

## Research Article

### The Research and Implementation of RFID Middleware Data Filtering Algorithm

<sup>1</sup>Ming Ke, <sup>1</sup>Jinlong Zhao, <sup>1</sup>Mingming Huo and <sup>1,2</sup>Xuhui Chen

<sup>1</sup>School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

<sup>2</sup>Internet of Things Lab, Xiamen University of Technology, Xiamen 361024, China

**Abstract:** RFID is an advanced non-contact automatic identification technology. It is widely used in location, real-time identification and tracking and it has become a hot topic in the field of research. A lot of redundant data and noise data greatly reduce the efficiency and performance of RFID system. This study puts forward a filtering algorithm based on dynamic adaptive time window. In order to ensure the sequence of acquainted RFID data correctly, the ordered filtering algorithm based on hash table is introduced. Experiments show that the improved filtering algorithm not only can effectively filter duplicate data, but also can ensure the correctness of the data reception.

**Keywords:** Data filtering, middleware, noise data, redundant data, RFID

## INTRODUCTION

RFID technology (You and Li, 2004) is a kind of real-time, fast non-contact recognition technology. Since its emergence in the 1990s, the applied research in logistics, production, medical, transportation and other fields has become the hotspot in the field of IT. RFID system is mainly composed of reader, electronic tags, application software and middleware. The middleware is the important part of the RFID technology and the application of which makes the operation of the underlying detail of the RFID system transparently. So the programming is easily implemented and the complexity of the many-to-many connection between equipment maintenance is reduced.

The effective and correct transfer of data is the key to achieve effective operation between modules in RFID system. The logic structure diagram of RFID system is shown in Fig. 1. But due to the effect of factors, such as the external environment, reader or tag, the raw data collected by readers often contain redundant data and noise data (Brusey, 2003). This will affect the efficiency of RFID system. So the original data filtering operation is necessary before the collected raw data directly applied. If the filtered data directly output is likely to cause the output data inconsistent with the read data. In this study a dynamic rolling window filtering algorithm is presented based on the analysis of the existing filtering algorithms. In order to solve the problem of data sequence and guarantee the consistency of input and output data, hash table algorithm is intruded in this study.

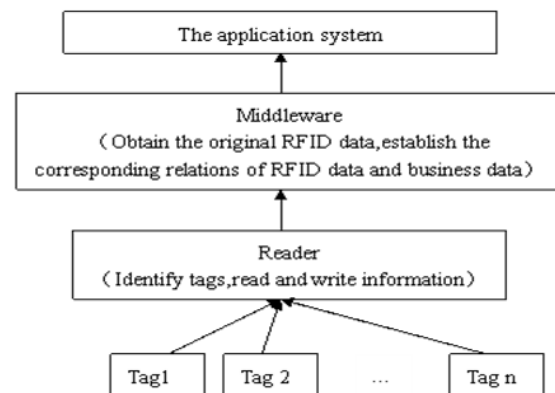


Fig. 1: The logic structure diagram of RFID system

## THE RFID MIDDLEWARE DATA FILTERING ALGORITHM

**Introduction of data filtering:** The original data often contains redundant data and noise data collected by RFID reader. The causes are as follows:

- **The causes of redundant data:** Tags are in the scope of radio frequency identification reader in multiple read/write cycle. So the tags are repeated many times. The tag is in radio frequency identification of multiple reader range which makes the tag read by multiple readers. A physical object is pasted more than one same tag information which makes the tags repeated read.
- **The causes of noise data:** The location of Tag is beyond the radio frequency range of reader which

**Corresponding Author:** Jinlong Zhao, School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China.

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

Table 1: The filter type description

Filter type	Type description
Based on the reader	Only to read data from the specified
Based on data and tag	Only to care for the specified Tag sets

makes tags missed. The abnormal or additional data caused by the reader or environmental is the noise data.

The original data is huge collected by RFID reader. Each reader can collect dozens or even hundreds of raw data every second. But the real part is less meaningful to the user data, so if the original data is passed directly to the RFID system without the filtering process, the influence of three aspects will be brought.

- A large amount of data transmission will affect the network transmission efficiency.
- A large amount of raw data processing will affect the processing efficiency of processor.
- A large number of additional data stored in the database will reduce the effective storage of database. Therefore, the raw data passed by the reader filtered is the priority of the RFID system.

**The data filtering algorithm:** In recent years the study of filtering algorithm has been closely concerned. The filtering operations are divided into two types generally: based on data filter, based on the reader and tags. The descriptions of the two types of filter are shown in Table 1. Deng and Li (2012) analyzed the causes of the noise data and redundant data and put forward a kind of noise removing algorithm-basic noise filtering algorithms. Set threshold and by the judgment of data within the window duplicate data to judge whether the number of occurrences of the noise. In order to ensure the sequence of the filtered output data accuracy, a hashtable filtering algorithm is proposed. The experiments verify the effectiveness of the two kinds of algorithm to filter data. In order to suppress the noise data and the purpose of redundancy, Qi (2009) adopted the method of weight accumulation to mark the repeated tags. The method of weight decreasing is adopted to mark the tag which did not appear. By setting threshold to trigger the corresponding tags incident, denoising and reducing the probability of false alarm events were realized. The simulation experiments proved the effectiveness and feasibility of the algorithm. Currently most of the data filtering algorithms are applied to solve the problem of the middleware data filtering. Chen *et al.* (2009) transplanted the filtering algorithm which is applied to the data of middleware into the reader. A binary tree data filtering strategy was proposed. The experiment showed that the binary tree strategy had better performance. Wang and Guo (2009) adopted the dynamic time Windows filtering algorithm from the application of RFID and applied it in warehouse items and the slow transit sites of RFID reader data filtering.

The experiments showed that the dynamic time filtering algorithm had a better performance to redundancy than fixed time window algorithm. Zhao (2006) used the tags state of 0 and 1 to encode. 1 expressed the disappearance of the tag. 0 expressed the appearance of the tag. And the time threshold was set. The data filtering was defined in the aspects of time dimension and the state definition. Although the above algorithms can realize the original RFID data filtered to a certain extent. In the actual production, due to the influence of external factors, may lead to misread the tag data or read leakage phenomenon which is the so-called noise data. Although some literatures have certain researches in filtering out noise data, it still be able to do better to improve and better applied to the actual production.

This study not only consider the further study of noise data filtering but also consider whether the order of the input of original useful data is consistency with the order of the filtered data based on the above filtering algorithm. An improved data filtering algorithm is put forward in this study. By the comparison experiments of basic filtering algorithm and the improved filtering algorithm and basic noise filtering algorithms show that the improved filtering algorithm has better performance.

### THE IMPROVE OF NOISE DATA FILTERING ALGORITHM

The adaptive operator and time window are introduced in this study on the basis of basic noise filtering algorithm. The window of the dynamic adaptive noise filtering algorithm is put forward. But the data after a noise filter algorithm processing can't achieve the goal of removing redundant. To solve this problem, a kind of based on hash table to redundancy filtering algorithm is put forward. This algorithm has more obvious advantage than basic redundancy filtering algorithm in data filtering average delay. The common problems in the RFID system is solved after the Tag data is processed by improved noise redundancy filtering algorithm and the data filtering algorithm based on hash table.

**The noise data filtering algorithm:** The recognition of the occasional noise data is the key to noise data filtering. To improve the identification readers to tags, many times read reader to tag is adopted. If fixed length window or a fixed cycle is used to determine noise data filtering when the original data transmitted by RFID reader, it will waste large memory space and greatly influence the efficiency of the system. To solve this problem, based on the ideas of the fixed window, a dynamic adaptive adjustment window is introduced and set the upper L and lower bounds U of the window. To determine whether the original RFID data is normal data, in this study, the data which the reader reads

marked in the state of flag when the flag is 1, it indicates the same data. If the flag is 0, it indicates another data. For each occurrence of flag is marked as 1 to weights of accumulation and calculate the number of occurrences of tag information. If the weight of flag is greater than the *threshold*, the raw RFID data is considered as normal data, otherwise is noise data.

- **The dynamic adaptive window:** Because of the different Settings of RFID system, the number of repeat read tags is also different. If the data window is set as fixed size, it would be a waste for repeated reading less system and greatly limits the performance of the system. In order to adapt to different systems and have stronger practicability and use value, this study establish the adaptive dynamic adjustment which has upper and lower limits of the data window inspired by the adaptive operator thought.

The initial window size is  $W_{length}^0$ . The length of data is  $D_{length}$ . Each tag is read  $D_{number}^0$  times. Namely the number of each group of data is  $D_{number}^0$ . When the data is repeated once, the weight adds 1. The flag of the first data that occurs in the window is 1. Comparing the data and the first time data is quite, if is equal to the current data marked as 1 and the data of weight add 1. When data is repeated  $D_{number}$  times and the weight meet  $threshold/2 + 1$ , the data is considered as normal data and output window data.  $D_{number}^0 - D_{number}$  Data which do not enter the window data delete directly. If  $D_{number}$  did not meet the conditions, the window is moved as 2 times of the data length  $D_{length}$ . Check the weight of the data in the window, until meet the termination conditions. The formula of adaptive dynamic window is as follows:

$$W_{length} = W_{length}^0 + 2nD_{length} \quad (1)$$

Among them,  $n = 1, 2, \dots, M$ . The M is a integer which meet  $W_{length} \leq U$ .

- **The dynamic adaptive filtering algorithm:** In this study, algorithm description is as follows:

**Step 1:** Create a window: the upper limit L, the lower limit U and the initial size of the window is  $W_{length}^0$ .

**Step 2:** Define the data number of each group  $D_{number}^0$ , data length  $D_{length}$ , size threshold *threshold*, set the weights of data, the initial weights are zero, set the flag, the first appeared data flag = 1.

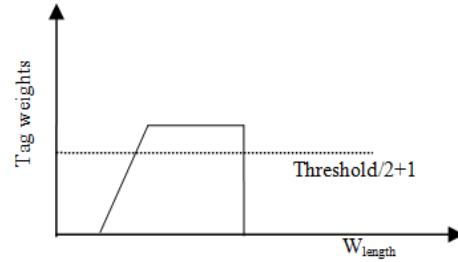


Fig. 2: The schematic diagram of tag weight changing

- Step 3:** Calculate the data weights of the window and compare whether the data value is greater than  $threshold/2 + 1$ , the size of *threshold* is  $D_{number}^0$ . If data weight is greater than or equal to  $threshold/2 + 1$ , the data is considered normal data, the output data to window and filter out the data that not in the window directly. If less than  $threshold/2 + 1$ , the window increases according to adaptive window formula (1) until meet the termination conditions.
- Step 4:** Repeat Step 2~3 until all raw RFID data is filtered.

In this algorithm, no matter whether the reader or environmental factors cause the aberrant or additional data, the data values is not greater than  $threshold/2 + 1$ , the noise data will not be triggered, so can remove the noise data effectively. The normal data label will be marked as the normal data as long as its threshold weight is greater than or equal to  $threshold/2 + 1$ , the data tag that not in the window of the group will automatically be removed, avoid repeating compare the data size, reduce the data redundancy and improve the efficiency of the RFID system. After dealing by the algorithm, the tag data consisted by multiple discrete point became a continuous change of weights. The graph is shown in Fig. 2.

**Redundancy filtering algorithm based on hash table:** Whether the basic noise filtering algorithm is applied to the reader or to the middleware, its advantage is that the algorithm idea is simple and easily realized, besides, it outputs results quickly. But once these basic noise filtering algorithms judge out the noise data, then delete them and directly output the normal data, thus cause a lot of redundant data, the waste of system memory and reduce the system performance.

Redundancy is that a tag is read only once in a certain prescribed period of time. Li and Li (2009) created a list (TagID, Time, TagData) for the tag information to solve the problem of duplicate data in the RFID system, can determine whether the two tag are the same by judging the time difference between the two tag information. Although Li Dexing can solve the

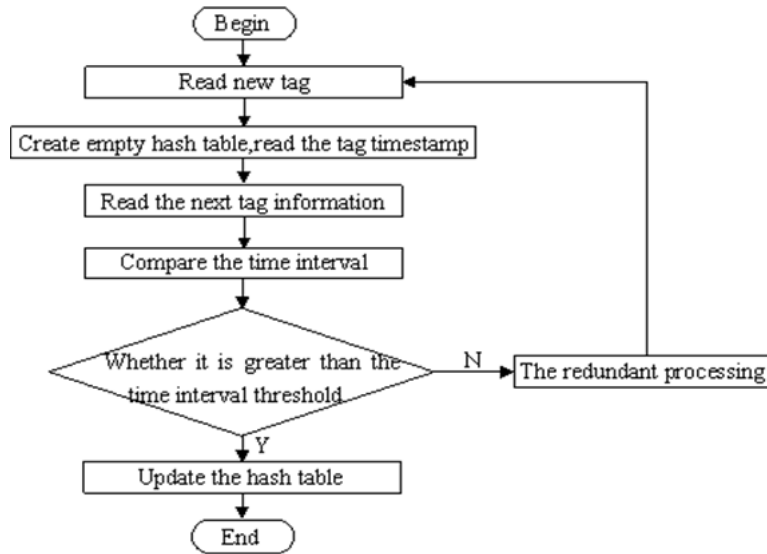


Fig. 3: The redundancy filtering algorithm flow chart

problem of data redundancy, but it need to construct multiple linked list in the algorithm, causing the waste of memory, the system time and influencing the efficiency of the system greatly. Inspired by Deng and Li (2012) this study uses hash table to process the normal data generated by the noise processing algorithm.

Each coming normal tag data will be set a timestamp; the timestamp carries the time of the reader tag, when a tag data arrives and then calculate the new tag data and its time interval, if their time interval is less than the setting threshold, then the arrived new tag is set as repeated tag, otherwise regarded as a different data. In this algorithm, the use of hash table to save the tag timestamp of data, the algorithm described as follows:

- Step 1:** Create an empty hash table: hashtable and set the time interval threshold.
- Step 2:** The arrived timestamp of the tag information is inserted into the hash table and reads the next tag information.
- Step 3:** Compare the time interval of two tag information, if the time interval is greater than the set threshold, then the two tags is considered for different tags and insert the arrival of the new tag into the hash table. Otherwise the two tags are considered as the same tag, redundant data and remove directly.
- Step 4:** Update the hash table.
- Step 5:** Repeat step 2-4 until all the tag information is processed.

The algorithm processing flow is shown in Fig. 3.

The above algorithm is a basic redundancy filtering algorithm; its time complexity is reduced to  $O(1)$  from  $O(n)$  because of the establishment of a hash table. At the same time, because we tag the timestamps for read tag information of every reader and finally avoid the problems of the output sequence and data sequence not matched for the reader read tags.

### THE SIMULATIONS AND ANALYSIS

In order to test the effectiveness of the dynamic adaptive window noise filtering algorithm, the experiment which the output delays of arrival rate of different label information is set.

Assuming that each label is read for 10 times, the noise rate is 5%, the adjacent label's reading time interval needs 200 msec, the average arrival rate is 1 piece/sec, 5 pieces/sec, 25 pieces/sec, 50 pieces/sec. Comparing the dynamic adaptive window noise filtering algorithm and the basic noise filtering algorithm, the simulation results are shown in Fig. 4.

It is shown in Figure 4 that when the average arrival rate of the label is low, the average filtering delay of label under the basic noise filtering algorithm do not change significantly. It indicates that the basic noise filtering algorithm ignores the label's output sequence. It may cause the inconsistencies between the order which reader reads tag and the outputting order of labels. However, the dynamic adaptive window noise filtering algorithm can output the label timely. Due to the introduction of the adaptive algorithm, the scan time of the label data is greatly shorted. The label data has low filtration delay time.

To verify filtering effect of the adaptive noise filtering algorithm based on the experiment 1, the

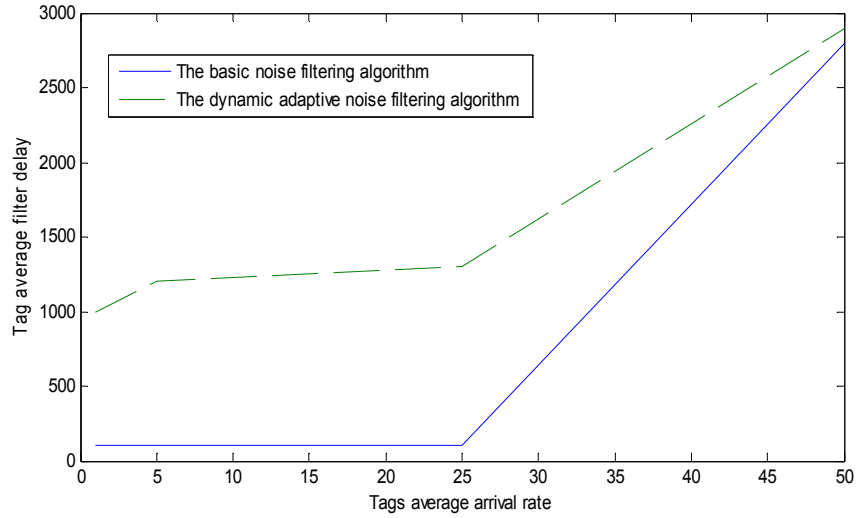


Fig. 4: The contrast figure of noise filtering algorithm (a)

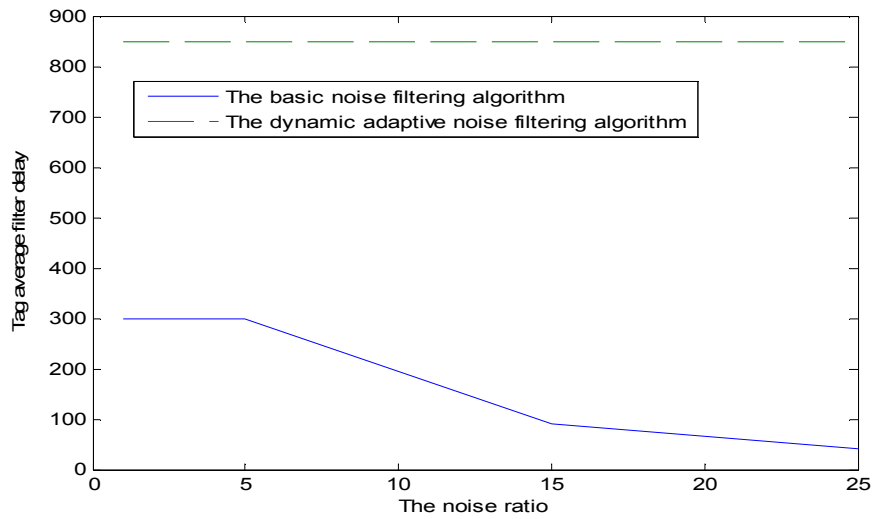


Fig. 5: The contrast figure of noise filtering algorithm (b)

Table 2: Average filter delay analysis table

Tag average arrival rate	Average filter delay	
	The basic redundancy filtering algorithm	The redundancy filtering algorithm based on hash table
10/sec	8/msec	8/msec
50/sec	8/msec	8/msec
250/sec	10000/msec	7/msec
1000/sec	10000/msec	6/msec

parameters are set as follows. Each tag is still read for 10 times, reading time interval of adjacent labels also is 200 msec. The noise rate is 1, 5, 15, 25%. The average arrival rate of the label is 1 piece/sec. The simulation results are in Fig. 5.

In Fig. 5, no matter how the noise ratio changes the label's average filtering delay has no change basically of the basic noise filtering algorithm. But the label's average filtering delay always maintains low

with noise ratio increasing of the dynamic adaptive noise filtering algorithm.

Based on experiment 1 and 2, comparing the redundant filtering algorithm based on the hash table with the basic redundant filtering algorithm, the parameters are set as follows. The reader read the label 10 times repeatedly. The label average arrival rate is 10 pieces/sec, 50 pieces/sec, 100 pieces/sec and 150 pieces/sec. The noise rate of label is 0%. The simulation results are shown in Table 2.

Table 3: All kinds of parameters of the items in the warehouse

	Object 1	Object 2	Object 3	Object 4	Object 5	Object 6
Rate (m/s)	1.3	1.5	1.0	0.8	1.5	2.0
The shortest distance from the boundary /m	2.0	2.5	2.0	2.3	5.0	2.5
The shortest time to leave the area /s	1.54	1.67	2.00	2.90	3.33	1.25

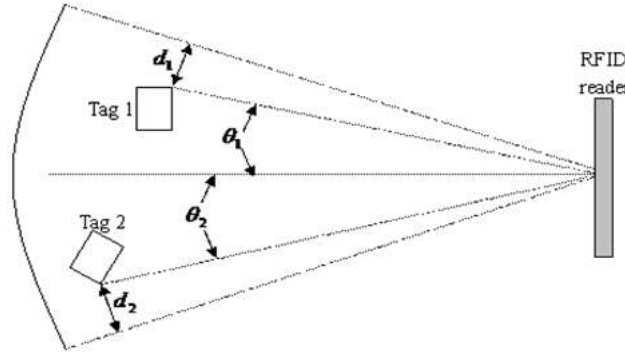


Fig. 6: The location diagram that electronic label in the radio frequency identification of RFID reader in the warehouse

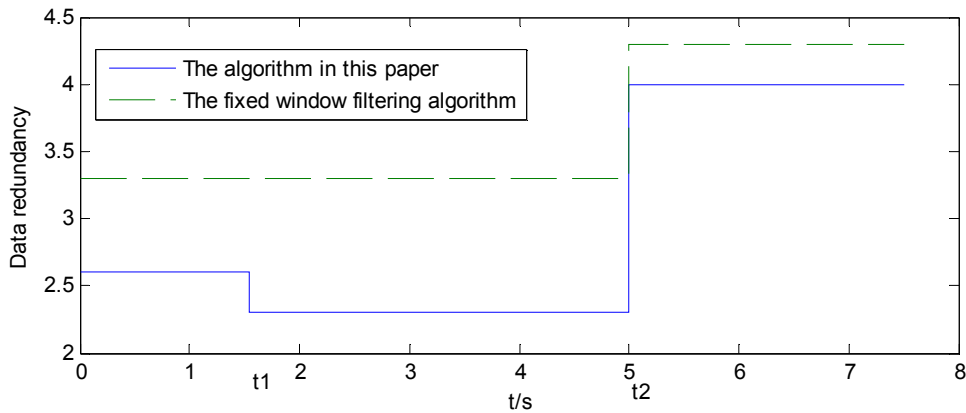


Fig.7: The contrast figure of the algorithm in this paper and the fixed size window algorithm

From Table 2, both the basic redundant filtering algorithm and redundant filtering algorithm based on the hash table can achieve the effect of removing redundancy. But the redundancy filtering algorithm based on the hash table has the obvious advantages in the aspect of data average filtering delay, while the data of average filtering delay increases significantly with the increasing of label 's arrival rate of the basic redundant filtering algorithm.

**APPLICATION OF THE IMPROVED DATA FITTING ALGORITHM**

Considering the application of RFID Middle ware in the actual production, the application of the RFID middle ware in the storehouse supervision is set as an example in this study. The position diagram of electronic label's radio frequency identification range of monitoring RFID reader is shown in Fig. 6. Take UHF RFID middle ware reader as example, the RF distance is 20 m~30 m. A 20 m×30 m×5 m<sup>3</sup> Warehouse

is selected and assuming that initially there are 4 goods (item 1~4) stored in the storehouse and all the goods in the storehouse are affixed with a electronic label. The position and velocity of the items are shown in Table 3.

At the moment t<sub>1</sub> item 1 is taken out and item 5 is put in the storage. At the moment t<sub>2</sub> item 6 is put in the storage. According to the items' parameters in Table 3, at the moment t<sub>1</sub>, the item of the four which left the storehouse first is item 1 and the window size is 1.54s. Between the moment t<sub>1</sub> and t<sub>2</sub>, the first one left is item 2 and the window size is 1.67s. After moment t<sub>2</sub>, item 6 leave first and the window size is 1.25s and so on. If the dynamic filtering window is used, to ensure that the data is not missed, the time window size is 1.2s.

This study compares the filtering algorithm which introduces the adaptive operator and hash table with the fixed window size. The simulation results are in Fig. 7. From Fig. 7, the introduction of adaptive operator can not only ensure the tag data read correctly, but also effectively avoid the data's leakage of reading and misreading caused by the goods in and out warehouse.

The introduction of hash table reduces the data redundancy greatly.

### CONCLUSION

Based on the analysis of the noise data and redundant data causes, this study introduces the adaptive operator and puts forward the data filtering algorithm for dynamic adaptive window, which can not only eliminate the noise, but also guarantee consistency of the label information in the input and output. The tag data which are filtered under the noise filtering method have great redundancy and with reference of the advantages of the hash table, this study put forward a kind of redundancy filtering algorithm which is based on the hash table and filter the redundant data effectively by setting a time stamp. Experiments prove that the dynamic adaptive window for noise filtering algorithm and redundancy filtering algorithm based on the hash table can not only filter noise and remove redundancy, but also save Memory time and space greatly, improve the performance of RFID systems. How to filter the noise data more effectively to improve the output of the effective data rate is one of the problems that need to make further study in this study.

### ACKNOWLEDGMENT

This study is financially supported by the National Natural Science Foundation of China under Grant No.61065007 and No.61263047. And it was also

supported by the Natural Science Foundation of Gansu under Grant No.0914ZTB098.

### REFERENCES

- Brusey, J., 2003. Reasoning about uncertainty in location identification with RFID [C]. Proceeding of Workshop on Reasoning with Uncertainty in Robotics (IJCAI, 2003), pp: 23-30.
- Chen, J., G. Liu, J. Wu and X. Zhou, 2009. Applications of data filtering in RFID systems [J]. *Stud. Opt. Comm.*, 4: 177-183.
- Deng, H. and J. Li, 2012. The research of RFID data stream filtering algorithm [J]. *Res. Comput. Technol. Filtering Algorithm*, 22(6): 26-34.
- Li, D. and G. Li, 2009. Research and realization of RFID middleware key part [J]. *Comput. Knowl. Technol.*, 5(34): 9846-9847, 9872.
- Qi, L., 2009. Filtration in RFID middleware data [J]. *J. Chongqing Inst. Technol. Nat. Sci.*, 23(10): 94-98.
- Wang, W. and F. Guo, 2009. Radio frequency identification middleware data filtering algorithm based on dynamic time window [J]. *Inform. Electron. Eng.*, 7(3): 177-183.
- You, Z. and S. Li, 2004. Radio Frequency Identification Technology (RFID) Theory and Application [M]. Electronic Industry Press, Beijing.
- Zhao, L., 2006. The Design and Implementation of RFID Middleware Event Management System [D]. Huazhong University of Science and Technology, Wuhan.