Optimization of Query Processing Time using Taguchi Method for RFID Data Management

V. Maria Anu*, R. Aroul Canessane and L. Mary Gladence

Faculty of Computing, Sathyabama University, Chennai - 600119, Tamil Nadu, India; mariaanu18@gmail.com, aroul-canessane@gmail.com, lgladence@gmail.com

Abstract

**Background:** RFID technology is being adopted by a number of applications including supply chain management. Due to the adoption of RFID technology challenging problems have appeared. The RFID database system handles huge amounts of path oriented, time dependent data. **Methods:** This paper focuses on minimizing the query processing time by using Taguchi optimization method. Simulations show how the factors considered impact RFID data processing efficiency. **Findings:** So the effectiveness of the system depends on the factors such as the number of tags accessed at a time, the data preprocessing techniques adopted, and selection of suitable indexing techniques and so on. However, few of the techniques adopted by existing methods focus on the efficiency of RFID data processing. **Applications/Improvements:** Taguchi method is applied to predict the best design combinations that achieve minimum query processing time.

**Keywords:** Data Compression, Data Indexing, RFID, Taguchi Method

1. Introduction

RFID technology is being used in the following fields. Supply Chain Management: Real-time inventory tracking, Retail: Active shelves, monitor product availability, Access control: Toll collection, transportation, Anti counterfeiting and security, Airline luggage management1-4. A good example of RFID technology being used is supply chain management. In supply chain management. An RFID tag is attached with every product. When the reader identifies an RFID tag it makes an entry in the form of (TAG_ID, LOC, and TIME).

If the product stays in a particular region for a long time it creates multiple entries. These multiple entries can be compressed into a single record of the form (TAG_ID, LOC, START_TIME, END_TIME) where TAG_ID represents the tag identifier, LOC the location, START_TIME the time when the tag enters the location, and END_TIME the time when the tag leaves the location.

RFID data has the following characteristics: Large volume, Inaccurate Data, Implicit Semantics, and Temporal Oriented.

Since RFID data have a temporal property, it is difficult to model RFID data by using the traditional ER model. In5 proposed a new warehousing model which uses compression in order to reduce the join cost. If products do not move together in large groups the compression technique cannot produce fruitful results.

An efficient storage scheme and query processing for RFID data in supply chain management was proposed by6,7, cannot be applied when the path length is large. Long paths are common, so the work of Lee and Chung has a practical difficulty that it requires on a special-purpose DBMS which supports very large numbers.

In8, proposed a method to divide a long path. Instead of modifying RDBMS or adopting a special purpose database engine the paths can be divided. But this results in computation overflow. The path encoding scheme proposed by Lee and Chung focused on ancestor, descendant relationships in a query.

Before processing query related to RFID data the volume data can be reduced. There are number of techniques are available for dimensionality reduction9. The
efficient dimensionality reduction techniques are used to reduce the size of the dataset. The main aim is to reduce the volume of massive amount of RFID data. The trace records are helpful in retrieving the data according to the type of retrieval queries. Such trace records are formed by prescribed in limited format. Despite the inclusion of semantics in query processing, the limited formats shall lead to complex processing. In contrast, it is practically infeasible to develop such a generalized encoding methodology and storing in the database. Unless a compromise is made between the two limits, the complexity reduction in processing such RFID data remains a big challenge\footnote{10}. Using orthogonal transformation the volume dataset is represented by principal components of the data and then grouped using Genetic Algorithm (GA). By preprocessing data the query processing time is reduced, which increases the efficiency of the RFID data processing system\footnote{11}. It is difficult to process the raw RFID data as such, so in data preprocessing the size can be reduced using Advanced Principal Component Analysis (APCA). The results are compared with PCA.

2. Proposed Work

When each product is tagged with RFID tags and these tags are recognized by the reader. And each RFID tagged product makes an entry in the reader, which is of the form (TAG_ID, LOC, and TIME). As the price of the RFID tag is very cheap, we can make use of it in supply chain management, which will be very useful for product tracking. Imagine when a container carries tons of products with more than 10000 RFID tagged products, and if this container crosses the reader it generates a huge amount of RFID data. This collection of RFID data, called as raw RFID data. (Figure 1. (a)) shows a sample collection of raw RFID data.

And if the container stays in a particular location for a long time, each RFID tag makes multiple entries for different time periods. Same TAG_ID and same Location. From the raw RFID data it is easy to generate stay records.

Each stay record of the form (TAG_ID, LOC, TIME-IN, TIME-OUT) (Figure 1. (b)). By finding the stay records the size of raw RFID data is reduced. Thus leads to data compression.

The path traveled by a particular Tag _Id can be represented using trace records (Figure 1. (c)).

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(a) Raw RFID data. (b) Stay records. (c) Trace records.}
\end{figure}

2.1 Architecture

In our work (Figure 2.) the RFID data is preprocessed using dimensionality reduction techniques like PCA and APCA. And query processing time is calculated for different data sets. Different types of queries are executed and their average query processing time is tabulated. And for the same dataset different data indexing techniques are applied and the query processing time is listed for different scenarios. Finally, using Taguchi method factors best suited for efficient RFID data processing are identified\footnote{10,11}.

2.2 Data Reduction using PCA

The benefit of PCA is to decrease the intricacy in translation that can be created by the large number, inter-related variables that can be diminished by utilizing just the initial few principal components. The dimensionality
reduction ought to be applied to the raw data to decrease the volume of the database. The orthogonal transformation given database size is decreased and subsequently the reduced database size values are used in the optimal path encoding process.\textsuperscript{12,13}

![Figure 2. Architecture of RFID data management.](image)

The essential steps to compute the PCA,

- The mean of the Dataset is computed by the given equation:
  \[ \overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \]  
  \[ (1) \]

- Compute the covariance matrix:
  \[ C_m = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T \]  
  \[ (2) \]

- Eigenvectors and Eigenvalues are computed using VC_{m,n} = \lambda V, V: Set of Eigenvectors \lambda: Eigenvalue.

- Arrange the computed Eigenvectors from bigger value to small value.

- Apply projection operation for the Eigenvector.

- Assume mean as a center and the Eigenspace is calculated using the given equation:
  \[ S_i = V_i^T (X_i - \overline{X}) \]  
  \[ (3) \]

- Populate the Database with the compressed values for further computation.

### 2.3 Data Reduction using APCA

The Advanced Principal Component Analysis (APCA) is an another compression technique used for dimensionality reduction. The modified projection and training method is used to diminish the response time. APCA is a straight forward technique and it is extremely versatile in processing and interpretation. Some of the initial steps are common for PCA and APCA.

The steps followed to perform APCA dimensionality reduction are,

Assume X be the training set of all database and X can be described as \( X = \{X_1, X_2, ..., X_p\} \)

The training sets of Kth databases
\( X_k = \{X_k^1, X_k^2, ..., X_k^p\} \)

Ni be the number of variables in the Kth database.

- The mean of the dataset is calculated by the given equation:
  \[ \overline{X}_k = \frac{1}{N_k} \sum_{m=1}^{N_k} X_m \quad m = \{1, 2, 3...p\} \]  
  \[ (4) \]

- The Covariance is computed by the given equation:
  \[ B_{sk} = \frac{1}{N_k} \sum_{m=1}^{N_k} (X_m^k - \overline{X}_k) (X_m^k - \overline{X}_k)^T \]  
  \[ (5) \]

- The Covariance Matrix is named as B_{sk} where v is the largest Eigenvalue \( e_j^k \) which takes the following values \( j = 1, 2, 3...v \).

- Transformation matrix is calculated by the given equation:
  \[ T_{sk} = \{e_1^k, e_2^k, ..., e_v^k\} \]

- The training data sets are projected to the corresponding to the Eigen subspace.
  \[ E_m^k = T_{sk}^T (X_m^k - \overline{X}_k) \]  
  \[ (6) \]

- Eigen matrix of the Kth Database is computed by the given equation:
  \[ W_k = \{W_1^k, W_2^k, ..., W_N^k\} \]

- Project the testing dataset to Eigen subspace for each database.
  \[ SS_k = \{SS_1^k, SS_2^k, ..., SS_N^k\} = T_{sk}^T (Q - \overline{X}_k) \]  
  (Assume \( k = 1, 2, ..., p \))

- SS_k is the projection vector in the kth Eigensubspace.

- Calculate the reconstructed Database of the kth position database.
  \[ Q_k = T_k E_k = \sum_{j=1}^{v} W_j^k e_j^k \]
The Euclidean distance between the testing database and reconstructed database is calculated using the given equation:

$$ED_k = \frac{|Q_s - Q_{s_k}|}{Q_s}, \quad k = \{1, 2, 3, \ldots p\}$$

After applying APCA the original database is replaced with reconstructed database. Comparison results show APCA compression technique is better than traditional PCA method.

### 2.4 Data Indexing for RFID Data

In RFID data management there are so many research problems during database storage also due to the high volume of data. Thus more computation and time is needed to take out valuable information from RFID data storage for the process of object tracking in supply chain management. In the RFID storage management many multidimensional dimensional indexing techniques are available. We have considered Bitmap indexes, R Trees and R+ Trees for our RFID data management system.

### 3. Taguchi Method

The Taguchi method helps us to design the high quality system. From the limited set of experiments Taguchi method finds the different parameters that affect the performance characteristics. The following are the phases of the Taguchi methodology:\(^1\)

#### 3.1 Planning Phase
- Determine the Factor.
- Identify the test condition.
- Identify noise and control factor.
- Orthogonal Array Design.
- Data Analysis Procedure.

#### 3.2 Experiments Phase
- Conduct designed experiments.

#### 3.3 Analysis Phase
- Individual factor distribution.
- S/N Analysis.
- Response table analysis.

Table 1 is the list of factors affecting the RFID data management system. We have considered the volume of tags as one of the factors. The volumes of tags are varied in three different levels small, medium and large. Second factor which affects the system is the data preprocessing method. We have compressed the data using Principal Component Analysis (PCA) and Advanced Principal Component Analysis (APCA). We have chosen the raw RFID data as one of the levels to find out the impact of raw RFID data and compressed RFID data. Third factor is the type of the indexing technique. There are various data indexing techniques are available. We have considered bitmap indexes, R tree and R+ tree indexing techniques from the literature review.

| Query Processing Time Parameters | Level 1 | Level 2 | Level 3 |
|---------------------------------|---------|---------|---------|
| Volume of Tags                  | Small   | Medium  | Large   |
| Data Compression Technique      | PCA     | APCA    | Raw     |
| Indexing                        | R+ tree | R tree  | Bitmap Indexes |

Table 2 describes the orthogonal array for Taguchi Method. We have considered three factors so we need to conduct 27 experiments. Table 2 describes L9 array for RFID data management.

The calculations to arrive S/N ratios are given in the following equation.

$$\eta_p = -10 \log_{10} \left( \sum y^2 \right)$$

$$\eta_p = -10 \log_{10} \left( 2.498^2 \right)$$

$$\eta_p = -7.952$$

Similarly, \( \eta_p \) is calculated and tabulated in Table 1. The effect of each factor on the response is calculated using following relation.

$$mA_{pi} = \frac{1}{3}(-7.952 + 0.896 + 4.013)$$

$$mA_{pi} = -1.014$$

Similarly, effect of factor A and B with different levels are calculated and the response graph is drawn as shown in Figure 3.

The optimum combination of minimizing the processing time is given by the following equation.

$$\eta_{opt} = \mu + (mA_A - \mu) + (mA_B - \mu) + (mA_C - \mu)$$

$$\eta_{opt} = -9.93 + (-17.832 + 9.93) + (-15.016 + 9.93) + (-11.052 + 9.93)$$

$$\eta_{opt} = -1.014$$
Table 2. Query processing time parameters and their levels with their responses

| Trial No. | Volume of Toys (A) | Data Compression Technique (B) | Indexing (C) | Response | $\mu P$ |
|-----------|--------------------|--------------------------------|--------------|----------|---------|
| 1         | 1 (A1)             | 1 (B1)                         | 1 (C1)       | 2.498    | -7.952  |
| 2         | 1 (A1)             | 2 (B2)                         | 2 (C2)       | 0.902    | 0.896   |
| 3         | 1 (A1)             | 3 (B3)                         | 3 (C3)       | 0.63     | 4.013   |
| 4         | 2 (A2)             | 1 (B1)                         | 2 (C2)       | 6.932    | -16.817 |
| 5         | 2 (A2)             | 2 (B2)                         | 3 (C3)       | 2.547    | -8.121  |
| 6         | 2 (A2)             | 3 (B3)                         | 1 (C1)       | 2.478    | -7.882  |
| 7         | 3 (A3)             | 1 (B1)                         | 3 (C3)       | 10.329   | -20.28  |
| 8         | 3 (A3)             | 2 (B2)                         | 1 (C1)       | 7.348    | -17.323 |
| 9         | 3 (A3)             | 3 (B3)                         | 2 (C2)       | 6.232    | -15.893 |

The average response in terms of the S/N ratio of each level of all factors for minimization is given in Table 3. The optimum level for all levels is selected on the maximum average S/N ratio only, irrespective of the problem (minimization).

Table 3. Average response in terms of S/N ratio for minimization

| Factor                  | Level 1       | Level 2       | Level 3       |
|-------------------------|---------------|---------------|---------------|
| Volume of Tags           | - 1.014       | -10.94        | (Min) 17.832  |
| Data Compression Technique | (Min) - 15.016 | -8.183        | - 6.587      |
| Indexing                | (Min) -11.052 | -10.605       | -8.129        |

Hence the best level to minimize the processing time, the Factors are A3, B1 and C1.

4. Results and Discussions

Experiments for managing RFID data are conducted in MATLAB with the following machine configuration

Processor : Intel Core i5
CPU speed: 3.20 GHZ.
Operating System : Windows 7.
RAM: 4GB.

Experiments for the different factors impacting system efficiency are conducted by generating four synthetic data sets. And the query performance is measured by the average execution time. According to the Taguchi method of optimizing performance, the following factors provide minimum query processing time. Even if the volume of tags is high when we apply Principal Component Analysis for data reduction along with a R+ Tree indexing technique we are able to achieve the minimum query processing time. In future number of factors can be increased so that the RFID data management system analyzed better.

5. References

1. Sundaram BR, Vasudevan SK, Aravind E, Karthick G, Harithaa S. Smart car design using RFID. Indian Journal of Science and Technology. 2015 Jun; 8(11):1–5.
2. Blessingson WJ, Jinila YB. Multi utility/tracing kit for vehicles using RFID technology. 2010 Proceedings of the Recent Advances in Space Technology Services and Climate Change, RSTCC; Chennai. 2010. p. 273–6.
3. Bollampally K, Dzever S. The impact of RFID on pharmaceutical supply chains: India, China and Europe Compared. Indian Journal of Science and Technology. 2015 Feb; 8(54):176–88.
4. Kumari S. Real time authentication system for RFID applications. Indian Journal of Science and Technology. 2014 Mar; 7(S3):47–51.
5. Gonzalez H, Han J, Li X, Klabjan D. Warehousing and Analysis of Massive RFID Data Sets. ICDE ’06, Proc 2006 Int’l Conference on Data Engineering; 2006. p. 1–18.
6. Lee C-H, Chung C-W. RFID data processing in supply chain management using a path encoding scheme. IEEE Transactions on Knowledge and Data Engineering. 2011; 23(5):742–58.
7. Lee C-H, Chung C-W. Efficient storage scheme and query processing for supply chain management using RFID. Proceedings of the ACM SIGMOD International Conference on Management of Data; 2008. p. 291–302.
8. Gonzalez H, Han J, Cheng H, Li X, Klabjan D, Wu T. Modeling massive RFID data sets: A gateway-based movement graph approach. IEEE Transactions on Knowledge and Data Engineering. 2010; 22(1):90–104.
9. Fodor IK. A survey of dimension reduction techniques. USA: Lawrence Livermore National Laboratory; 2002.
10. Anu VM, Mala GSA. RFID data encoding scheme in supply chain management with aid of orthogonal transformation and Genetic Algorithm (GA). IRECOS. 2013; 8(11):2562–9.
11. Anu VM, Mala GSA, Mathi K. Comparison of RFID data processing using dimensionality reduction techniques. The Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies, ICCICCT 2014; Kanyakumari. 2014. p. 265–8.
12. Smith LI. A tutorial on principal components analysis. USA: Cornell University; 2002.
13. Ghodsi A. Dimensionality reduction a short tutorial. Ontario, Canada: University of Waterloo; 2006.
14. Tsui KL. An overview of Taguchi method and newly develop statistical methods for robust design. IIE Transactions. 1992; 24(5):44–57.