of activity by tracking location information, threatening their person and property.

(3) Resistance to Replay Attacks. In an extreme case, the attacker may obtain relevant information. It prevents it from being reproduced by using this information, thereby illegally passing the authentication

(4) Mutual Authentication. It aims to achieve the tag, reader, and the back-end server mutual authentication between the three.

Security design will be embedded above the various subfunctional modules and unified security design to ensure that all functional modules are safe. The next step will be a detailed analysis of key technologies and the authentication protocol in the security module.

3. Key Technologies

As mentioned, the system is divided into monitoring subsystem and service subsystem from the vertical aspect. The area location module is very important and supports the other modules. And that diet data and exercise data are the important parts of service subsystem. So in this section we will focus on the key technologies on these modules: area location, diet data, and exercise data collection.

3.1. Area Location Based on RFID. According to the structure characteristics of the monitoring area in nursing homes, it can be divided into different areas, such as indoor gymnasium area, diet area, indoor-recreation area, and access and entrance area. The system is designed to provide a more flexible, easily configurable deployment model. So we can deploy the location node according to actual size of areas to be monitored and other requirements. The actual physical location is known when the location node has been deployed. In this section we will talk about the process of data acquisition of one location node and the location algorithm.

The location nodes mainly work together with RFID wristbands; then the workflow of one single location node and RFID wristband is shown in Figure 3.

The process of data acquisition consists of two phases: namely, activate phase and communication phase. In the trigger phase, the location node sets control parameters and initiates readers at first. Then the RFID reader (location node) will scan RFID tags (RFID wristbands) in its coverage area. Finally, when the searched tag ID matches the ID stored in memory, the communication can begin. In the communication phase, the RFID reader starts to receive data after authenticating successfully. If the data is a distress signal or warning signal, then give priority to transmit; otherwise transmit the location data in the order queue.

The location data mainly contains location array, which comes from the virtual routing location algorithm. In the following parts, we will simply describe the basic concepts about the algorithm. The algorithm uses ZigBee and RFID technology to form the RFID wireless network; the coverage is far away. As we known, the read-write distance of RFID reader is relatively close. So we assume that the tag to be located has the same physical location with the RFID reader. As the resident wearing RFID wristband moves in the RFID wireless network, the wristband will transmit a location array to back-end server. The principle of the algorithm is shown in Figure 4.

Assuming that the coordinates of all readers are already known, then the coordinate of the tag can be gotten by the algorithm. As shown in Figure 4, the solid line is the actual route the tag selected and the dotted line is the virtual route, which is calculated by the algorithm. For example, the actual route of the tag can be the same as the virtual route as shown below:

$$(0,0) \longrightarrow (1,1) \longrightarrow (1,2) \longrightarrow (1,3) \longrightarrow (1,4) \longrightarrow (2,4) \longrightarrow (3,3) \longrightarrow (4,2) \longrightarrow (3,1) \longrightarrow (3,0) \quad (1)$$

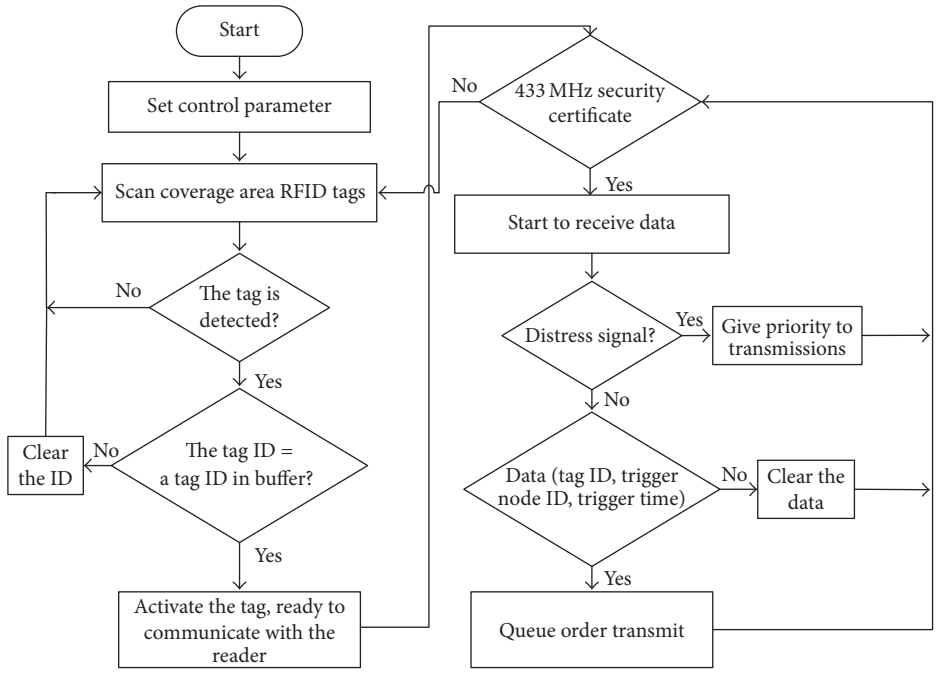


Figure 3: Workflow for one location node.

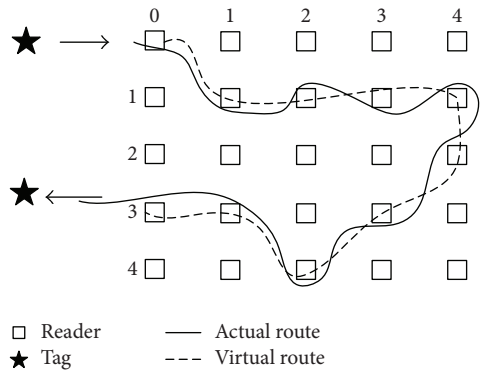


Figure 4: Schematic diagram of the algorithm.

A location array is chosen to realize the progress of the algorithm. It contains three parameters: the ID of the tag, the reading time, and the ID of the reader. The following formula can be used.

$$\left\langle T_i, t_j, R_k \right\rangle = \left\langle \text{tag } j, \text{time } j, \text{reader } k \right\rangle. \quad (2)$$

As shown is formula (2), the location array is expressed by $\langle T_i, t_j, R_k \rangle$; T_i is the ID of the ith tag. R_k is the ID of the kth reader. And t_j is the time when the kth RFID reader starts to communicate with the ith tag. Location arrays are firstly categorized according to T_i and then according to t_j. Finally, we can get the coordinate of kth RFID reader as the coordinate of ith tag at the time of t_j. The algorithm is simple and effective, fully in line with the needs of area location for the nursing homes.

3.2. Health Data Based on RFID. In this section we will build a module to collect diet and exercise data automatically based on RFID and do some statistical analysis. Diet and exercise data is usually very complex, but there still has certain regularity and periodicity for collective life environment in nursing homes. Based on this application background, a simplified model for diet and exercise data collection is designed as follows.

Diet area is divided into selection-area and settlement-area. In selection-area, plates and bowls sticking with tags establish the one-to-one relationship between each tag and the food in its plate or bowl. The settlement-area mainly contains RFID readers and displays. With the help of sticking tags, RFID readers, and wristbands, the subsystem can realize settlement and diet data collection quickly and automatically. Next, we will introduce the progress of data analysis in the

Table 1: Exercise data based on RFID.

Categories	Indoor-entertainment	Outdoor-activities	Athletic facilities
Methods	Accumulate the continuous time by indoor-RFID readers	Accumulate the continuous time by entrance-RFID readers	Accumulate the exercise time on athletic facilities by RFID readers
Time (hours/day)	T_1	T_2	T_3
The weight	w_1	w_2	w_3

back-end server. Five categories of the essential nutrients in body are chosen as the main parameters to analyze, which are proteins, fats, carbohydrates and trace elements, vitamins, and minerals. Based on a nutrient criterion proposed by China and the special requirements for the elderly, a daily meal nutrition supplement standards are designed. Compared with the standards, the system will judge whether the daily diet is reasonable and then put forward some suggestions. Finally, a mathematical model is built to detail the progress. Let the amount of common food in nursing homes as i. k take 1 to 5, respectively, as protein, fat, vitamins, energy, minerals, and inorganic salts. a_{ki} is the value of kth nutrient in ith food, and $\{a_{ik}\} \in A$; u_{ik} is the value of daily intake of kth nutrient in ith food, and $\{u_{ik}\} \in B$. $u_{ik} = 1$; the formula indicates that the kth nutrient contained in the ith food is ingested. The daily intake of nutrients is shown in formula (3).

$$Y = A * U = \begin{pmatrix} a_{11} & \cdots & a_{1i} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{ki} \end{pmatrix} \begin{pmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{i1} & \cdots & u_{ik} \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1i} \\ \vdots & \ddots & \vdots \\ y_{k1} & \cdots & y_{ki} \end{pmatrix}. \quad (3)$$

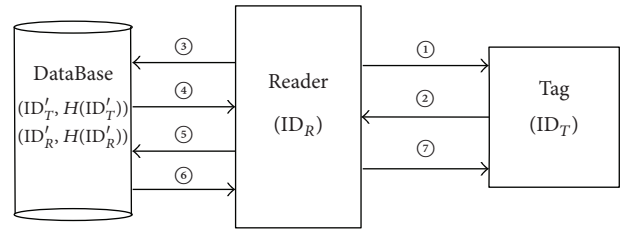
Y is the daily intake of nutrients, y_{ki} express the kth nutrient intake from the ith food. Add the elements in the same column for array Y, and get a new array $V = (v_1, v_2, \ldots, v_k)$; v_k is the kth nutrient you has been absorbed. $M = (m_1, m_2, \ldots, m_k)$ is the standard value of nutrients. Comparing the V with M, a result that whether you get the sufficient nutrients or not will be offered to the residents and staff.

The system classifies exercises as indoor-entertainment, outdoor-activities, and exercise on athletic facilities. Formula (4) as below is used to calculate the amount of exercises daily simplify.

$$S = \sum_{i=1}^{n} (T_i * w_i). \quad (4)$$

As shown in formula (4), $n = 3$, $i = \{1, 2, 3\}$ represent indoor-entertainment, outdoor-activities, and exercise on athletic facilities, respectively. w_i is the weight value of the amount of the whole exercises. The sum amount of ith exercises daily is T_i, and the sum of all kinds of exercise is calculated by formula (4), noted by S. The specific progress of calculation is shown in Table 1.

According to the health status of the elderly and special needs, daily exercise standards are designed when the elderly checked in the nursing home. And the standard for comparison, to determine the amount of the daily exercise, is appropriate or not.



① Reader ⟶ Tag: Query, r_R
② Tag ⟶ DataBase: M_1, M_2, M_3, r_T
③ Reader ⟶ DataBase: Query
④ DataBase ⟶ Reader: r_D
⑤ Reader ⟶ DataBase: M_2, M_3, M_4, r_T, r_R
⑥ DataBase ⟶ Reader: M'_3, M'_4
⑦ Reader ⟶ Tag: M'_3

Figure 5: The authentication flow of the protocol.

4. Security and Privacy

As before, security and privacy are very important for the wireless system, and security mode is embedded into every subsystem. So a mobile authentication protocol based on Hash function is designed in this section. The procedure and implementation of the protocol are discussed as follows.

4.1. Proposed Protocol. Based on a one-way Hash function, this paper proposed a mutual authentication for information protection. It is depicted in Figure 5. The symbols in the protocol are described as follows: $H(x)$ is the Hash function of x. ID_T is the identification number of the tag and is stored in the tag. ID_R is the identification number of the RFID reader and is stored in the reader. $(ID'_R, H(ID'_R))$ and $(ID'_T, H(ID'_T))$ are stored in the DataBase. The authentication flow of the protocol is shown in Figure 5.

(1) The RFID reader generates a random number r_R and sends $(Query, r_R)$ to the tag.

Table 2: Protocol performance analysis.

Performance actor	Hash-Lock	Radom Hash-Lock	Hash-Chain	The proposed protocol
Computational complexity				
Tag	$1H$	$1H + 1R$	$2H$	$3H + 1R$
Reader	—	$(\sum(n/2))\,H$	—	$4H + 1R$
DataBase	—	—	$(\sum(n/2))\,H$	$4H + 1R$
Security performance				
Resistance to counterfeiting attacks	×	×	×	√
Resistance to tracking attacks	×	×	√	√
Resistance to replay attacks	×	×	×	√
Two-way authentication	×	×	√	√

(2) The tag receives the data, generates a random number r_T, and then uses the received number and calculates $M_1 = H(r_R \parallel r_T)$, $M_2 = H(\mathrm{ID}_T) \oplus M_1$, and $M_3 = H(\mathrm{ID}_T \parallel r_R \parallel r_T)$. M_3 is stored in the tag and (M_1, M_2, M_3, r_T) are sent to the reader.

(3) Using the received random number r_T and its generated random number r_R, the reader calculates $M_1' = H(r_R \parallel r_T)$. Then it makes a judge whether M_1' is equal to the received variable M_1. if $M_1' = M_1$, the tag is authenticated. And then Query is sent to the DataBase.

(4) After receiving the Query, the DataBase generates a random number r_D and sends it to the reader.

(5) Using the received r_D and its own numbers r_R and ID_R, the reader calculates the following numbers: $M_4 = H(\mathrm{ID}_R) \oplus H(r_R \parallel r_D)$ and $M_5 = H(\mathrm{ID}_R \parallel r_R \parallel r_D)$. M_5 is stored in the reader and $(M_2, M_3, M_4, r_T, r_R)$ are sent to the DataBase.

(6) When the DataBase receives the data, it will carry on the following three steps. The first step is to authenticate the reader: it calculates $H''(\mathrm{ID}_R) = M_4 \oplus H(r_R \parallel r_D)$ to meet the requirement of $(\mathrm{ID}_R', H(\mathrm{ID}_R'))$, which are stored in the DataBase. And if $H''(\mathrm{ID}_R) = H(\mathrm{ID}_R')$, the reader is authenticated. In the second step, it calculates $H''(\mathrm{ID}_T) = M_2 \oplus H(r_R \parallel r_T)$. If $H''(\mathrm{ID}_T) = H(\mathrm{ID}_T')$, the tag is authenticated. The third step calculates $M_5' = H(\mathrm{ID}_R' \parallel r_D \parallel r_R)$ and $M_3' = H(\mathrm{ID}_T' \parallel r_T \parallel r_R)$ and sends (M_3', M_5') to the reader.

(7) The reader compares the received data M_5' with the data M_5, which is stored earlier. If $M_5' = M_5$, then the reader authenticates the DataBase and sends M_3' to the tag.

(8) The tag receives the data M_3' and compares it to the data M_3 which is stored in tag earlier. If $M_3' = M_3$, then the tag authenticates both the Reader and the DataBase.

4.2. Protocol Performance Analysis. The following will analyze security performances of the proposed protocol from four aspects.

(1) Resistance to Counterfeiting Attacks. The protocol can effectively exploit the one-way of Hash function. The attackers cannot analyze the identification numbers of the tag or the reader by intercepting data. So the system has the ability to resist counterfeiting attacks.

(2) Resistance to Tracking Attacks. The tag, the reader, and the DataBase will generate random numbers; the response data are changing in each certification process. So attackers are unable to obtain location information, thus avoiding tracking attacks.

(3) Resistance to Replay Attacks. The random numbers of the tag, the reader, and the DataBase are changed during each authentication process, so that the previous authentication information cannot be used to complete the replay attacks.

(4) Two-Way Authentication. Firstly, the reader authenticates the tag by judging whether $M_1' = M_1$. Then the DateBase verifies the security of the tag and the reader by the received $(M_2, M_3, M_4, r_T, r_R)$. Finally, the reader authenticates the DataBase by the formula $M_5' = M_5$. And the tag verifies that the formula $M_3' = M_3$ is true and authenticates the Reader and the DataBase.

On the basis of security considerations, the protocol proposed also effectively reduces the calculation of the tag, reducing tag costs and the use of energy consumption, as shown in Table 2. In Table 2, H is the Hash function, R is the random function, n is the total number of tags, and "—" is not Hash function, random function, and so on [18–20].

As shown in Table 2, this authentication protocol fully considers the new problems brought by the wireless transmission of the mobile RFID system compared with the classical protocol. The proposed protocol balances the computational protocol at the tag, the reader, and the DataBase while providing strong security which can resist various types of attacks, obtain dual-authentication, and decrease computational complexity. All these performances make the proposed the system be appropriate for the nursing homes usage.

5. Performance Testing and Related Work

The following will analyze the performance and related work to verify the feasibility and superiority of the system. Firstly,

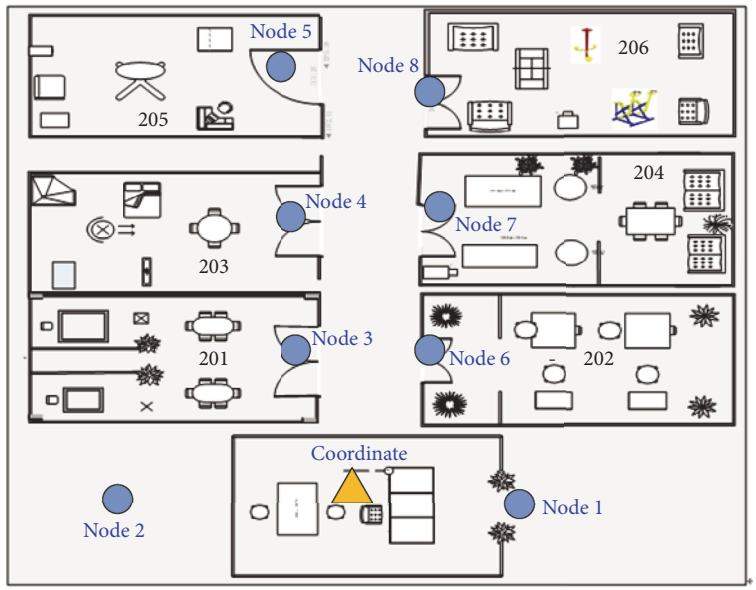


Figure 6: Location nodes layout diagram.

we will build the test environment, testing the system's network delay characteristics and then analyze the related work of the system and compare it with the common monitoring system.

5.1. Performance Testing. Because the indoor layout of the nursing home is similar to the laboratory, we choose the second floor laboratory to carry out system testing. The monitoring rooms are divided into 201 room, 202 room, 203 room, 204 room, 205 room, 206 room, the lobby, and server room, as shown in Figure 6.

As shown in Figure 6, the RFID node 1 and node 2 are located in the lobby area. The node 3 is arranged at the entrance of the room 201. Moreover, node 4 is located at the 203 entrance of the room 203, node 5 at the room 205, node 8 at the room 206, node 7 at the room 204, and node 6 at the room 202. The coordinate node is arranged at the server room and connected with the back-end server. The distance between the location nodes is 6–8 meters, and the nodes that are far away from the server communicate with their nearest nodes as the parent node. We use packet sniffer and hardware timers to test the system. Assume that the data acquisition interval is 5 s and 100 times for each node during the test. The average network delay for each node is shown in Figure 7.

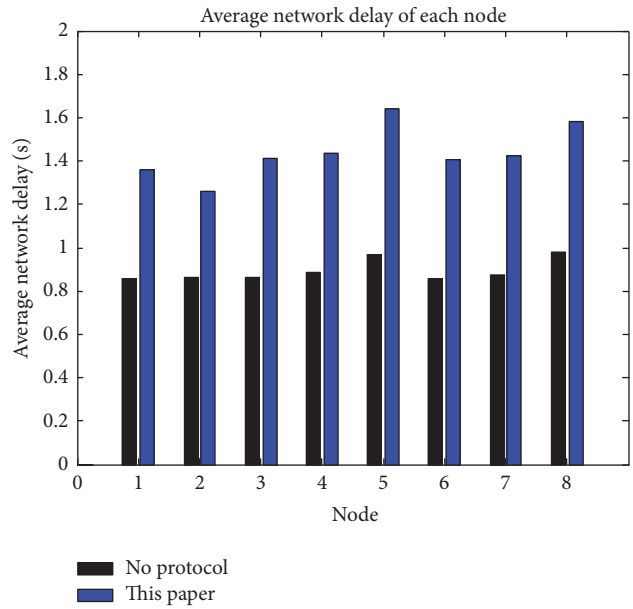


Figure 7: The average network delay of each node.

From Figure 7, we can see that the network latency of nodes 1, 2, 3, and 6 is close to each other. In the network topology, these nodes are single-hop nodes, and the actual physical location is closer to the coordinator node. Node 4 and node 7 are close to each other. In the network topology, they are two-hop nodes. Node 5 and node 8 are three-hop nodes and the network delay is relatively large. Therefore, the system topology design has some impact on the network delay, but overall, adding a security authentication protocol increases the system's average network latency. [19, 20] and the simulation show that the implementation of the protocol time is about 500–550 ms, but many monitoring systems data acquisition time is about 5–60 s. So here 5 s is chosen as the sampling interval; we discuss the relative network latency and the average network delay divided by the sampling interval, as shown in Figure 8.

Figure 8 shows that, compared with the sampling interval, the relative network delay for the proposed system with lightweight security and privacy protocol is less than 35%, so the authentication protocol does not affect the system real-time data collection but can greatly improve system security and privacy.

Table 3: Week of diet data from a tester.

Data/appliance	Week of diet data from a tester				
	Protein (g)	Fat (g)	Vitamins (mg)	Minerals and inorganic salts (mg)	Energy (KJ)
02/10/2016 (Sun)	73.1	23.7	81.32	5828	6307.5
03/10/2016 (Mon)	76.1	45.5	130.75	6273.33	6953.81
04/10/2106 (Tue)	95.2	60.3	86.3	7832.85	7200.67
05/10/2016 (Wed)	107.5	71.8	84.08	7409.73	9790.56
06/10/2016 (Thu)	87.1	45	89.35	6899.35	7284.34
07/10/2016 (Fri)	92.8	63.7	113.55	7541.83	8037.46
08/10/2016 (Sat)	66.1	29.6	46.36	5282.9	5514.51
Mean	85.41	48.51	90.24	6724	7298.4
Reference value	65~75	25	130.5~131.4	6530	7100~9200
Conclusion	High	High	Low	High	Normal

Table 4: Week of exercise data from a tester.

Data/appliance	Week of exercise data from a tester				
	Indoor-entertainment ($w_1 = 0.2$)	Outdoor-activities ($w_2 = 0.35$)	Athletic facilities ($w_3 = 0.45$)	Sum (hour)	Reference value (hour)
02/10/2016 (Sun)	2 hours	0.5 hour	0.3 hour	0.71	0.6~1
03/10/2016 (Mon)	3 hours	0.8 hour	0.2 hour	0.97	
04/10/2106 (Tue)	2.5 hours	1 hour	0.2 hour	0.94	
05/10/2016 (Wed)	4 hours	0.1 hour	0 hours	0.835	
06/10/2016 (Thu)	3.6 hours	0.3 hour	0.3 hour	0.96	
07/10/2016 (Fri)	2.8 hours	0.4 hour	0.1 hour	0.745	
08/10/2016 (Sat)	3.6 hours	0.3 hour	0.3 hour	0.96	

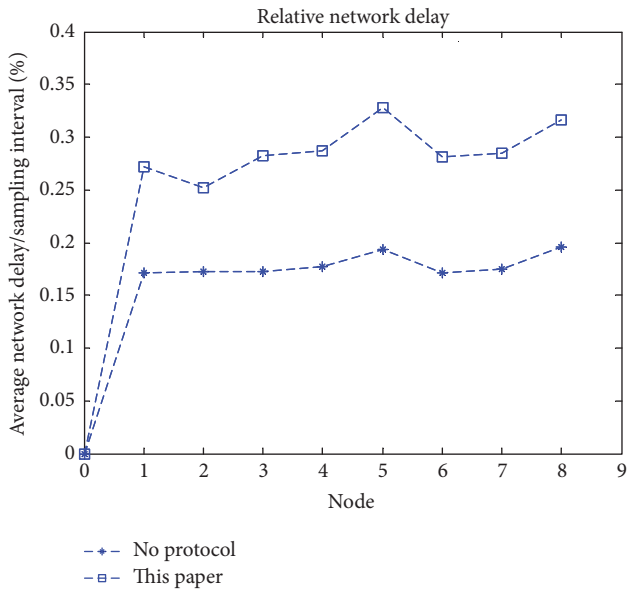


Figure 8: Relative network delay.

In the experimental environment, we collected a week of diet and exercise data from a tester, as shown in Tables 3 and 4.

In Table 3 the reference value comes from daily intake of dietary nutrient values for Chinese residents aged 60 years and over. The nursing home provides a basic diet menu similar to the one provided each week, which facilitates the collection of diet data. In Table 4 the reference value can be adjusted according to the previous exercise and health status. We can see that week of diet and exercise data have certain regularity and periodicity to reflect the elderly's diet and exercise status, which can monitor the health of the elderly.

5.2. Related Work. There are a lot of monitoring systems, but few ones are in line with the needs of nursing homes in all aspects. Reference [21] introduces a remote monitoring system for the tower clusters, which focuses on industrial data transmission and is not suitable for the elderly monitoring. In [22], a remote monitoring system based on intelligent fiber structure is proposed. The system uses the liquid core optical fiber structure based on ARM and GPRS to communicate. The cost is high and cannot be applied to the nursing home monitoring system. References [23, 25] all use video for data acquisition. Reference [23] uses the pyroelectric infrared sensor and the video monitor to carry on the multitarget tracking. But its calculation and communication complexity are high, and the data acquisition is not comprehensive. In [24], a wireless network life-monitoring system for the nursing home is proposed, which uses the wireless sensor

Table 5: Comparison of related system performance.

Monitoring systems	System performance					Main technical shortcomings
	Real-time	Data comprehensiveness	Security and privacy	Computational and communication costs	Suitable for nursing homes	
Reference [21]	Medium	Industrial data	Low	medium	No	Lack of security
Reference [22]	Medium	Industrial data	Low	High	No	Communication costs are high
Reference [23]	High	Location data	Medium	High	Yes	Lack of security Computational costs are high Data are not comprehensive
Reference [24]	Medium	Health data	Low	Low	Yes	Data are not comprehensive
The system proposed	High	Comprehensive data	High	Low	Yes	No

network to collect the basic health data but lacks the consideration of security and privacy protection. Specific analysis and comparison are shown in Table 5.

Compared with the monitoring system shown in Table 5, the system has four characteristics: high real-time monitoring, small network delay; comprehensive data collection, involving location information, diet, and exercise data collection; security and privacy protection mechanism embedded in the module design; the use of lightweight mobile security architecture, computing, and low communication cost.

6. Conclusion

According to the characteristics and needs of the nursing home, this paper designed a health monitoring system with lightweight security and privacy protection, which is focused on vertical and horizontal aspects: health data collection and security and privacy protection. From the vertical aspect, RFID dual-frequency band, virtual route location algorithm, and diet and exercise data acquisition based on RFID are adopted in the health data collection. From the horizontal aspect, a lightweight RFID authentication protocol based on Hash function is embedded into each collection module, which has high security and low computation cost. Through the performance analysis and testing, we can see that the system has characteristics of high security, high real-time, and high data comprehensive and low computational and communication complexity and fully meets the needs of health monitoring system for the nursing homes.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by Educational Commission of Anhui Province under Grant 2015jyxm237 and in part by the Foundation of University Research and Innovation Platform Team for Intelligent Perception and Computing of Anhui Province.

References

[1] M. H. Y. Shum, V. W. Q. Lou, K. Z. J. He, C. C. H. Chen, and J. Wang, "The 'Leap Forward' in nursing home development in Urban China: future policy directions," *Journal of the American Medical Directors Association*, vol. 16, no. 9, pp. 784–789, 2015.

[2] N. Kumar, K. Kaur, S. C. Misra, and R. Iqbal, "An intelligent RFID-enabled authentication scheme for healthcare applications in vehicular mobile cloud," *Peer-to-Peer Networking and Applications*, vol. 9, no. 5, pp. 824–840, 2016.

[3] N. K. Suryadevara and S. C. Mukhopadhyay, "Wireless sensor network based home monitoring system for wellness determination of elderly," *IEEE Sensors Journal*, vol. 12, no. 6, pp. 1965–1972, 2012.

[4] J. E. Morley, G. Caplan, M. Cesari et al., "International survey of nursing home research priorities," *Journal of the American Medical Directors Association*, vol. 15, no. 5, pp. 309–312, 2014.

[5] J. E. Morley, "Under nutrition: a major problem in nursing homes," *Synthetic Communications*, vol. 44, no. 8, pp. 1019–1042, 2014.

[6] H. Makimura, K. Watanabe, H. Igarashi, and H. Waki, "Monitoring system of railway using passive RFID in UHF band," *IEEJ Transactions on Electronics, Information and Systems*, vol. 132, no. 5, pp. 691–696, 2012.

[7] L. M. Wang and S. M. Xiong, *Introduction to the Internet of Things Engineering*, Tsinghua University Press, Beijing, China, 2011.

[8] J. C. Chen and T. J. Collins, "Creation of a RFID based real time tracking (R-RTT) system for small healthcare clinics," *Journal of Medical Systems*, vol. 36, no. 6, pp. 3851–3860, 2012.

[9] N. Li and B. Becerik-Gerber, "Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment," *Advanced Engineering Informatics*, vol. 25, no. 3, pp. 535–546, 2011.

[10] K. Liu and Z. C. Ji, "Research on RFID reader network tracking algorithm," *Computer Engineering*, vol. 38, no. 18, pp. 248–250, 2012.

[11] M. Arebey, M. A. Hannan, H. Basri, R. A. Begum, and H. Abdullah, "Solid waste monitoring system integration based on RFID, GPS and camera," in *Proceedings of the International Conference on Intelligent and Advanced Systems (ICIAS '10)*, pp. 1–5, June 2010.

[12] Y. Tao, X. Zhou, Y. Ma, and F. Zhao, "Mobile mutual authentication protocol based on hash function," *Journal of Computer Applications*, vol. 36, no. 3, pp. 657–660, 2016.

[13] S.-J. Zhou, W.-Q. Zhang, and J.-Q. Luo, "Survey of privacy of radio frequency identification technology," *Journal of Software*, vol. 26, no. 4, pp. 960–976, 2015.

[14] A. Wickramasinghe, D. C. Ranasinghe, C. Fumeaux, K. D. Hill, and R. Visvanathan, "Sequence learning with passive RFID sensors for real time bed-egress recognition in older people," *IEEE Journal of Biomedical and Health Informatics*, p. 1, 2016.

[15] B. Fabian, T. Ermakova, and C. Muller, "SHARDIS: a privacy-enhanced discovery service for RFID-based product information," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 3, pp. 707–718, 2012.

[16] X. L. Liu, H. Qi, Q. K. Li et al., "Efficient detection of cloned attacks for large-scale RFID systems," in *Algorithms and Architectures for Parallel Processing*, vol. 8630 of *Lecture Notes in Computer Science*, pp. 85–99, Springer International Publishing, Berlin, Germany, 2014.

[17] A. Arbit, Y. Oren, and A. Wool, "A secure supply-chain RFID system that respects your privacy," *IEEE Pervasive Computing*, vol. 13, no. 2, pp. 52–60, 2014.

[18] H. Jannati and B. Bahrak, "Security analysis of an RFID tag search protocol," *Information Processing Letters*, vol. 116, no. 10, pp. 618–622, 2016.

[19] X. Y. Wang, F. X. Jing, and Y. Q. Wang, "An improved hash-based RFID security authentication algorithm," *Journal of Shangdong University*, vol. 49, no. 9, pp. 154–159, 2014.

[20] D.-Z. Sun and J.-D. Zhong, "A hash-based RFID security protocol for strong privacy protection," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1246–1252, 2012.

[21] H. M. Zheng, Y. J. Wang, K. Chen, and J. T. Zhang, "Design of wireless remote security monitoring system for tower crane fleet," *Journal of Electronic Measurement and Instrumentation*, vol. 28, no. 5, pp. 520–527, 2014.

[22] L. Shen, Z. Zhao, and X. Yu, "Design of remote monitoring internet of things system for new optical fiber smart structure," *Journal of Nanjing University of Aeronautics and Astronautics*, vol. 3, no. 47, pp. 453–458, 2015.

[23] F.-M. Li, N. Jiang, J. Xiong, and J.-Y. Zhang, "Multi-object tracking scheme with pyroelectric infrared sensor and video camera coordination," *Acta Electronica Sinica*, vol. 42, no. 4, pp. 672–678, 2014.

[24] Y.-J. Chang, C.-H. Chen, L.-F. Lin, R.-P. Han, W.-T. Huang, and G.-C. Lee, "Wireless sensor networks for vital signs monitoring: application in a nursing home," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 685107, 12 pages, 2012.

[25] G.-X. Han and C.-R. Li, "Improvement on moving object tracking method for network video surveillance," *Journal on Communications*, vol. 35, pp. 160–164, 2014.