

The Challenges and Issues Facing the Deployment of RFID Technology

Peter Darcy, Prapassara Pupunwiwat and Bela Stantic
*Institute of Integrated and Intelligent Systems, Griffith University
Australia*

1. Introduction

Radio Frequency Identification refers to wireless technology that uses radio waves to automatically identify items within certain proximity. This process involves tagging items with a transmitter which will emit bursts of information including, but not limited to, the identification of the tag. There are three main varieties of tags: Active, Semi-active and Passive. Active tags rely solely on a battery for its power source resulting in the maximum integrity rate and reading range but, also, a limited lifespan and higher cost. Semi-Active tags use batteries to extend the range of the tag only resulting in a higher reading rate than passive tags, a longer lifespan than the active tags, but also higher cost. The passive tag uses the electromagnetic pulse from readers as a power source to transmit its identifier. Due to its lack of a battery, passive tags are the most cost effective and theoretically have an unlimited lifespan. However, due to their lack of the power source, passive tags also have a limited range and produce the largest amount of data anomalies. The RFID Reader is used to interrogate the zone to discover tags within proximity of the reader range. If a tag is discovered, its identification along with the reader's ID and the timestamp of the observation are recorded. This information is then passed through the Middleware where initial filtration is done to avoid data anomalies being recorded. Finally, the information will then be stored within a database ready to be queried for future analysis.

Due to the benefits of the technology, RFID is currently employed in various commercial sectors to provide automated assistance for mundane tasks. There are hospitals which have employed tagged bracelets to ensure maximum care is given to surgical patients. At various airports around the world, RFID is being utilised to track passengers' bags to ensure that the location of the luggage will be known at all times. In various cities around the world, pets have had RFID chips implanted to ensure that, when lost, the authorities can find their owners' information by simple scanning the tag. Various countries have also introduced the RFID-enabled toll system designed for cars at RFID-enabled toll booths which allow drivers to continue on their journey and avoid the necessity of stopping to pay.

Despite the advantages gained from RFID technology integration, various drawbacks prevent the wide-scale adoption into the majority of the commercial sector. There are three main issues concerning the integration of the architecture. The first issue is security when using the technology as tags are prone to various physical and virtual attacks upon the system. The second concern stems from the need of privacy surrounding the data collected as the observations recorded can be used for breaches in privacy. The third issue is that the

data collected among systems, in particular where passive tags are utilised, produces data characteristics that make the systems harder to use.

With regard to the data characteristics issue of RFID, there are four main problems. The first is that the data collected only contains two identifiers and a timestamp making the low-level data useless without context of other information. The large amounts of data gained in short periods of time is the second complication that arises from the use of RFID technology resulting in the database storing massive amounts of observations, some of which are useless. The third obstacle found among the integration of RFID systems is the complex spatial and temporal dimensions resulting from handheld readers and other advanced devices. The final difficulty is the tags generating ambiguous and incorrect observations resulting in duplicate, wrong and missing anomalies.

Various methodologies have been mentioned in literature to address the current problems with RFID data anomalies. We have categorised these solutions into three main groups: Physical, Middleware and Deferred approaches. Various physical solutions have been proposed in past studies to avoid missed readings in particular such as metallic-proof tag pads, tag orientation and multiple tagging. Smoothing Filters and Anti-Collision Protocols are Middleware solutions proposed to correct anomalies found within the Reader at the point of scanning. Finally, there have been several rule-based and classification algorithms proposed in past methodologies to be utilised at a deferred stage of the scanning cycle to correct various anomalies already stored in the database.

Unfortunately, each of the proposed solutions has drawbacks that prevent it from eliminating all problems found within RFID systems. With regard to the physical solutions, most have been designed to eliminate a specific problem (i.e. the metallic padding) or it will generate additional and unforeseen complications (multiple tags introducing duplicate reads). Middleware solutions have been intended to be applied at the edge of the device when the scanning is conducted which results in a limited amount of analytical information for correction allowing ambiguous anomalies to persist. The Deferred approaches have the advantage of having access to additional information in the database. However, they cannot be applied in real-time and rely on user-specified rules or probabilistic algorithms that may result in additional artificial anomalies.

We have examined RFID technology and its current uses in various applications. We have also examined three core issues stopping the mass integration of RFID in the systems including security, privacy and problematic data characteristics. We have further explored the data characteristics issue to find that it contains low-level nature, large data gathering, complex spatial and temporal aspects, and data anomalies. There have been various methodologies proposed in the past to cope with the various data anomalies which we have categorised into physical, middleware and deferred solutions. Unfortunately, due the various drawbacks such as application-specified solutions, lack of analytical information or reliance on user-specified/probabilistic algorithms, current approaches do not provide the adequate support needed in RFID systems to be adopted in commercial sectors. In this work, we have identified the importance of RFID, the shortcomings of existing approaches designed to correct its issues, and have recommended solutions to these methodologies.

2. Radio Frequency Identification

Radio Frequency Identification (RFID) has had a long history commencing with its utilisation during the Second World War to its modern usage. The basic architecture of RFID itself

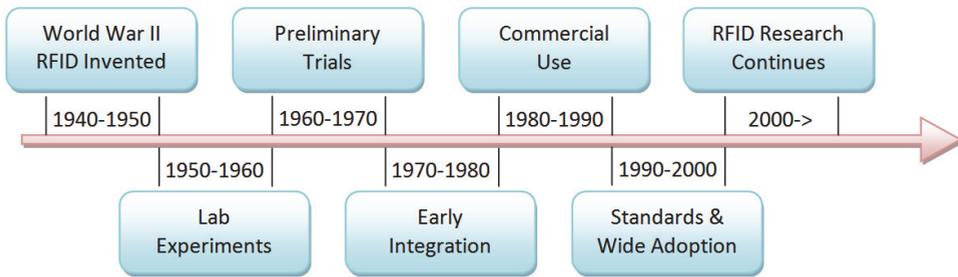


Fig. 1. The timeline of recent RFID history from the 1940s through to the present day (Landt, 2001).

consists of a tag, reader and middleware to perform advanced analysis on the data which makes it practical for use in many applications with beneficial outcomes. There are several problems which arise when using the passive tags due to the nature of the system, in particular, the amount of unreliable readings in the raw data.

2.1 History of RFID

For a general overview of RFID's historical achievements, please see the timeline illustrated in Figure 1. The physical birth of RFID would not come until the fusion of two technologies was achieved approximately around the era of the World Wars. The first technology was the Continuous Wave radio generation which was created in 1906 by Ernst F. W. Alexanderson. The second technology was the Radar device which is thought to have been developed in 1922 and was utilised extensively in World War II (Landt, 2001). The combination of these two devices resulted in the concept of RFID which was first academically proposed in theory by Harry Stockman in 1948. During this time, RFID was employed as a means to distinguish between enemy and allied aircrafts in the war. Unfortunately, as Stockman notes, technology had not progressed to the point that the complete potential of RFID technology could be realised (Stockman, 1948).

RFID research continued to be pursued in both the academic community and the military aircrafts' division who were attempting to develop "Identification Friend or Foe" (IFF) technology throughout the 1950s. It was not until the late 1960s that a Sensormatic and Checkpoint developed the first commercial RFID product in the form of EAS or "Electronic Article Surveillance" which consisted of a security system incorporating RFID tags that only stored an "on or off" command to prevent theft in stores. RFID's focus throughout the 1970s was in the tracking of animals and vehicles and, also, within the automation of factories. This adoption of the technology eventually led to the first RFID integrated road toll which was established in Norway in 1978. It was employed later in various other locations world-wide, the second notable one having been set up in 1989 at the Dallas North Turnpike in America (Landt, 2005).

In the 1990s, RFID had been integrated into people's daily activities. An example of this includes the utilisation of RFID key cards for enhanced security to enable a higher level of integrity for secure locations (Chawathe et al., 2004). In its most recent history from 2000-2010 and onwards, RFID has received the majority of its attention from various commercial sectors adopting its technology (Derakhshan et al., 2007). Some of these industries include Wal*Mart (Engels, 2005) where it has been used to enhance the supply chain, the US Department of

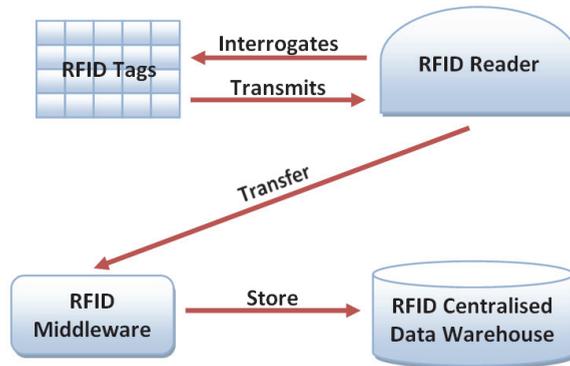


Fig. 2. The flow of information between the different components of the RFID System Architecture

Defence which has developed smarter tags (Collins, 2005) and the Aviation Industry which attaches tags to identify different parts when shipping out items (Collins, 2004). For a more comprehensive analysis of current RFID applications please see Section 3.

2.2 System Architecture

The System Architecture of an RFID system contains four important components (Chawathe et al., 2004): an RFID Tag, an RFID Reader, the RFID Middleware and the Database Storage. For a diagram representing the flow of information in this System architecture, please see Figure 2.

The RFID Tag is the simplest, lowest level component of the RFID System Architecture. These tags come in three types - Passive, Semi-Passive and Active. The Tag itself is made up of three different parts: the Chip which holds the information the tag is to dispense, the Antenna which is used to transmit the signal out and the Packaging which houses the Chip and Antenna and may be applied to the surface of other items. The Passive Tags are the most error-prone, but due to not needing a battery, also the most cost-effective and long-lasting. Electromagnetic pulses emitted from the Readers allow the Passive Tag enough energy to transmit its identification back. In comparison, the Semi-Passive Tag has a battery. However, it is only utilised to extend the readability scan resulting in a shorter life-span but increased observation integrity. The final tag is the Active Tag which utilises a battery to, not only extend its range, but also to transmit its identification number. From its heavy reliance of the battery, the Active Tag has the highest cost and shortest life-span of all the tags currently available (Chawathe et al., 2004). Even today, there are novel and emerging technologies to reduce the production cost even further such as the *Chipless RFID System Tags* and Readers (Preradovic et al., 2008; Preradovic & Karmakar, 2009).

The RFID Readers are the machines used to record the Tag identifiers and attach a timestamp of the observation. It does this by emitting a wave of electromagnetic energy which then interrogates the Tags until they have responded. These devices have a much greater purpose when needing to interrogate Passive and Semi-Passive Tags as they also provide the power necessary to transmit the information back. Readers, like the Tags, come in a variety of types such as the Hand-held reader and the Mounted Reader. The mobile hand-held tags are used for mainly determining which objects are present within a group, for example, when needing

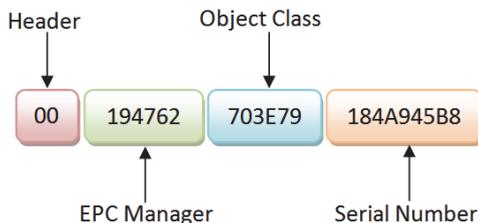


Fig. 3. The various parts of a Electronic Product Code (EPC) stored on RFID Tags.

to stocktake several items within a supermarket. In comparison, the Mounted Readers are static in geographical locations and used primarily to track items moving through their zones such as mounted readers to observe all items on a conveyer belt.

The Middleware, also commonly known as the Savant or Edge Systems, is the layer at which the raw RFID readings are cleaned and filtered to make the data more application-friendly. It receives information passed into it from the Readers and then applies techniques such as Anti-Collision and Smoothing Algorithms to correct simple missing and duplicate anomalies (Jeffery et al., 2006; Shih et al., 2006). The filtrated observational records, including the Tag and Reader Identifiers along with the Timestamp the reading was taken, are then passed onto the Database Storage.

The final destination of all the observational records is to be placed within a collection of readings taken from all connected RFID Readers. This component is known as the Database Storage and is used to hold all information which is streamed from the Readers. In most cases, due to the massive amount of interrogation undertaken to read all Tags at all times, this can result in massive floods of data, for example, 7TB of data generated daily (Schuman, 2005). Having all information stored in a central database also allows for higher level processes such as data cleaning, data mining and analytical evaluations.

EPC	Reader	Timestamp
030000E500023C000431BA3	001	2008-07-29 14:05:08.002
030000E500023C000431BA3	003	2008-07-29 14:32:12.042
030000E500023C000431BA3	002	2008-07-29 14:45:54.028
030000E500023C000431BA3	004	2008-07-29 15:02:06.029
030000E500023C000431BA3	007	2008-07-29 15:18:49.016

Table 1. A table populated with sample RFID Data containing the information of EPC, Reader and Timestamp.

2.3 Format of observations

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 which made the observation, and the Timestamp which contains the time the reading occurred. Table 1 contains information typically found stored in the Database Storage.

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. The number itself, as seen in Figure 3, is made up of a Header for 8 bits, EPC Manager for 28 bits, Object Class for 24 bits and Serial Number for 36 bits (Ward et al., 2006). Ward and Kranenburg state that a possible alternative to using the

EPC is to employ IPv6 which is the advanced version of internet addresses. These will take over the current system which is IPv4 (Ward et al., 2006). It is estimated that, since IPv6 will have 430 quintillion internet addresses as opposed to the current 4 billion address limit, there will be enough addresses for all items being tracked with RFID.

The EPC Class 1 Generation 2 is widely used in the Ultra High Frequency (UHF) range for communications at 860-960MHz. The passive RFID tag is sometime referred to as EPC Gen-2 tag, where the standards have been created by EPCGlobal (EPCGlobal, 2006), (EPCGlobal, 2005), (EPCGlobal, 2008). The most common encoding scheme with 96 bits encoding currently used includes: the General Identifier (GID-96), the Serialised Global Trade Item Number (SGTIN-96), the Serialised Shipping Container Code (SSCC-96), the Serialised Global Location Number (SGLN-96), the Global Returnable Asset Identifier (GRAI-96), the Global Individual Asset Identifier (GIAI-96), and the DoD Identifier (DoD-96).

In order to manage and monitor the traffic of RFID data effectively, the *EPC pattern* is usually used to keep the unique identifier on each of the items arranged within a specific range. The EPC pattern does not represent a single tag encoding, but rather refers to a set of tag encodings. For instance, the General Identifier (GID-96) includes three fields in addition to the 'Header' with a total of 96-bits binary value. *25.1545.[3456-3478].[778-795]* is a sample of the EPC pattern in decimal, which later will be encoded to binary and embedded onto tags. Thus, within this sample pattern, the Header is fixed to 25 and the General Manager Number is 1545, while the Object Class can be any number between 3456 and 3478 and the Serial Number can be anything between 778 and 795.

Within each EPC, the Uniform Resource Identifier (URI) encoding complements the EPC Tag Encodings defined for use within RFID tags and other low-level architectural components. URIs provide an information for application software to influence EPC in a way that is independent of any specific tag-level representation. The URI forms are also provided for pure identities, which contain just the EPC fields which are used to distinguish one item from another. For instance, for the EPC GID-96, the pure identity URI representation is as follows: `urn:epc:id:gid:GeneralManagerNumber.ObjectClass.SerialNumber`

In this representation, the three fields `GeneralManagerNumber`, `ObjectClass`, and `SerialNumber` correspond to the three components of an EPC General Identifier (EPCGlobal, 2008). There are also pure identity URI forms defined for identity types corresponding to certain encodings, the URI representations corresponding to these identifiers are as shown in Table 2.

Encoding Scheme	Uniform Resource Identifier
GID	<code>urn:epc:id:gid:GeneralManagerNumber.ObjectClass.SerialNumber</code>
SGTIN	<code>urn:epc:id:sgtin:CompanyPrefix.ItemReference.SerialNumber</code>
SSCC	<code>urn:epc:id:sscc:CompanyPrefix.SerialReference</code>
SGLN	<code>urn:epc:id:sgln:CompanyPrefix.LocationReference.ExtensionComponent</code>
GRAI	<code>urn:epc:id:grai:CompanyPrefix.AssetType.SerialNumber</code>
GIAI	<code>urn:epc:id:giai:CompanyPrefix.IndividualAssetReference</code>
DoD	<code>urn:epc:id:usdod:CAGECodeOrDODAAC.serialNumber</code>

Table 2. The Uniform Resource Identifier encoding complements the EPC Tag Encodings defined for use within RFID tags and other low-level architectural components

An example encoding of GRAI is demonstrates as follows:

`urn:epc:id:grai:0652642.12345.1234`

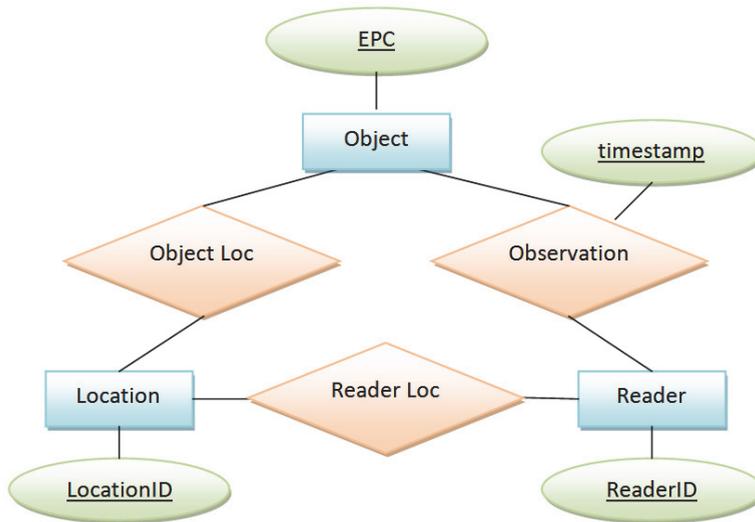


Fig. 4. An example RFID scheme which could be used to house the captured information generated from a RFID system.

From the above example, the corresponding GRAI is 06526421234581234. Referring to Table 2, the CompanyPrefix, AssetType, and SerialNumber of GIAI are represented as 0652642, 12345, and 1234 respectively.

The Reader Identifier attribute is the unique identifier of the Reader so that the analyser will be informed of which reader took the EPC reading. If the Reader is static in its location as well, such a position of the reading may be derived from a simple query in the database later using this value. Knowledge of the geographical location of each unique Reader identifier may also provide additional information needed in future business processes.

The Timestamp contains a temporal reading used to identify the date and time that the Tag passed within vicinity of the Reader. For example, 2008-07-29 14:05:08.002 would be stored as a timestamp.

2.4 Storage of RFID data

In its rawest form, RFID data is recorded in a temporal stream of data consisting of EPC, Reader and Timestamp. After the burst of information is recorded from the reader, the RFID Savant or RFID Middleware modifies data to represent a higher level description of the events that took place. For example, the Siemens RFID Middleware extracts the data and loads it into a Dynamic Relationship Entity Relationship Model (Wang & Liu, 2005). Figure 4 depicts the Entity Relationship Diagram (ERD) used as a basic Database Storage for RFID events. As seen in the diagram, there are three prime entities that must be known, the Object, the Reader, and the Location of the Reader. Each entity has an identifying tuple attached including the Observation weak entity that also attaches the timestamp of a recorded event. Additionally, more advanced systems will only record the start and end time that an Object is within a Location, thereby saving memory so that observations are not recorded as frequently (Wang et al., 2010).

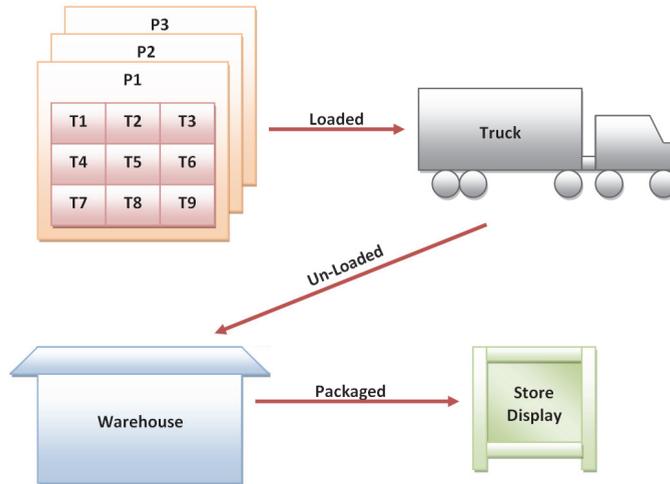


Fig. 5. The various stages taken when transporting various RFID-enabled items in a supply chain.

2.5 RFID advantages

The main advantage of RFID technology is that it is not necessary to have a line-of-sight between the object and the reading device (Derakhshan et al., 2007). In comparison to object scanners currently employed in various commercial sectors such as supermarkets, an object is needed to be taken out, place on a conveyor belt, rotated until the barcode is within the position and then placed back into the shopping trolley. If RFID is employed within this scenario, all items would automatically be recorded when the customer approaches the register and the cost tallied in one scan without the need of moving the items outside the trolley, thus saving the company time, money and physical labour. Specifically in relation to Passive Tags, there are two main advantages found when integrating RFID technology (Chawathe et al., 2004). The first is that the manufacture of the RFID Passive tag is extremely cheap. It is estimated that it only costs 5 cents per tag when bought in bulks of billions. The second advantage of the Passive RFID System is that, due to the ingenuity of the tag itself, it is not application-specific and may be applied to almost any domain. With regard to the variety of uses of RFID, as stated by Polniak - "Uses of automatic identification are manifold, limited only by one's imagination" (Polniak, 2007).

3. Current uses of RFID

From investigating the current uses of RFID, we have discovered that each utilisation may be placed into two different categories of RFID applications. The first, which we have labelled "RFID Integrated Applications", includes already existing systems which have been enhanced and made more effective and efficient using RFID technology. We have labelled the second category "RFID Specific Applications" in which prototype machines have been built from the bottom-up to incorporate RFID technology in its very make up.

3.1 Integrated RFID Applications

We have defined Integrated Applications as scenarios in which originally existing business operations have been augmented with the integration of RFID technology. The most common use of RFID integrated applications is the generic supply-chain example of RFID integration commonly employed by commercial stores such as Wal-Mart. In the example illustrated in Figure 5, tagged Objects (T1-T9) are added to specific Pallets (P1-P3), which are then loaded onto a Truck. The Truck will then transport the Pallets to their Warehouse destination at which point the items are then packaged for display at their retail stores. Additionally, as described by Derakhshan, Orlowska and Li, there are several other applications which have integrated RFID technology into their business models (Derakhshan et al., 2007) such as:

- **Defense and Military:** The US Department of Defence (DOD) is investigating a new active tag which has the ability to access and communicate via satellites. This new tag, known as the "Third Generation Radio Frequency Identification with Satellite Communications (3G RFID w/SATCOM)", is expected to be used to increase the visibility of the DOD's supply chain and, in turn, increase the confidence of shipments to various war-torn regions (Collins, 2005).
- **Postal Package Tracking:** The postal service has been found to incorporate RFID world-wide with the primary goal of increasing the effectiveness of tracking packages and parcels thereby increasing customers' property security (Harrop, 2005).
- **Aviation Industry:** Two major aircraft manufacturers, Boeing and Airbus, have started ensuring that the supplying factory parts for the aircraft use RFID tags for identifications resulting in an easier process to locate and identify needed parts (Collins, 2004).
- **Health Care:** The Taiwanese Chang-Gung Memorial Hospital has been monitoring surgical patients with RFID wristbands in order to ensure maximum care is given where needed. The features available in the wristbands include the ability to decrypt data, obtain read-only static fields (such as blood-types) and read/write dynamic fields which may be updated and modified by medical staff (Swedberg, 2005).
- **Baggage/Passanger Tracing:** The Boston Logan International Airport and the Boston Engineering Incode Corporation have integrated RFID technology within the Secure Environment for Airport Terminal Systems (SEATS) which passengers and their baggage with passive RFID tags to track all movements from their arrival at the airport to boarding the flight (Ferguson, 2005). This technology ensures not only that passengers will be able to make their flight easier, but that their baggage location will always be known.

3.2 Specific RFID applications

We have categorised applications specifically designed and built with the integration of RFID technology as Specific Applications. Four such examples which have been developed in the recent years include the Magic Medicine Cabinet, the Multipurpose Smart Box, the Augmentation of Desktop Items and the Smart Shelves (Brusey et al., 2003; Floerkemeier, 2004).

The Magic Medicine Cabinet, as described in (Wan, 1999), is a bathroom cabinet which is used to assist in bridging the gap between the informational and physical aspects of the medical world. The Magic Medicine Cabinet will allow RFID based tracking systems to describe the content of what is being placed into and removed out of storage by the user. Through a combination of Facial Recognition, Vital Sign Monitors, Voice Synthesisers and

RFID technologies, the Cabinet can intelligently decide whether or not the person currently interacting with it should be taking the medicine. This, in turn, would bring the action to the owner's attention if necessary.

As discussed in (Floerkemeier et al., 2003; Lampe & Floerkemeier, 2004), an automatic content monitoring application called the "Smart Box", similar to the Magic Medicine Cabinet, has been designed to monitor the RFID-enabled contents placed inside. The Smart Box may also be set up in different configurations to suit the context to which it will be applied such as a Smart Surgical Kit for hospitals and a Smart Toolbox for mechanics (Floerkemeier et al., 2003). The Augmentation of Desktop Items is a means of combining physical objects with virtual interfaces using the inexpensive power of RFID tags and readers (Want et al., 1999). In a typical scenario, an office object such as a book would be tagged and then read by a Reader connected to a computer to allow the user additional functionality. For example, when someone scans a book by the reader, the computer would use stored information relating to the office to identify the book's title and would begin to provide additional internet-features such as summaries, discussions or would allow the user to order the book from Amazon.com. The Smart Shelf is an RFID enabled device which tracks all items placed on it to accurately determine the location of the said object (Decker et al., 2003; TecO & SAP/CEC, 2003). The Smart Shelf was designed specifically with the secondary goal of obtaining the unobserved events of a person handling an item at retail outlets and, subsequently, returning it to the shelf thereby allowing business analysts further glimpses into the decision-making of the consumers. From this information, it would be possible to detect if a shopper mentally debates over the decision to purchase the product.

4. RFID issues

Before RFID can be utilised to its maximum potential, as opposed to the fraction in which it is currently employed, certain issues need to be understood by the users, and corrected if possible. The three core obstacles include the concerns of security, the problems surrounding the privacy of the data captured and the characteristics associated with the nature of RFID. Additionally, we will further examine the specific problems associated with anomalies present within the captured observational records which are regarded as a characteristics of RFID. When all of these issues are rectified to provide maximum security, privacy and integrity, RFID will be able to realise its full potential in massive wide-scale adoptions.

4.1 RFID security

The issues associated with RFID Security, also known as Intrusion Detection, refers to the discovery of foreign attacks upon the system usually utilising the tags that hinder the overall integrity of the data. The following five issues are some of the most dominant with regard to RFID security (Mitrokotsa et al., 2010; Thamilarasu & Sridhar, 2008):

- **Eavesdropping:** The act of setting up an additional reader to record tag data.
- **Unauthorised Tag Cloning:** Copying tag data onto an additional tag to gain the same privileges.
- **Man-in-the-Middle (MIM) Attack:** When an external object pretends to be either a tag or reader between actual tags and readers.
- **Unauthorised Tag Disabling:** When an external reader disables a tag not allowing it to be utilised again.

- **Unauthorised Tag Manipulation:** Manipulating the tag data using an external reader.

Until these security issues existing in the current architecture, it becomes difficult for facilities to employ RFID as a means of combatting unauthorised actions such as safe-guarding sensitive or expensive objects or restrict personnel access into various locations. Currently, there are techniques and approaches such as Tag Deactivation and Encryption (Karygiannis et al., 2007), Mutual Authentication (Konidala et al., 2007), Detections in Tag Ownership (Mirowski & Hartnett, 2007), Reader Analysers (Thamilarasu & Sridhar, 2008) and certain data cleaners (Darcy, Stantic, Mitrokotsa & Sattar, 2010) to reduce the difficulties associated with RFID Security.

4.2 RFID privacy

Privacy within the context of an RFID-enabled facility refers to either unknowingly releasing critical information (deriving specific knowledge or tracking meaningless data) (Langheinrich, 2009), or compiling a list of all items currently found on a person (Juels, 2006). There have been several methodologies proposed in the past to ensure maximum privacy of an individual, including the general approaches of Encrypting/Rewriting and Hiding/Blocking Tags (Langheinrich, 2009). In addition to these general solutions, there have been more specific and advanced approaches suggested such as killing/sleeping the Tags, carrying around a privacy-enforcing RFID device, releasing certain information based solely on distance from the reader and introducing Government Legislations (Juels, 2006).

4.3 RFID characteristics

There are certain characteristics associated with the nature of RFID technology (Cocci et al., 2008; Derakhshan et al., 2007). These challenges include Low Level Data, Error-Prone Data, High Data Volumes and its Spatial and Temporal Aspects. Low Level Data refers to the raw observational readings being taken by the RFID Reader; Error-Prone Data is the problem which RFID has with capturing the data; High Data Volumes refers to the ongoing obstacle with managing exponential RFID data streams and Spatial and Temporal Aspects alludes to the aspects of RFID's freedom in being capable of being used in all situations.

As previously discussed in Section 2.3, the format of the data at the time of scanning is very low level and lacks crucial information needed later for analysing the information captured. The core problem with these observations is the lack of associations between the readings and other information such as what the tags are attached to or the locations of the readers thereby making captured data useless on its own. Humans must find significant information extracted from these low level observations such as high level RFID Events (Khoussainova et al., 2007) which are the transformed state of the raw readings into meaningful milestones. For example, if a certain tag "202" is read at the reader "794" at timestamp "25/05/08 07:30:04", there is not enough information to comprehend the significance of the observation. By using relational information such as reader locations and tag information, these low level observations may be transformed into a high level event depicting the person named John being at the Front Door of location at 7:30:04 on the 25/05/08.

RFID Data integrity is constantly lowered to the point of questioning its authenticity especially when utilising passive tags due to errors captured within the observational data. These errors include Missed Reads in which a tagged item is present but not recorded, Wrong Reads in which data is captured where it should not resulting in the data set not reflecting events which are actually taking place, and duplicate reads in which a tagged item is stored twice in the

database where it should only be stored once. Section 4.4 further expands the error-prone nature of RFID where an analysis together with each of these errors are given.

Due to the continuous stream of information and the need to constantly interrogate tags, readers record massive amounts of data over long periods of time. It has been estimated that Wal*Mart currently generates about 7TB of information daily due to its RFID integration (Raskino et al., 2005). Additionally, it is estimated that by the year 2015, with a steady increase of RFID presence but lack of content management of the data generated, the information collected will be a serious problem for integrated systems. This may ultimately lead to a decrease of RFID usability and waste of information already gained unless either the management of data collected is properly attended to or the technology currently employed greatly increases its storage capacity.

As previously discussed in Derakhshan et al. (2007); Wang & Liu (2005), the exponential growth of smaller hardware RFID solutions coupled with the cost reduction in manufacturing these units results in RFID applications becoming increasingly dynamic in both spatial and temporal properties. For example, there are hand-held RFID Readers which are carried by people to scan groups of RFID tags in various locations. However, these scans will never be able to be placed into a geographical context thus limiting the potential of analytical processes that may be performed. Unless properly managed, the dynamic properties of RFID's spatial and temporal aspects may result in increasingly complex ambiguity ultimately resulting in the data losing significance, context and usability.

4.4 RFID anomalies

RFID observational data suffers from three main anomalies which are recorded with the correct RFID readings. The first is a Wrong Reading in which data is captured where it should not be. The second is Duplicate Readings in which a tag is observed twice rather than once. The third is the Missed Readings which occur when a tag is not read when and where the object it is attached to should have been physically within proximity. Figure 6 contains an example of a RFID-enabled shelf which has also generated the three anomalies, the recorded data may be seen in Table 3.

What is Recorded		
Tag EPC	Timestamp	Reader ID
T1	13/10/2010 14:31:05	R1
T2	13/10/2010 14:31:05	R3
T3	13/10/2010 14:31:05	R3
T3	13/10/2010 14:31:05	R4
T3	13/10/2010 14:31:05	R5
What is meant to be Recorded		
Tag EPC	Timestamp	Reader ID
T1	13/10/2010 14:31:05	R1
T2	13/10/2010 14:31:05	R2
T3	13/10/2010 14:31:05	R3
T4	13/10/2010 14:31:05	R5

Table 3. The recordings that took place from the example in Figure 6 and the observations that should have been recorded.

Wrong Readings, also known as Unreliable Readings or Ghost Reads falling into the False Positives category, refer to observations found in the data storage of tag which were not physically present in the location or time. These false readings may be produced when

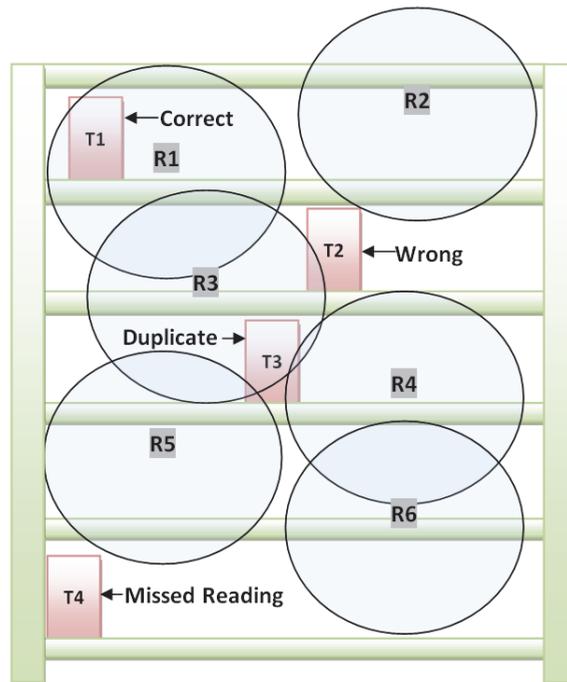


Fig. 6. A graphical representation of a RFID-enabled bookshelf with the data anomalies that may occur highlighted.

tags outside the normal Reader range are captured or where there is a problem with the environmental setup (Bai et al., 2006). As discussed previously (Embry, 2005; Engels, 2005), this problem has been identified as one of the two main technical problems with RFID. It may also result in additional unnecessary labor to continually monitor the objects where the locations of the tagged items is vital to the business process, for example, the tracking of livestock. Within the example in Figure 6 and Table 3, Tag T2 is read by Reader R3 when in reality it is closer to the area which should be scanned by Reader R2.

Duplicate Readings refer to an RFID tag which has been scanned twice in the database as opposed to just one scanning. Like the Wrong Readings, Duplicate anomalies also fall into the category of False Positive observations as they record the data which do not accurately depict reality. This may occur in several situations such as the situation in which there is more than one Reader covering an area and a tag happens to pass within overlapped region (Carbunar et al., 2005). This can be hazardous and redundant as the tag is represented in two areas during the same time period. Other duplicate reading situations occur when a scanned item stays in the reader range for a long period of time or when the owners of the RFID system attach multiple tags in order for an item to enhance its read rate (Bai et al., 2006). Ultimately, not only does this anomaly cause contradicting observations where tags may appear in two locations at the same time but it also leads to memory being wasted where it could be utilised to store factual information. In the sample scenario depicted in Figure 6 and Table 3, Tag

T3 is captured by not only the correct Reader R3, but also R4 and R5 resulting in T3 having duplicate entries in the recorded data set.

Missed Readings, also known as False Negative observations, refers to tagged objects not being scanned when, in actuality, they were present. The exact percentage of tags that are read remain only at 60%-70% under certain circumstances (Floerkemeier & Lampe, 2004). Reasons for these anomalies stem from problems such as Tag Collisions, Tag Detuning, Water/Metal Interference and Misalignment of the Tags. The missed reads anomaly has been identified as the second major problem in RFID deployment by an array of researchers (Engels, 2005; Floerkemeier & Lampe, 2004; Rahmati et al., 2007). The result of this anomaly may cause the users to believe that all items which are meant to be present are not, thereby hindering the overall process it was designed to make more efficient. Tag T4 in the example Figure 6 is shown to be a missed read due to it being placed slightly outside the scanning range of Reader R5 resulting in it not being recorded with the other tags in Table 3.

5. Current state-of-the-art approaches

In this section, we will provide a brief summary of all the current state-of-the-art approaches we have investigated to correct the RFID issues. We have divided the methods into three categories: Physical Approaches in which methods attempt to correct RFID anomalies by improving the environment around the scanners, Middleware Approaches in which algorithms attempt to correct the anomalies at the time of capturing and Deferred Approaches which attempt to correct RFID data when it is stored in the Database. Table 4 provides a list of each of the techniques examined in this section and the Corrected and Potentially Generated Anomalies.

	Methodology	Anomalies Corrected			Anomalies Generated		
		Wrong	Duplicate	Missed	Wrong	Duplicate	Missed
Physical	Tag Orientation	-	-	X	X	-	-
	Weighing	-	X	X	-	-	-
	Multiple Tags/Cycles	-	-	X	X	X	-
	Eccopad	-	-	X	-	-	-
Middleware	Edge Filtering	X	X	X	X	X	X
	Anti-Collision	-	-	X	-	-	-
	Thresholds	-	-	X ¹	-	-	-
	Statistical Approx.	X	X	X	X	X	X
Deferred	P2P Collaboration	X	X	X	-	-	-
	Proximity Detection	X	-	X	-	-	-
	Cost-Conscious Cleaning	X	X	X	X	X	X
	Data Mining Techniques	X	X	X	X	X	X
	Probabilistic Inference	-	-	X	X	-	-
	Event Transformation	X	X	X	X	X	X
	Intelligent Classifiers	X	X	X	X	X	X

Table 4. A table depicting which anomalies are corrected and generated by the various methodologies proposed. The 'X' denotes where the methodology either corrects or generates the anomaly. Note¹: The Thresholds methodology does not actually correct the missing data but, rather, alerts the user to a False-Negative anomaly.

5.1 Physical approaches

One common solution to improve the tag reads in RFID systems is to employ Physical Approaches. This enhances the environment where the scanning is conducted. We define Physical Approaches as any solution which requires interaction with the equipment as opposed to virtual interaction used at the middleware or at a deferred warehouse stage to correct the captured data.

Rahhmati, Zhong, Hiltunen and Jana have conducted a study into the effects of reader rate when positioning the RFID tag in different positions (Rahmami et al., 2007). The research found that the Reader may scan the tags on an object most effectively when the Tag is positioned at the front.

Potdar, Hayati and Chang have formulated a novel and simple solution designed to detect missing RFID tags through the use of weights (Potdar et al., 2007). This method was created for applications in situations in which items are required to be tracked while being transported to various venues. It requires all tagged items to be weighed at the start of the transportation route. The items are then weighed at the end of the trip to determine any difference in the cargo weight. The system will determine if there are any missed reads coupled with an attempt to find any missing weight. If there are missing readings but a constant weight, the system will scan the cargo again until all items have been recorded.

As described by both (Bai et al., 2006; Vogt, 2002), a common solution to deal with RFID anomalies is to either install multiple readers or to attach multiple tags. Multiple readers are installed in the environment in an attempt to enhance the reader rate by covering a more substantial amount of ground. Another method of dealing with the enhancement of the read rate is to attach multiple tags housing identical EPC numbers to the same object in an effort for at least one of these tags to be read by the reader. Unfortunately, drawbacks arise from both methods in the form of duplicate readings and tag collision occurrences.

Emerson & Cuming Microwave Products (Emerson & Cuming Microwave Products, 2008) provides a device known as the Eccopad which is designed to enhance the reading rate of tags placed on metal. As described in (Floerkemeier & Lampe, 2004), metallic objects within a certain proximity will affect the reading potential of a passive RFID tag causing missed readings. The Eccopad insulates the RFID tag in a discrete manner which enables the maximum potential reading rate with little or no change in the spatial properties occurring.

5.2 Middleware approaches

Middleware Approaches refers to employing an algorithm to eliminate anomalies found in systems to correct the data before storing it. This can refer to any program used at the middleware stage of the RFID capture cycle to correct the raw incoming streamed data.

Edge Filtering refers to the cleaning being completed at the edge of the RFID system, that is, at the point of raw observations being read. Jeffery, Garofalakis and Franklin have proposed a method analysing the usage of an adaptive sliding window to correct unreliable readings within an RFID system (Jeffery et al., 2006). A sliding window is used to smooth out the raw data in order to accommodate both false positive and false negative readings. The problem associated with this technique is that the result of utilising a small sliding window will be false negatives whereas the large window may result in false positives being introduced. Thus, Jeffery et al, proposed a solution to create a declarative and adaptive smoothing window named SMURF (Statistical sMoothing for Unreliable RFid data) which they have continually improved (Jeffery et al., 2008).

RFID Anti-collision protocols are algorithms used at the edge to avoid missed readings. When an RFID scan is performed on several RFID tags, there are many relaying messages sent back and forth between the tag and the reader. If there are a large number of tags to be scanned in a certain read, these messages may collide in the air between their source and destination resulting in the information not arriving at the correct time if at all. Certain protocols are also designed to handle other forms of hazards such as instances where readers placed within a certain proximity interfere with each other's interrogation cycle causing collisions (Shih et al., 2006).

The various types of anti-collision methods for *collision* can be reduced to two basic types: *probabilistic* and *deterministic* methods. In a probabilistic method, tags respond at randomly generated times. If a collision occurs, colliding tags will have to identify themselves again after waiting for a random period time frame. From past literature, there have been several methods proposed such as: Basic Framed-Slotted ALOHA (Lee et al., 2008); Dynamic Framed-Slotted ALOHA (Ding & Liu, 2009); Enhanced Dynamic Framed-slotted ALOHA (Lee & Lee, 2006); and Probabilistic Cluster-Based Technique (Pupunwiwat & Stantic, 2010d), to enhance the performance efficiency of the data capturing process. In addition, several *Frame Estimation* approaches have been suggested to improve the accuracy of *frame-size* prediction including the Schoute method (Schoute, 1983), the Lowerbound method, the Chen1 and Chen2 methods (Chen, 2006), the Vogt method (Vogt, 2002), the Bayesian method (Floerkemeier, 2007), and the Precise Tag Estimation Scheme (Pupunwiwat & Stantic, 2010b), (Pupunwiwat & Stantic, 2010a).

The deterministic method operates by asking for the first EPC string of the tag until it gets matches for the tags, it will then continues to ask for additional characters until all tags within the region are found. There have been several methods proposed in literature in order to improved quality of the captured data such as: the Query Tree (Myung & Lee, 2006a); the Adaptive Splitting Tree (Myung & Lee, 2006b); the Hybrid Query Tree (Ryu et al., 2007); and the Joined Q-ary Tree (Pupunwiwat & Stantic, 2009), (Pupunwiwat & Stantic, 2010c).

Tan, Sheng and Li have proposed in their research the utilisation of a threshold to identify an excessive amount of missing RFID readings (Tan et al., 2008). By using two different protocols, the trusted reader and un-trusted reader protocols, the methodology analyses a RFID data set and finds missing data without the need for ascertaining tag identifiers. The system will then consult a threshold defined by the owner as to the number of missing tags which are tolerable in a given situation with the system alerting the user if this threshold is breached. It will not however replace the missed readings.

Statistical Approximations refer to the use of a Model-Based Querying system to return approximate readings found from the sensor networks (Deshpande et al., 2004; Deshpande et al., 2005). Although this method is not used primarily for RFID technology, the method is applied to wireless sensors which provide additional functionality that RFID tags do not (i.e. Temperature Sensors). This approach is designed to capture a query from the User, find the values from the sensor readings, and return approximate values to the User.

5.3 Deferred approaches

We have defined Deferred Approaches as methodologies applied at a deferred stage of the capturing cycle when the observational data is stored in the database. This includes P2P Networks, Probabilistic Tag Proximity Detection, Cost-Conscious Cleaning, Data Mining Techniques, Probabilistic Inference and Probabilistic Event Extraction.

The P2P Collaboration method, proposed by Peng, Ji, Luo, Wong and Tan (Peng et al., 2008), is an approach utilising Peer-to-Peer (P2P) networks within the RFID data set to detect and remove inaccurate readings. The system works by breaking the readings into detection nodes, which are constantly sending and receiving messages. From these transmitted messages, false negatives and false positives are able to be detected and corrected resulting in a cleaner data set.

Ziekow and Ivantysynova have presented a method designed to correct RFID anomalies probabilistically by employing maximum likelihood operations (Ziekow & Ivantysynova, 2008). Their method utilises the position of a tag which may be determined by measuring properties associated with the Radio Frequency signal.

The Cost-Conscious cleaning method is a cleaning algorithm which utilises a Bayesian Network to judge the likelihood that read tags correctly depict reality when based upon the previously read tags (Gonzalez et al., 2007). The Cost-Conscious cleaning approach houses several different cleaning algorithms and chooses the least costly algorithm which would offer the highest precision in correcting the raw data. A similar approach has also been proposed that utilises a Bayesian Network to judge the existence of tags scanned (Floerkemeier, 2004). It lacks, however, the cost-saving analysis that would increase the speed of the clean.

Data Mining Techniques refer to the use of mining past data to detect inaccuracies and possible solutions to raw RFID readings. A study which has used data mining techniques extensively to correct the entire data set table is the Deferred Rule Based Approach proposed in (Rao et al., 2006). The architecture of the system is reliant on the user defining rules which are utilised to determine anomalies in the data set and, possibly, to correct them.

Probabilistic Inference refers to a process by which the in-coming data node will be evaluated. This is primarily based upon the weight of its likelihood and the weight of the remainder of the readings (Cocci et al., 2007; 2008). The cleaning algorithm utilises several techniques to correct that data such as Deduplication, Time conversion, Temporal Smoothing and Anomaly Filtering, and, additionally, uses a graph with probabilistic weights to produce further inferences on the data.

Probabilistic High Level Event Transformations refers to the process of observing the raw partial events of RFID data and transforming these into high level probable events. It has been primarily used in a program entitled Probabilistic Event EXtractor (PEEX) which has evolved from several publications. In its embryonic phase, Khoussainova, Balazinska and Suciu published a paper detailing the use of an algorithm called StreamClean which employ probabilistic inference to correct incoming data (Khoussainova et al., 2006).

A year after this article, the first papers for PEEX were published. This described the method which enabled high level event extraction based upon probabilistic observations (Khoussainova et al., 2007; Khoussainova, Balazinska & Suciu, 2008). The system architecture deciphers the raw RFID information searching for evidence which a high level event transpired. The system uses a Confidence Learner, History Lookup and Event Detector to enhance the reliability of the returned events. By transferring these low level readings into high level events, PEEX engages in cleaning as the process of probabilistically by categorising the results of these events, and in the process, caters for missed and inaccurate readings.

Currently, PEEX is being incorporated into a new system named Cascadia where it will be utilised to help perform high level management of RFID tracking in a building environment (Khoussainova, Welbourne, Balazinska, Borriello, Cole, Letchner, Li, Ré, Suciu & Walke, 2008; Welbourne et al., 2008). Bayesian Networks have also been implemented in several studies to infer high level behaviour from the raw readings. The specific application was first

demonstrated on a traveller moving through an urban environment (Patterson et al., 2003) and the second using RFID tags to track the activities of daily living (Philipose et al., 2004). In previous work, we have proposed the concept of using high level classifiers coupled with intelligent analysis to correct the various anomalies found in RFID data. First, we examined the potential of employing a simple algorithm that corrects a simple missed reading (Darcy et al., 2007). We then proposed the utilisation of highly intelligent analytical processes coupled with a Bayesian Network (Darcy et al., 2009b;c), Neural Network (Darcy, Stantic & Sattar, 2010a) and Non-Monotonic Reasoning (Darcy et al., 2009a; Darcy, Stantic & Sattar, 2010b) to correct missing RFID Data. Following this, we applied our Non-Monotonic Reasoning approach to both false-negative and false-positive data anomalies (Darcy, Stantic & Sattar, 2010d). We then also introduced a concept to extract high level events from low level readings using Non-Monotonic Reasoning (Darcy, Stantic & Sattar, 2010c). Finally, we proposed a methodology that considers and differentiates between a false-positive anomaly and breach in security using Non-Monotonic Reasoning (Darcy, Stantic, Mitrokotsa & Sattar, 2010).

6. Drawbacks and proposed solutions for current approaches

In this section, we highlight several drawbacks we have found associated with the various methodologies currently employed to correct RFID captured data. We also supply our suggested solutions to these problems where possible in an effort to encourage further interest in this field of research. Finally, we conclude with an overall analysis of these methodologies and their respective drawbacks.

6.1 Physical drawbacks and solutions

With regard to Physical Approaches, we have highlighted three main drawbacks and our suggested solutions to correct these issues where possible:

- **Problem:** The main problem that we foresee with the utilisation of Physical Approaches is that it usually only increases the likelihood that the missed objects will be found.
Solution: We do not have a solution to the problem of physically correcting wrong or duplicate anomalies other than suggesting to utilise Middleware and/or Deferred solutions.
- **Problem:** Physical Approaches generates artificial duplicate anomalies in the event that all the tags attached are read.
Solution: Specific software tailored to the application to automatically account for the artificially generated duplicate anomalies could be used for correction filtering at the edge.
- **Problem:** Physical Approaches suffer from additional cost to the user or more labour to purchase extra tags, equipment or time to move the objects.
Solution: We do not believe there is a solution to this as Physical Approaches demand additional labour for the user to correct the mistakes as opposed to Middleware or Deferred Approaches.

6.2 Middleware drawbacks and solutions

We found three major drawbacks to the Middleware Approaches that prevent these from acquiring their maximum integrity. These issues include:

- **Problem:** Correcting incoming data at the edge of the RFID capture process will not provide the cleaning algorithm with adequate information needed to deal with highly

ambiguous and complex anomalies.

Solution: We believe that to correct this drawback, the user must employ a Deferred methodology in addition to the Middleware Approach to utilise all stored readings. This would result in more observational data eliminating highly ambiguous anomalies.

- **Problem:** When utilising probabilistic algorithms such as Bayesian Networks to correct anomalies, there is a risk of the methodology introducing artificially generated anomalies. This may occur in cases such as the training set not reflecting the reality of the scenarios or the system probabilistically choosing the incorrect action to take in a situation.

Solution: To correct this issue, the user may be able combine various probabilistic techniques together or to employ a deterministic approach in order to enhance the method of cleaning the database.

- **Problem:** RFID data streams that are captured by readers can be accumulated quickly resulting in data collisions. Simultaneous transmissions in RFID systems will also lead to collisions as the readers and tags typically operate on the same channel. There are three types of collisions possible to occur: Reader-Tag collision, Tag-Tag collision, and Reader-Reader collision.

Solution: It is crucial that the RFID system must employ *anti-collision* protocols in readers in order to enhance the integrity of the captured data. However, the step of choosing the right *anti-collision* protocol is also very important, since we cannot depend solely on the capability of anti-collision protocol itself, but also on the suitability of each selected technique for the specific scenario. The user may employ decision making techniques such as both the Novel Decision Tree and the Six Thinking Hats strategy for complex selective technique management to determine the optimal anti-collision protocol. The novelty of using complex selective technique management is that we will get the optimal outcome of *anti-collision* method for the specific scenario. This will, in turn, improve the quality of the data collection. It will also help over long period of use when these captured data are needed for transformation, aggregation, and event processing.

6.3 Deferred drawbacks and solutions

While reviewing the Deferred Approaches to correct RFID anomalies, we have discovered that there are certain shortcomings when attempting to clean captured observational data.

- **Problem:** Similar to the Middleware Approaches which utilise probabilistic calculations, a major problem in the Deferred Approaches is that due to the nature of probability, false positive and negatives may be unintentionally introduced during cleaning.

Solution: As stated previously, the inclusion of multiple probabilistic techniques or even deterministic approaches should increase the intelligence of the methodology to block artificial anomalies from being generated.

- **Problem:** Specifically with regard to the Data Mining technique, it relies on the order the rules appear as opposed to using any intelligence to decipher the correct course of action.

Solution: It is necessary to increase the intelligence of the order of the rule order by integrating high level probabilistic or deterministic priority systems.

- **Problem:** With regard to the Cost-Conscious Cleaning method, due to the fact that the method only utilises immediate previous readings and focuses on finding the least costly algorithm, accuracy may be lowered to ensure the most cost-effective action.

Solution: In the event that this algorithm is applied at a Deferred stage, it will not require

the data to be corrected as fast as possible. Therefore in this situation, the emphasis on cost-effectiveness is not relevant as is usually the case and other actions could be examined to derive the highest accuracy.

- **Problem:** As a general constraint of all Deferred Approaches, it is necessary to apply the correction algorithm at the end of the capture cycle when the data is stored in the Database. The main problem with this characteristic is that the methodologies will never be able to be applied as the data is being captured and, therefore, cannot correct in real-time.
Solution: As most of the Deferred Approaches, especially the Data Mining and Highly Intelligent Classifier, requires certain observational data to correct anomalies, we propose the use of a buffering system that runs as the data is being captured and takes snapshots of the read data to correct any anomalies present. Unfortunately, due to the need that the methodology is run in real-time, it may not be able to include all the complexities of the current Deferred Approaches such as dynamic training of the classifiers.

6.4 Drawback analysis

In this research, we evaluated the current state-of-the-art approaches designed to correct the various anomalies and issues associated with RFID technology. From our findings, we have found that, while Physical Approaches do increase the chances of a tag being captured, it does generate duplicate anomalies and places cost in both time and labour onto the user that may not be beneficial. With regard to Middleware Approaches, we found that most anomalies are corrected through these techniques. However, due to the limited scope of information available, the more complex procedures such as dealing with highly ambiguous errors or transforming the raw observations into high-level events is not possible. In contrast, Deferred Approaches have an advantage to correct highly ambiguous anomalies and transform events. Its main issue, however, is not being available to process the observational information in real-time limiting its cleaning to a period after the records have been stored.

Overall, we have found from our research that a truly robust RFID system that eliminates all possible natural and artificial anomalies generated will require the integration of most approaches we have recognised. For example, various real-time anomalies are best filtered at the edge while increasingly ambiguous anomalies can only be corrected at a deferred stage of the capture cycle. Additionally, we found that there is a need to, not only employ probabilistic techniques, but also deterministic where possible as it theoretically should reduce the artificial anomalies produced. We, therefore, recommend the inclusion of all methods where possible, at least one of the Middleware and Deferred categories, and, where applicable, the inclusion of both deterministic and probabilistic techniques.

7. Conclusion

In this study, we have examined RFID technology and its current uses in various applications. We have also examined the three various issues among the integration of the systems including security, privacy and data abnormalities. Furthermore, we have examined the data abnormality issue to find that four problems exist including low-level nature, large intakes, data anomalies and complex spatial and temporal aspects. There have been various methodologies proposed in the past to address the various problems in the data abnormalities categorised into physical, middleware and deferred solutions. Unfortunately, due the various drawbacks such as application-specified solutions, lack of analytical information or reliance

on user-specified/probabilistic algorithms, current approaches do not provide the adequate support needed in RFID systems to be adopted in commercial sectors.

Specifically, we contributed the following to the field of RFID study:

- We provided a detailed survey of RFID technology including how it was developed, its various components and the advantages of integrating its technology into business operations.
- We highlighted the current usages of RFID categorising it into either “Integrated RFID Applications” and “Specific RFID Applications”.
- We examined the various issues preventing the adoption of RFID technology including the concerns of security, privacy and characteristics. We also focused on the specific Anomalies generated by the capturing hardware including wrong, duplicate and missing errors.
- After examining the issues surrounding RFID, we investigated the state-of-the-art approaches currently employed for correction. We categorised these methodologies into Physical, Middleware or Deferred Approaches.
- Finally, we explored the drawbacks found in currently employed Approaches and suggested several solutions in the hope of generating interest in this field of study.

With regard to future work, we specifically would like to extend our previous studies discussed in Section 5.3 by allowing it to function in real-time. We would do this through the creation of a buffer system discussed in Section 6.3 by taking snapshots of incoming data and correcting anomalies where found. We also firmly believe that this sincerely is the next step of evolution of our approach to allow it to be employed as the observational records are read into the Middleware.

8. References

- Bai, Y., Wang, F. & Liu, P. (2006). Efficiently Filtering RFID Data Streams, *CleanDB*, pp. 50–57.
- Brusey, J., Floerkemeier, C., Harrison, M. & Fletcher, M. (2003). Reasoning About Uncertainty in Location Identification with RFID, Workshop on Reasoning with Uncertainty in Robotics at IJCAI.
- Carbunar, B., Ramanathan, M. K., Koyuturk, M., Hoffmann, C. & Grama, A. (2005). Redundant Reader Elimination in RFID Systems, *SECON*.
- Chawathe, S. S., Krishnamurthy, V., Ramachandran, S. & Sarma, S. E. (2004). Managing RFID Data, *VLDB*, pp. 1189–1195.
- Chen, W. T. (2006). An Efficient Anti-Collision Method for Tag Identification in a RFID System, *IEICE Transactions* 89-B(12): 3386–3392.
- Cocci, R., Diao, Y. & Shenoy, P. (2007). SPIRE: Scalable Processing of RFID Event Streams, *5th RFID Academic Convocation*.
- Cocci, R., Tran, T., Diao, Y. & Shenoy, P. J. (2008). Efficient Data Interpretation and Compression over RFID Streams, *ICDE*, IEEE, pp. 1445–1447.
- Collins, J. (2004). Boeing Outlines Tagging Timetable, *RFID Journal*. Available from: <<http://www.rfidjournal.com/article/view/985/1/1>>.
- Collins, J. (2005). DOD Tries Tags That Phone Home, *RFID Journal*. Available from: <<http://www.rfidjournal.com/article/articleview/1458/1/1/>>.
- Darcy, P., Stantic, B. & Derakhshan, R. (2007). Correcting Stored RFID Data with Non-Monotonic Reasoning, *Principles and Applications in Information Systems and Technology (PAIST)* 1(1): 65–77.

- Darcy, P., Stantic, B., Mitrokotsa, A. & Sattar, A. (2010). Detecting Intrusions within RFID Systems through Non-Monotonic Reasoning Cleaning, *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2010)*, pp. 257–262.
- Darcy, P., Stantic, B. & Sattar, A. (2009a). A Fusion of Data Analysis and Non-Monotonic Reasoning to Restore Missed RFID Readings, *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2009)*, pp. 313–318.
- Darcy, P., Stantic, B. & Sattar, A. (2009b). Augmenting a Deferred Bayesian Network with a Genetic Algorithm to Correct Missed RFID Readings, *Malaysian Joint Conference on Artificial Intelligence (MJCAI 2009)*, pp. 106–115.
- Darcy, P., Stantic, B. & Sattar, A. (2009c). Improving the Quality of RFID Data by Utilising a Bayesian Network Cleaning Method, *Proceedings of the IASTED International Conference Artificial Intelligence and Applications (AIA 2009)*, pp. 94–99.
- Darcy, P., Stantic, B. & Sattar, A. (2010a). Applying a Neural Network to Recover Missed RFID Readings, *Australasian Computer Science Conference (ACSC 2010)*, pp. 133–142.
- Darcy, P., Stantic, B. & Sattar, A. (2010b). Correcting Missing Data Anomalies with Clausal Defeasible Logic, *Advances in Databases and Information Systems (ADBIS 2010)*, pp. 149–163.
- Darcy, P., Stantic, B. & Sattar, A. (2010c). Intelligent High-Level RFID Event Transformation Utilising Non-Monotonic Reasoning, *Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference on*, pp. 1–4.
- Darcy, P., Stantic, B. & Sattar, A. (2010d). X-CleLo: Intelligent Deterministic RFID Data Transformer, *International Workshop on RFID Technology*, pp. 59–68.
- Decker, C., Kubach, U. & Beigl, M. (2003). Revealing the Retail Black Box by Interaction Sensing, *ICDCS Workshops*, pp. 328–333.
- Derakhshan, R., Orłowska, M. E. & Li, X. (2007). RFID Data Management: Challenges and Opportunities, *RFID 2007*, pp. 175 – 182.
- Deshpande, A., Guestrin, C., Madden, S., Hellerstein, J. M. & Hong, W. (2004). Model-Driven Data Acquisition in Sensor Networks, *VLDB*, pp. 588–599.
- Deshplande, A., Guestrin, C., Madden, S., Hellerstein, J. & Hong, W. (2005). Model-based Approximate Querying in Sensor Networks, *VLDB*, pp. 417–443.
- Ding, J. & Liu, F. (2009). Novel Tag Anti-Collision Algorithm with Adaptive Grouping, *Wireless Sensor Network (WSN) 1(5)*: 475–481.
- Embry, W. (2005). Are you ready for RFID? The promise and challenges of implementing Radio Frequency Identification (RFID) systems across the extended supply chain, *Technical report*, SAS.
- Emerson & Cuming Microwave Products (2008). Emerson & Cuming Microwave Products – ECCOPAD [online], Emerson & Cuming Microwave Products. Available from: <<http://www.eccosorb.com/europe/english/page/66/eccopad>>.
- Engels, D. W. (2005). On Ghost Reads in RFID Systems, *Technical Report AUTOIDLABS-WP-SWNET-010*, Auto-ID Labs.
- EPCGlobal (2005). EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860MHz-960MHz Version 1.0.9. Available from: <<http://www.epcglobalinc.org/standards/>>.
- EPCGlobal (2006). EPCGlobal Tag Data Standards Version 1.3: Ratified Specification. Available from: <<http://www.epcglobalinc.org/standards/tds/>>.
- EPCGlobal (2008). EPCGlobal Tag Data Standards Version 1.4: Ratified Specification. Available from: <<http://www.epcglobalinc.org/standards/tds/>>.

- Ferguson, R. B. (2005). Logan airport to demonstrate baggage, passenger rfid tracking, eWeek. Available from: <<http://www.eweek.com/c/a/Mobile-and-Wireless/Logan-Airport-to-Demonstrate-Baggage-Passenger-RFID-Tracking/>>.
- Floerkemeier, C. (2004). A Probabilistic Approach to Address Uncertainty in RFID, *Auto-ID Labs Research Workshop*.
- Floerkemeier, C. (2007). Bayesian Transmission Strategy for Framed ALOHA Based RFID Protocols, *RFID, 2007. IEEE International Conference on RFID Gaylord Texan Resort, Grapevine, TX, USA*, pp. 228–235.
- Floerkemeier, C. & Lampe, M. (2004). Issues with RFID Usage in Ubiquitous Computing Applications, in A. Ferscha & F. Mattern (eds), *Pervasive Computing: Second International Conference, PERVASIVE 2004*, pp. 188–193.
- Floerkemeier, C., Lampe, M. & Schoch, T. (2003). The Smart Box Concept for Ubiquitous Computing Environments, *Proceedings of sOc'2003 (Smart Objects Conference)*, pp. 118–121.
- Gonzalez, H., Han, J. & Shen, X. (2007). Cost-Conscious Cleaning of Massive RFID Data Sets, *ICDE*, pp. 1268–1272.
- Harrop, P. (2005). RFID in the Postal Service, MoreRFID. Available from: <http://www.infowars.com/articles/big_brother/rfid_trackers/rfid_in_the_postal_service.html>.
- Jeffery, S. R., Franklin, M. J. & Garofalakis, M. N. (2008). An Adaptive RFID Middleware for Supporting Mtagphysical Data Independence, *VLDB Journal* 17(2): 265–289.
- Jeffery, S. R., Garofalakis, M. N. & Franklin, M. J. (2006). Adaptive Cleaning for RFID Data Streams, *VLDB*, pp. 163–174.
- Juels, A. (2006). RFID Security and Privacy: a Research Survey, *Selected Areas in Communications, IEEE Journal on* 24(2): 381–394.
- Karygiannis, T., Eydt, B., Barber, G., Bunn, L. & Phillips, T. (2007). Guidelines for Securing Radio Frequency Identification (RFID) Systems. National Institute of Standards and Technology.
- Khousainova, N., Balazinska, M. & Suci, D. (2006). Towards Correcting Input Data Errors Probabilistically using Integrity Constraints, *MobiDE*, pp. 43–50.
- Khousainova, N., Balazinska, M. & Suci, D. (2007). Probabilistic RFID Data Management, *UW CSE Technical Report UW-CSE-07-03-01*.
- Khousainova, N., Balazinska, M. & Suci, D. (2008). PEEK: Extracting Probabilistic Events from RFID Data, *International Conference on Data Engineering (ICDE)*, pp. 1480–1482.
- Khousainova, N., Welbourne, E., Balazinska, M., Borriello, G., Cole, G., Letchner, J., Li, Y., Ré, C., Suci, D. & Walke, J. (2008). A Demonstration of Cascadia through a Digital Diary Application, *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 1319–1322.
- Konidala, D. M., Kim, Z. & Kim, K. (2007). A simple and cost-effective RFID tag-reader mutual authentication scheme, *International Conference on RFID Security (RFIDSec)'07*, pp. 141–152.
- Lampe, M. & Floerkemeier, C. (2004). The Smart Box Application Model, In: Alois Ferscha, Horst Hoertner, Gabriele Kotsis (Eds.): *Advances in Pervasive Computing, Austrian Computer Society (OCG)*, pp. 351–356.
- Landt, J. (2001). *Shrouds of Time The history of RFID*, The Association for Automatic Identification and Data Capture Technologies.
- Landt, J. (2005). The history of RFID, *Potentials, IEEE* 24(4): 8–11.

- Langheinrich, M. (2009). A Survey of RFID Privacy Approaches, *Personal Ubiquitous Comput.* 13: 413–421.
- Lee, J. G., Hwang, S. J. & Kim, S. W. (2008). Performance Study of Anti-collision Algorithms for EPC-C1 Gen2 RFID Protocol, *Information Networking. Towards Ubiquitous Networking and Services*, Vol. 5200/2008, Springer Berlin/Heidelberg, Estoril, Portugal, pp. 523–532.
- Lee, S. R. & Lee, C. W. (2006). An Enhanced Dynamic Framed Slotted ALOHA Anti-collision Algorithm, *Emerging Directions in Embedded and Ubiquitous Computing*, Springer Berlin/Heidelberg, Seoul, Korea, pp. 403–412.
- Mirowski, L. & Hartnett, J. (2007). Deckard: A System to Detect Change of RFID Tag Ownership, *International Journal of Computer Science and Network Security* 7(7): 89–98.
- Mitrokotsa, A., Rieback, S. & Tanenbaum, A. (2010). Classifying RFID Attacks and Defenses, *Special Issue on Advances in RFID Technology, Information Systems Frontiers* pp. 491–505.
- Myung, J. & Lee, W. (2006a). Adaptive binary splitting: a RFID tag collision arbitration protocol for tag identification, *Mob. Netw. Appl.* 11(5): 711–722.
- Myung, J. & Lee, W. (2006b). Adaptive splitting protocols for RFID tag collision arbitration, *MobiHoc '06: Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*, ACM, New York, NY, USA, pp. 202–213.
- Patterson, D. J., Liao, L., Fox, D. & Kautz, H. A. (2003). Inferring High-Level Behavior from Low-Level Sensors, *Ubicomp*, pp. 73–89.
- Peng, X., Ji, Z., Luo, Z., Wong, E. C. & Tan, C. J. (2008). A P2P Collaborative RFID Data Cleaning Model, *The 3rd International Conference on Grid and Pervasive Computing GPC 2008*, pp. 304–309.
- Philipose, M., Fishkin, K. P., Perkawitz, M., Patterson, D. J., Fox, D., Kautz, H. & Hahnel, D. (2004). Inferring Activities from Interactions with Objects, *Pervasive Computing* pp. 10–17.
- Polniak, S. (2007). *The RFID Case Study Book RFID application stories from around the globe*, Abhisam Software.
- Potdar, V., Hayati, P. & Chang, E. (2007). Improving RFID Read Rate Reliability by a Systematic Error Detection Approach, *RFID Eurasia, 2007 1st Annual*, pp. 1–5.
- Preradovic, S., Balbin, I., Karmakar, N. & Swiegers, G. (2008). A Novel Chipless RFID System Based on Planar Multiresonators for Barcode Replacement, *RFID, 2008 IEEE International Conference on*, pp. 289–296.
- Preradovic, S. & Karmakar, N. (2009). Design of Short Range Chipless RFID Reader Prototype, *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2009)*, pp. 307–312.
- Pupunwiwat, P. & Stantic, B. (2009). Unified Q-ary Tree for RFID Tag Anti-Collision Resolution, in A. Bouguettaya & X. Lin (eds), *Twentieth Australasian Database Conference (ADC 2009)*, Vol. 92 of CRPIT, ACS, Wellington, New Zealand, pp. 47–56.
- Pupunwiwat, P. & Stantic, B. (2010a). A RFID Explicit Tag Estimation Scheme for Dynamic Framed-Slot ALOHA Anti-Collision, *Wireless Communications, Networking and Mobile Computing (WiCOM 2010)*, IEEE, Chengdu, China, pp. 1–4.
- Pupunwiwat, P. & Stantic, B. (2010b). Dynamic Framed-Slot ALOHA Anti-Collision using Precise Tag Estimation Scheme, in H. T. Shen & A. Bouguettaya (eds), *Twenty-First Australasian Database Conference (ADC 2010)*, Vol. 104 of CRPIT, ACS, Brisbane, Australia, pp. 19–28.

- Pupunwiwat, P. & Stantic, B. (2010c). Joined Q-ary Tree Anti-Collision for Massive Tag Movement Distribution, in B. Mans & M. Reynolds (eds), *Thirty-Third Australasian Computer Science Conference (ACSC 2010)*, Vol. 102 of CRPIT, ACS, Brisbane, Australia, pp. 99–108.
- Pupunwiwat, P. & Stantic, B. (2010d). Resolving RFID Data Stream Collisions using Set-Based Approach, *The Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2010)*, IEEE, Brisbane, Australia, pp. 61–66.
- Rahmati, A., Zhong, L., Hiltunen, M. & Jana, R. (2007). Reliability Techniques for RFID-Based Object Tracking Applications, *DSN '07: Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, IEEE Computer Society, Washington, DC, USA, pp. 113–118.
- Rao, J., Doraiswamy, S., Thakkar, H. & Colby, L. S. (2006). A Deferred Cleansing Method for RFID Data Analytics, *VLDB*, pp. 175–186.
- Raskino, M., Fenn, J. & Lenden, A. (2005). Extracting Value From the Massively Connected World of 2015, *Technical Report G00125949*, Gartner Research.
- Ryu, J., Lee, H., Seok, Y., Kwon, T. & Choi, Y. (2007). A Hybrid Query Tree Protocol for Tag Collision Arbitration in RFID systems, *MobiHoc '06: Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*, IEEE Computer Society, Glasgow, UK, pp. 5981–5986.
- Schoute, F. C. (1983). Dynamic Frame Length ALOHA, *IEEE Transactions on Communications* 31(4): 565–568.
- Schuman, E. (2005). Will Users Get Buried Under RFID Data?, Enterprise Technology News and Reviews. Available from: <<http://www.eweek.com/c/a/Enterprise-Applications/>>.
- Shih, D.-H., Sun, P.-L., Yen, D. C. & Huang, S.-M. (2006). Taxonomy and Survey of RFID Anti-Collision Protocols, *Computer Communications* 29(11): 2150–2166.
- Stockman, H. (1948). Communication by Means of Reflected Power, *Insistute of Radio Engineers (IRE)*, pp. 1196–1204.
- Swedberg, C. (2005). Hospital Uses RFID for Surgical Patients, *RFID Journal*. Available from: <<http://www.rfidjournal.com/article/articleview/1714/1/1/>>.
- Tan, C. C., Sheng, B. & Li, Q. (2008). How to monitor for missing RFID tags, *IEEE ICDCS*, Beijing, China, pp. –.
- TecO & SAP/CEC (2003). Projects - SmartShelf, TecO. Available from: <<http://www.teco.edu/projects/smartshelf/>>.
- Thamilarasu, G. & Sridhar, R. (2008). Intrusion Detection in RFID Systems, *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pp. 1–7.
- Vogt, H. (2002). Efficient Object Identification with Passive RFID Tags, *Pervasive '02: Proceedings of the First International Conference on Pervasive Computing*, pp. 98–113.
- Wan, D. (1999). Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare, *Handheld and Ubiquitous Computing*, pp. 352–355.
- Wang, F. & Liu, P. (2005). Temporal Management of RFID Data, *VLDB*, pp. 1128–1139.
- Wang, F., Liu, S. & Liu, P. (2010). A temporal RFID data model for querying physical objects, *Pervasive and Mobile Computing* 6(3): 382–397.
- Want, R., Fishkin, K. P., Gujar, A. & Harrison, B. L. (1999). Bridging Physical and Virtual Worlds with Electronic Tags, *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pp. 370–377.

- Ward, M., van Kranenburg, R. & Backhouse, G. (2006). *RFID: Frequency, standards, adoption and innovation*, JISC Technology and Standards Watch.
- Welbourne, E., Khoussainova, N., Letchner, J., Li, Y., Balazinska, M., Borriello, G. & Suciu, D. (2008). Cascadia: A System for Specifying, Detecting, and Managing RFID Events, *Mobile systems, applications, and services (MobiSys)*, pp. 281–294.
- Ziekow, H. & Ivantysynova, L. (2008). A Probabilistic Approach for Cleaning RFID Data, *RFDM'08 Workshop in conjunction with ICDE 2008*, pp. 106–107.