

# Design of Abnormal Data Analysis and Processing System Based on RFID Supply Chain

Chunfeng Wang<sup>1</sup>, Shuiming Zhong<sup>2</sup> and Jin Wang<sup>2</sup>

<sup>1</sup>Modern Education Technology Center, Yancheng Institute of Technology, Yancheng 224051, China

<sup>2</sup>School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044, China

## Abstract

*With the deep application of RFID, the deployment of the RFID system is gradually to have the large-scale, networked, distributed development trend. The previous abnormal data analysis and processing way can not meet with the need of data processing. According to the survey of international famous consulting company (Gartner) for the "information overload", results show that more than 90% of the enterprises think the enterprise competitiveness is closely related to its information capacity. In order to obtain the useful information, we must analyze, process the data, and exclude "exception" data. So, it is very important to realize the detection of abnormal data. This paper presents the RFID supply an abnormal data analysis and processing system of RFID supply chain based on distance, rules and middleware technology. This system is divided into three levels: data processing layer, anomaly analysis lay and graphical display layer. Finally, through the performance analysis of the system, the effectiveness of the system has been verified.*

**Keywords:** RFID, anomaly detection, middleware technology, data processing

## 1. Introduction

In recent years, with the development of large-scale integrated circuits, radio frequency communication, information security technology, RFID technology has been widely used in logistics, manufacturing, transportation, medical, security, tracking, management and other public information service industry [1]. In supply chain management, RFID reader is arranged in a different place, and the purpose is to detect tagged objects in the supply chain flow. If it is found that the state or data of the object is abnormal, then it is can timely report the abnormal situation and makes timely treatment.

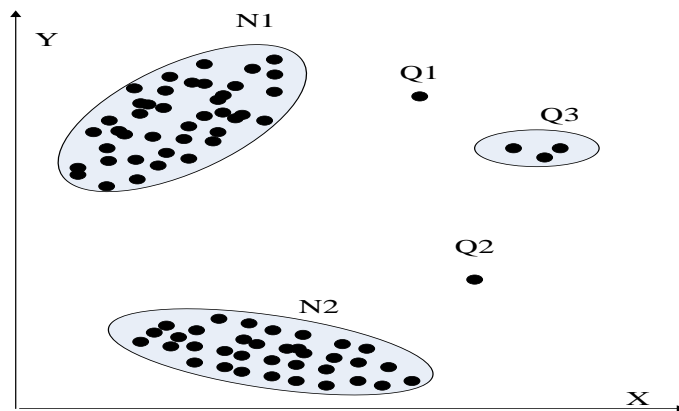
RFID data process is the problem which must be solved by the RFID application, and has been explored by a lot of domestic and foreign literature. To sum up, RFID data process mainly divides into two kinds: process of reading RFID data and process of submitting RFID data. Process of reading RFID data is mainly for RFID raw data [2]. Middleware is responsible for the process of this layer, and mainly implement the abnormal filtering of the raw data. Process of submitting RFID data is after the process of the middleware layer and it can find rules or abnormal situation by data mining or other means, then the final result will help management to generate a valid decision [3].

Supply chain management is one of the important application fields based on RFID. It is Significant to analyze and deal with the abnormal data of RFID the supply chain for security and reliability [4]. Based on the summary of the existing literature, this paper has researched and designed the abnormal Data Analysis and Processing System based on RFID Supply

Chain. Through the application of the system, the enterprise can obtain goods information of the supply chain flow in a more accurate, timelier and more detailed way, and can enhance the ability of enterprise daily management and the level of decision-making power.

## 2. Technology of Abnormal Data Analysis

The origin of the abnormal data usually lies in two aspects. On one hand, it dues to the err of the computer input, network data transmission, and human, etc. This type of abnormal must be removed or modified; otherwise it will affect the results of data analysis. On the other hand, the abnormal data may reflect the true nature, and it can obtain more information through analysis [5]. Abnormal example of simple two-dimensional data sets is shown in Figure 1.



**Figure 1. Abnormal Example of Simple Two-dimensional Data Sets**

The example of simple two-dimensional data sets has two normal data areas (N1 and N2). The area of O1, O2 and O3 are away from the normal areas, so they are abnormal.

At present, the research of abnormal data analysis technology has been attracted more and more attention from academic and industry. The direct abnormal data analysis method defines a normal boundary, and puts the other data beyond the boundaries to abnormal data range [6-7]. But the reason of the following respects makes this work become very challenging:

- Normal boundaries are hard to be defined, and it is difficult to make all possible normal behavior to be included in the normal boundary.
- Exceptions can not be accurately defined, and the abnormal definition in different application areas is different.
- The data itself contains "noise".
- Degree of confidence of Heterogeneous data sources is usually different.
- It is difficult to obtain the signature data for the training or validation.

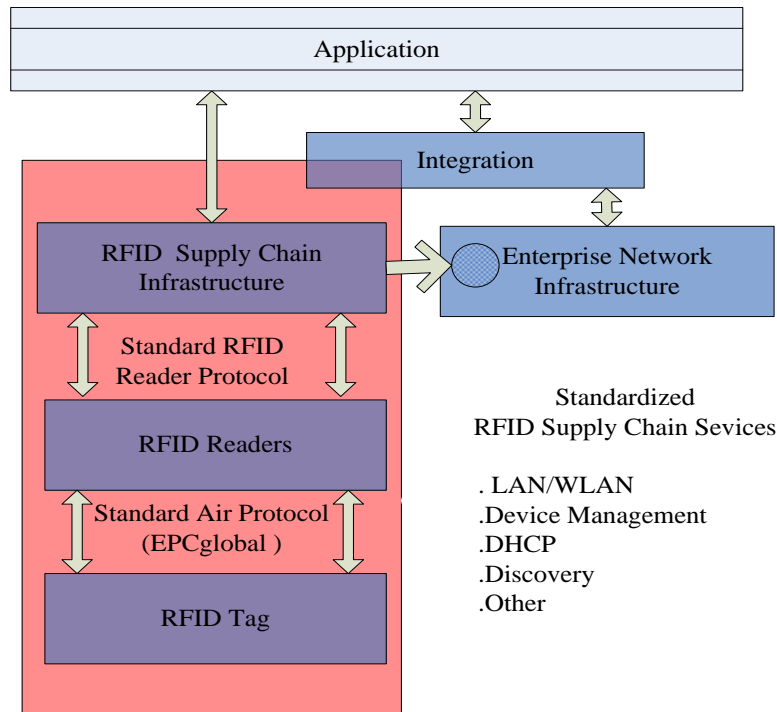
In view of above analysis, abnormal analysis is not an easy thing to be solved. The problem can be solved according to the following rules. Firstly, we should need to determine

the application areas, such as intrusion detection, fraud detection, anomaly detection and other supply chain [8]. Secondly, according to the characteristics of application itself, we can determine the anomaly detection methods, such as nature of the data, abnormal type and other factors [9]. Finally, we will select the appropriate method to solve the problem, such as data mining, machine learning, and statistics etc.

### 3 Technology of RFID Middleware Data Processing

The scheme realizes the management of the RFID reader by the electronic label processor (Tag Acquisition Processor, TAP) and management system. On one hand, TAP is the infrastructure equipment based on RFID network, and it can realize the automation management in the form of hardware equipment. On the Other hand, TAG can process the tag data oriented location, and it is a standard data interface oriented application [10].

For the handling of RFID tag data, the system is mainly related to the classification, filtering and related data identified and so on. The processing results can provide more accurate data for the application of enterprise; and facilitate the development of new applications [11]. The method really reflects the concept of hierarchical optimization. The architecture of RFID middleware data processing is shown in Figure 2.



**Figure 2. Architecture of RFID Middleware Data Processing**

In addition, the data management function of RFID middleware also supports the relevant international standards and the common enterprise standard interface [12]. The system can flexibly express and transmit the data, and realize redundancy backup. In a word, the system can ensure the reliability of the RFID Middleware Data.

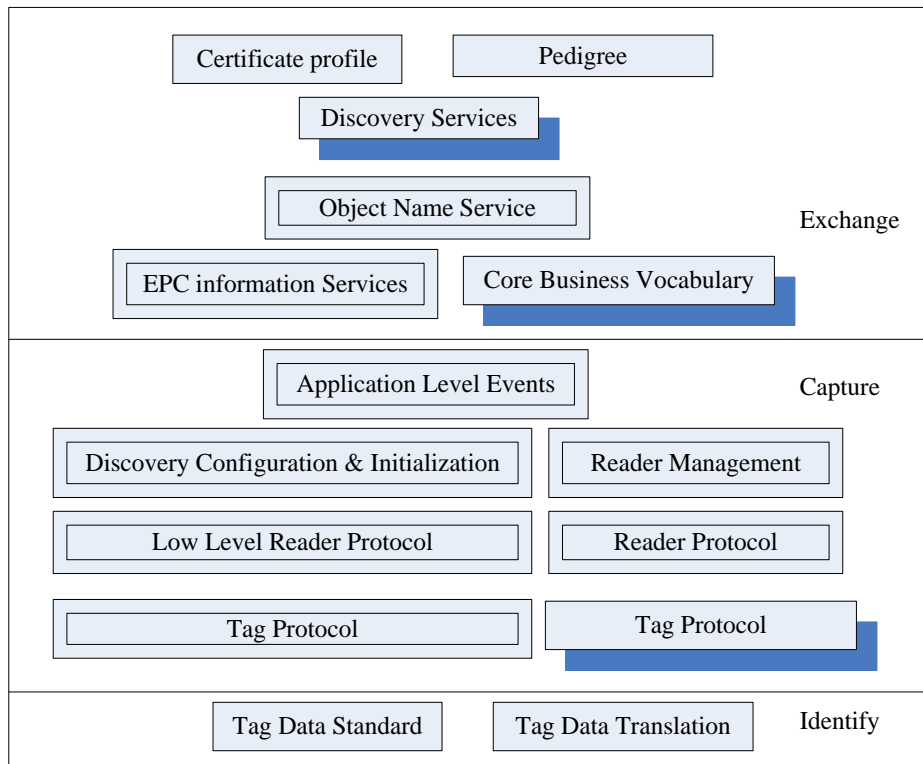
## 4. Outline Design of System

- Ideas of Research

Anomaly detection method based on statistics mainly appeals to a single attribute, and need know the distribution model and parameter information of data set. The method is not suitable for anomaly detection of RFID data.

Anomaly detection method based on distance need calculate the distance between the calculation of normal data and the abnormal data. The method has wide range of application, but there are the problems of efficiency and error.

Analysis method based on rules is more effective, can identify the abnormal data. But some supply chain system can not provide specific business rules, so the range of application is limited.



**Figure 3. Architecture of EPCglobal**

To sum up, based on the summary of the existing abnormal data discrimination method and considering the characteristics of RFID data itself, the paper has constructed the anomaly detection system of RFID supply chain data based on EPC. Combined with the detection method based on distance and detection method based on rules, the system plays their respective advantages and provides the graphical display, so that managers can easily browse exception information of the supply chain.

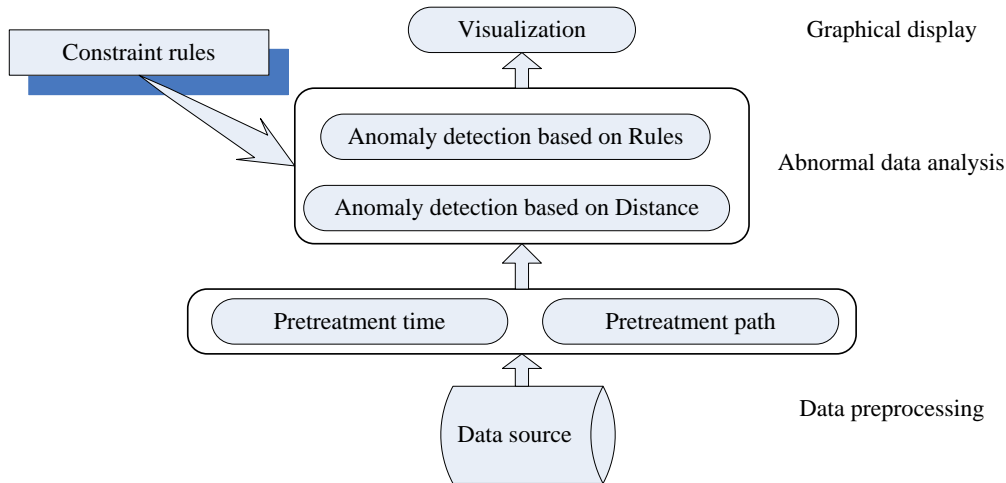
- Architecture of EPCglobal

EPCglobal is the non-profit Organization for Standardization of neutral [13]. It is made up of two standardization organization (EAN and UCC). Its major responsibility is to establish

and maintain the EPC network in the global scope of each industry, and to ensure the automatic, real-time recognition technology. The architecture of EPCglobal is shown in Fig.3 above.

## 5 Architecture of System

The architecture of the system is shown in Figure 4.



**Figure 4. Architecture of the System**

This paper has designed three layers of anomaly detection system, such as data preprocessing layer, anomaly analysis layer and graphical display layer.

- Data preprocessing layer

This layer is mainly to pretreat the event data of collecting from EPCIS based on two aspects of time and route. By the appropriate transformation of time data, this layer is convenient for the analysis of subsequent anomaly detection, and is more direct to analyzing the abnormal data without further processing.

- Abnormal Analysis layer

The layer can realize the anomaly detection from the global and local aspects by detection method based on distance and detection method based on rules. Detection method based on distance method transforms the temporal information to frequency domain information, and uses the path similarity of frequency space to exclude abnormal. Detection method based on rules can exclude those data which does not meet the rules, using some predefined rules, such as the flow, speed, Residence time. This layer can analyze the large amounts of data for analysis for the purpose of finding abnormal phenomenon.

- Graphical display layer

This layer displays the data flow by anomaly detection based on approach, and can provide a visual interface and convenient managers to view the information of supply chain circulation and exceptions.

## 6. Detailed Design of System

In the system of abnormal data analysis and processing based on RFID supply chain, the data preprocessing can extract useful and important data from a large amount of data, filter out redundant and irrelevant data [14]. Its goal is to reduce the amount of data processing and ensure the data analysis requirements. In this paper, the preprocessing of the data is divided into time pretreatment and path pretreatment from two aspects.

### Time pretreatment

The form of a typical RFID records is (EPC, location, time). EPC expresses the only code of goods. Location items express the observation sites of goods read and written. Time items express the recognition time of goods read or written.

Firstly, in the supply chain system, each node of supply chain contains the action of goods shipped and received, so we will expanded the format of RFID record, express RFID record as four tuple (EPC, location, time, BizStep), in order to facilitate abnormal data analysis of the subsequent RFID supply chain.

Secondly, according to goods' received and transported movements in the supply chain, we have converted raw data to four tuple (EPC, location, time\_in, time\_out) [15]. The four tuple represents two meanings: one is the same place of goods retention time; another is the moving time of adjacent nodes.

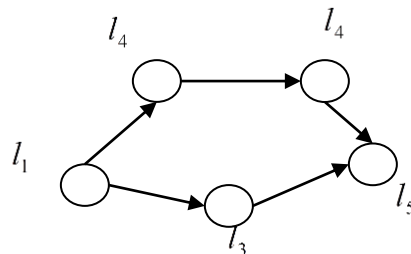
Thirdly, assuming that goods path sequence is known, we calculate the moving time of two adjacent nodes. Afterwards RFID mobile time data set pre processed is shown in Table 1.

**Table 1. RFID Mobile Time Data Set after Pre Processing**

After cleaning dataset(EPC, location, shipping_time, receiving_time)
$(i_1, l_1, shipping\_time_1, receiving\_time_2)$
$(i_1, l_2, shipping\_time_2, receiving\_time_3)$
.....
$(i_1, l_{m-1}, shipping\_time_{m-1}, receiving\_time_m)$

- Path pretreatment

RFID data path is the position sequences caused by goods moving in the RFID supply chain application, and it is composed the sequence of node [16]. There are precedence relations between nodes. The node adjacent matrix is a matrix, which represents all possible node path sequence adjacent to each node achieved. An example is shown in Figure 5.



**Figure 5. Path Map**

The node adjacent matrix is shown as follows.

$$B = \begin{bmatrix} & l_1 & l_2 & l_3 & l_4 & l_5 \\ l_1 & 0 & 2 & 3 & 0 & 0 \\ l_2 & 0 & 0 & 0 & 4 & 0 \\ l_3 & 0 & 0 & 0 & 0 & 5 \\ l_4 & 0 & 0 & 0 & 0 & 5 \\ l_5 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Finally, in all the path sequence, the path node of repetitive path sequence is removed in order to obtain complete transportation route of goods.

● Abnormal analysis

After proper pretreatment of the original RFID event data, the next is to encode the RFID data stream, compare the coding sequence obtained with path sequence completed and corrected, and then obtain the quantitative indicators of distance.

(1) Analysis and process of abnormal based on distance

Methods of analysis and processing abnormal data based on distance consider the data that deviate some distance from the normal value as the category of abnormal data. Considering the massive RFID data in the supply chain, the efficiency of processing data is particularly important and the selection of algorithm is very important [17]. The method of Discrete Fourier Transform has unique advantages in similarity search, so this paper uses the analysis and processing method of Discrete Fourier Transform. Specific implementation process is shown in Figure 6.

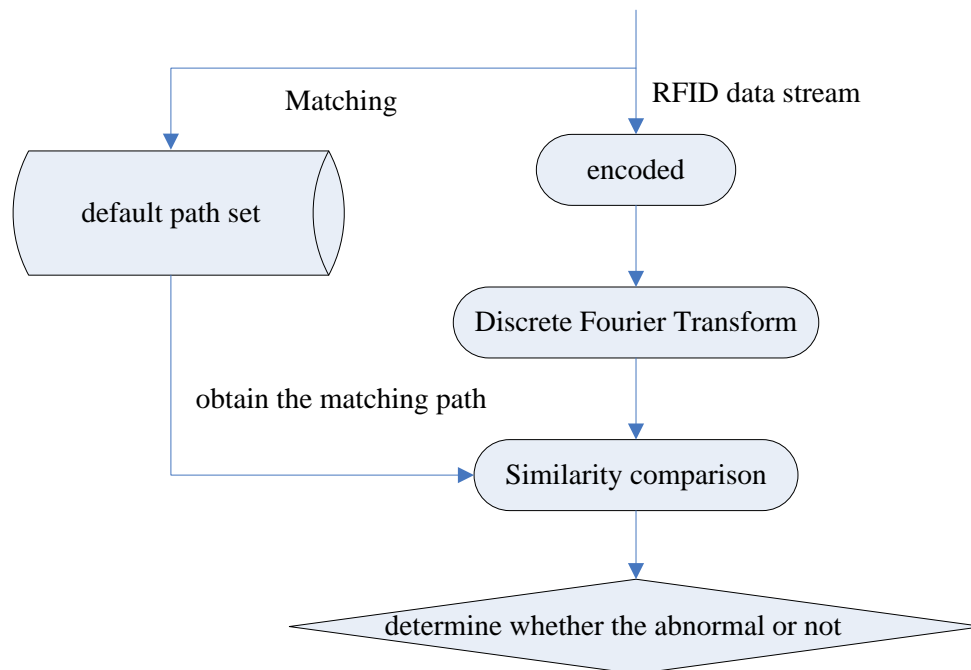


Figure 6. The Flow Chart of Distance-based Method

Generally speaking, for the supply chain system is complex, there are many possible paths in the predicted path concentration and each path length is longer [18]. Using Euclidean Distance should take longer calculation time and has the weakness of noise sensitive. But using Discrete Fourier Transform can achieve similar path matching with the complete matching and subsequence matching, which not only reduces the dimension of time series data, but also reduces the computational complexity.

## (2) Analysis and process of abnormal based on rules

Methods of analysis and processing abnormal data based on distance must ensure the consistency of flow and speed for the RFID supply chain. Flows express the value of goods occurring within a certain period time of the supply chain [19]. The flow is measured in a certain period of time and its size has a time dimension. The flow of nodes in the supply chain is an important Anomaly detection index. In general, the outflow of the former node is the inflow of next node. In addition, when a node stacks too many goods, the abnormal condition may have occurred in the node. Therefore, the supply chain flow tracking can effectively avoid the possible abnormal condition. The formula is as follows.

$$\text{Beginning inventory} + \text{inflow} = \text{End inventory} + \text{outflow}$$

An important parameter associated with speed consistency is the moving time of goods. The moving time is the required time of goods moving from one node to the next node in the supply chain. Basing on the analysis of moving goods, we can find the transport anomaly in the process of transportation or the counterfeit goods in the supply chain. For example, the supply chain system shows a piece of goods is in the Shanghai at the 10 o'clock, while the goods appear in the American after one hour [20]. We can find that the moving time beyond the fastest transport means by analysis of using the moving time rule. So, the judgment of the goods is very likely the counterfeit goods. It is necessary to monitor the moving time of goods in the supply chain system. Use the following formula to determine.

$$\alpha < \text{moving time} < \varepsilon$$

Here, we set the maximum of moving time  $\varepsilon$  and the minimum of moving time  $\alpha$ . Setting the threshold is to avoid exceeding the license of transport mechanisms during the transportation. As for the situation beyond the limit condition is often the result of errors in circulation. When the same labels appear differently, it shows that the circulation speed is too quick, such as fake commodities above. If the circulation speed is too slow, some abnormal things are possible to happen, such as truck fault *etc...*

At the same time, the slow circulation speed is likely to be happened some abnormal things in the process of transportation, such as truck fault *etc...*

In the supply chain system, there are usually three kinds of mobile objects: mass movement, decentralized mobile and non mobile. According to observation, goods are moving in large quantities at the beginning stage of transportation. The specific algorithm is shown as follows:

```
Sequence: empty
shipBiz= "shipping"
receiveBiz="receiving"
for nid←1 to N // N batch goods
// Get moving path of a batch goods
```

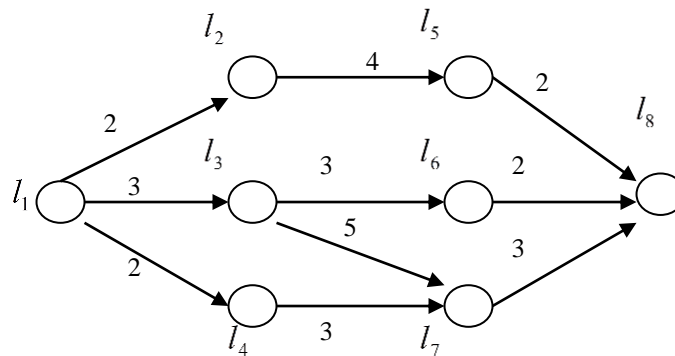


```

Sequence=getLocationSequence(nid)
// from the first node to the end node sequence
For location← 1 to sequencelast-1
Shipping =get (nid, location, shipBiz)
// “location+1” express behind node of location sequence
Receiving = get(nid, location+1, receiveBiz)
// determine whether meet the constraint condition or not
If shipping == null or receiving== null
// Abnormal display
alarm(nid, location, location+1)
end if
end for
end for
    
```

## 7. Performance analysis of System

We simulated the RFID path data sets as shown in Figure 6.



**Figure 6. Simulation of Supply Chain Path Map**

The minimum residence time is 0 days, the minimum movement time is the number of inter node line, maximum residence time = (minimum residence time +2) days, maximum moving time = (minimum move time +2) days. The specific simulation rules are shown as follows:

- The loop is not allowed in the path graph.
- According to the order, the residence time and movement time of goods, the system can simulate the data of RFID supply chain.

We have simulated 50000 EPCIS data, then Classified data set to different path, and expressed as A (1258), B (1368), C (1378) and D (1478). According to the similarity calculation formula to calculate the similarity of each path, the four path similarity calculation result is shown in Table 2.

**Table 2. The Result of Four Path Similarity Calculation**

path	the similarity of each path
A	0.8124
B	0.7929
C	0.7338
D	0.7561

We will take the above results as threshold. When the similarity of the tested path sequence and the best matching route sequence is less than the threshold, it means abnormal. In addition, when the detection method based on distance is used for the transport delay, theft case, its performance is shown as follow:

- It is staying a long time in some place and route sequence is not complete.
- The calculation of the similarity is generally less than the threshold.

As a result, in the anomaly detection, this method is effective.

## 8. Conclusions and Future Work

Supply chain management is the important application fields of RFID system [21]. Analysis of the abnormal data is important to the supply chain security and reliability. Through the analysis of RFID supply chain business needs and summary of the existing literature, the paper has presented an anomaly detection scheme for RFID data supply chain based on EPC network. The main innovation of the system is:

- Put forward the architecture of overall system, and improve the efficiency of system.
- Construct the anomaly detection system of RFID supply chain data based on EPC. The system can be combined with the detection method based on distance and detection method based on rules to play their respective advantages.

Due to the abnormal detection of RFID supply chain data is a challenging subject, and abnormal forms also emerge in an endless stream, so there are some problems need to be studied in the future. For example, we can detect anomalies or rules in the batch movement of the containers according to the supply chain nodes in the specific implementation process.

The detection rules are different depending on the type of application, so the rule of abnormal data analysis and processing system based on RFID supply chain is a field needed to dig deeper.

## Acknowledgements

This paper is a revised and expanded version of a paper entitled “Research on Anomaly Detection of RFID Supply Chain Data Based on EPC” presented at CST 2014, Indonesia, 19-22 June, 2014. This work was supported by the 12th Five Year Planning Foundation of

Jiangsu Province (Grant No.B-b/2013/01/012), and by the National Natural Science Foundation of China (61402234). It was also supported by the Modern Education Technology Foundation of Jiangsu Province (Grant No.2013-R24773).

## References

- [1] S. Leaver, "Evaluating: RFID middleware", Forrester Research, (2004).
- [2] "Oracle and BEA Systems", BEA WebLogic RFID Enterprise Server Product Overview Version 2.0. BEA Systems, Inc., (2006).
- [3] H. Gonzalez, J. Han and X. Shen, "Cost - conscious cleaning of massive RFID data sets", Proc of ICDE, (2007).
- [4] "EPCglobal Inc. Application Level Events (ALE) Standard", Version 1.0. EPCglobal Inc., (2008).
- [5] Q. Y. Dai, R. Y. Zhong, M. L. Wang, X. D. Liu and Q.Liu, "RFID-enable Real-time Multi-experiment Training Center Management System", International Journal of Advanced Science and Technology (IJAST), vol. 7, (2009), pp. 27-48.
- [6] P. S. Huang and D. Wang, "Research on supply chain abnormal event detection based on the RFID technology", Applied Mechanics and Materials, vol. 513-517, (2014), pp. 3309-3312.
- [7] R. Li, J. Han, Z. Wang, J. Zhao, Y. Gong and X. Zhang, "Tracer: Taming anomalous events with crfid tags for trajectory management", International Journal of Distributed Sensor Networks, (2013).
- [8] J. Park, K. Lee, W. Ryu, J. Kwon and B. Hong, "An efficient processing scheme for continuous queries involving RFID and sensor data streams", Communications in Computer and Information Science, vol. 186, (2011), pp. 187-192.
- [9] S. Piramuthu, S. Wochner and Grunow, "Should retail stores also RFID-tag 'cheap' items?", European Journal of Operational Research, vol. 233, no. 1, (2014), pp. 281-291.
- [10] S. Piramuthu, P. Farahani and Grunow, "RFID-generated traceability for contaminated product recall in perishable food supply networks", European Journal of Operational Research, vol. 225, no. 2, (2013), pp. 253-262.
- [11] L. Feinbier, Y. Yaslar and H. Niehues, "RFID slab logistics", ThyssenKrupp techforum, no. 1, (2011), pp. 68-71.
- [12] Q. Lei, Z. Sheng and Y. Ling, "Robust optimization of RFID's investment in supply chain with inaccurate inventory", Applied Mechanics and Materials, vol. 457-458, (2014), pp. 1661-1664.,
- [13] S. Yoon, T. W. Chang and H. Kim, "Adoption of RFID-based common standard platform for supply chain management", Source: ICIC Express Letters, Part B: Applications, vol. 5, no. 2, (2014), pp. 491-496.
- [14] S. Y. Li, H. Y. Kung, C. H. Chen and W. H. L. Hsu, "An Efficient RFID Data Processing Scheme for Data Filtering and Recognition", International Journal of u - and e - Service, Science and Technology (IJUNESST), vol. 5, no. 2, (2012), pp. 59-76.
- [15] I. P. Vlachos, "A hierarchical model of the impact of RFID practices on retail supply chain performance", Source: Expert Systems with Applications, vol. 41, no. 1, (2014), pp. 5-15.
- [16] S. J. Chuu, "An investment evaluation of supply chain RFID technologies: A group decision-making model with multiple information sources", Source: Knowledge-Based Systems, (2014).
- [17] T. Irrenhauser and G. Reinhart, "Evaluation of the economic feasibility of RFID in the supply chain", Source: Production Engineering, (2014).
- [18] L. Pfahl and C. Moxham, "Achieving sustained competitive advantage by integrating ECR, RFID and visibility in retail supply chains: A conceptual framework", Source: Production Planning and Control, vol. 25, no. 7, (2014), pp. 548-571.
- [19] S. P. Huang and D. Wang, "Research on supply chain abnormal event detection based on the RFID technology", Source: Applied Mechanics and Materials, vol. 513-517, pp. 3309-3312, (2014).
- [20] J. Leung, W. Cheung and S. C. Chu, "Aligning RFID applications with supply chain strategies", Source: Information and Management, vol. 51, no. 2, (2014), pp. 260-269.
- [21] R. Derakhshan, M. E. Orłowska and L. Xue. "RFID Data Management: Challenges and Opportunities, In Proceeding of IEEE International Conference on RFID, (2007), pp. 26 – 28.

## Authors



**Chunfeng Wang**, she received the M.S. degree from Hohai University in Nanjing in Computer Science & Application in 2006. Now, she is a lecturer in Modern Education Technology Center, Yancheng Institute of Technology. She research interests mainly include Network security, virtual reality and data mining.



**Shuiming Zhong**, he received the M.S. and Ph.D. degrees from the Hohai University, China, in 2007 and 2011 respectively. He currently is a lecturer in the School of Computer and Software, Nanjing University of Information Science and Technology, China. His research interests involve artificial neural networks, machine learning, and pattern recognition.