A novel integrated approach for handling anomalies in RFID data

A.Anny Leema,

PhD Scholar, Karpagam University, Coimbatore, India

annyleema@bsauniv.ac.in

Dr.Hemalatha.M,

Head of MSc. Software Systems, Karpagam University, Coimbatore, India

hema.bioinf@gmail.com

ABSTRACT:

Radio Frequency Identification (RFID) is a convenient technology employed in various applications. The success of these RFID applications depends heavily on the quality of the data stream generated by RFID readers. Due to various anomalies found predominantly in RFID data it limits the widespread adoption of this technology. Our work is to eliminate the anomalies present in RFID data in an effective manner so that it can be applied for high end applications. Our approach is a hybrid approach of middleware and deferred because it is not always possible to remove all anomalies and redundancies in middleware. The processing of other anomalies is deferred until the query time and cleaned by business rules. Experimental results show that the proposed approach performs the cleaning in an effective manner compared to the existing approaches.

Keywords:

Issues in data cleaning, RFID data, simulation, Deferred approach

1 INTRODUCTION

The immense attractiveness of RFID technology is ability to store and read data without the reader coming in contact with or in line of sight with the tag [3]. The RFID techniques are appropriate for in-door object monitoring applications [2]. Transponder (tag), Interrogator (reader) and Antenna are the three basic components of RFID. Data stored on a tag is transferred via radio frequency link by RFID tagging which is a form of Automatic Identification and Data Capture (AIDC) technology. By communicating with the tag, the RFID reader infers the identity of the object to which the tag is attached. RFID has many advantages over the more familiar bar codes which transfer data optically, like capability of accumulating huge quantities of data and ability to read several tags concurrently, though their principles are identical [4]. It has its own attractiveness because of no line of sight is required between the reader and the antenna [1]

RFID has penetrated into all the sectors like asset tracking, supply chain automation, medical/Health Care applications, retail, people tracking, Warehouses, Manufacturing, Livestock Timing and the Healthcare sector due to its increased functionality and easy-to-use capabilities. Of these, Medical/Health care applications are of more importance because minute errors in it can cost heavy financial and personal losses. The risk factors of human error are reduced by RFID in

15

the healthcare industry [6]. However, huge quantities of data [7] are created by these devices at high production rates because of the streaming nature of RFID readings [7]. So, an important problem with great potential benefits in the health care sectors is the warehousing and mining of RFID data.

2 DATASET OF RFID

The format of the data recorded in the database after a tag has been read consists of three primary pieces of information: the Electronic Product Code, the Reader Identifier and the Timestamp which contains the reading time. The Electronic Product Code (EPC) is a unique identification number introduced by the Auto-ID Center and given to each RFID Tag which is made up of a 96 bit, 25 character-long code containing numbers and letters. It is made up of a Header for 8 bits, Object Class for 24 bits, EPC Manager for 28 bits and Serial Number for 36 bits. Data quality has become increasingly important to many firms as they build data warehouses and focus more on customer relationship management .Data quality was concerned with timelines, precision, and accuracy.

3 ISSUES IN DATA CLEANING

RFID tags can be used to identify each individual item and enormous amounts of location-tracking data are generated. [8] Time-dependency, dynamic changes, huge quantities and large numbers of implicit semantics are some of the characteristics of the RFID data. Data cleaning is necessary for improving the quality of data so that it becomes "fit for use" by users. It improves the documentation and presentation of data by reducing the errors present in it. Auditing data for discovering discrepancies, selecting transformations to correct these, and employing the transformations on the dataset are the three components present in data cleaning [5].

- Lack of Completeness
  - RFID readers capture only 60-70% of all tags that are in the vicinity
  - Smoothing of data is done to rectify the loss of intermediate messages
- Temporal Nature of data or tag dynamics
  - RFID tags are in motion and that is what makes them more difficult to handle
  - But motion of a tag causes dropping of messages

4 RFID ANOMALIES

There are three types of errors in RFID data reading. They are misreads, unexpected readings and duplicate readings. False negatives (missed readings), tags are mistakenly assumed to be absent in the reader's detection range; False positives, tags are mistakenly assumed to be present in the reader's detection range. Duplicate readings are classified into reader duplicates and data duplicates. The former occurs when a tag is present in vicinity of more than one reader which is simultaneously sending signals to it; the latter occurs when a reader reads a large amount of non-difference information at a time interval. The reliability of current RFID systems is far from optimal. Unless they are filtered and cleaned it cannot be applied for high end applications. Therefore novel data cleaning techniques are proposed adapting to the diversity of conditions found in a large RFID implementation and able to clean large data sets with high accuracy.

5 PROPOSED SCENARIO

We have intended our methodology for a scenario in which readers R1 to R10 are mounted around a healthcare environment and tags are passing through the area to be scanned. Doctors, staff nurses, infants, surgical patients, equipments and records are monitored. It is crucial that a known environment is used in the scenario as the geographical locations of each of the readers and their proximity to one another must be recorded in the system.

6 SIMULATIONS IN HEALTHCARE

Simulation 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.[6] Experimentation of different workflows, staffing decisions and what-if analysis are all promising applications of simulation in healthcare.[9,10] So our work consists of three steps given below and sample parameters are framed for all the services. For each services assumptions and set of rules are framed.

1. The first step is to simulate the entire operation in hospital and to provide the following services.
   (i)Patient Identification & Tracking
   (ii) Patient Record Maintenance
   (iii)Work flow Management
   (iv)Resource management
    (v) Avoiding Equipment Theft
   (vi) Prevent Infant Theft
    (vii)Inventory management

2. The second step deals with the secure fetching and management of online tag associated information.
3. The Third area is related to removal of errors that are present in RFID because of error inaccuracies.

7 ARCHITECTURAL DESIGN

Raw RFID data stream is a spontaneous and very complex data to use for any analysis. The Readers are the detection nodes and are deployed in different locations across various places in the Hospital. Each detection node is identified by a unique ID that serves as the location ID. Edges between the detection nodes record the history of item transitions between locations. RFID tags in different locations are detected by these readers. The user must register the information about the tag and reader in the database. Each tag contains an unique identifier. The information that has to be stored in the database about the tag are Tag ID and to whom it is assigned. The information in the tag can be accessed only by the reader. So the database also includes the details about the reader such as Reader name and loc id(reader id). The architectural design of our proposed work is depicted in fig:1
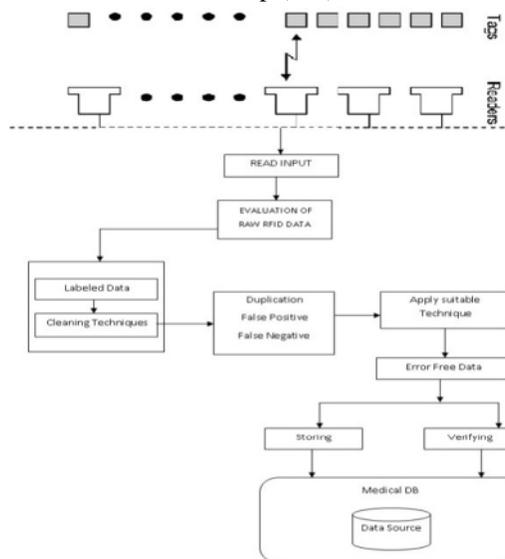
Fig:1 Depicts the architectural design of our proposed work

8. SOFTWARE IMPLEMENTATION

The entire project is divided into four modules. The four modules include Admin User Interface Design, Code Logic Development, Database Management and Report Generation.

8.1 Admin User Interface Design:

Admin User Interface is the first point of interaction for anybody coming to the Hospital. It has all the information of the patients, doctors, departments and activities of the Hospital. In the enhanced mode this facility will be available on web. All information available here are in real time like client registration and any enquiry about the patient status, Room Status or Doctors availability since the data is constantly updated.

8.2. Code logic Development:

Code Logic focuses on providing the most important aspects of code analysis to allow a developer to realize productivity gains immediately. Here we propose a new effective and accurate data cleaning technique. We assume that there is a hidden process that determines the true location of a tag, but we only see noisy observations of the hidden process.

8.3. Database Management:

A database is an integrated collection of data records, files, and other objects [1]. A DBMS allows different user application programs to concurrently access the same database. DBMSs may use a variety of database models, such as the relational model or object model, to conveniently describe and support applications. It typically supports query languages, which are in fact.

8.4. Report Generation:

The construction of a healthcare document consists of the formatting and/or structuring of captured information through RFID technology. It is the process of analyzing, organizing, and presenting recorded patient information from the RFID data for authentication and inclusion in the patient's healthcare

9 DATA CLEANING APPROACHES
9.1 Physical approaches

The success of an RFID implementation depends largely on the extensive preparedness made on all environmental factors at the time of planning. This includes determining area surveys, Feasible Reader positions, RF signal coverage, Obstructions etc,. This enhances the entire RFID environment. Focus needs to be on flawless hardware installation and seamless connectivity. It is found that the Reader may scan the tags on an object most effectively when the Tag is positioned at the front.
The common solution suggested to deal with RFID anomalies is to either install multiple readers in a specific location or to attach multiple tags with same EPC code.

9.2. RFID Middleware place an important role in any RFID implementation.

RFID middleware, simply put, is a software layer residing between the RFID hardware and the existing back-end system or application software. It extracts data from the RFID interrogators (readers), filters it, aggregates it and routes it to enterprise applications. Middleware components of RFID is depicted in fig:2
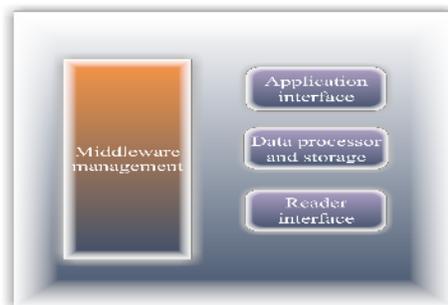

Fig 2. Middleware components of RFID

9.3. Deferred Approach

In deferred approach cleaning is limited after storing the data into the database and anomalies are not cleaned properly.

10 PROPOSED APPROACH

The conventional data cleaning approach is to remove all anomalies upfront and to store only the cleaned data in database. Such cleansing methods can potentially reduce the amount of data that have to be managed by the application and helps to avoid repeated cleansing of the data at query

time. However, it is not always possible to remove all anomalies and redundancies beforehand. One reason is that the rules and the business context required for cleansing may not be made available at the data loading time. On the other hand the HRMS application defines its own set rules and process specifications (say..on how long the tag is allowed to stay in a particular location). Maintaining multiple cleaned versions of RFID data are prohibitive and worthless when the rule sets defined by the application is dynamic in nature. These reasons make the physical and middleware approach infeasible.

We propose the integrated approach of middleware and deferred depicted in fig:3. Known anomalies whose detection and correction are handled by the RFID middleware and 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 when the application issues queries. In this approach, although an application pays cleaning overheads, it gains flexibility of being able to evolve as a dynamic and effective cleaning approach.
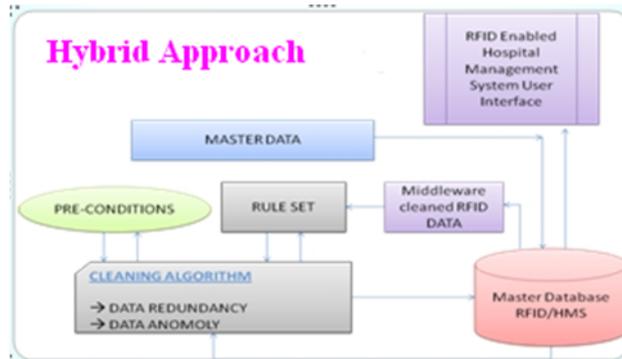

Fig: 3 Proposed Integrated Approach

11. EXPERIMENTAL RESULT

In this section, we have presented the experimental results of the proposed deferred approach. The proposed approach is implemented in C# with SQL server 2008 as backend. The RFID reader is modeled based on the design features of SkyeTek's M1 Mini RFID reader. This reader operates from a Lithium rechargeable battery which has 0.48 KJ of energy. The tag to reader data rate is taken as 26 Kbps as per ISO 15693.

Simulation has long been used as a decision support tool in various sectors. There are several examples of simulation studies of healthcare facilities in the literature. Simulation 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. [5] 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.

The entire process in hospital is simulated using RFID, then the raw RFID tag readings are collected and low complex anomalies are cleaned in the middleware. As a final step using our

proposed algorithms the high complex anomalies like false positive and false negative are predicted. Using human intervention final decision can be taken whether to retain the data for future reference or cleaning can be done to avoid unnecessary values.

11.1 R-PFP- Algorithm for Prediction of False positive in RFID data

```
Alg PFP(Number: m, Dimensionality: k)
begin
S = Initial Seed Population of p strings;
BestSet = null;
while not(termination criterion) do begin
S = Selection(S);
S = CrossOver(S);
S = Alteration(S; loc_id,tag_id);
Update BestSet  with most negative parity coefficients;
end;
O = Set of data points covered by BestSet;
return(BestSet, O); end;
```

11.2 Algorithm-Prediction of false positive and false negative based on Preconditions

```
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
```

11.3. Sample Coding

```
d1 = "select * from RFID_READING where Tag_ID='" + TagIDDropDown.SelectedValue + "' and
Loc_ID='" + TextBox1.Text + "' and  Time between '" + t1 + "' and '" + t2 + "'"; cmr = new
SqlCommand(d1, con);
  dgr = cmr.ExecuteReader();
  if (dgr.HasRows)
      {
        //dg.Read();
        dgr.Close();
```

```
        d2 = "update RFID_READING set Status='Normal' where Tag_ID='" +
TagIDDropDown.SelectedValue + "' and Loc_ID='" + TextBox1.Text + "' and Time between '" + t1 + "'
and '" + t2 + "'";
        cm8 = new SqlCommand(d2, con);
        cm8.ExecuteNonQuery();
      }
    if (dgr.HasRows)
      {
        //dg.Read();
        dgr.Close();
        d2 = "update RFID_READING set Status='Crossover' where Tag_ID='" +
TagIDDropDown.SelectedValue + "' and Loc_ID='" + TextBox1.Text + "' and Time not between '" + t1 +
"' and '" + t2 + "'";
        cm9 = new SqlCommand(d2, con);
        cm9.ExecuteNonQuery();
      }
dgr.Close();
```

11.4 Sample Output

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

The chart (fig:4) depicts the percentage of occurrence of false positive and fig:5 shows the percentage of false negative in the generated RFID readings.

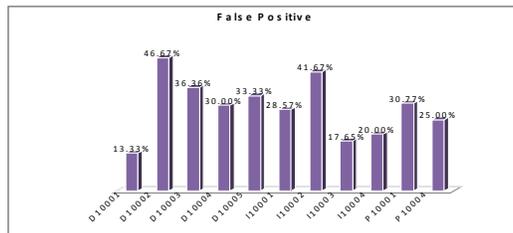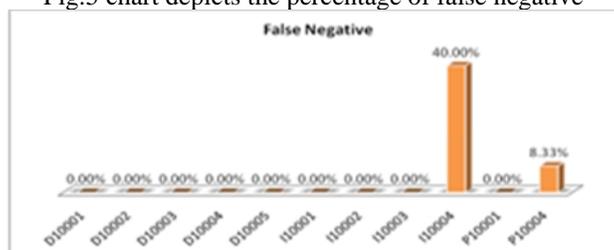Fig: 4 chart depicts the percentage of false positive.

Fig:5 chart depicts the percentage of false negative

12 CONCLUSION:

The effectiveness in cleaning the RFID data in healthcare sectors remains a concern, even though a number of literary works are available. The effective data warehousing involves collecting data without any losses. If any data is missed out, there is a possibility for the decisions taken by that warehouse to go wrong. The cleaned data is applied for any high end applications.

In this paper first we have simulated the entire process in hospital using RFID, second we focused on the dirty data read by the readers and performed the cleaning to deal with the anomalies false positive and false negative.

Major barriers to RFID adoption are identified as prohibitive costs, technological limitations, and privacy concerns. By improving perceptions, safety, reliability, user-friendliness, accessibility, and privacy in healthcare or any organizations it is expected that the demand for RFID will grow.

Integration of middleware and deferred approach gains the flexibility of being able to evolve as a dynamic and effective cleaning approach. Finally, solutions are validated with a common RFID application and demonstrated the advantages of the proposed approach through extensive simulations. Our proposed algorithm implemented using C# and tested in the healthcare environment to increase the operational efficiency and to improve the patient safety.

In future the existing system is easily enhanced to launch in online so that the public is benefited based on the requirements.

13 REFERENCES

[1]   Barjesh Kochar, Rajender Chhillar. 2012. An Effective Data Warehousing System for RFID Using Novel Data Cleaning, Data Transformation and Loading Techniques. The International Arab Journal of Information Technology.9(3):208-216.

 [2]   Tseng, Vincent S., Lu, Hsueh-Chan, Tsai, Chia-Ming, Wang and Chun-Hung, "A Hybrid Data Mining Approach for Analysis of Patient Behaviors in RFID Environments", 2008.

[3].   H. Gonzalez, J. Han, X. Li and D. Klabjan, "Warehousing and Analysis of Massive RFID Data Sets", In Proceedings of 2006 Int. Conf. Data Engineering (ICDE'06), Atlanta, Georgia, April 2006.

[4].   Angela M. Wicks, John K. Visich and Suhong Li, "Radio Frequency Identification Applications in Hospital Environments", heldref publication, Vol. 84, No.3, pp.3-9, 2006.

[5]. Vijayshankar Raman and Joseph M. Hellerstein, "Potter'sWheel: An Interactive Data Cleaning System", In Proceedings of the 27th International Conference on Very Large Data Bases, pp.381 - 390, 2001.

[6]. S. Shepard, "RFID Radio Frequency Identification", The McGraw-Hall Companies, Inc. USA, 2005.

[7] Bettina Fazzinga, Sergio Flesca, Elio Masciari and Filippo Furfaro, "Efficient and effective RFID data warehousing", In proceedings of the 2009 International Database Engineering & Applications Symposium,pp. 251-258 , 2009.

[8]. Hector Gonzalez, Jiawei Han, Hong Cheng, Xiaolei Li, Diego Klabjan, Tianyi Wu, "Modeling Massive RFID Data Sets: A Gateway-Based Movement Graph Approach," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 1, pp. 90-104, Jan. 2010.

[9] Anny Leema, A. and M. Hemalatha, 2011.Optimizing operational efficiency and enhancing data reliability using effective and adaptive cleaning approach for RFID in healthcare. International Journal of Computer Applications.3(7):26-29.

[10] Anny Leema and M.Hemalatha, 2011. An effective and adaptive data cleaning technique for colossal RFID data sets in healthcare. WSEAS transactions on Information Science and Applications .8(6) : 243-252. SCOPUS indexed