Egyptian Informatics Journal (2013) 14, 135–145

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

ORIGINAL ARTICLE

Proposed prediction algorithms based on hybrid approach to deal with anomalies of RFID data in healthcare

A. Anny Leema [a,*], M. Hemalatha [b]

[a] *Karpagam University, Coimbatore, India*
[b] *Dept. of Software Systems, Karpagam University, Coimbatore, India*

Received 2 November 2012; revised 24 January 2013; accepted 17 February 2013
Available online 26 March 2013

KEYWORDS

Issues of RFID data;
RFID in healthcare;
Methodology;
RFID detection model;
Effectiveness

Abstract The RFID technology has penetrated the healthcare sector due to its increased functionality, low cost, high reliability, and easy-to-use capabilities. It is being deployed for various applications and the data captured by RFID readers increase according to timestamp resulting in an enormous volume of data duplication, false positive, and false negative. The dirty data stream generated by the RFID readers is one of the main factors limiting the widespread adoption of RFID technology. In order to provide reliable data to RFID application, it is necessary to clean the collected data and this should be done in an effective manner before they are subjected to warehousing. The existing approaches to deal with anomalies are physical, middleware, and deferred approach. The shortcomings of existing approaches are analyzed and found that robust RFID system can be built by integrating the middleware and deferred approach. Our proposed algorithms based on hybrid approach are tested in the healthcare environment which predicts false positive, false negative, and redundant data. In this paper, healthcare environment is simulated using RFID and the data observed by RFID reader consist of anomalies false positive, false negative, and duplication. Experimental evaluation shows that our cleansing methods remove errors in RFID data more accurately and efficiently. Thus, with the aid of the planned data cleaning technique, we can bring down the healthcare costs, optimize business processes, streamline patient identification processes, and improve patient safety.

1. Introduction

RFID is a technology which uses radio communication between tags and readers to automatically identify the locations of items. In a networked environment of RFID readers, enormous data are generated from the proliferation of RFID readers [1]. The raw data generated from the readers cannot be directly used by the application because it consists of enormous volume of data duplication, false positive, and false negative. Thus, the RFID data repositories must cope with a

* Corresponding author. Tel.: +91 9884208328.
E-mail addresses: annyleema@gmail.com (A.A. Leema), hema.bioinf@gmail.com (M. Hemalatha).
Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

1110-8665 © 2013 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.
http://dx.doi.org/10.1016/j.eij.2013.02.003

number of quality issues. These data quality issues include data redundancy, false positive, and false negative. Poor data quality has adverse effects at the operational, tactical, and strategic levels of an organization. This is especially true in the healthcare field where cost pressures and the desire to improve patient care drive efforts to integrate and clean organizational data.

2. Issues in RFID data

RFID data acquisition and transmission are unreliable. Applications dynamically generate readings that change along the time and are temporal in nature. The huge amount of raw RFID data stream is generated and it is very complex to analyze [2,3]. The information can be read and sometimes written at distances of up to 30 feet, depending on the system. The observed read rate in real world RFID deployments is often in the 60–70% range. Duplicate or missing data may cause wrong or even ambiguous statistics. The success of these applications depends heavily on the quality of the data stream generated by the RFID readers [4,5]. The reader should detect the tags that are present within its vicinity and cannot detect those which are present outside its scope and those that are present due to its business rule cannot be detected. The effectiveness in cleaning the RFID data remains a concern, even though a number of literary works are available. Existing cleaning techniques work under a wide set of conditions but are disregarded due to high cost and complexity. Hence, effective prediction and cleaning algorithms are essential for the correct interpretation and analysis of RFID data.

3. Objectives

The readers are the detection nodes and deployed in different locations. Each detection node is identified by a unique ID that serves as the location ID. RFID tags are in different locations and can be detected by these readers. Data inaccuracies are inevitable in the RFID system considering the complexity of deployment and diverse business needs it caters to. The raw data cannot be directly used by the high-end applications unless they are filtered and cleaned. The objectives are as follows:

- To integrate middleware and deferred approach to overcome the drawbacks that exists in the existing approaches.
- To develop effective data cleaning techniques using hybrid approach to deal with anomalies false positive, false negative, and duplication of high accuracy and less complexity.
- To use the error free data for high-end applications.
- Simulate healthcare environment and test the anomaly cleaning algorithms in the healthcare data set [6].

4. Problem definitions

Probabilities of errors and redundancies are high in the RFID data which results in the limited deployment of RFID technology [7]. There are three types of errors in RFID data reading. They are unexpected readings, misread, and duplicate readings.

An RFID reader periodically sends out RF signals to its range. When an RF tag that moves within the range of the reader receives the signals, it will send a response signal along with its unique identifier code, timestamp, and location ID. The reader receives the response signal and registers the data stream as one entry. There would be some RF tags which are not supposed to be detected by the reader and may be read due to the spatial divergence of RF signals sent by the reader. Such readings are termed as false positive readings [8,9].

A significant number of tags which are within the reader's read range are not consistently read by the reader either due to their orientation with respect to the reader, distance from the reader, presence of metal, dielectric or water material close to the tag and other factors. Practically, few of the tags might not be read in every cycle though present in the effective detection range called as false negative or missed readings.

Duplicate readings are classified into reader duplicates and data duplicates. The former occurs when a tag is present in the vicinity of more than one reader which is simultaneously sending signals to it. Consider a scenario where readers R1, R2, and R3 are redundant since the tag T1 is read by all three readers at the same time, thus responsible for reader level redundancy.

The latter occurs when a reader reads a large amount of non-difference information at a time interval. For instance, in Hospital Management System, a tagged entity (Say a doctor) may move to his consulting room and sit the whole day and send the data to the RFID management system constantly through the reader placed in his vicinity. But, from the management point of view, the most useful information for event detection is when the tagged entity (Say a Doctor) enters and exits his consulting room. Therefore, it is necessary to reduce RFID data redundancy before processing.

5. Existing approaches

The first step in our work is to analyze the three existing approaches physical, middleware, and deferred. In the existing physical approach, cost is increased, the cycle is increased, tag collision occurs and duplication arises, whereas the middleware approach deals with low complex anomalies and is not very efficient in noisy environment. In deferred approach, cleaning is limited after storing the data into the database and anomalies are not cleaned properly. In these three approaches, anomalies are cleaned to some extent but results in some other anomalies. Fig. 1 depicts the three different existing data cleaning approaches and its limitations.

6. Proposed approach

- The proposed approach shown in Fig. 2 is a hybrid approach of middleware and deferred based Cellular model that can be used for detecting out of the range readings.
- The RFID readers have Omni-directional antenna, and hence, there are possibilities for the adjacent regions to overlap with each other.
- Middleware is deployed between the readers and applications to correct the captured readings and for correct interpretation of data [10].

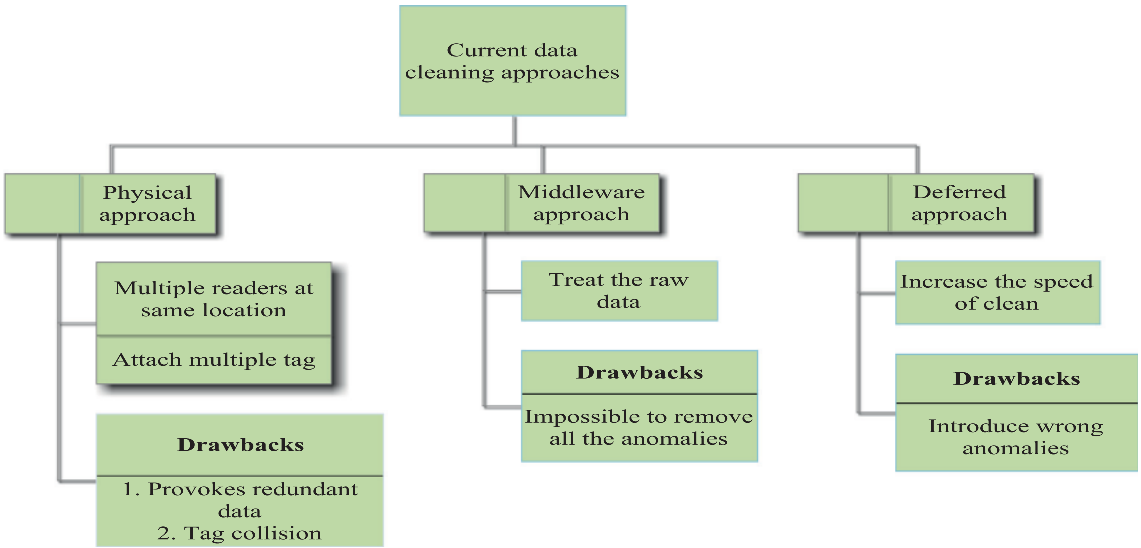

Figure 1 Existing approaches and its limitations.

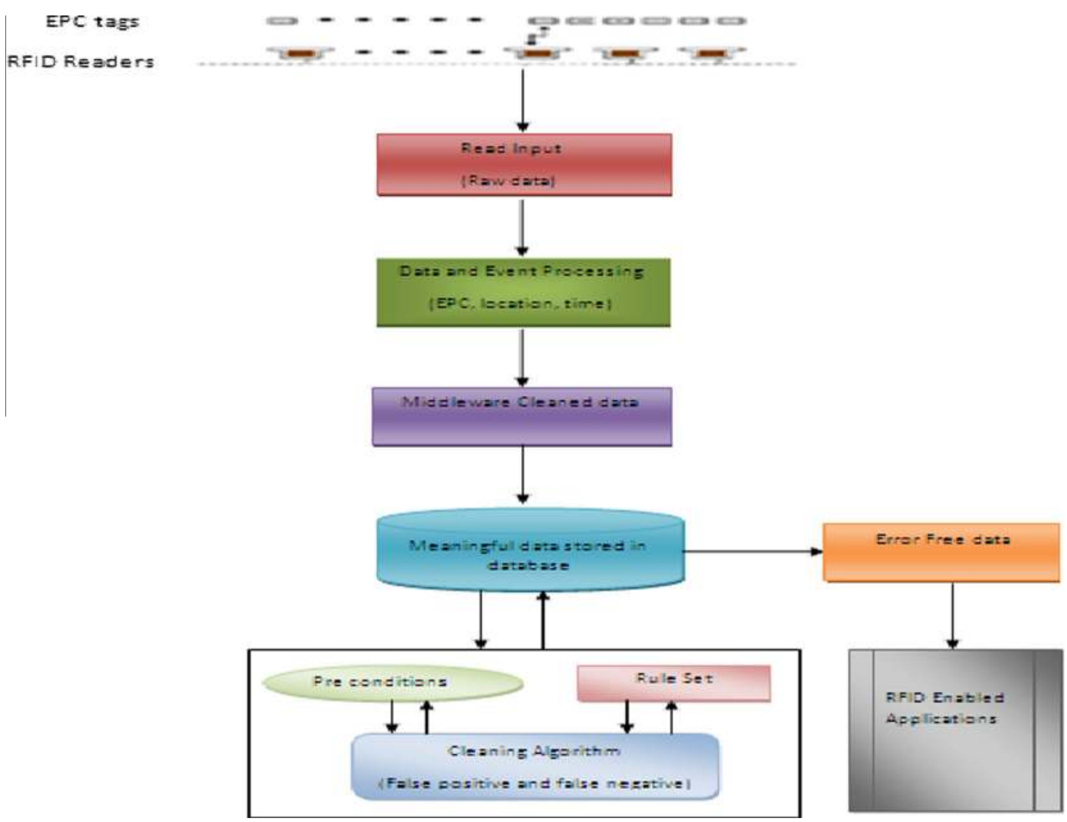

Figure 2 Architecture diagram for proposed approach.

- It is not always possible to remove all sorts of anomalies and redundancies in this approach [11]. Raw data are treated to some extent before entering the master tables of the database [12].
- Validating data at different levels to ensure data consistency, monitor incoming data stream, provide real-time integration with the existing hospital management system, mapping data onto the relevant database table, and redefining and executing business rule set are the various prime functions done here.
- The business context is dynamic and it is not even framed during the loading of data.
- Known anomaly duplication is handled in RFID middleware, but the processing of other anomalies is deferred until the query time. Each application specifies its own anomalies by defining cleansing rules.
- The rules do not change the database contents directly, but are evaluated only during application issue queries

6.1. Advantages of proposed system

The main property of the proposed system is as follows:

- Highest expected cost reduction.
- Clean large number of tag readings with minimum resources.
- Efficient and accurate data cleaning techniques.
- Easy to maintain data and Update.
- Less time and staffing.

7. RFID in hospital management

RFID can be applied in all applications, but the role of RFID in healthcare are of more importance. It is the place where a small mistake can cost a human life. However, human beings are more prone to errors. It is the necessity to minimize the human intervention, so that the chances of error are minimized. Technological evolution of Radio-frequency identification (RFID) is just starting to make inroads into healthcare. RFID (Radio Frequency Identification Devices) is the best technology to provide a better solution in healthcare industry by reducing medication errors and improving patient care. Technological applications for radio frequency identification in the healthcare field seem to grow rapidly. Modernized RFID is also beginning to provide more extensive patient identification than traditional bar coding can and also to track and locate capital equipment within the hospital. As years to come, RFID technology could be used for a variety of applications, thus including tracking and matching blood for transfusions, tracking pharmaceuticals, and combating the counterfeiting of

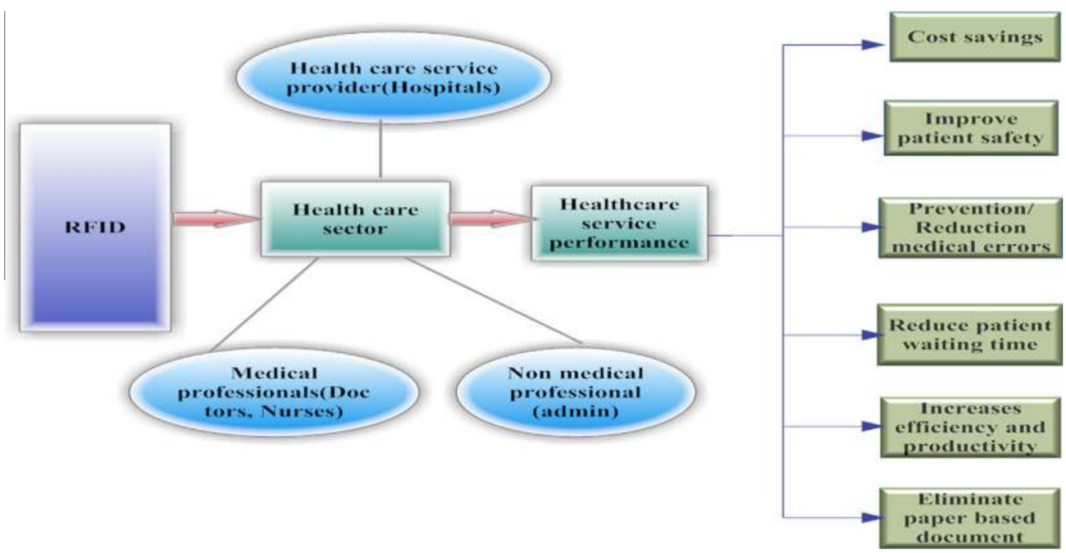

Figure 3 Benefits of using RFID in healthcare sector.

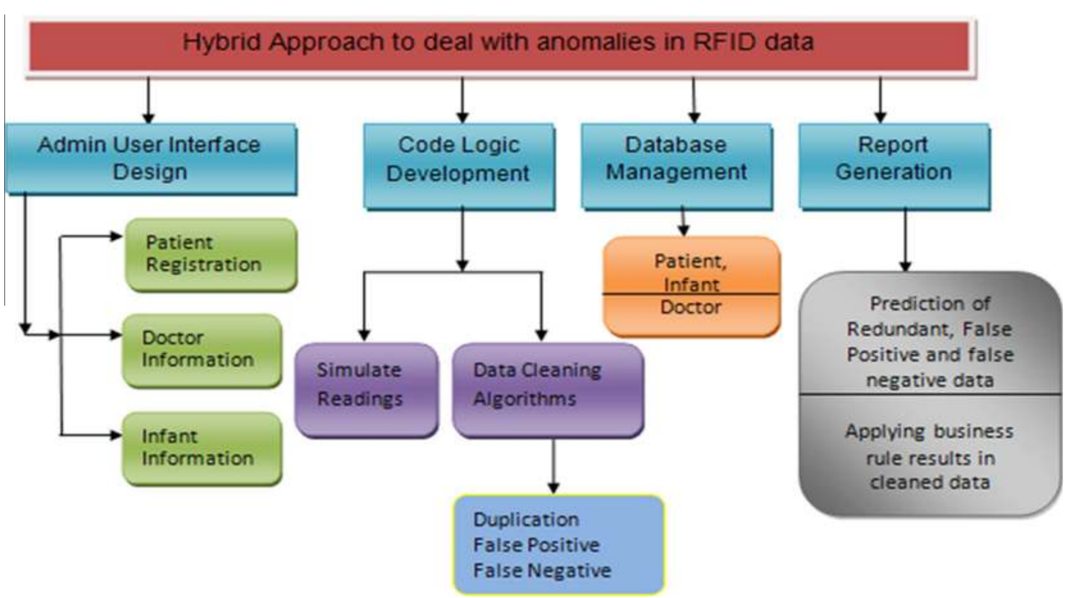

Figure 4 Conceptual diagram of proposed approach.

medical products. RFID system capabilities are powerful and accurate in capturing data. These can easily be integrated into the hospital's wireless infrastructure.

Potential benefits improve patient's safety and cost savings, eliminate paper-based document, increase efficiency and productivity, reduce patient waiting time, prevent/reduce medical errors, and so on, by using RFID technology within healthcare sector. The real value of RFID can be realized when it is integrated with existing HIS. It can provide valuable process-integrated decision support through current medical knowledge. In addition, it can comprehensively use patient data for research and healthcare reporting. Benefits of using RFID in healthcare are given in Fig. 3.

7.1. Advantages of RFID in healthcare industry

RFID systems offer many advantages compared to other identification technologies:

- Deploy RFID in healthcare to build an elegant hospital environment.
- Improves patients' safety by preventing errors.
- Streamlines patient identification process
- Reduces healthcare cost
- Enhances security
- Increases operational efficiency.
- The proposed work predicts and cleans the anomalies in an effective manner.
- Avoids equipment theft and infant theft.
- Facilitates effective work flow management.

7.2. Methodology

The methodology adopted in this paper is a top-down approach also known as step-wise design, essentially the breaking down of a system to gain insight into its compositional subsystems shown in Fig. 4. In a top-down approach, an overview of the system is formulated, specifying but not detailing any first-level subsystems. Each subsystem is then refined in greater detail, sometimes in many additional subsystem levels, until the entire specification is reduced to base elements.

8. RFID detection model

The reader detection model is based on the RFID tag-reader detections regions depicted in Fig. 5 and the three distinct regions of operations of a passive RFID reader tag system are the following:

- Strong-in-field
- Weak-in-field
- Out-of-field regions.

8.1. Strong-in-field region

The tag responds to almost all of the attempts from the reader. Thus, the response rate in the strong-in-field region is very high.

8.2. Weak-in-field region

The tag responds to most of the attempts from the reader and the tag performance then degrades gradually with increasing distance in this field.

8.3. Out-of-field region

The tag hardly responds to any of the attempts from the reader. The response rate tends to become negligible. The detection range in RFID deployment environment plays a significant role where the reader location is complex and overlapping.

9. Our premise

The proposed data cleaning algorithms can be applied to any kind of applications. Role of RFID in healthcare are of more importance because minute errors in it results in heavy financial and personal losses. The readers are the detection nodes deployed in different wards in the Hospital depicted in Fig. 6. Each detection node is identified by a unique ID that serves as the location ID. Reader will monitor the tags within its frequency range [13]. The reason to deploy RFID is

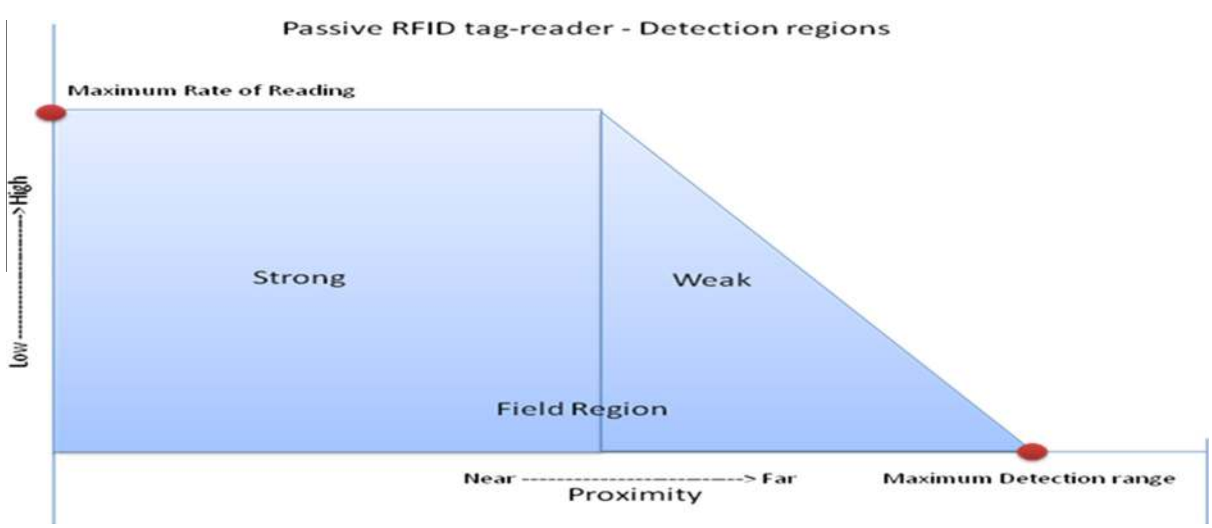

Figure 5 Reader detection model.

Figure 6 RFID system design.

identification, authentication, location, or automatic data acquisition (ADA) [14]. Read/write range is a distance of communication between the reader and tag. The read range is a maximum distance to read data out from the tag and write range is the maximum distance to write data from the reader to tag [15]. RFID tags in different locations are detected by these readers. One of the biggest challenges of the RFID data is the data volume. Sending terabyte data into a centralized system for data cleaning requires a high performance server as well as a high speed network, which will inevitably increase the total hardware cost. Some of the data cleaning methodologies apply to data fetched by the readers, some requires an RFID middleware and others require a centralized data processing server to handle the raw data. The server level data observations include data validations, data inconsistencies, and identification of anomalies before entering the enterprise application database. The business context is dynamic and it is not even framed during the loading of data. Known anomaly duplication is handled in RFID middleware, but the processing of other anomalies is deferred until the query time. Each application specifies its own anomalies by defining cleansing rules.

10. Proposed CBADE – cellular based approach algorithm for duplication detection and elimination

The proposed CBADE algorithm checks whether the tagged object is read by more than one reader at the same timestamp, then duplication is occurred and it is termed as adjacent

Table 1 Sample readings observed by the reader.

Case study	Tag Id	Location	Loc_Id	Date	Time
1	P10004	General ward	102	10/4/2012	5:00:00
2	P10004	Visitors area	107	10/4/2012	5:10:00
3	P10004	General ward	102	10/4/2012	5:10:00
4	P10004	General ward	102	10/4/2012	5:15:00
5	P10004	Consulting area	109	10/4/2012	5:20:00
6	P10004	General ward	102	10/4/2012	5:20:00
7	P10004	Infant ward	105	10/4/2012	5:20:00
8	P10004	General ward	102	10/4/2012	5:25:00
9	P10004	General ward	102	10/4/2012	5:30:00
10	P10004	Pharmacy	110	10/4/2012	5:40:00
11	P10004	Infant ward	105	10/4/2012	6:00:00
12	P10004	Consulting area	109	10/4/2012	6:30:00

Table 2 Analysis of RFID tag readings.

Case study	Status of the reading	Data analysis
1	Normal	Reading for P10004 is from the allotted location at the allotted time and allotted date
2	Crossover parallel	107 is not an adjacent cell but P10004 is read by both 107 and allotted location at the same time and date
3	Normal	Reading for P10004 is from the allotted Location at the allotted Time and date
4	Normal	Reading for P10004 is from the allotted Location at the allotted time and date
5	Adjacent parallel	109 is an adjacent cell and P10004 is read by both 109 and allotted location at the same time and date
6	Normal	Reading for P10004 is from the allotted Location at the allotted time and date
7	Adjacent parallel	105 is an adjacent cell and P10004 is read by both 109 and allotted location at the same time and date
8	Normal	Reading for P10004 is from the allotted location at the allotted time and date
9	Normal	Reading for P10004 is from the allotted location at the allotted time and date
10	Crossover	110 is not an adjacent cell and p10004 is read by 110 and not read by allotted location(102)
11	Adjacent crossover	105 is an adjacent cell and P10004 is read by 105 and not read by allotted location(102) at the same time and date
12	Adjacent Crossover	109 is an adjacent cell and P10004 is read by 105 and not read by allotted location(102) at the same time and date

parallel. This is because of overlapping in the reading vicinity of multiple readers and it is termed reader level duplication. Duplicate readings at the data level occur when an RFID reader keeps reading the same object repeatedly. The proposed CBADE predicts and cleans the duplication in middleware approach. The middleware cleaned data are stored into the database

```
Algorithm CBADE (Reader[ ], Tag [ ])
// Input: Reader R1, Reader R2
// Input: Tag 1, Tag2, Tag3...Tag n
Begin
For (every tag in reader X (X = A, B)) do
if count (Selected Tag_Id) in all Tag_id > 1
Sub (Each Similar Tag_Id timestamp - Select TagId
Timestamp) = 0;
return Duplication detected and anomaly is cleaned;
Select Max (Tag Id Timestamp)
delete other Tag_Ids; // retain current values and delete other
duplicated tuples
else
return No Duplicates;
end for
End
```

11. Proposed R-PFP algorithm for false positive detection and elimination

The middleware cleaned data are stored into the database. Here, the precondition based algorithm R-PFP checks whether the tag is in the allotted location at the specified time. If it is exactly true, then there is no anomaly. With the help of this algorithm, the presence of false positive is detected and cleaned in deferred approach. Precondition is the assumption or set rules that specify an RFID tag and its mobility. It defines the list of all allowed tag-location combinations. Timestamp here defines the assumption or set rule that specify an RFID tag and its validity in a specific region mentioned in precondition with a time bound limit. It defines the list of all allowed time window for a specific tag-location combination.

```
Algorithm False positive(Reader[ ], Tag [ ])
Set = Initial Location of all Tag_Id
NewSet = Null;
While not (Ta_detection) do begin
M= Choose (M); // Finding Adjacent set for the related Loc_Id
M = Crossover (M); // Checking whether Tag_Id is in the Adjacent
Set
M = Alteration(Loc_Id, M, Tag_Id);Mutating the set (Original)
and Adjacent Set
Update NewSet with the Most –ve Coeff;
End;
S1 = Set of data covered by NewSet with return (Newset/ S1);
End
```

12. Proposed R-PFN algorithm for false negative detection and elimination

In the proposed R-PFN algorithm, an initial set of estimates for the parameters is obtained. Given these estimates and the training data as input, the algorithm then finds the missing data. For eliminating false positives and false negatives, business layer checks the status column of tags. If the previous value is 0 and current's value is 1 and next value is 0, then this data are false positive and should be eliminated. Also if previous value is 1 and current value is 0 and next value is 1, then this data are false negative and should be eliminated.

Figure 7 Simulator of healthcare.

- Precondition = 0 → the set rule of tag-location combination is violated
- Precondition = 1 → the set rule of tag-location combination is not violated
- Time stamp = 0 → the set rule of time limit is violated
- Time stamp = 1 → the set rule of time limit is not violated

P10004 HAS 2 REDUNDANT READING(S) DUE TO ADJACENCY + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	Consulting Area	109	10/4/2012 12:00:00 AM	05:20:00
Patient	P10004	Hari	Infant Ward	105	10/4/2012 12:00:00 AM	05:20:00

P10004 HAS 3 CROSSOVER READING(S) + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	Pharmacy	110	10/4/2012 12:00:00 AM	05:40:00
Patient	P10004	Hari	Infant Ward	105	10/4/2012 12:00:00 AM	06:00:00
Patient	P10004	Hari	Consulting Area	109	10/4/2012 12:00:00 AM	06:30:00

P10004 HAS 6 NORMAL READING(S) + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	General Ward	102	10/5/2012 12:00:00 AM	05:00:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:10:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:15:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:20:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:25:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:30:00

P10004 HAS 3 CROSSOVER READING(S) + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	Pharmacy	110	10/4/2012 12:00:00 AM	05:40:00
Patient	P10004	Hari	Infant Ward	105	10/4/2012 12:00:00 AM	06:00:00
Patient	P10004	Hari	Consulting Area	109	10/4/2012 12:00:00 AM	06:30:00

P10004 HAS 6 NORMAL READING(S) + Show - Hide

Tag_Type	Tag_ID	Tag_Name	Loc_Name	Loc_ID	Date	InTime
Patient	P10004	Hari	General Ward	102	10/5/2012 12:00:00 AM	05:00:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:10:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:15:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:20:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:25:00
Patient	P10004	Hari	General Ward	102	10/4/2012 12:00:00 AM	05:30:00

Figure 8 Sample output of the proposed algorithms.

```
False Negative Elimination
Status Flag 0 – not available
Status Flag1 – Available
Input: Reader R1, Reader R2
Input: Tag 1, Tag2, Tag3...Tag n
Begin
For (every tag in reader X (X = A, B) at Time t) do
If (t – 1 = 0; t + 1 = 0; t = 1 & Precondition = 0 OR
Timestamp = 0)
Return "Wrong Data – To be eliminated";
Else if (t – 1 = 1; t + 1 = 1; t = 0 & Precondition = 1 OR
Timestamp = 1)
Return " False negative - To be eliminated";
Endif
Endfor
End
```

13. Simulations

Simulation has long been used as a decision support tool in various sectors. It is especially suited to the analysis of healthcare organizations due to its ability to handle high complexity and variability which is usually inherent in this sector. It also acts as continuous quality improvement framework by integrating with the software agent developed via a database structure. Experimentation of different workflows, staffing decisions, and what-if analysis are all promising applications of simulation in healthcare, and it is practically infeasible in a healthcare environment. Simulation study requires deliberate data collection effort over a considerably long period of time.

14. Experimental results and evaluation

Case Study: For an example, assume a patient Hari is supposed to be in General ward (102) from 5.00 p.m. to 5.30 p.m. and the time reading is captured every 5 min. The sample readings observed by the reader of the patient Hari are given in Table 1. The observed readings are tested with our proposed algorithms and the status of the tag readings is analyzed and it is depicted in Table 2.

14.1. Sample output

RFID data set has been used in our work of which 80% is treated as training data and 20% is considered as testing data. Simulator designed in C# with SQL server 2008 as backend to evaluate the performance of the proposed algorithm is shown in Fig. 7. The outcome of the proposed algorithm is given in the Fig. 8. The occurrence of anomalies in our RFID data is depicted in Fig. 9. Proposed method has been evaluated using the four metrics Precision, Recall, *E*-measure, and Accuracy.

Precision is the ratio of the number of relevant tag readings observed by the reader to the total number of irrelevant and relevant records retrieved. Mathematically,

$$\text{Precision} = \text{True positive } / (\text{True positive} + \text{false positive}) \tag{1}$$

Recall is the ratio of the number of relevant tag readings read by the reader to the total number of relevant records in the database.

$$\text{Recall} = \text{True positive(Normal)}/(\text{True positive(normal)} + \text{false negative(missed reading)}) \tag{2}$$

The Error measure is *E*-measure and it is calculated using the formula,

$$E(p, r) = 1 - 2/([1/p] + [1/r]) \tag{3}$$

Accuracy of an experiment is a measure of how closely the experimental results agree with a true or accepted value. It is calculated as

$$\text{Accuracy} = \text{True positives} + \text{True negative} / \text{True positives} + \text{True negative} + \text{False positive} + \text{False negative}. \tag{4}$$

14.2. Result analysis

We have analyzed the performance of our proposed algorithms with the existing algorithms in terms of precision and recall. We simulate 10,000 samples with 2000 wrong data and see the results of cleaning algorithm.

The proposed CBADE algorithm cleans the data to 97% and the result shows that the proposed algorithm (CBADE) has better execution time and good percentage of cleaned data than SMURF, Bspace, WSTD, and BBS shown in Fig. 10

The proposed R-PFP algorithm is found to have 71% precision and recall 95%. The obtained results shows that the proposed R-PFP algorithm performs best in terms of average precision, recall, *E* – measure and accuracy, which is

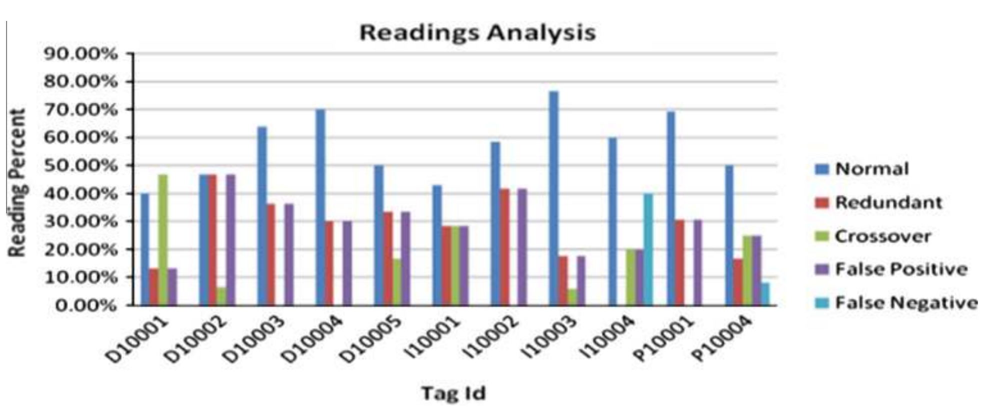

Figure 9 Occurrence of anomalies false positive, false negative, and duplication.

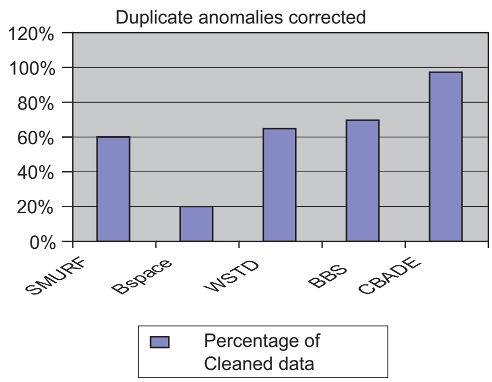

Figure 10 Comparative study of CBADE with existing algorithms.

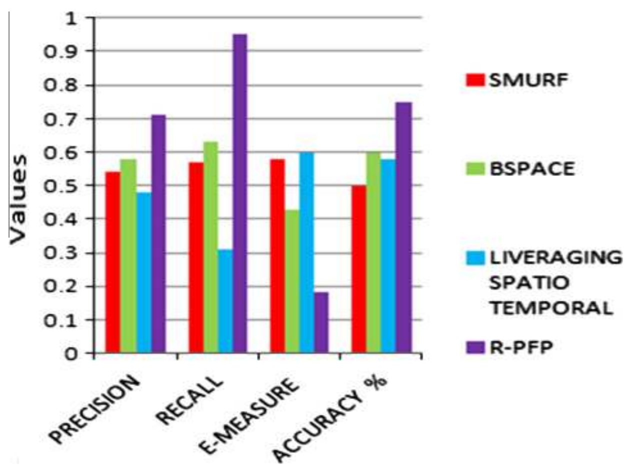

Figure 11 Comparative study of R-PFP with existing algorithms.

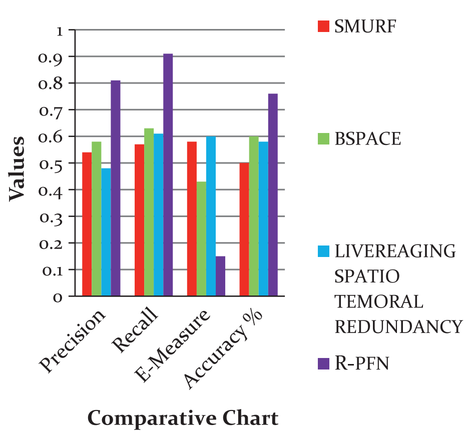

Figure 12 Comparative study of R-PFN with existing algorithms.

significantly superior with all other cleaning algorithms shown in Fig. 11

The proposed R-PFN algorithm is found to have 81% precision, 91% recall, and 76% accuracy shown in Fig. 12. Hence, the obtained results shows that the proposed R-PFP algorithm performs best in terms of average precision, recall, E – measure, and accuracy which is significantly superior to all other clean algorithms.

15. Conclusion

RFID plays an essential role in all the subdomains of the applications in healthcare applications. The effectiveness in cleaning the RFID data in healthcare sectors remains a concern, even though a number of literary works are available. To a maximum, the dirty data that are read may even leads to patients' death. The errors need to be cleansed in an effective manner before they are subjected to warehousing. Current solutions to correct missed readings usually use time window filtering. A serious issue is that a single static window size cannot compensate for missed readings while capturing the dynamics of tag motion. An adaptive time window filtering cannot deal with the condition that tags are always moving. In this paper, we have proposed algorithms to clean the anomalies false positive, missed readings, and duplications. It is decided to record all the values associated with each tag event for future reference otherwise too much valid data will be lost. Finally, the management can analyze the data and filter by applying business rules based on the requirement. The proposed algorithms predict and clean the anomalies based on the integration of middleware and deferred. Our experimental result proved that our algorithms predicts and removes the anomalies in an effective manner compared to the existing works. Thus, it will pave the way for an effective means of data warehousing system that will keep the RFID data safe for future mining.

References

[1] Shoewu O, Badejo O. Radio frequency identification technology: development, application, and security issues. Pac J Sci Technol 2006;7(2).

[2] Kochar Barjesh, Chhillar Rajender. An effective data warehousing system for RFID using novel data cleaning, data transformation and loading techniques. Int Arab J Inf Technol 2012;9(3): 208–16.

[3] Massawe Libe Valentine, Kinyua Johnson DM, Vermaak Herman. Reducing false negative reads in RFID data streams using an adaptive sliding-window approach. Open Access Int J Sens 2012:4187–212.

[4] Meng Lingyong, Yu Fengqi. RFID data cleaning based on adaptive window. In: IEEE 2010 2nd international conference; 2010. p. 746–9.

[5] Manage Data Successfully with RFID Anywhere EdgeProcessing. <http://www.ianywhere.com/developer/rfid_anywhere/rfidanywhere.ed geprocessing.pdf>.

[6] Leema Anny, Hemalatha M. An effective and adaptive data cleaning technique for colossal RFID data sets in healthcare. WSEAS Trans Inf Sci Appl, ACM Digital, Library 2011;8(6): 243–52.

[7] Jeffrey Shawn R, Garofalakis Minos, Franklin Michael J. Adaptive cleaning for RFID data streams. In: ACM 32nd International Conference on Very large data bases; 2006. p. 163–74.

[8] Darcy Peter, Pupinwiwat Prapassara, Stantic Bela. Challenges and Issues Facing the Deployment of RFID Technology. Australia: Institute of Integrated and Intelligent Systems, Griffith University; 2011, p. 1–26.

[9] Hua Fan, Quanyuan Wu, Yisong Lin. Behavior-based cleaning for unreliable RFID data sets. Sens J 2012;12(8):10197–207.

[10] Libe Valentine Massawe, Johnson Kinyua DM, Herman Vermaak. Reducing false negative reads in RFID data streams using an adaptive sliding-window approach. Open Access Int J Sens 2012:4187–212.

[11] Chhillar Rajender Singh, Kochar Barjesh. A new efficient approach for effective Warehousing of RFID data: readers load sentient scheme. Am J Sci Res 2009;4:85–95.

[12] Rao Jun, Duraisamy Sangeetha, Thakkar Hetal, Colby Latha S. A Deferred Cleansing Method for RFID Data Analytics. ACM Very Large Data Base Endowment; 2006. p. 175–86.

[13] Chhillar Rajender, Kochar Barjesh. A new efficient approach for effective warehousing of RFID data. Readers load sentient scheme. Am J Sci Res 2009;4:85–95.

[14] Sabbaghi Asghar, Vaidyanathan Ganesh. Effectiveness and efficiency of RFID technology in supply chain management: strategic values and challenges. J Theor Appl, Electronic Comm Res 2008; 3(2):71–81.

[15] Srinivasa Rao S, Rajan EG, Lalkishore K. Human activity tracking using RFID tags. IJCSNS Int J Comput Sci, Netw Secur 2009;9(1):387–94.