Reducing False Reads in RFID-Embedded Supply Chains

Yu-Ju Tu[1] and Selwyn Piramuthu[2]

[1] Information Systems, University of Illinois at Urbana-Champaign, tuyuju@ufl.edu
[2] Information Systems and Operations Management, University of Florida, selwyn@ufl.edu

Abstract

RFID tags are increasingly being used in supply chain applications due to their potential in engendering supply chain visibility and reducing tracking errors through instantaneous location of objects as well as the tagged object's immediate ambient conditions. During its life-time, an RFID tag is highly likely to physically pass through different organizations that have disparate information systems with varying information needs. We consider management of data in these applications with respect to volume, data and query characteristics, privacy and security, among others, and propose means to reduce false positive and false negative reads. The beneficial properties of RFID tags can be realized only when they are readable with a high degree of accuracy. The read rate accuracy in turn depends on several factors including the medium and distance between tag and reader, presence of objects that are impenetrable by RF signal, orientation of tag with respect to the reader, among others. We propose three algorithms that address read rate accuracy, and illustrate their relative performance using an example scenario.

Key words: RFID, Supply chain management, Object identification, False reads, Read accuracy

1   Introduction

Radio Frequency Identification (RFID) tags have been used since World War II for tracking and recording ambient conditions of tagged objects. RFID tags are increasingly being used to supplant bar codes, although their use spans a much broader spectrum of applications. Barcodes have also been in existence since the 1940s, and are clearly more widely used compared to RFID tags simply because of their characteristics including cost and the ability to print bar codes on virtually any substrate material.

Both RFID tags and bar codes have been in commercial use since the early 1970s [8]. Their use has simplified the process of automatically identifying, tracking, and processing data about the presence of entities in a system. While barcodes have their advantages such as low unit cost and their physical profile, they are inherently limited by their low information content. Even the basic RFID tag, on the other hand, has memory to store more information than barcodes and enough minimal processing power to perform minor computations when necessary. RFID tags have other beneficial properties including the ability to be read as a group without direct line of sight by a reader that need not be in close proximity. Barcodes, unlike RFID tags, are generally placed on the outside of objects and therefore are vulnerable to destruction in hostile environments. In addition to being used to supplant barcodes, RFID tags are used in applications where barcodes were never intended to be used such as for sensing ambient temperature, locating the presence of tagged objects without direct line-of-sight (e.g., Siemens' RFID tagged sponge inadvertently left inside humans in operating rooms), preventing theft, and identifying counterfeit items by tagging genuine items, among others [5], [30], [31], [35], [43]. These characteristics are oftentimes used to justify the high cost involved in building and maintaining necessary infrastructure such as tags, readers, back-end servers, and middle-ware to support RFID-enabled systems [9]. A majority of applications involving RFID tags are related to effective management of supply chains [14], [15], [32], [24], [36], [42]. A tagged entity can be easily tracked from beginning to end of a supply chain. Moreover, recent decrease in unit price of RFID tags has prompted resurgence in interest for these tags [31].

Michael and McCathie [23] study RFID-embedded supply chain applications and evaluate the pros and cons of such a setting vis-à-vis a system with barcodes and conclude that the pros outweigh the cons in the long run to improve supply chain efficiency. As beneficial properties, they list automatic non line of sight scanning, labor reduction due to increased automation, enhanced visibility of supply chain, improved asset tracking and inventory management, item level tracking, traceable warranties and targeted product recalls, improved reliability, quality control and regulation, improved utilization of resources, security against product shrinkage, durability, and capacity to hold more information. As concerns, they list software and equipment costs, lack of standards, resistance to cooperation among different layers in the supply chain required to maintain information transparency, interferences that thwart signal transmission between tag and reader, and privacy concerns. Cost and privacy concerns have generally been recognized as major factors influencing the success of RFID applications [11], [13], [18], [25], [29], [33], [44]. After conducting a Delphi study with four experts, Viehland and Wong [37] conclude that although cost is a major factor, privacy concern is not a major issue at present since most applications do not have item-level RFID tags.

The ability to identify entity instances, as opposed to entity class using bar codes, using RFID tags essentially translates to an exponential increase in data generated in systems with RFID tags. This is a direct result of the ability and need to store information about each tagged object instance instead of an object class. The problems associated with handling huge amounts of data is exacerbated even further by the need or ability to store more information about an entity instance such as its unique identification number, tagged object characteristics, among others. The process of planning and collecting these huge amounts of data translates to a useless and expensive exercise unless the data thus gathered can be used for decision making purposes. Generally speaking, as far as decision making with such data is concerned, they are solely for the purpose of tracking over time, identifying its presence at any location, etc. of the tagged object of interest. Given that the read rate in an appreciable number of applications do not come anywhere close to a 100%, a side effect or unintended use of these data can be to help determine the presence or absence of any given tag at any given location at any given time.

Based on their source of power, RFID tags can be classified as passive, semi-passive, or active. Passive tags, the most commonly used RFID tags do not have an internal power source. The active tags have their own (internal or external) power source. The semi-passive tags have internal power only for restricted (e.g., collect sensor data) purposes. The passive tag gets its power through inductive coupling frequency of the reader. With very limited capacities, the passive tag is generally used only for identification purposes. However, because of its low cost, passive tags are used in a majority of cases including both personal and organizational applications. Borriello et al. [6] demonstrate an application system that adopts the passive tag and can remind users if they mistakenly leave their personal belongings behind. Deterministic rules are used in the RFID application system to verify whether an object that is supposed to be with a person at a given time does indeed exist with that person. This system also has its own language to help users express reminders. For example, *reminder (day=tuesday, destination=work, location=home, starttime=12:00, endtime=1:00, items=[keys, wallet])*. They develop a prototype and implement it to show the feasibility of this system with existing RFID tag technology.

To boost operational performance, organizations have utilized RFID tags in disparate applications. Stanford [34] lists some of the most common RFID applications: access control, baggage ID, keyless entry and immobilization systems, document tracking, express-parcel tracking, book management, livestock or pet tracking, container and product tracking, and toll payment service. Apte et al. [2] use the case method to study how RFID applications can be further exploited in both civilian and military organizations with different needs. They claim that in the evaluation of RFID technology, the real-options analysis is more practicable than traditional net-present value (NPV) analysis as well as return on investment since the former takes knowledge and contingencies that are revealed in later time periods into consideration for investment decisions, which can be completely altered at a later time if necessary. Moreover, they list the general operational needs of RFID technology to be as follows: read distance, read rate, real-time asset location, process security, single-or multi-party access to information, information richness, and medium of concern. On the basis of these features, organizations can determine the most suitable RFID application system.

Clearly, there are several facets in RFID implementations and each facet has its corresponding pros and cons. We consider the data generated as a result of incorporation of RFID tags. Specifically, we consider some characteristics of RFID-generated data and how they relate to overall system performance. Although a majority of literature in this area mistakenly assumes that data read from RFID tags are highly accurate, variations in accuracy can and do occur due to several reasons. While not significantly altering the findings in these literatures, incorporating the actual read rate accuracy can precipitate with the need to reexamine the results. These accuracy variations are generally measured in terms of the difference between observed or measured state and the true state. Type I and Type II errors are commonly used as measures for this purpose. In this context, a Type I error (or, false positive) occurs when a tagged object is confirmed to be present when it really is absent. Similarly, a Type II error (or, fast negative) occurs when a tagged object is confirmed to be absent when it really is present. Depending on the circumstance as well as performance objective, either or both Type I and II errors can result in overly optimistic or overly pessimistic decisions in the supply chain. There is, therefore, a clear need to reduce these read error rates. We propose algorithms to reduce false positive and false negative readings of RFID tags, and illustrate their relative performance in terms of read rate accuracy. Specifically, we utilize data that are already present in RFID-embedded systems to improve the accuracy of determining the presence or absence of individual RFID tags in these systems.

The remainder of this paper is organized as follows. We list general characteristics of data generated in RFID applications in the next section. We then review related literature in Section 3. We present our proposed algorithms aimed at reducing false positives and false negatives in Section 4. We experimentally evaluate the proposed algorithms, and illustrate their relative performance in Section 5. We discuss the importance of using RFID data, the proposed algorithms, and their effects in Section 6. Section 7 concludes this paper.

2   RFID Data Characteristics

We consider some of the characteristics that distinguish analyses of RFID generated data, specifically as it relates to the volume of data, typical query characteristics, and privacy/security issues.

2.1   Data Volume

Compared to typical barcode applications, RFID applications generate more data. RFID-enabled systems generate about ten to hundred times the data volume of typical bar code applications [21]. In large retail and supply chain organizations, this translates to an estimated daily RFID data generation that is measured in terms of terabytes, an amount that far exceeds input data rate in most current systems and data warehouses.

The order of this magnitude depends on the requirement of the system of interest. For example, in a supply chain scenario that is comparable for both barcodes and RFID tags, decision makers may be content with scanned data just indicating the presence of the entity of interest at different locations (e.g., manufacturer, warehouse, and retailer). On the other extreme, data could be gathered more frequently along the supply chain not only about the presence of the entity of interest but also about ambient conditions (e.g., temperature, humidity) surrounding it. Even if the polling frequencies for barcode and RFID data are the same, the number of parameters that is captured and used by applications in these categories is not the same. In other words, even if the number of records about an entity in captured and stored data is the same, the number of attributes captured may not necessarily be the same. Therefore, compared to the barcode case the resulting database is larger in size for the RFID case even when data are gathered at the same frequency. This implies that as the polling frequency increases, the size of the database as measured by both the number of records as well as the number of attributes increases.

Chawathe et al. [7] discuss the challenges associated with RFID data management. For example, they consider how to determine the right number, type, and placement of readers. These clearly depend on the application requirements. They then discuss means to transform raw data into a form that can be used by organizational applications. Since sensor data is oftentimes incomplete, they then consider how inferences could be made from such data. Answering these questions reveal that it is necessary to study such issues in extensive detail to further discover how to efficiently clean, filter, and augment the vast amount data generated in such RFID applications.

This paper is available online at
www.jtaer.com

Angeles [1] also raises some concerns with regard to RFID data management. They first consider how the high volume of RFID data can be rendered consistent across the firm and its value chain participants. Additionally, the level of granularity as well as interpretation speed is also considered. Thirdly, the application software has to be able to incorporate the business manager's expertise and experience in coping with routine decisions. For example, a reader on a smart shelf that tracks items may not trigger actions based on available information unless the inventory count has dropped below the Economic Order Quantity (EOQ) level. They conclude that the database system used in the infrastructure should be capable of reliably dealing with high volume data generated by various RFID applications.

2.2   Query Characteristics

As data are gathered more frequently, there is a direct tendency for more frequent queries. This is because frequent data gathering results in finer granularity information about the tags. The presence of finer granularity information essentially triggers more queries or more complex queries to take advantage of the increased availability in wealth of information. For example, consider a basic scenario where an entity is loaded in a truck at a warehouse to be delivered to a retailer after five days. Without the presence of a tracking mechanism between the warehouse and the retailer, there is really no need for queries based on instantaneous tracking information during those five days. However, with the presence of reader(s) in the truck that transmit scanned information frequently, it is natural to expect more queries based on instantaneous tracking information.

The frequency of queries is also bound to increase with the presence of non-passive tags that generate sensor data, depending on volume of sensor data generated per unit time and the tag's data storage constraints. Since the tags are not queried every time a query is processed at the database, frequency of polling tags depend on their (1) mobility, (2) authentication frequency, and in case of non-passive tags (3) frequency and volume of generated sensor data, if any. By mobility, we mean the physical movement of the object to which the tag is attached, either by itself or by being carried by other entities. Although clearly context dependent, the more mobile a tag is, the increase in need to identify its location and therefore a corresponding increase in the number of times it is polled. However, even if a tag is not mobile, it may have to be authenticated several times over time to ensure its continued presence at a given location. An increase in frequency and volume of sensor data generated may also translate to more frequent polling of tags.

To collect these types of sensor data, Madden [22] proposes a query processing system architecture using a declarative approach to data collection. In this data-centric approach, users can easily specify the data that they are interested in collecting through simple declarative queries. Examples of such queries are: "tell me the average temperature on the 4th floor of this building," or "tell me the location of the sensor with the least remaining battery capacity." Since such queries are of logical nature, the data user is freed from concerns over the underlying data collection and processing details.

2.3   Privacy/Security Issues

Using the Delphi method, Viehland and Wong [37] used a panel of four experts to evaluate the future of RFID. They considered seven issues including standardization, system cost, business process reengineering, integration, privacy, lack of RFID-skilled professionals, and warehousing. Based on their study, they concluded that standardization was the most important issue and privacy the fifth most important issue. Standardization is an evolving process, and is being addressed by organizations such as EPCglobal [41]. Despite their results, security and privacy issues are of paramount significance in most consumer goods (e.g., [26], [27]). Privacy and security issues generally fall under a few major categories: (1) privacy of customers who are in physical possession of the object, and thus the tag that is attached to it, (2) security of tags from malicious adversaries who retrieve information from tags without the knowledge of owners and later use such information to clone tags, spawn DoS (Denial of Service) attacks, impersonate tags to valid readers, etc., and (3) privacy of generated data. Customer privacy has been a major concern, and several approaches to addressing this have been proposed over the years (e.g., clipped tags where tag antenna can be clipped by the customer upon purchase of the tagged item, blocker tags where signal between tag and reader is selectively blocked). Data generated by RFID tags are rich in the sense that they contain an enormous amount of information about the tags and their ambient surroundings across time. An unintended recipient of this data can possibly use it for malicious purposes. The data thus generated by RFID tags must therefore be protected.

There have been numerous instances where opposition from customers has resulted in companies scaling back the extent of spread of RFID deployments. For example, a widely known case is that of Benetton, a famous Italian clothing retailer that was planning to deploy RFID tags in its clothing line. However, there was no information to the ultimate customer instructing them on how to disable the tags after the sale. This resulted in massive consumer reaction because the RFID tags can carry enough data bits to identify the individual garment instead of only the type. Therefore, it is possible to track the movement of each consumer by any tag reader. This concern finally resulted in a boycott against Benetton (Site 1). To address these types of concerns, Garfinkel's [10] paper discusses approaches to ensure personal privacy and technologies that can be used to prevent abuse of RFID tags and the information

they transmit. Garfinkel also discusses how someone can abuse this technology by using covert tag readers to track items that are associated with individuals [10].

3   Related Literature

Liu et al. [20] explore data characteristics that determine how data are incorporated in a general data modeling framework for RFID applications (DMRA) accounting for variations in aggregation-, location-, temporal-, and history-oriented data. They consider several scenarios such as those with fixed reader as well as mobile reader with or without its own operator. The tags considered are those that allow read only operations and those that allow both read and write operations by either a reader, the sensor itself, or both. They extend the traditional Entity-Relationship (ER) model to capture temporal properties generated through interactions among entities including state changes (e.g., changes in object location, aggregation relationships, beginning and end of operations, and reader location) and events (e.g., observations, sensor measurements, property values, and transacted items). Again, given the existence of several means to address the temporal nature of data in databases (e.g., [12]) and the "temporal" component in RFID applications is not disparate compared to those in other applications, these extensions are hard to justify.

Wang and Liu [39] use Dynamic Relationship Entity-Relationship (DRER) model, which is an extension of ER model, to represent temporal data in RFID applications. They illustrate the proposed modeling extensions using a prototype of Siemens RFID middleware in a healthcare context. Although extensions to ER model are minimal, justifications of these modifications seem ad hoc since there are clearly ways to accomplish the same without these extensions. For example, the SQL/OLAP functions that are part of the SQL99 standard can be utilized for processing temporal (or, sequential) data.

Bai et al. [3] propose means to filter and clean data streams from RFID applications that contain false (i.e., false positive, false negative) reads and duplicates, while preserving arrival sequence. False negative readings could be a result of metal shielding or RF waves causing signal interference or due to collisions that occur when a reader attempts to simultaneously read multiple tags. Collision occurs whenever signal from multiple tags are simultaneously read by a reader, and it is not uncommon for a reader to simultaneously attempt to confirm the presence of multiple tags. Although not specific to RFID applications, the phenomenon of signal collision has been studied extensively (e.g., [38]). False positive readings or noise could be a result of the reader's detection field being spread wider than necessary that signals from tags farther than its intended scope are captured. Duplicate readings occur when tags remain in the reader's detection field for a longer duration, when tags are simultaneously present in the field of multiple readers, or when multiple tags attached to an object are simultaneously scanned. Problems associated with false negative and false positive readings are alleviated through multiple repeated readings. However, this process generates a significant number of duplicate readings.

Bai et al. [3] propose means to reduce noise and remove duplicates in data using sliding time window, and through elimination of readings with frequency count below a pre-determined threshold respectively. They propose lazy de-noising and eager de-noising methods based on when a data point that is determined to be valid is saved in the database, with the former ignoring signal sequence constraints. Choosing the window size involves balancing between being able to identify the presence of a tag and at the same time being able to capture its dynamics, where the former requires a larger window size while the latter requires a smaller window size. A given window size is not ideal under all circumstances. Window size should be updated periodically depending on the dynamics of the system [17]. The benefits gained from using a sliding time window scheme seems fairly minimal. Record keeping is expensive, and needs to be justified based on potential benefits. Just keeping track of sliding time window and the tags that enter and leave the window at any given point in time necessitates allocation of resources including computational as well as storage space. The benefits are those that relate to preventing false positives, false negatives, and duplicate readings. Duplicate readings are not a serious problem in most realistic RFID applications since they are often easily corrected at the edge of the network by edge devices which often have this capability (e.g., OAT Systems offers error detection and correction capabilities in their edge systems; SAP's Auto-ID infrastructure includes data filtering, enriching, and aggregation components).

False positives can be reduced by calibrating the reader's detection field appropriately. As for false negatives, given the time lag between successive readings from a given tag, a tag that signals false negative reading at any given point in time has a high probability of repeating the same during successive readings. Therefore, commonly suggested means to alleviate problems associated with false negative readings are suspect unless they address the main cause (e.g., metal or other radio-wave-impenetrable shielding between tag and reader, collision due to the presence of several tags). For instance, the problems associated with the presence of metal shielding will not change regardless of the number of attempts at bouncing signals off of tags unless there is a change in the orientation of tag with respect to the reader.

Bai et al. [4] propose a data stream management system (DSMS) language for processing stream query data for temporal event detection, extending the declarative rule based language and event engine proposed in Wang et al. [40]. Specifically, they present example scenarios where they claim that standard SQL does not have the expressive modeling power and show how their proposed extension accommodates these cases.

This paper is available online at
www.jtaer.com

Hu et al. [16] propose a bitmap data type and illustrate its beneficial characteristics using a prototype implemented in Oracle DBMS. They claim that tracking related groups (e.g., based on physical proximity or shared property) of items is more efficient than individual items. The bitmap representation thus is used to model a collection of RFID tags. They show that in addition to its loss-less transformation, their bitmap representation results in compact data storage. The bitmap data type translates to expensive incremental modifications, and is therefore better suited for applications where updates occur infrequently. Hu et al. [16] propose keeping an item-level table in addition to those with bitmap representation to keep track of modifications that occur in-between updates at the aggregate level. Although this seems as a work-around, this results in duplicating information that could potentially increase storage space requirements, and thus defeating the purpose and the claims supporting compact representation. Moreover, maintaining an item-level table involves additional processing and storage resources. One of the beneficial characteristic of RFID tags is the ability to identify and track individual tags. Although tracking of aggregates (e.g., pallets) works in principle in a supply chain, it is inappropriate in other applications (e.g., hospital where a tagged item is taken out of a room). Even in a supply chain, it may be critical to record ambient conditions for individual objects (e.g., objects on a pallet might not all be subject to the same ambient conditions depending on the size of the pallet and the location of the object of interest on the pallet), and an aggregate bitmap clearly would not suffice in such applications.

Most data cleansing approaches for RFID data are done before data are stored in databases. Rao et al. [28] propose deferred data cleansing of data in RFID systems since eager cleansing may not be feasible (e.g., rules and business context for cleansing data may change over time and across applications) nor appropriate (e.g., pharmaceutical e-pedigree tracking that precludes cleansing of data) in all applications. They use the functionalities present in SQL/OLAP for processing sequential data.

Generally speaking, data reliability and ambiguity are the two major issues in extraction of information from RFID data. Most existing methods used for RFID event detection either ignore issues related to data ambiguity and data input error or deterministically clean the data before processing. However, in certain situations, input data errors can lead deterministic approaches to sacrifice large significant events. Khoussainova et al. [19] use a probabilistic method to approach this sort of problem. With probabilistic properties, their approach provides the application with the flexibility to build its own balance between detection and precision rate. Through experimentation, they also show their approach to have better event detection rates than traditional deterministic approaches.

4   Proposed Methods

As seen in Section 3, several studies have attempted to propose means to reduce false (positive and negative) reads. False positives and false negatives are a problem in RFID tag-embedded systems since the ensuing decisions depend crucially on those results. Such a decision not only affects the quality of immediate local decisions since several levels or layers of decision-making that follow directly or indirectly depend on results from lower levels. Any inaccuracy in such read rates would result in cascading propagation of erroneous decisions that magnify its deleterious effects and consequences.

There is, therefore, a clear need to address this problem where it begins. We propose means to reduce false positives and false negatives through a variation of triangulation where we consider the presence/absence of a related tag. Specifically, the proposed algorithms are based on the premise that most RFID tags generally are not present alone. In other words, at any point in the supply chain, a tag is usually present with at least one other tag. We use this property to develop algorithms that decrease false positive and false negative readings.

In the system we model, two readers are simultaneously used to read information from each RFID tag that passes through the system. We assume that tag T1, which is the tag of interest, and another tag (say, T2, which can be different tags at different points in the supply chain) are expected to be present together. Therefore, we have two tags (T1 and T2) and two readers (R1 and R2) in this scenario. Our goal is to reduce the read rate error for the tag of interest (i.e., T1). We propose three algorithms: the base case algorithm and two variations of the base case. While Algorithms 1 and 2 utilize information about the tag of interest in determining its presence or absence, Algorithm 3 utilizes ambient information about another (i.e., the second) tag to determine the presence of the tag of interest.

4.1   Algorithm 1 (the base case)

The base case makes use of information only from the tag of interest (i.e., T1). Here, the tag of interest (T1) is confirmed to be present only when both the readers (R1 and R2) identify its presence. I.e., when both R1 and R2 identify it as being present, tag T1 is confirmed to be present. When either R1 or R2 or both identify it as being absent at any given point in time, tag T1 is confirmed to be absent. We use this as the base case to compare the other two algorithms that we propose.

4.2   Algorithm 2

This algorithm is similar to Algorithm1 when both the readers agree. Here, the tag (T1) is confirmed to be present when both the readers (R1 and R2) identify its presence just like in the base case. As in the base case, when both the readers (R1 and R2) identify it to be absent, the tag T1 is confirmed to be absent in the field of the readers. However, unlike the base case where only one of the readers identifies the tag to be present, we use a sliding window approach to determine the presence or absence of the tag. We use a window size of 15 tag reads. For the first 15 tag reads, we assume the tag is present with a probability of 0.5. After the first 15 tag reads, we consider the immediate past 15 tag reads to determine the tag's presence at that point in time. If the number of reads indicating the tag's presence is more than those indicating the tag's absence, we confirm the presence of the tag at that time point. Otherwise, we confirm the absence of the tag at that location and time point.

4.3   Algorithm 3

Unlike Algorithms 1 and 2, Algorithm 3 utilizes information about a second tag (i.e., T2) that is assumed to be simultaneously present with the tag of interest (i.e., T1). Here, the tag of interest (T1) is assumed to be present when both the readers (R1 and R2) identify its presence just like in the base case. Moreover, when both the readers (R1 and R2) identify it to be absent, tag T1 is confirmed to be absent in the field of the readers just like in the base case. However, when one of the readers identifies it to be present while the other does not, we utilize information from reads corresponding to the other tag (i.e., T2) in determining the presence/absence of the tag of interest (T1).

In the scenario where both the readers present conflicting evidence for tag T1, when both the readers identify tag T2 to be present, we confirm the tag T1 to also be simultaneously present. When neither of the readers (R1 and R2) identifies the tag T2 to be present, we confirm the tag T1 to be absent as well. When only one of the readers (R1 or R2) identifies the tag T2 to be present, we utilize a sliding window approach similar to that in Algorithm 2. Similar to Algorithm 2, we use a window size of 15. For the first 15 tag (T2) reads, we confirm the presence of tag (T1) with probability 0.5. After the first 15 tag reads, we consider a sliding window of size 15 and take a majority vote. I.e., If a majority of tag reads indicate the presence of the tag (T2), we conform tag T1's presence. Similarly, if majority of tag reads indicate the absence of the tag (T2), we confirm tag T1's absence at that point in time. By utilizing additional information about the presence or absence of a related tag (i.e., T2) to determine the presence or absence of the primary tag of interest (i.e., T1), we decrease the probability of false readings due to unavoidable circumstances such as tag orientation, presence of impenetrable object in the field of the reader, among others.

4.4   Implementation Details

To illustrate the performance of the three proposed algorithms and to compare their relative performance, we implemented and simulated the three algorithms 10 times with 10,000 reads per run with the following assumptions: We assume the probability that tag T1 is present, P(T1)=0.5. When the tag T1 is present, P(reader R1 confirming T1 is present)=0.8 and P(reader R2 confirming T1 is present)=0.8. Similarly, for tag T2, P(reader R1 confirming T2 is present)=0.8 and P(reader R2 confirming T2 is present)=0.8. We include these probabilities to account for signal noise and other incidental reasons resulting in false readings. For the sliding window part, we use 15 tag reads as a basic unit to implement our approach. For example, we take the majority vote of every read for the first fifteen time points, and from then on we use a sliding window of 15 to determine the presence or absence of a tag of interest. Therefore, whenever both readers observe inconsistent identification results for a tag, that tag's presence or absence will be determined either by chance with probability 0.5 for the first time point or by majority voting for the next 14 time units until a threshold read of 15 is reached. Beyond that point in time, we use majority voting with the appropriate sliding window of 15 reads.

5   Results

Results from these simulation runs, in the form of confusion matrices, are given in Tables 1-3. In these tables, the numbers given are the average of ten runs with 10,000 readings per run. The standard deviation values are given in parentheses. The true positives are in the upper left quadrant and the true negatives are in the lower right quadrant. Our goal is to improve the combined values in these two cases. Of the false cases, the false positives are in the upper right quadrant and the false negatives are in the lower left quadrant. Our goal here is to reduce the combined values in these two cases. Depending on the application, false positives and false negatives may have different negative utilities.

As can be seen from Tables 1-3, the number of true positives is the lowest for the base case and highest for algorithm 3. However, the number of true negatives is highest for the base case and lowest for Algorithm 2. The number of false positives is highest for Algorithm 2 and lowest for the base case, whereas the number of false negatives is highest for the base case and lowest for Algorithm 3. Overall, the true cases (true positives and true negatives) are highest for Algorithm 3 and lower and comparable for Algorithms 1 and 2 and the false cases (false positives and false negatives) are comparable and higher for Algorithms 1 and 2 and lower for Algorithm 3. The standard deviation values are lowest for the best among the three Algorithms for the different cells of interest (true

This paper is available online at
www.jtaer.com

positive, true negative, false positive, and false negative). Results also indicate that Algorithm 1 has the lowest scores for test results that are positive and the highest scores for test results that are negative. Whereas results from Algorithm 3 show that it performs the best for the true positive case, it performs the worst for the false negative case. However, results for Algorithm 2 show that it performs the worst for the false positives case as well as true negative case.

Table 1: Results (confusion matrix) for Algorithm 1 (base case)

	ACTUAL		
	Present	Absent	
TEST RESULT	324.5 (10.34)	18.9 (4.70)	Present
	182.3 (14.24)	474.3 (17.68)	Absent

Table 2: Results (confusion matrix) for Algorithm 2

	ACTUAL		
	Present	Absent	
TEST RESULT	422.2 (30.05)	111.4 (20.46)	Present
	84.6 (22.16)	381.8 (29.76)	Absent

Table 3: Results (confusion matrix) for Algorithm 3

	ACTUAL		
	Present	Absent	
TEST RESULT	454.6 (16.26)	52.5 (8.42)	Present
	52.2 (8.66)	440.7 (19.55)	Absent

Results from Tables 1-3 also indicate that, overall, using more ambient information leads to improved tag identification. This is especially critical in environments where a tag is invisible to the reader due to unforeseen circumstances (e.g., RF shielding) and the only means to determine its presence is through determining the presence or absence of a related tag that is required to be simultaneously present with the tag of interest. Without such supplemental information from other tags, it is virtually impossible to obtain any information about a tag of interest that is invisible to the reader. Moreover, such data are collected in most RFID tagged systems as a part of their regular operations. The only difference is in the decision to use such data for tag locating purposes. As we have shown using three proposed algorithms, this decision does not necessarily translate to processing overload on the

system due to their minimal additional processing requirement. Any information that could be obtained with minimal processing is multiple-fold better than utilizing incorrect tag reads for decision-making (e.g., inventory) purposes. The base case (Algorithm 1) represents the scenario where no assumption with respect to the presence of a tag is made, and a conservative approach is taken. Results based on the three algorithms indicate that as more information about tag reads are introduced, for the tag of interest as well as a related tag, the read rate accuracy can be improved with minimal additional processing resources.
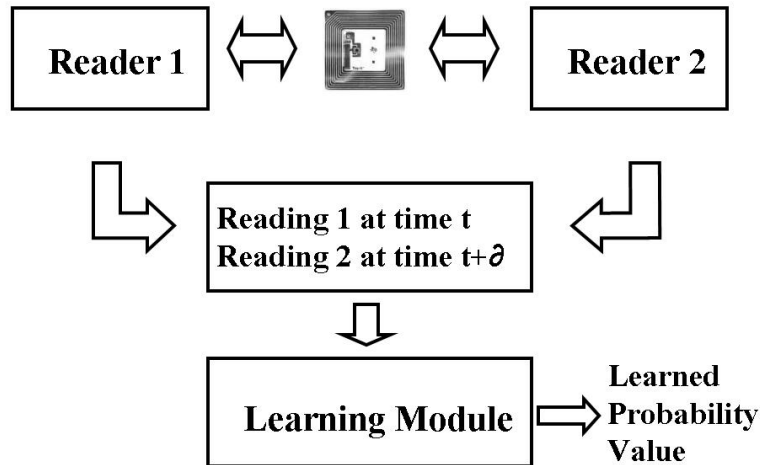
5.1   Extension

Figure 1: Extended Framework

A possible extension to Algorithm 3 is to dynamically vary the probability as more data are gathered over time. A schematic of this extension is provided in Figure 1. Here, we have two readers and a tag of interest that is read twice within short periods of time at each location in the supply chain. This can be done in addition to including information about the presence or absence of a related tag. With experimentation, we learn the probability of co-occurrence of pairs of different types of tagged objects over time as more data are gathered. Learning itself occurs in the learning module, and the learned knowledge can be represented using production rules in IF-THEN form.

6   Discussion

False readings are a serious problem in RFID tag embedded systems. Although most proponents of RFID tag embedded systems generate their recommendations based on error-free reads of these tags, the reality does not support this assumption. Error-free reads are possible only in ideal scenarios where the tag and reader are in close proximity to each other with no interference from other objects in the reader's field and no deleterious effects due to the medium separating the reader and tag, which is optimally oriented with respect to the reader. Such ideal environments are a luxury in most realistic scenarios, and are available only in a small fraction of RFID tag embedded systems. This reality clearly dictates necessary measures to alleviate any resulting deleterious effects. We presented a means to approach this through improving the read rate accuracy in these systems.

The algorithms proposed in this paper aim to reduce false positives and false negatives while identifying the presence/absence of an RFID tag in the field of a reader. We proposed and illustrated three algorithms using an example scenario. Although our results are based on simplified assumptions, they show that false positives and false negatives can indeed be reduced through minor modifications to input data analysis. Traditionally, data read from RFID tags have not been used to improve their read rates. This is a recent development, as is seen from the brief review of existing literature presented in Section 4.

Among the algorithms we studied, Algorithm 3 is clearly the best and Algorithm 1 the worst overall. The proposed methods have several advantages over the moving time window algorithms that have been previously proposed. This includes reduction in processing and the absence of need to wait for a given threshold number of readings of a tag to been taken before a data point can be validated. Moreover, in most applications, a given tag is scanned a minimal number of times at any location and the windowing algorithms are not applicable in these scenarios.

The results from this study are preliminary since they are specific to the probability values used. Further research is necessary to validate the findings across a wider range of probability values. Regardless of the availability of extensive results, preliminary results from this study are promising. It clearly shows the need to utilize any available information, even from other tags, to leverage the odds for increasing the read rate accuracy. Since data gathering accuracy has been touted as a major beneficial property of RFID tag-enabled systems, false readings can drastically

This paper is available online at
www.jtaer.com

reduce the utility of RFID tags. There is, therefore, an urgent need for studies that attempt to improve true (positive and negative) reads. Given the additional costs that are associated with any RFID tag-enabled system, justifications for additional investment can only be made when the read rates can be guaranteed to be close to error-free. This study is a step in that direction. We are in the process of extending this study to include more ambient information in determining the presence/absence of RFID tags in the field of a reader.

7   Conclusion

Competing technologies (e.g., bar codes, sensors), although not formidable from a purely functional perspective, evoke resistance among prospective RFID tag users due to their relative cost, privacy/security issues, etc. RFID tags clearly have their strengths and weaknesses, and capitalizing on their strengths necessitates RFID tag reads to be as close to error-free as is possible. Given the recent interest and widespread use of RFID tag technology there have been relatively few attempts to mitigate problems associated with false tag reads. Moreover, any such attempt must ensure that it does not overly burden an existing system. Data collected from RFID tags are used for identification, authentication, as well as other purposes. The algorithms we propose use these data for yet another purpose – that of reducing effects due to false reads. A majority of existing RFID tag applications does not compensate or account for false reads. The proposed algorithms utilize data that are already collected in these systems to improve read rate accuracy with minimal additional burden on the system.

Websites List

Site 1: Boycott Benetton - No RFID tracking chips in clothing!
http://www.boycottbenetton.com

References

[1]   R. Angeles, RFID Technologies: Supply-Chain Applications and Implementation Issues, Information Systems Management, vol. 22, no. 1, pp. 51-65, 2005.
[2]   U.M. Apte, N. Dew and G. Ferrer, (2006, January). What is the Right RFID for your Process? [Online]. Available: http://www.acquisitionresearch.org.
[3]   Y. Bai, F. Wang, and P. Liu, Efficiently Filtering RFID Data Streams, CleanDB, 2006.
[4]   Y. Bai, F. Wang, P. Liu, C. Zaniolo, and S. Liu, RFID Data Processing with a Data Stream Query Language, in Proceedings of the 23nd International Conference on Data Engineering, ICDE 2007, pp. 1184-1193, 2007.
[5]   G. Bell and P. Dourish, Yesterday's Tomorrows: Notes on Ubiquitous Computing's Dominant Vision, Personal and Ubiquitous Computing, vol. 11, no. 2, pp. 133-143, 2007.
[6]   G. Borriello, W. Brunette, M. Hall, C. Hartung, and C. Tangney, Reminding About Tagged Objects Using Passive RFIDs, Proceedings of Ubicomp, pp. 36-33, 2004
[7]   S.S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. Sarma, Managing RFID Data, In Proceedings of the International Conference on Very Large Data Bases (VLDB). Toronto, Canada. August 2004.
[8]   B. Eckfeldt, What Does RFID Do for The Consumer?, Communications of the ACM, vol. 48, no. 9, pp. 77-79, 2005.
[9]   K. Finkenzeller, RFID Handbook, second edition, Wiley, 2002.
[10] S.L. Garfinkel, Adopting Fair Information Practices to Low Cost RFID Systems, Privacy in Ubicomp'2002 workshop, September 2002.
[11] S.L. Garfinkel, A. Juels, and R. Pappu, RFID Privacy: An Overview of Problems and Proposed Solutions, IEEE Security and Privacy, vol. 3, no. 3, pp. 34-43, 2005.
[12] H. Gregersen and C.S. Jensen, Temporal Entity-Relationship Models – A Survey. IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 3, pp. 464-497, 1999.
[13] O. Gunther and S. Spiekermann, RFID and the Perception of Control: the consumer's view, Communications of the ACM, vol. 48, no. 9, pp. 73-76, 2005.
[14] Z. Guo, F. Fang, and A.B. Whinston, Supply Chain Information Sharing in a Micro Prediction Market, Decision Support Systems, vol. 42, no. 3, pp. 1944-1958, 2006.
[15] H.V.D. Heijden, Mobile Decision Support for In-store Purchase decisions, Decision Support Systems, vol. 42, no. 2, pp. 656-663, 2006.
[16] Y. Hu, S. Sundara, T. Chorma, and J. Srinivasan, Supporting RFID-based Item Tracking Applications in Oracle DBMS Using a Bitmap Datatype, in Proceedings of the 31st VLDB Conference, pp. 1140-1151, 2005.
[17] S.R. Jeffery, M.N. Garofalakis, and M.J. Franklin, Adaptive Cleaning for RFID Data Streams, in Proceedings of the 32nd VLDB Conference, pp. 163-174, 2006.
[18] A. Juels, RFID Security and Privacy: A Research Survey, IEEE Journal on Selected Areas in Communications, vol. 24, no. 3, pp. 381-394, 2006.
[19] N. Khoussainova, M. Balazinska, and D. Suciu, Probabilistic RFID Data Management, UW CSE Technical Report UW-CSE-07-03-01, 2007.
[20] S. Liu, F. Wang, P. Liu, Integrated RFID Data Modeling: An Approach for Querying Physical Objects in Pervasive Computing, in Proceedings of CIKM06, pp. 822-823, 2006.

[21] D. Lyle. (2004, August). Understanding and Solving the RFID Data Dilemma. [Online]. Available: http://www.cioupdate.com/article.php/3391791.

[22] S. R. Madden, The Design and Evaluation of a Query Processing Architecture for Sensor Networks, PhD Dissertation, University of California, Berkeley, 2003.

[23] K. Michael and L. McCathie, The Pros and Cons of RFID in Supply Chain Management, in Proceedings of the International Conference on Mobile Business, pp 623-629, IEEE Press, 2005.

[24] F. Niederman, R.G. Mathieu, R. Morley, and I.W. Kwon, Examining RFID applications in supply chain management, Communications of the ACM, vol. 50, no. 7, pp. 77-79, 2007.

[25] M. Ohkubo, K. Suzuki, and S. Kinoshita, RFID Privacy Issues And Technical Challenges, Communications of the ACM, vol. 48, no. 9, pp. 66-71, 2005.

[26] S.Piramuthu, Protocols for RFID Tag/Reader Authentication, Decision Support Systems, vol. 43, no. 3, pp. 897-914, 2007.

[27] S.Piramuthu, Lightweight Cryptographic Authentication in Passive RFID-Tagged Systems, IEEE Transactions on Systems, Man, and Cybernetics Part-C, vol. 38, no. 3, pp. 360-376, 2008.

[28] J. Rao, S. Doraiswamy, H. Thakkar and L.S. Colby, A Deferred Cleansing Method for RFID Data Analytics, in Proceedings of the 32nd VLDB Conference, pp. 175-186, 2006.

[29] M.R. Rieback, B. Crispo, and A. S. Tanenbaum, The Evolution of RFID Security, IEEE Pervasive Computing, vol. 5, no. 1, pp. 62-69, 2006.

[30] K. Romer, T. Schoch, F. Mattern, and T. Dubendorfer, Smart Identification Frameworks for Ubiquitous Computing Applications, Wireless Networks, vol. 10, no. 6, pp. 689-700, 2004.

[31] F. Siegemund, C. Floerkemeier, and H. Vogt, The Value of Handhelds in Smart Environments, Personal and Ubiquitous Computing, vol. 9, no. 2, pp. 69-80, 2005.

[32] N. Singh, K. Lai, and T.C.E. Cheng, Intra-organizational Perspectives on IT-enabled Supply Chains, Communications of the ACM, vol. 50, no. 1, pp. 59-65, 2007.

[33] F. Stajano, RFID is X-ray Vision, Communications of the ACM, vol. 48, no. 9, pp. 31-33, 2005.

[34] V.M. Stanford, IEEE Pervasive Computing: Applications - Pervasive Computing Goes the Last Hundred Feet with RFID Systems, IEEE Distributed Systems Online, vol. 4, no. 6, 2003.

[35] D. R. Thompson, RFID Technical Tutorial, Journal of Computing Sciences in Colleges, vol. 21, no. 5, pp. 8-9, 2006.

[36] D. Veeramani, J. Tang, and A. Gutierrez, A Framework for Assessing the Value of RFID Implementation by Tier-One Suppliers to Major Retailers, Journal of Theoretical and Applied Electronic Commerce Research, vol. 3, no. 1, pp. 55-70, 2008.

[37] D. Viehland and A. Wong, The Future of Radio Frequency Identification, Journal of Theoretical and Applied Electronic Commerce Research, vol. 2, no. 2, pp. 74-81, 2007.

[38] H. Vogt, Efficient Object Identification With Passive RFID Tags, in Proceedings of the International Conference on Pervasive Computing, Pervasive2002, pp. 98-113, Springer-Verlag, 2002.

[39] F. Wang and P. Liu, Temporal Management of RFID Data, in Proceedings of the 31st VLDB Conference, pp. 1128-1139, 2005.

[40] F. Wang, S. Liu, P. Liu, and Y. Bai, Bridging Physical and Virtual Worlds: Complex Event Processing for RFID Data Streams, EDBT, 2006.

[41] S.F. Wamba, and H. Boeck, Enhancing Information Flow in a Retail Supply Chain using RFID and the EPC Network: A Proof-of-Concept Approach, Journal of Theoretical and Applied Electronic Commerce Research, vol. 3, no. 1, pp. 92-105, 2008.

[42] L. v.d. Wijngaert, J. Versendaal, and R. Matla, Business IT Alignment and Technology Adoption: The Case of RFID in the Logistics Domain, Journal of Theoretical and Applied Electronic Commerce Research, vol. 3, no. 1, pp. 71-80, 2008.

[43] R.C. Yoder, Using RFID in the Classroom to Teach Information Systems Principles, Journal of Computing Sciences in Colleges, vol. 21, no. 6, pp. 123-129, 2006.

[44] T. Zhang, Y. Ouyang, and Y. He, Traceable Air Baggage Handling System Based on RFID Tags in the Airport, Journal of Theoretical and Applied Electronic Commerce Research, vol. 3, no. 1, pp. 106-115, 2008.