

Efficient Method for Identifying Location and Removal of Data Redundancy for RFID Data

Prof. K Srinivas Rao¹, Dr. K R Chandran²

¹Hyderabad Institute of Technology and Mgt. Department of CSE, Hyderabad, India

Email: ksreenu2k@gmail.com

²PSG College of Technology, Department of IT, Coimbatore, India

Email: chandran_k_r@yahoo.co.in

Abstract—Radio Frequency Identification (RFID) Technology has gained a significant momentum, in the past few years. This technology plays an essential role in object tracking in supply chain management applications and library management applications. RFID Technology holds the promise to simplify baggage handling, clinical laboratory procedures and other tasks by providing the automatic identification and data capture. In near future, it is expected that RFID will replace barcode system. The large volume of data, resulting from the fast capturing multiple RFID readers poses challenges for data management. The captured data may be having some errors such as data duplication or miss reads. The techniques which are available are unable to eliminate RFID data duplication. In this paper we propose an efficient method for eliminating duplicates for RFID data from multiple RFID readers. This is proposed using set theory fundamentals. Our experimental results demonstrate that the proposed method has better quality in terms of utility and feasibility.

Index Terms—Location Identification, Multiple RFID Readers, Relative Complement Algorithm.

I. INTRODUCTION

RFID is a technology that allows a sensor (RFID reader) to read, from a distance and without line of sight, a unique identifier that is provided (via a radio signal) by an inexpensive tag attached to an item. RFID offers a possible alternative to barcode identification systems and it facilitates applications like item tracking and inventory management in the supply chain. The core RFID technology is not new, and it can be traced back to World War II where it was used to distinguish between friendly and enemy aircrafts. [10], [11], [12], [14], [16], [19], [21].

However, captured data has significant percentage of errors particularly as a result of miss reads. To reduce the miss reads multiple readers can be applied, which can result in duplication of data. To eliminate duplication of data several techniques for filtering RFID data at the edge (before it gets stored in to the database) have been proposed in the literature [6],[9],[17],[18].

The most common error is duplication of data. These duplicated data can be removed but the problem is which data should be kept and stored into the data base to indicate the right location. To reduce a false reading and to increase the accuracy of data, it is advised that at least two readers need to cover one area [20], which increases the probability that the tag will be recorded.

In this study we focus on the library scenario, where in line with recommendations [20] and to increase the data

capture accuracy each bookshelf is covered by at least two readers. Because each tag is captured by multiple readers, which cover several shelves, tag must be allocated to the right location to the specific shelf where it is actually located, in this work, an efficient method for indentifying exact location and removal of data redundancy for RFID data streams are proposed.

Our method will promise that the output data has better quality and at the same time ensures that the correct location of the tag can be identified. In experimental study we show that our method can provide the non-redundant data of better quality of output data comparing to previously used methods.

The rest of the paper is organized as follows. Section 2, background information related to error types and their causes is discussed. Section 3, Identifying location methodology is introduces. Sections 4, duplication of data removal for RFID data streams algorithms have been proposed. Section 5 we present experimental performance results. We discuss analysis and conclude our study in Section 6.

II. BACKGROUND

RFID is a technology that allows a sensor to automatically identify objects. RFID data, can be accumulated very fast and does not carry much information as it is raw. These data are not meaningful and are inaccurate. They need to be filtered before they get stored in database in order to reduce data volume and improve the quality of data.

A. Errors in RFID Data Streams

Due to the low power and low-cost characteristics of RFID tags, reliability of RFID readings is of concern in many circumstances [8]. There are four main types of errors in RFID data reading they are Unreliable readings, misreads, additional – unexpected readings and duplicate readings.

Unreliable readings: Individual sensor readings are often unreliable. For example, RFID readers tend to fail but continue to output faulty values. In an RFID deployment, environmental interference such as metal or water can sometimes cause unreliable readings, moving tags is also a common issue that causes unreliable readings.

Miss-reads: These problems are very common in RFID applications and often happen in a situation of low-cost and low- power hardware, which result in frequently dropped reading [13]. Miss readings usually happen when multiple tags are to be detected and Radio Frequency collisions occur and signals interfere with each other, preventing the reader from correctly identifying tags [6].

Additional unexpected readings: These are caused by the RFID tag outside the normal reading scope of a reader, captured by the reader or unknown reasons from the reader or environment [6]

Duplicate Readings: As stated in [13], the redundancy problem is recognized as a serious issue in RFID and sensor networks. Duplication at reader level occurred when there is more than one Reader deployed to cover a specific area. The RFID data can be captured very fast several times. Because raw level of RFID data does not have meaningful information, same tags can be captured more than once therefore it is necessary to identify and eliminate those data before they posted in database.

B. Related works

Several methods related to the RFID data duplication elimination has been presented in the literature. In [6] two algorithms for duplication elimination have been proposed, base-line merge and hash merge. These algorithms are not focused on identifying the exact location and they consider that all data streams are from the same reader.

In [17], [18] duplication elimination technique has been introduced. This can only eliminate duplicate tags. In [9], an algorithms called Redundant Reader Elimination (RRE) has been proposed, which is randomized, decentralized, and localized approximation algorithm for the redundant reader elimination problem. This technique will disable some redundant reader when the other reader is activated.

III. EXACT LOCATION IDENTIFICATION AND DUPLICATION ELIMINATION

Our methodology uses the set theory in order to solve a location and date duplication problems. Firstly, a simple library scenario is set up where there are several bookshelves, books and readers. It is assumed that there will be at least two but no more than three readers to cover one bookshelf. If only one reader per shelf is deployed and if this reader fails to read some EPC tags from that shelf, the result will have a false error.

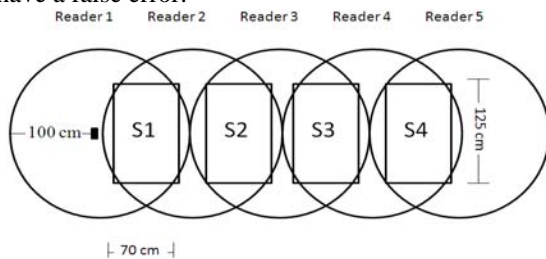


Fig. 1. An example of Library Scenario

In Figure 1, running scenario is shown where there are four bookshelves covered by five readers. As displayed, bookshelf 1 laid within the intersection of reader 1 and reader 2.

The methodology for location identification is derived from a set theory.. However, reader failure is unavoidable and should be handled once it occurred. There are four rules:

- 1) No reader failure 2) One reader failure
- 3) Two readers failure 4) Three readers failure

We will implement these rules to handle up to three readers failure since we assume that out of five readers, there should not be more than half failure in readers.

1) Rule One: No reader failure

The first rule applies to the case where none of the five readers from Figure 1 had failed. The intersection of two readers where a specific shelf lays can be represented as in equation 1.

$$A \cap B := (x \in A) \wedge (x \in B) = (x \in A : x \in B) = (x \in B : x \in A) \quad (1)$$

Where: \in = Belong to, \cap = Intersect.

Equation 1 shows that, where Both $x \in A$ And $x \in B$, then A and B are intersected. For the area of the intersection between the two readers reader A and reader B , a bookshelf will be placed within this area to ensure that all tags attached to each book will be captured by both readers. This area of intersection where the bookshelf lays can be represented as in equation 2 where *Reader A* count as one set and *Reader B* count as another set.

$$\text{Shelf } S - \text{Reader } A \cap \text{Reader } B \quad (2)$$

For example, $S1$ is covered by $R1$ and $R2$, therefore any tags captured by both $R1$ and $R2$ belong to $S1$.

2) Rule Two: One reader failure

There are five possibilities that one reader may failed. For example $R1$ failed while other readers remain function. Therefore, $R1, R2, R3, R4$, and $R5$ may failed but not at the same time.

The intersection area of two readers can be used to identify the right shelf where a specific EPC tag belong to, even though there is a failure in some readers. For example, if $R2$ in Figure 1 failed, we are still able to identify which tags belong to $S2$ using a relative complement of A and B as shown in equation 3.

$$A - \in B := (x : (x \in A) \wedge \neg (x \in B)) = (x \in A : \neg x \in B) \quad (3)$$

Equation 4 shows that *Shelf S* is the shelf where reader had failed to capture any data. *Reader C* is one reader to the left, if the failed reader is on the left of *Shelf S*; *Reader C* is one reader to the right, if the failed reader is on the right of *Shelf S*. *Reader D* is two readers to the left, if the failed reader is on the left of *Shelf S*; or two readers to the right, if the failed reader is on right of *Shelf S*.

$$\text{Shelf } S = \text{Reader } C - \in \text{Reader } D \quad (4)$$

By using Equation 4, we can identify EPC data that belong to $S2$ by eliminating those data within the intersection between $R3$ and $R4$. This can be represented as:

$$S2 = R3 - \in R4$$

When $R2$ fails and we need to know which data belong to $S2$ we need both data from $R3$ and $R4$ ($\text{Reader } C = R3, \text{Reader } D = R4$) since $S2$ is on the right side of $R2$. We already know that the intersection between $R3$ and $R4$ is where $S3$ stands, therefore, any data are not within this intersection belongs to $S2$. In the case that only one reader

exist to the left or to the right of the failed reader, we will know for certain that tags captured by that reader belong to one shelf. For example, **R2** failed and we need to know that EPC belongs to **S1**, since there is only one reader to the left, we know that EPC captured by this reader (**R1**) belong to **S1**.

3. Rule Three: Two readers failure

In Table 1 we show all ten possibilities when two readers may fail. However, we only focus on six possibilities, where two readers that fail are not subsequent to each other. This is because if two subsequent reader failed at the same time, EPC data from **S1** will not be captured at all. As mentioned earlier if there is a need to cover scenario where two readers that cover the same shelf. In case of two subsequent readers failure the percentage of successful allocation of tags to **S1** is equal to zero and therefore, it is unnecessary to consider those cases. Thus option No. 1, 5, 8, 10 from the Table 1 will not be considered.

Table 1. Ten possibilities of two readers failure at the same time(“X” represents a failed reader).

Possibility	R1	R2	R3	R4	R5
1	X	X			
2	X		X		
3	X			X	
4	X				X
5		X	X		
6		X		X	
7		X			X
8			X	X	
9			X		X
10				X	X

The intersection area of two reader can be used to identify in which shelf the specific EPC tags belong, even though there is a failure in some readers. This rule was represented earlier in Equation 4. However, this equation cannot handled the case when two readers separated by one reader in the middle, failed at the same time. For example it is impossible to tell if the EPC data belong to either **S1** or **S2** if **R1** and **R3** failed together. In order to handled this situation, we use a randomization function to allocate the specific EPC data to either **S1** or **S2**. In Table 1 we can see that the possibility that falls to the above mentioned case are options number 2, 6, and 8.

$$Shelf A^c := (x \in Reader E : x \notin Shelf A) \quad (5)$$

Where c : = Complement.

Equation 5 shows a complement of **Shelf A** (in **Reader E**) where **Shelf A^c** is a set of all members of **Reader E**, which are not member of **Shelf A**.

$$Shelf B = Shelf A^c \quad (6)$$

Where : **Shelf B** is a set of all members of **Reader E**, which are not member s of **Shelf A**.

Equation 6 can be used to allocate tags randomly, to either **Shelf A** or **Shelf B** using data from **Reader E**. For example, **R1** and **R3** has failed to capture any tags. The only active reader, which cover **S1** and **S2** is **R2**. Therefore, by using Equation 6, **Shelf A** is **S1**, **Shelf B** is **S2** and any data from belong **R2**, that do not belong to **Shelf A**, belong to **Shelf B**.

The reason for using Equation 6 to randomly allocate tags to two different locations is because ; if in this case we only allocate all data from **Reader E** to **Shelf A**, **Shelf B** will be empty. In the real world deployment, if tag from **Reader E** get allocated to the wrong location, we can easily reallocate these tags to their correct location by searching through two shelves, one to the left and one to the right of their current location.

4. Rule Four: Three readers failure

There are ten possibilities that three readers may fail, in Table 2 we shows all ten possibilities. However, we only focus on options where three failures are not subsequent to each other. We will not implement any algorithm to handle the case of **R1**, and **R2** and **R3** or **R1** **R2** and **R4** failed together. This is because, if the two or three readers next to each other failed at the same time, EPC from **S1** and /or **S2** will not be captured n any reader. The percentage to allocate tags successfully to **S1** and /or **S2** is equal to zero, therefore, it is unnecessary to handle that case. Thus, possibility number 1-4 and 6-10 from the Table 2 will not be considered.

Table 2. Ten possibilities of three readers failure at the same time(“X” represents a failed reader).

Possibility	R1	R2	R3	R4	R5
1	X	X	X		
2	X	X		X	
3	X	X			X
4	X		X	X	
5	X		X		X
6	X			X	X
7		X	X	X	
8		X	X		X
9		X		X	X
10			X	X	X

For possibility number 5 from shown in Table 2, it is certain that this case can be handled by using Equation 6. However, this equation is required to be applied twice at first, it must be applied to **S1** and **S2** and then, it must be applied to **S3** and **S4**.

IV EXACT LOCATION IDENTIFICATION AND DUPLICATE ELIMINATION ALGORITHMS

Following algorithms are based on the methodology presented in the previous section where we described how to allocate EPC data to their right location. The first algorithm is *Intersection algorithm*, which compares EPC data between two readers. If the same data exist in both readers, this specific data will be allocated to the specific shelf. The algorithm is based on Equation 2. For example, **R1** and **R2** both has ten EPC as their members. By using the *Intersection algorithm*, each member of **R1** will be compared to each members of **R2**. If the member of **R1** matched one of the member of **R2**, this member will be allocated to the **S1**

Algorithm 1 Intersection Algorithm

Input : Reader A, Reader B
Output: Shelf S
Begin
 for (Every member of Reader A) do
 for(every member of Reader B) do
 if (Reader A = Reader B) then
 Put member of Reader A into Shelf S
 end if
 end for
end for
end

The second algorithm *Relative complement Algorithm*, which also compare EPC data between two readers. However, if there are duplications between both readers, these data will be ignored. Any data based on Equation 4. For example **R3** and **R4** both hasten EPC as their member. By using the *Relative complement Algorithm* each member of **R3** will be compared to each members of **R4**. If the member of **R3** matched one of the member of **R4**, this member will be discarded. The rest of EPC within **R3** will be allocated to the **S2**.

Algorithm 2 Relative Complement Algorithm

Input : Reader C, Reader D
Output: Shelf S
Begin
 for (Every member of Reader C) do
 for(every member of Reader D) do
 if (Reader C != Reader D) then
 Put member of Reader C into Shelf S
 end if
 end for
end for
end

Finally, the last algorithm is *Randomization Algorithm*. This algorithm will randomize values between "0" and "1" If the outcome is equal to "0", the specific EPC data will be allocated to the shelf SA. Otherwise, if the outcome is equal

to "1" the EPC data will be allocated to the shelf SB. A Randomization Algorithm is based on Equation 6. For example, R4 has ten EPC as its members. By using the Randomization Algorithm, each member of R4 are randomly allocated to the S3 or S4. In this instance, five EPC are allocated to the S3 and the other five EPC are allocated to the S4. However, in the real randomization, there may not be equal number of tags allocated between two shelves.

Algorithm 3 Randomization Algorithm

Input : Reader E
Output: Shelf SA, Shelf SB
Begin
 for (Every member of Reader E) do
 Randomize number between 0 AND 1
 if (Reader E = 0) then
 Put member of Reader E into Shelf SA
 else
 if(Reader E = 1) then
 Put member of Reader E into shelf SB
 end if
 end if
end for
end

By using these three algorithms, any duplication that exists between two readers can be automatically eliminated, since only one EPC from two readers will be allocated to the specific shelf.

V EXPERIMENTAL EVALUATION

To show the practical relevance of our method of duplication elimination and location filtering extensive experimental evaluation has been conducted.

A. Environment

The experimental setup is shown in Figure 1. Five RFID readers are used and allocated opposite to bookshelves with their middle point of sensors between two bookshelves. Four bookshelves are allocated 20cm, apart from each other facing their readers. Each shelf is stocked with books represented passive tags. Tags are positioned in the same pane as readers oriented such that their antennas were directly facing readers. Algorithm proposed in section 4 are developed in Net beans 4.1 and built in as an extension to the sun Java RFID software [5].

B. Data Set

For both experiments, the EPC data are simulated according to the intersection of two readers. For example, R2 will have both data from S1 and S2 while R1 will only have data from S1. In experiment one, the correct percentage of successful tags allocation will be compared versus the size of the data set. In experiment two, the correct percentage of successful tags allocation will be compared between twelve possibilities of reader/readers failure. We will also compare the results of Location Identification and Duplication

Elimination approach to the traditional previously approach as explained in subsection 2.2 where one reader covers only one bookshelf.

Experiment One

For the first experiment, four data sets of 10EPC, 25EPC, 50EPC, and 100EPC per bookshelf are simulated. The EPC data are in standard format (their type, size and identification). As explained earlier in chosen scenario there is five readers. Each bookshelf will be covered by two readers, for example, R2 will cover data from S1 and S2, while R3 will cover data from s2 and S3 etc. For this experiment, there is no failure in data capturing process. It is assumed that all data for each shelf are captured by both readers. In order to eliminate effect of miss reads smoothing algorithm, which is part of the Sun Java RFID software was used. All captured data represent smoothing of five reads to ensure that no miss read occurs. The expected results from this experiment are that all EPC data captured by R1 to R5 will be allocated correctly into their right location.

Experiment Two

For the second experiment, the efficiency of the algorithm will be test by disabling some readers and assuming that these readers have failed during the experiment. The data set will be similar to experiment one except that there will be no data for the disabled readers. For example, R2 failed to read any data, which means that any EPC data from S1 and S2 will only get captured by R1 and R3. In this experiment, there are some readers' failures in data capturing process. It is assumed that at least one reader will always capture all data and according to the assumption that two readers, which cover the same area, will not fail at the same time as explained in subsection 3.3. The challenge will be to locate tags captured by one reader (no intersection found) to their right location. The expected results from this experiment is that all EPC data from all possibilities, where no subsequent reader or two readers from left and right of active reader failed at the same time, will be allocated to their right location. In Table 3, possibility 0 has no reader failure, possibilities 1-5 has one reader failure, possibilities 6-11 has two readers failure and possibility 12 has three readers failure.

C. Algorithms Required

In the first experiment, we assume that there is no reader failure during the process, therefore the only algorithm needed for this situation is *Intersection Algorithm*, where all tags from every bookshelf can be allocated using intersection rule.

In the experiment two, twelve possibilities of reader failure are shown in Table 3. In the case of one or two readers failure at the same time, but no two readers on the left and right of active reader failed together, we need a

Possibility	R1	R2	R3	R4	R5
0					
1	X				
2		X			
3			X		
4				X	
5					X
6	X		X		
7	X			X	
8	X				X
9		X		X	
10		X			X
11			X		X
12	X		X		X

Table 3. Possibilities of none, one, two or three readers failure at the same time. "X" Represents a failed reader.

Relative Complement Algorithm. For example, R1 and R4 fail to captured any data but R2 and R3 are still active at that time, by using a relative complement rule, we are still able to allocate all tags to their right location. However, in the case when two readers on the left and right of one active reader have failed, we need a *Randomization Algorithm*. For example, R1 and R3 fail to capture any data and only R2 remains active at that time, but using a Randomization Algorithm, it is still possible to allocate some tags to their right shelf correctly.

Experiment One:

No reader failed = Intersection Algorithm

Experiment Two:

R1 or R2 or R3 or R4 or R5 or R1 & R5 or R2 & R5 failed = Intersection Algorithm and Relative Complement Algorithm

R1 & R3 or R2 & R4 or R3 & R5 failed = Intersection Algorithm, Relative Complement Algorithm and Randomization Algorithm

R1 & R3 & R5 failed = Randomization Algorithm

D. Results and Analysis

For experiment one, we will not show the results comparison graph due to space limitation because results show that there are no differences caused by the size of data set and the allocation of all tags are 100 percent correct. The reason that the size of data set does not affect the result is because, the Intersection Algorithm does not rely on the size of data set, but only rely on the accuracy of readers and data captured from them.

Results for experiment two for all twelve possibilities of reader/readers failure is presented in Figure 2. The figure 2 shows that the percentage of success of possibilities from Table 3 number 1,2,3,4,5,7,8, and 10 are 100 percent. This proves that the Intersection Algorithm along with Relative Complement Algorithm has successfully allocated all tags to

their right shelf. However, possibilities number 6,9,11 and 12 has not achieved 100 percent allocation. This is because; it is required to use a Randomization Algorithm to randomize tags from one reader to two shelves within its range. Possibilities number 6, 9 and 11, have only about 70 percent success in the allocation. This is because a Randomization Algorithm needs to be applied to two out of the four shelves. Therefore, two shelves have achieved 100 percent allocation, while the other two shelves have achieved around 50 percent allocation each. This gives the overall percentage of about 70 percent depending on randomization. On the other hand, possibility number 12 from Table 3 has the lowest percentage of success of about 50 percent; this is because three of five readers failed to captured tags, and all data that are captured need to be randomly allocated to all four bookshelves.

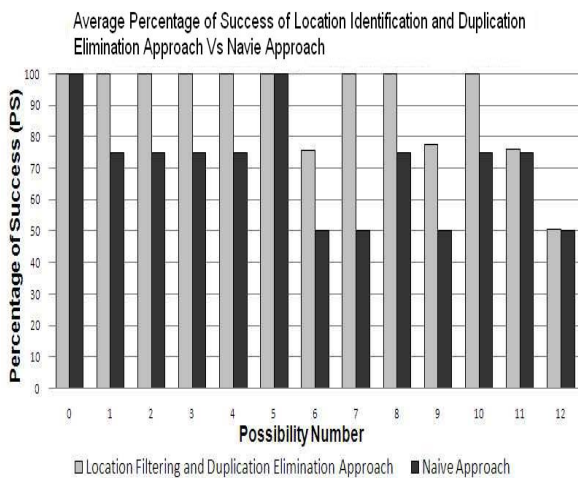


Fig.2. A result of experiment two, where percentage of successes is shown versus four different size of data set.

VI CONCLUSION

In this study, we identified the significance of RFID data allocation and duplication elimination and developed efficient method to eliminate duplication and at the same time to ensure that the RFID tags can be allocated to their correct location Identification. In experimental evaluation we showed that our method ensures identification of correct location even in case of reader failures and more importantly ensures that the RFID tags can be read regardless of some reader failure, which means no miss reads will happen.

Our approach of location identification and duplication elimination is essential to provide correct RFID data to their right location before they can be further processed, transformed, and integrated for RFID-enabled applications.

REFERENCES

- [1] Sybase RFID Solutions. <http://www.sybase.com/rfid>
- [2] Alien RFID Passive Tag. www.acccorp.com.hk/Products/Alien/,2007.
- [3] Alien ALR-9780 Reader.<http://www.alientechnology.com/products>
- [4] Oracle Sensor Edge Server. <http://www.oracle.com/technology/products/iaswe/.2007>
- [5] Sun RFID Technology. <http://www.oracle.com/technology/products/iaswe/.2007>
- [6] Y. Bai, F. Wang, and P.Liu. Efficiently Filtering RFID Data Streams. In clean DB, 2006.
- [7] C. Bornhovd, T. Lin, S. Haller, and J. Schaper. Integrating Smart Items with Business Processes An Experience Report. In HICSS '05: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 8, Washington, DC, USA, 2005. IEEE Computer Society.
- [8] J. Brusey, C. Floerkemeier, M. Harrison, and M.Fletcher. Reasoning about Uncertainty in Location Identification with RFID. Acapulco,Mexico. Aug 2003.
- [9] J. B. Carbutar, M. K. Ramanathan, M. Koyuturk, C.Hoffmann,and A. Grama. Redundant Reader Elimination in RFID Systems. 2005 CERIAS TR 2005-63
- [10] S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. Sarama. Managing RFID data. Pages 1189-1195, 2004.
- [11] J. Collins. Boeing Outlines Tagging Timetable. RFID Journal, 2006.
- [12] R. Collins. DOD Tries Tags That Phone Home. RFID Journal 2006.
- [13] R. Derakhshan, M.E. Orłowska, and X.Li. RFID Data Management: Challenges and Opportunities. Pages 175-182,2007.
- [14] R.B. Ferguson. Logan Airport to Demonstrate Baggage, Passenger RFID Tracking. Eweek,2006.
- [15] K.P. Fishkin, B. Jiang, M.Philipose , and S. Roy. I sense a Disturbance in the Force: Unobtrusive Detection of Interaction with RFID-tagged Objects. In Ubicomp,pages 268-282, 2004.
- [16] P. Harrop. RFID in the Postal Services. More3 RFID, 2005.
- [17] S.R. Jeffery,G. Alonso, M. J. Franklin, W. Hong, and J. Widom. A Pipelined Framework for Online Clearing of Sensor Data Streams. Technical Report UCB/CSD-05-1413,ECS Department, University of California, Berkeley,2005.
- [18] S.R. Jeffery, G. Alonso, M.J Franklin, W. Hong,na dJ. Widom. Declarative3 Support for Sensor Data Cleaning. In Pervasive, pages 83 -100,2006.
- [19] M. Lampe and C. Floerkemeier. The Smart Box Application Model. Vienna, Austria, Apr 2004.
- [20] T. Sun L-J Chen C.-C Han, and M. Gerla. Improving Data Reliability via Exploiting Redundancy in Sensor Networks. Technical Report TR040037, UCLA CSD, 2004.
- [21] C.Swedberg. Hospital User RFID for Surgical Patients. RFID Journal, 2005.