Using Low-Level Reader Data to Detect False-Positive RFID Tag Reads

Thorben Keller*, Frédéric Thiesse†, Jens Kungl‡ and Elgar Fleisch‡
*IBM Deutschland GmbH
Frankfurt, Germany
Email: thorben.keller@de.ibm.com
†University of Wuerzburg
Wuerzburg, Germany
Email: frederic.thiesse@uni-wuerzburg.de
‡METRO Group Information Technology GmbH
Duesseldorf, Germany
Email: jens.kungl@mgi.de
§University of St. Gallen & ETH Zurich
St. Gallen & Zurich, Switzerland
Email: elgar.fleisch@unisg.ch

***Abstract*—Radio Frequency Identification (RFID) can be used in various ways for the optimization of supply chain management processes. However, there are technological constraints that delay a reliable and productive use of the technology. One of these constraints is the problem of false-positive RFID tag reads i.e., tags that have been read unintentionally by an RFID reader. We propose a machine learning based approach that makes use of the low-level reader data collected when reading tags to detect such false-positives. We evaluate our approach by verifying it with data collected in a productive RFID enabled distribution center, where it is necessary to distinguish between pallets that are loaded onto trucks and pallets that are in range of the reader by accident only. Furthermore, we identify several attributes which are expected to reveal characteristics within the low-level reader data that is typical to such false-positive reads.**

I. Introduction

Radio Frequency Identification (RFID) is a wireless communication technology that can be used for the automatic identification (Auto-ID) of physical objects. The use of the technology is expected to grow significantly in the next years and it is predicted that someday RFID tags will be as pervasive as bar codes [1]. It can be used in various ways for the optimization of supply chain management and especially for the optimization of processes in distribution centers [2], [3].

The long-term objective is the automatic identification, tracing and verification of goods along their way from the producer to the customer. Prerequisite for such an all-embracing RFID process is that the producer, as the starting point of the supply chain, attaches RFID tags to all his products. This would allow the unique identification of every single product from production to consumption. The Pros and Cons of RFID in the supply chain are discussed in [4].

One of the key processes in many supply chains is the shipping from distribution centers where a highly efficient warehouse management is essential to ensure a smooth flow of goods to the stores. If using a bar code solution, a warehouseman has to manually scan each pallet which is time consuming and error prone. In contrast, shipment dock doors equipped with RFID portals can automatically register all goods leaving the warehouse by identifying the transponder attached to the pallet. The data is then transmitted to the warehouse management system and faulty deliveries are recognized immediately [5]. The business value of RFID data in various industrial settings is discussed for example in [6] and [7].

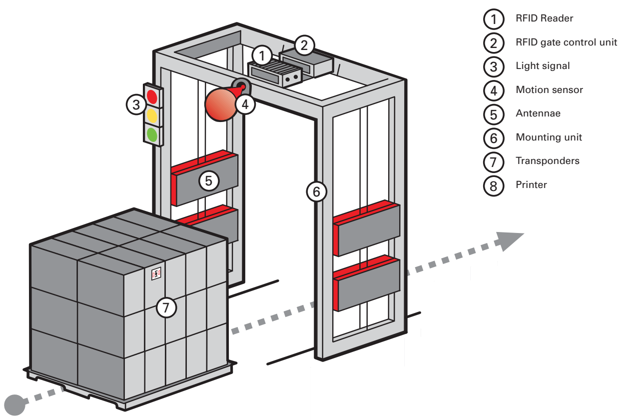


Fig. 1. RFID Portal. Source: METRO Group

Figure 1 shows the main components of such an RFID portal. It is installed right in front of the shipment dock door so that every pallet entering or leaving the truck unavoidably passes through it. In the case of outgoing goods, a warehouseman approaches the truck with an RFID tagged pallet (7) on his forklift in order to load it onto the truck. This is recognized by the motion sensor (4) which then signals the RFID reader (1) to activate its antennas (5) and detect any transponder in range. The data gathering of the reader runs for a period of about 5 seconds and is called a *gathering-cycle*. After the end

978-1-4244-7414-1/10/$26.00 ©2010 IEEE

of the gathering-cycle the warehouseman requires immediate feedback about the loading in order to correct any mistakes. This is done by the use of a color signal light (3) installed at the portal.

- If the pallet[1] has been recognized correctly and the warehouse management system (WMS) confirms that it is actually meant for the destination of the container the light flashes green.

There are occurrences of faulty deliveries that at this point are automatically detected and can immediately be dealt with.

- In the case that the pallet was not intended for the destination market the signal flashes yellow and the warehouseman has to unload it again and continues with another pallet.
- If the pallet is unknown to the warehouse management system the light flashes red. This might happen for various reasons such as incorrect tag registrations to the WMS.
- In some cases it might happen that no tags are recognized by the reader. In this case the signal light does not flash at all, telling the warehouseman that he has to reload the pallet by passing through the portal with it again.

Using this procedure, the number of faulty deliveries, i.e., pallets that are shipped to the wrong market, is expected to be minimized.

However, the use of Radio Frequency Identification to support outgoing goods yields the same problem as many other RFID applications as they are all limited by technological constraints. Since the antennas are not directed and have a read range of several meters, they do not only recognize the pallet moving through the portal but also any other in range. The problem is that one cannot distinguish between the pallet that was loaded onto the truck and the ones that just appear in the reading field by accident. The reasons why other pallets are detected by the RFID reader are manifold:

- Someone might have buffered another pallet temporarily near the portal.
- Another warehouseman is just passing by the portal with another pallet at the moment of data gathering.
- A pallet is loaded at an adjacent RFID portal.

If all the detected pallets are reported to the warehouse management system as being shipped, this would lead to incorrect invoices and stores are going to pay for goods that they have not received. In the following the pallet that was loaded onto the truck during a gathering-cycle is called a *moved pallet*, the ones that have been read by accident are called *static pallets*, or *moved* and *static* tags respectively.

The above problem, generally referred to as the problem of false-positive RFID tag reads, can be observed in many practical applications and is described in many book dealing with RFID-Technology [8], [9], [10], [11].

A *positive read* represents a transponder that is present in the reading field and is detected by the reader. A *negative read* represents a transponder present in the reading field that is not detected by the reader, for whatever reason. Consequently a *false-positive read* is a tag that has been detected by the reader but is not the one of interest. Referring to the above scenario of outgoing goods, the false-positive reads during a gathering-cycle correspond to the pallets that stand around a portal and are not the ones being currently loaded onto the truck. As long as this problem remains unsolved a reliable productive use of the RFID technology in distribution center processes like incoming or outgoing goods is rather questionable.

The paper is organized as follows. Chapter II discusses available approaches that have been proposed and how we contribute to them. The general idea of detecting false-positives is demonstrated in Chapter III using samples of low-level reader data. Chapter IV explains how the data was collected in a real world productive system for the verification of our approach. The actual approach is introduced in Chapter V followed by the results and an outlook on future work in this direction.

[1]The terms *pallet* and *tag* will be used interchangeably throughout the paper

II. Related Work

A. Literature Review

The detection of false positive RFID tag reads is a complex challenge to the technology in many applications as stated for example in [9] and [11]. As of today, only few approaches were presented in the literature to deal with this problem but they often consider it together with the occurrence of false-negative tag reads, for example in [12]. In contrast to false-positive RFID tag reads false-negatives are tags that should be read but can't for whatever reason. In the above scenario of RFID enabled outgoing goods the latter are pallets that have not been read although they were loaded onto trucks during that gathering-cycle. Several authors have discussed requirements and design alternatives for the implementation of specialized RFID middleware components to handle large amounts of raw data collected from distributed RFID readers (e.g., [13], [14]).

In [15] and [16], a sliding window approach is used to filter the RFID data stream. If a tag is read less often than a given threshold during the sliding time window it is considered be a false-positive. Because individual tags are filtered out on the low-level reader level this might lead to an increasing number of false-negatives. Thus, a balance has to be found between the readability of tags and the detection of false positives as studied in [17].

The use of additional hardware is proposed in [18] and [19]. The authors propose to determine the presence or absence of tagged objects at arbitrary locations in the supply chain. For this purpose 2 readers (i.e., one additional) are installed at each read point. If a tag is read at both readers then it is classified as present and absent if none of the readers recognized it. If only one of the readers reads the tag, a sliding window approach is used to determine the presence or absence of the tag. An advanced approach makes use of an additional tag attached to the object that has to be read by the readers as well.

B. Research Gap

The available approaches can be divided into two different groups. The first one examines the timestamps of the individual tag reads and classifies tags using rules like *the less often a tag has been read the more likely it is to be a false-positive*. The second group uses multiple readers and rules like *if two different readers have read the tag it is unlikely to be a false-positive*.

First of all, there is more information available from the reader than solely the timestamps of the individual tag reads. For each tag read also the signal strength of the tag answer and the particular antenna where the read occurred is accessible. Note that there is a difference between reader and antenna, since a reader can use multiple antennas. A detailed overview of the available data is given in Chapter IV.

Secondly, the attempt to simply use more than one reader appears rather simplistic and crude. Furthermore, this approach entails additional costs and hence is often not a valid option.

All approaches have in common that they make unrealistic assumptions since the data basis used (if any) is obtained under lab conditions. A generalization of the results is consequently questionable. Eventually none of the approaches covers a real-world scenario comparable to ours. This paper focuses on avoiding the shortcomings of prior research by using the full range of information available from the reader. Additionally, an extensive dataset obtained from a real life productive system is used to ensure a high degree of generalization.

III. From Low-Level Reader Data to Detection of Movement

During a gathering-cycle (i.e., the loading of a pallet) each tag that is in range of the antennas responds several times by reporting its Electronic Product Code (EPC) to the RFID reader. On the reader side additional information about each individual tag answer (in the following called *tag event*) is available that could possibly be used to tell moved and static pallets apart. Strictly speaking, a tag event can be interpreted as a tuple of the following form:

$$tagEvent = (EPC, RSSI, SinceStart, Antenna)$$

A. Information in the Low-Level Reader Data

During each gathering-cycle the individual tags that were read usually give multiple answers. Consequently, each tag can be described by a list of temporally ordered tag events. The meaning of the information contained in a single tag event is described in the following.

1) EPC: The Electronic Product Code (EPC) is an identification scheme for the universal identification of physical objects via RFID tags and other means. An EPC construct consists of an object class identification and a serial number used to uniquely identify the instance of the pallet along with some other information [20].

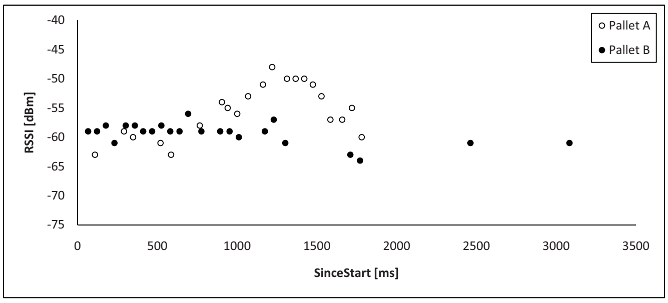


Fig. 2. Best-case gathering-cycle

2) RSSI: The Received Signal Strength Indication (RSSI), measured in dBm, is a measurement of the power of the received radio signal the tag emits and can intuitively be interpreted as *how loud* the tag has been heard by the antennas. By nature the RSSI value becomes higher the closer a tag is to the antennas and lowers the further away it is.

3) SinceStart: For each tag event a timestamp is stored. It is relative to the time that has passed since the beginning of the gathering-cycle and is measured in microseconds. Since each gathering-cycle runs for 5 seconds only this is the theoretical maximum although the vast majority of the tag events can be observed within the first 3.5 seconds after the beginning of the pallet loading.

4) Antenna: Each RFID reader has a number of antennas attached to it that try to read all tags in range. The information at which antenna the tag event occurred is stored as an integer. In our setting, each reader is equipped with 4 antennas, so the numbers stored range from 1 to 4.

B. Best Case

An example of the tag events that occurred during the first 3.5 seconds of a gathering-cycle is shown in Figure 2. In this case, there were two pallets present in the reading field of the antennas, pallet A was the one that moved through the RFID portal and pallet B was one placed nearby. On the y-axis the RSSI values are depicted, on the x-axis the timestamp of the tag event. Not shown in the figure are the antennas at which the individual tag events occurred. It can be observed that the static pallet shows more or less constant RSSI values, which is understandable since it does not change its distance to the antennas. The moved pallet however shows an increase of RSSI values after 500ms, reaching the maximum at around 1250ms, and then decreasing again. This is an indication of a pallet approaching the antennas, passing them and finally moving away from them.

C. Normal Case

In most cases it is not that easy to tell moved and static pallets apart. Figure 3 shows 4 examples of gathering-cycles randomly chosen from our sample data where different effects can be observed. Figure 3a shows a moved pallet that has been read only within the first 0.5 seconds with RSSI values dropping from -40dBm to -50dBm. The static pallet has a

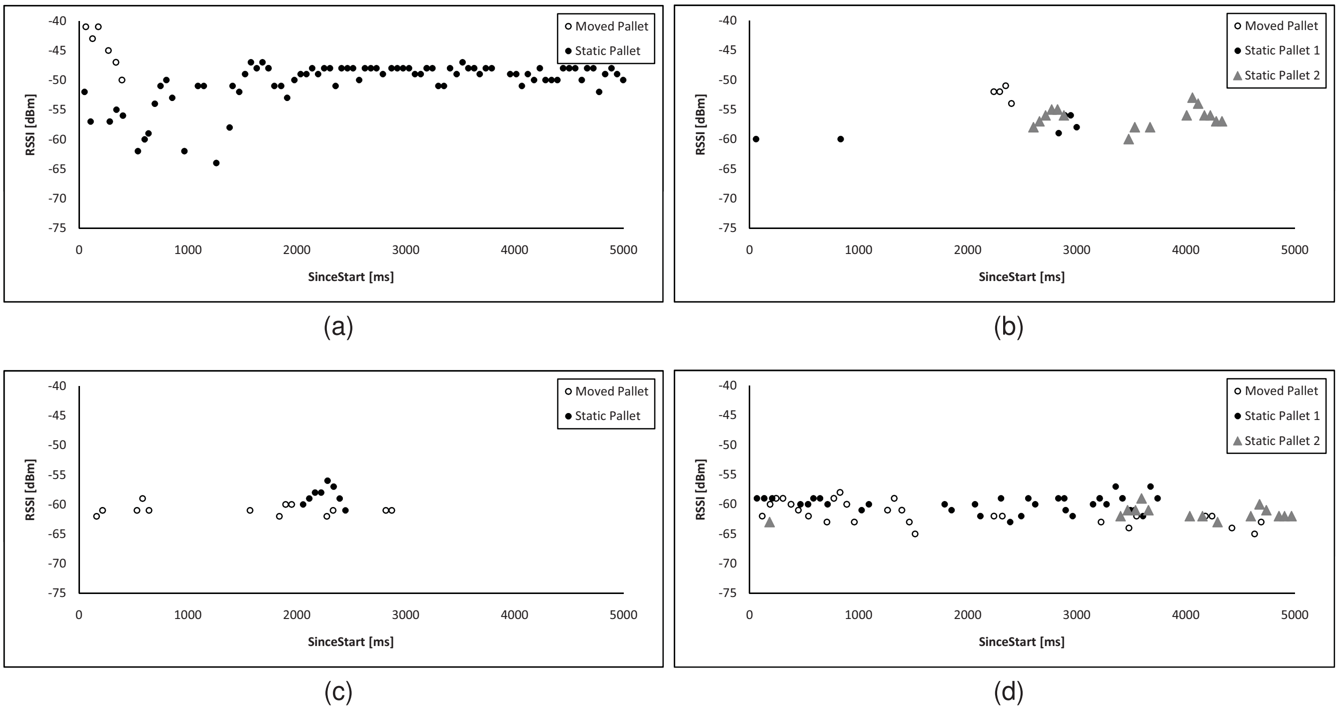


Fig. 3. Sample gathering-cycles

lot of variety within the first 1.5 seconds and then shows a constant RSSI value level. Figure 3b shows 3 pallets with only a few tag reads from each tag. While the moved pallet has been read only over a very short period of time (around 0.2 seconds) the static pallets appear to be in range of the antennas between 2 and 3 seconds. Figure 3c shows a moved pallet that has been read over a period of 3 seconds with more or less constant RSSI values and a static pallet that shows the moved-tag-like bump at the middle of the gathering-cycle. Figure 3d shows another gathering-cycle where 3 tags have been read. All of them have been read over a period of several seconds with similar RSSI values. These example gathering-cycles demonstrate how difficult it is in the normal case to identify which pallet has actually been loaded onto the truck and which ones are false-positives.

D. Tag Characteristics

The next step is to identify certain characteristics derived from the low-level reader data that can be considered as typical for a moved or a static tag. Once these characteristics are available it would be possible to determine which pallets have been moved or not. Based on the domain knowledge acquired after observing hundreds of pallet loadings the following characteristics, in the following called attributes (Table I, II and III), were identified. They are aggregations of the set of all tag events for a specific tag. So for example the minimum RSSI value of a tag is the minimum RSSI value of all tag events associated with that tag.

1) RSSI Attributes: The RSSI values depend on the distance between tag and reader. Higher RSSI values indicate that a pallet was close to the antenna when read and lower RSSI values indicate that a pallet was further away from the antenna.

TABLE I
EXPLANATION OF RSSI ATTRIBUTES

Attribute Name	Description
RSSI_MIN	The minimum of all RSSI values
RSSI_MAX	The maximum of all RSSI values
RSSI_DIFF	The RSSI value range (RSSI_MAX - RSSI_MIN)
RSSI_MEAN	The average of all RSSI values
RSSI_STDDEV	The standard deviation of all RSSI values

TABLE II
EXPLANATION OF SINCESTART ATTRIBUTES

Attribute Name	Description
SINCESTART_MIN	The time passed before the first tag event
SINCESTART_MAX	The time passed before the last tag event
SINCESTART_DIFF	The time period during which a tag has been read

2) SinceStart Attributes: It is reasonable to consider the time that has passed since the beginning of the gathering-cycle to find out whether moved and static tags are read at different timestamps. The following knowledge derived from the low-level reader data can be examined.

3) Antenna Attributes: The tags are usually attached to the same side of the pallet. Thus, it is interesting to find out if certain antennas are more likely to read either moved or static tags.

TABLE III
EXPLANATION OF ANTENNA ATTRIBUTES

Attribute Name	Description
ANT1_CNT,...,ANT4_CNT	Number of reads by the respective antenna
CNT	The sum of all antenna tag reads

IV. DATA COLLECTION

In order to analyze the various attributes, a dataset containing a large number of tag events for both moved and static pallets is required. The data collection took place in the central distribution center of METRO Group Cash & Carry markets in Unna, Germany. METRO Group is one of the largest retailers in the world and currently one of the key players in RFID adoption along the supply chain. In the distribution center all of the 70 shipment dock doors were equipped with RFID portals to automatically detect pallet shipments. Over a period of 7 months students were assigned to monitor the loading of pallets. They kept track of which pallets that have been read during a gathering-cycle were actually moved and which were static. This was done using custom developed software that immediately after the ending of a gathering-cycle shows them a list of all pallets scanned. The students now simply mark each of them as either moved or static. All observations were made at any of the 70 shipment dock doors without having an impact on the warehousemen or their way of working. During the 7 months of data collection 53,998 pallets have been monitored, where 40,743 of them were static and 13,245 were moved through the outgoing goods portal. This dataset, collected in a productive real world scenario, constitutes the foundation for the research and allows for a much better insight than any simulation under lab conditions.

V. ALGORITHM OF OUR APPROACH

The general idea is that moved and static tags are expected to show a different behavior with respect to the previously defined characteristics (i.e., attributes) based on the low-level reader data.

A. Different Tag Distributions

Figure 4a exemplarily shows the distribution of the standard deviation of the RSSI values for both tag types based on the data that was collected at the METRO Group distribution center. The y-axis depicts the number of tags that had a specific standard deviation during a gathering-cycle. The key information in this figure is that the received signal strength indication of moved tags is most likely to have a standard deviation between 4 and 8. The static tags, i.e., the false-positives, are likely to have a standard deviation between 0 and 3. The reason for this different behavior is that moved tags have by definition a significant change in distance to the antennas during the gathering-cycle in contrast to the static tags. Therefore the variation of the RSSI values is accordingly higher. Another import information in this histogram is the remarkable number of static tags having a standard deviation of RSSI values of 0 (other than depicted this corresponds to even more than 9000 static tags and not only 3000, this was done for presentation purposes only). This is because it has turned out that a large number of static tags responded only a single time to the reader thus having a standard deviation of 0.

An example of an attribute that doesn't show such a meaningful distribution is given in Figure 4b. Depicted is the time that has passed since the start of the gathering-cycle until the tags gave their first answer to the RFID reader. Similar to the above example of the RSSI standard deviation, the number of tags that answered within the first 0.4 seconds is significantly higher but has been cut down for presentation purposes. It can be seen directly that both moved and static tags usually answer within the first second. However, in contrast to the average RSSI value, there is no clear distinction between the two distributions that could help determine their tag class, as the ratio of moved and static tags is about the same for all buckets in the histogram.

B. Determining the Optimal Threshold Value

Given an attribute as shown in Figure 4a that allows to perform a useful classification, we now have to determine the optimal threshold value to separate the two distributions. For our running example, Figure 5 shows the distribution of the average RSSI values for static and moved tags is depicted. The key information in this figure is similar as in Figure 4a, i.e., moved tags usually have a significantly higher average RSSI value than static tags. But additionally a graph representing the information gain [21] (also called Kullback-Leibler divergence) is depicted. It can be interpreted as a measure of how well the two tag types can be separated using a specific value as threshold and is a common method in machine learning [22]. Illustrated below is how we use the information gain obtained for specific values to determine the optimal threshold that separates moved tags from the false-positives.

Generally, the entropy H is a measurement of the information associated with a dataset D and is defined as

$$H(D) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

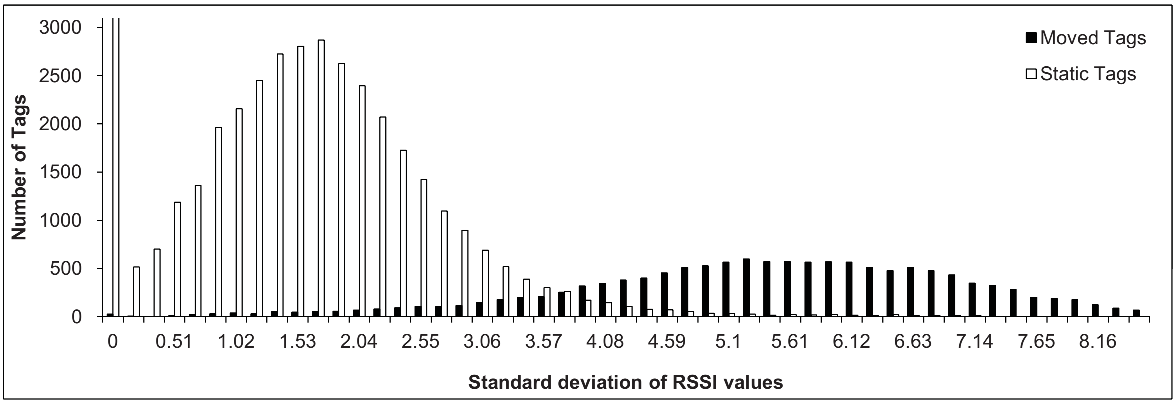
where $p(i)$ is the probability that an object in the dataset is of class i. We have two different classes in our scenario with 75.8% static and 24.2% moved tags. Thus, the entropy of our dataset equals

$$H(D) = -(0.758 \log_2 0.758 + 0.242 \log_2 0.242) = 0.798.$$

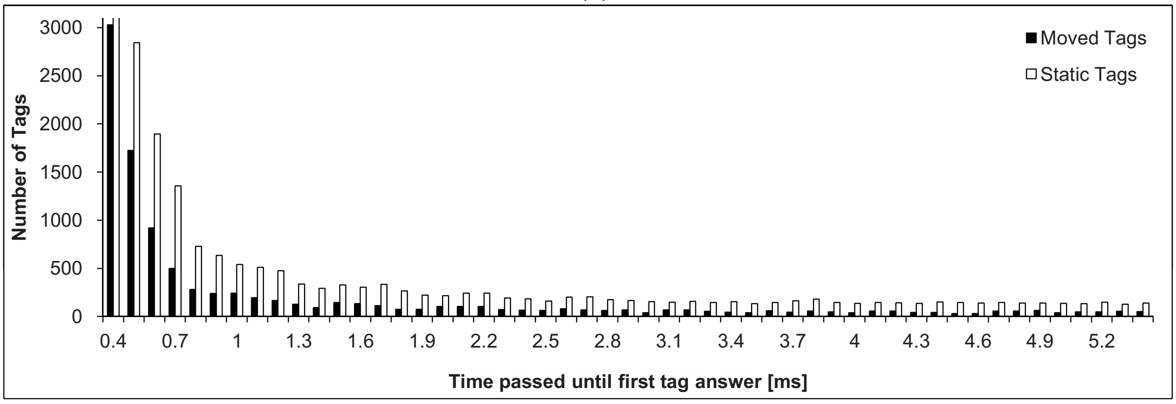
Let A be an attribute and x a threshold under consideration. Then $D_{\leq}$ is the set containing all tags having a value for attribute A less than or equal x and $D_{\geq}$ is the set of all tags with an attribute value greater than x:

$$D_{\leq}(A, x) := \{Tag \in D | Tag(A) \leq x\}$$

$$D_{\geq}(A, x) := \{Tag \in D | Tag(A) \geq x\}$$



(a)



(b)

Fig. 4. Distribution of different RSSI values

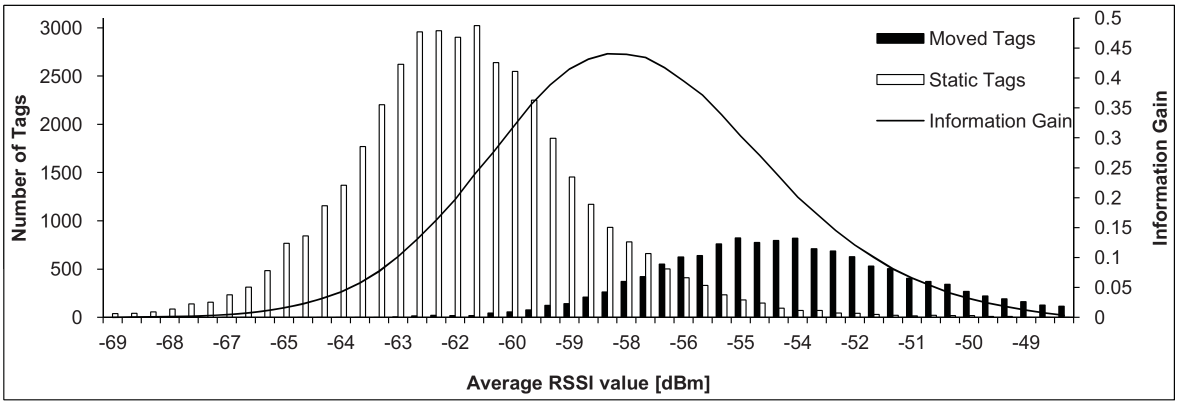


Fig. 5. Information Gain of possible Split Points for Moved and Static Tags

The information gain I obtained by splitting attribute A on x is then defined as

$$I(A,x) = H(D) - \left(\frac{|D_{\leq}(A,x)|}{|D|} \cdot H(D_{\leq}(A,x)) + \frac{|D_{\geq}(A,x)|}{|D|} \cdot H(D_{\geq}(A,x))\right).$$

The optimal threshold o to separate moved and static tags is the one where the information gain maximizes, i.e.,

$$o = \max_{x \in X}\{I(A,x)\}.$$

If there would be a perfect attribute, i.e., one with a threshold value perfectly splitting static and moved tags, the maximum information gain of the attribute is equal to the entropy of the dataset, 0.798.

Figure 5 shows the information gain calculated for all possible split points using the average RSSI value as an

TABLE IV
RSSI ATTRIBUTE CLASSIFICATION RESULTS

Attribute Name	Class Recall		Class Precision		Overall
	Static	Moved	Static	Moved	
RSSI_MIN	100.00%	0.00%	75.47%	0.00%	75.47%
RSSI_MAX	96.60%	92.88%	97.66%	89.89%	95.69%
RSSI_DIFF	96.69%	85.53%	95.36%	89.35%	93.95%
RSSI_MEAN	91.29%	91.42%	97.03%	77.33%	91.32%
RSSI_STDDEV	97.01%	87.64%	96.02%	90.49%	94.71%

attribute. As intuitively expected, the split point is located near the intersection of the two distributions (at -58.1dBm) having an information gain of 0.440.

The algorithm to decide whether a pallet has been moved or not based on the average RSSI value is now as simple as follows. For every tag that has been read during a gathering-cycle we calculate the average RSSI value. If this value is less than -58.1dBm then we assume it to be a static (i.e., a false-positive) tag otherwise we assume it has been loaded onto the truck. The feedback for the warehouseman via the signal light can now be altered accordingly. If he moves through the portal and no moved pallet was detected by the reader the lights flashes red telling the warehouseman to repeat the pallet loading. A different signal can be given when more than one tag was classified as moved (sometimes a pallet has more than one tag attached to it).

VI. RESULTS

Using the information gain approach presented above, the optimal threshold to separate moved and static tags is calculated for all considered attributes. The quality of an attribute is primarily measured by the overall classification rate C_{ALL} which is defined as

$$C_{ALL} := \frac{\text{Number of Tags correctly classified}}{\text{Number of Tags}},$$

thus corresponding to the percentage of correctly classified tags. Additionally, two more quality measures commonly used in machine learning, namely class recall R and class precision P, are calculated as well. The class recall tells us how many tags of a specific class have been classified correctly and the class precision indicates how many of the classifications of a specific class are actually correct.

$$R_{Static} := \frac{\text{Number of Tags correctly classified as Static}}{\text{Number of Static Tags}}$$

$$P_{Static} := \frac{\text{Number of Tags correctly classified as Static}}{\text{Number of Tags classified as Static}}$$

Calculation of class recall and class precision for moved tags is done analogously. Tables IV-VI present class recall and class precision for both static and moved tags as well as the overall classification rate is presented that was achieved after splitting on the pre calculated optimal threshold.

TABLE V
SINCESTART ATTRIBUTE CLASSIFICATION RESULTS

Attribute Name	Class Recall		Class Precision		Overall
	Static	Moved	Static	Moved	
SStart_MIN	100.00%	0.00%	75.47%	0.00%	75.47%
SStart_MAX	100.00%	0.00%	75.47%	0.00%	75.47%
SStart_DIFF	100.00%	0.00%	75.47%	0.00%	75.47%

TABLE VI
ANTENNA ATTRIBUTE CLASSIFICATION RESULTS

Attribute Name	Class Recall		Class Precision		Overall
	Static	Moved	Static	Moved	
CNT	100.00%	0.00%	75.47%	0.00%	75.47%
ANT1_CNT	79.56%	75.74%	90.98%	54.64%	78.62%
ANT2_CNT	80.32%	67.86%	88.49%	52.84%	77.26%
ANT3_CNT	100.00%	0.00%	75.47%	0.00%	75.47%
ANT4_CNT	100.00%	0.00%	75.47%	0.00%	75.47%

A. *RSSI Attributes*

The minimum RSSI value does not carry any information that could be used to detect false-positives. Because no useful split point could be identified there is no other possibility than classifying all tags as static which results in a classification rate of only 75%. The other RSSI attributes appear to be a lot more useful as they are all able to classify more than 90% of all tags correctly. The maximum RSSI value yields the best results with an overall classification rate of almost 96%. The RSSI value range and the standard deviation perform quite similar, only the average RSSI value performs notably worse. These results can be explained by the assumption made above, namely that tags moving through the RFID portal have a significant change in distance to the antennas. Static tags on the other hand are assumed to stand still and have not much variation in their RSSI values.

B. *SinceStart Attributes*

As can be seen from Table V, all attributes that are based on the relative timestamps of the individual tag events (Table V) have a class recall and class precision for moved tags of 0% and allow no differentiation of the tag types at all as can be seen in Figure 4a. Consequently all tags are classified as static, as there is a 75% chance that a tag is static based on the static / moved ratio.

C. *Antenna Attributes*

Similar to the attributes based on the timestamps, the antenna attributes do not appear to be helpful for distinguishing between moved and static tags. However, it is notable that there are two different antennas, namely antenna 1 and antenna 2 that have a slightly higher overall classification rate than the others. This is due to the fact that they can be used to identify a couple of moved tags.

VII. Summary & Future Work

We have shown that it is possible to detect false-positive RFID tags in distribution center processes by the use of the low-level reader data. This is possible because moved and non-moved tags show different characteristics with respect to the 3 introduced data dimensions *received signal strength indication (RSSI)*, *timestamp of the individual tag events* and *number of reads per antenna*. We generated various aggregated attributes based on the low-level reader data that we expected to be suitable for the discrimination of static and moved tags. In the scenario we presented, those static tags were attached to pallets located in the range of an RFID enabled outgoing goods portal and are considered to be false-positives. Our algorithm that uses the information gain criterion to tell static and moved pallets apart was able to classify more than 95.5% of the real world data correctly. Furthermore it was shown that the RSSI values, which can be interpreted as a measure of how loud a tag has been read by a reader, are the most suitable tag characteristics for this task.

Using the algorithm framework presented above it is now possible to identify the loaded pallet just after the data gathering has finished. In contrast to the process of RFID enabled outgoing goods used before, this yields a much more reliable loading process as false-positives are detected immediately and faulty deliveries are minimized.

There are basically two different fields of interest we are going to concentrate our attention on in the future.

- Advanced algorithms: The proposed algorithm relies on the examination of a single attribute value. It is reasonable to take additional attributes into account to identify possible relationships between different attributes. A more advanced machine learning approach is probably going to result in an even better classification performance than presented above.
- Alternative application scenarios: The data used above was acquired by monitoring the loading of pallets at RFID enabled outgoing goods in a distribution center. However, there are various applications that deal with the same problem of false-positive tag reads. In the case of electronic article surveillance (EAS) there is a need to identify tagged articles that have been stolen by a theft the moment he leaves the store by passing through an EAS portal. The same applies to an automatic point of sale where articles are automatically charged to the customer when leaving the store.

Acknowledgment

The authors would like to thank METRO Group and METRO Group Cash & Carry for making their monitored low-level reader data available. Furthermore we would like to thank the team of IBM Global Business Services for valuable input and feedback.

References

[1] D. V. Hunt, A. Puglia, and M. Puglia, *RFID: A Guide to Radio Frequency Identification*, ser. Wiley Interscience. Hoboken, NJ, USA: John Wiley & Sons Inc., 2007.

[2] M. T. Hompel and T. Schmidt, *Warehouse Management: Automation and Organisation of Warehouse and Order Picking Systems*. Berlin, Germany: Springer, 2006.

[3] D. McFarlane and Y. Sheffi, "The Impact of Automatic Identification on Supply Chain Operations," *International Journal of Logistics Management*, vol. 14, no. 1, pp. 1–18, 2003.

[4] K. Michael and L. McCathie, "The Pros and Cons of RFID in Supply Chain Management," in *Proceedings of the 4th International Conference on Mobile Business (ICMB)*, Sydney, Australia, July 2005, pp. 623–629.

[5] METRO Group, "METRO Group RFID Innovation Center - An Information and Development Platform for the Future of Commerce," 2007. [Online]. Available: http://www.future-store.org

[6] F. Thiesse, J. Al-Kassab, and E. Fleisch, "Understanding the Value of Integrated RFID Systems: A Case Study from Apparel Retail," *European Journal of Information Systems*, vol. 18, no. 6, pp. 592–614, 2009.

[7] D. Delen, B. C. Hardgrave, and R. Sharda, "RFID for Better Supply-Chain Management through Enhanced Information Visibility," *Production and Operations Management*, vol. 16, no. 5, pp. 613–624, 2007.

[8] S. R. Jeffery, G. Alonso, M. J. Franklin, W. Hong, and J. Widom, "Declarative Support for Sensor Data Cleaning," in *Proceedings of the 4th International Conference on Pervasive Computing*, Dublin, Ireland, May 2006, pp. 83–100.

[9] S. B. Miles, S. E. Sarma, and J. R. Williams, *RFID Technology and Applications*. Cambridge, UK: Cambridge University Press, 2008.

[10] J. Brusey, C. Floerkemeier, M. Harrison, and M. Fletcher, "Reasoning about Uncertainty in Address Identification with RFID," in *Workshop on Reasoning with Uncertainty in Robotics at International Joint Conferences on Artificial Intelligence (IJCAC)*, Acapulco, Mexico, August 2003, pp. 23–30.

[11] S. S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. Sarma, "Managing RFID Data," in *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)*, Toronto, Canada, 2004, pp. 1189–1195.

[12] C. Floerkemeier and M. Lampe, "Issues with RFID Usage in Ubiquitous Computing Applications," in *Proceedings of the 2nd International Conference on Pervasive Computing*, Vienna, Austria, April 2004, pp. 188–193.

[13] L. Ye, "A Research on General Software Architecture on RFID," in *Proceedings of the 3rd International Conference on Convergence and Hybrid Information Technology (ICCIT)*, Busan, South Korea, November 2008, pp. 1054–1057.

[14] C. Floerkemeier and M. Lampe, "RFID Middleware Design - Addressing Application Requirements and RFID Constraints," in *Proceedings of the Joint Conference on Smart Objects and Ambient Intelligence (soC-EUASI)*, Grenoble, France, October 2005, pp. 219–224.

[15] Y. Bai, F. Wang, and P. Liu, "Efficiently filtering RFID Data Streams," in *Proceedings of the 1st International Very Large Data Base Workshop on Clean Databases (CleanDB)*, Seoul, South Korea, September 2006, pp. 50–57.

[16] B. Jiang, K. Fishkin, S. Roy, and M. Philipose, "I Sense a Disturbance in the Force: Unobtrusive long-range Detection of passive RFID Tag Motion," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 1, pp. 187–196, 2006.

[17] S. R. Jeffery, M. Garofalakis, and M. Franklin, "Adaptive Cleaning for RFID Data Streams," in *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, Seoul, South Korea, September 2006, pp. 163–174.

[18] Y.-J. Tu and S. Piramuthu, "Reducing False Reads in RFID-Embedded Supply Chains," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 3, no. 2, pp. 60–70, 2008.

[19] ——, "A Decision Support Model for Filtering RFID Read Data," in *Proceedings of the 16th International Conference on Advanced Computing and Communications (ADCOM)*, Chennai, India, December 2008, pp. 221–224.

[20] EPCGlobal, "EPCGlobal Tag Data Standards Version 1.4," http://www.epcglobalinc.org.

[21] E. B. Hunt, J. Marin, and P. J. Stone, *Experiments in Induction*. New York, NY, USA: Academic Press, 1966.

[22] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, ser. Wiley Interscience. Hoboken, NJ, USA: John Wiley & Sons Inc., 2005, ch. Decision Trees, pp. 107–127.