Robust Statistical Methods for Detection of Missing RFID Tags

Popovski, Petar; Nielsen, Karsten Fyhn; Jacobsen, Rasmus; Larsen, Torben

Published in:
I E E E Wireless Communications Magazine

DOI (link to publication from Publisher):
10.1109/MWC.2011.5999767

Publication date:
2011

Document Version
Accepted author manuscript, peer reviewed version

Robust Statistical Methods for Detection of Missing RFID Tags

Petar Popovski, *Senior Member, IEEE*, Karsten Fyhn, *Student Member, IEEE*, Rasmus Melchior Jacobsen, *Student Member, IEEE*, and Torben Larsen, *Senior Member, IEEE*

Abstract

The technology of Radio Frequency IDentification (RFID) is being deployed in many applications, such as logistics and inventorying. However, a generic problem in all RFID systems is to ensure that the RFID readers can reliably read a set of RFID tags, such that the probability of missing tags stays below some acceptable value. This paper introduces statistical methods to deal with the problem of missing RFID tags. These methods are applied at the *reliability* layer, which initiates multiple *reader sessions* by invoking certain MAC-layer arbitration (anti–collision) protocols to collect tag responses. The reliability layer obtains a running estimate of the probability of having at least one tag missing. This estimate is used to detect if an additional reader session is required. We present several estimators, which can be used at the reliability layer to obtain an estimate of the probability of missed tags after R reader sessions have been carried out. These estimators are derived under idealized assumptions. However, when tested under more realistic conditions, which violate these ideal assumptions, the estimators exhibit high robustness and provide a very close approximation of the true probability of missing tags.

Index Terms

Missing tag problem, reliable RFID readings, tree–based arbitration protocol, RFID networks

I. Introduction

RFID technology features a growing set of applications for identification of various objects. The applications span from simply identifying objects, serving as more informative barcodes, gathering of sensory data and holding private/confidential information [1][2][3]. Passive RFID tags are powered from the signal transmitted from the reader, and information is communicated via backscattering [4]. The proliferation of passive RFID tags and their integration with sensors is expected to be one of the main enabling technologies for the "Internet of Things" paradigm.

The communication paradigm in passive RFID systems is based on request/response: in the first step, the reader sends an interrogation signal to the tags within its range. In the second step the tags send their response to the

The authors are with the Department of Electronic Systems, Aalborg University, Denmark, e-mails {petarp,kfyhn}@es.aau.dk, rmj@kamstrup.dk, tl@es.aau.dk

reader by backscattering the signal. If multiple tags simultaneously reply to the reader, the reader experiences tag collision. Hence, the reader should run a certain anti–collision protocol (also called collision resolution or arbitration protocol) in order to successfully resolve each tag in its proximity. There are various anti–collision protocols, which are in general divided into two groups: ALOHA–based [5][6] and tree–based [7][8].

Regardless of the actual arbitration protocol used, *reader session* is defined as a protocol run that is sufficient to collect the ID of all the tags in the reader's proximity *when there are no errors*. However, in practice errors do occur during a reader session if either the query from a reader is not received correctly at a tag or the tag reply is not received at the reader. In principle, if a tag is at a blind spot [9], the communication between the tag and the reader is always in error. The probability that a tag is at a blind spot can be substantial and is primarily determined by the physical disposition of the tag, but also by the material to which the tag is affixed. In [10] test results indicate, that if a tag is attached to solar cream, the probability of not resolving a tag is 30%, and with mineral water it is 67%. The error probability can vary a lot, increasing the probability of missing one or more tags. In summary, if during an arbitration protocol run the link between a reader and a tag is in error, then this tag is not read by the end of the reader session. We define this as the problem of *missing RFID tags* [11].

A. Possible Approaches to the Missing Tag Problem

There are multiple approaches to minimize the probability of missing a tag. In [12], a method for determining group completeness in an RFID network is described, based on each tag storing one or more references to surrounding tags. The resolved tags and the references are compared, and if not all references are resolved, the reading/comparison is repeated. Thereby the reader knows with high probability if tags are missing. This method is targeting rather static constellations of tags, e.g. goods on pallets. In [9] one sample is gathered by a shelf equipped with RFID readers in a retail store, while the other sample is taken by the RFID readers at the point of sales. These samples are used in the classical capture–recapture model [13] to derive estimators for the tag set cardinality.

The approach presented here is termed *multi–capture–recapture*, and it achieves reliability by using multiple (> 2) reader sessions in a capture–recapture model. We introduce a *reliability layer*, which runs on top of the anti–collision algorithm. When a reader session is finished, the reliability layer estimates the probability that there are tags that have not been read. This estimation is based on the reading results of all sessions conducted so far. If this probability is below an acceptable threshold, the reader sessions are stopped, otherwise a new session is initiated. Hence, the reading process is *reliable* if, through a sequence of several readings, it can be guaranteed that the probability of having unread tags remaining, stays under a certain tolerable value. This approach has first been proposed in [11]. In this paper we present other types of estimators and demonstrates the robustness of the used statistical methods when the idealized assumptions from [11] are violated.

Relevant target scenarios for our proposed approach are depicted on Fig. 1. The *first application example* is a turning table where goods are wrapped in plastic before they are shipped. As the table turns, the box with goods changes the position with respect to the reader. In addition, if the goods are not affixed within the box, then when the table turns, the position of a given object and the associated RFID tag is changed, which may change the

readability of that object. A question of interest is — how many times should the table be turned and read until we can be certain that the chances that there is a missed tag is e.g. less than 0.001%? In this scenario, instead of a turntable, there can be a person that uses a handheld reader and manually changes the position of the reader for each reader session; when the reliability requirements are met in a given session, the reader signals to the person that the reading process is completed and the probability that there are missed tags is acceptable. In the *second application example* on Fig. 1, the tagged goods are put in boxes that are moving on a conveyor belt. The reader has multiple antennas, distributed along the conveyor belt. The question is: how many antennas to deploy in order to guarantee certain reliability of the reading process?

The paper is organized as follows. In the next section we introduce the basic statistical mechanisms used in the reliability layer. Specific estimators are proposed in Section III under idealized assumptions. In the next sections we investigate the robustness of the estimators by considering realistic conditions, in which some of the idealized assumptions are violated. Section V concludes the paper and provides directions for future extensions.

II. A Simple Illustration of the Multi–Capture–Recapture Approach

Consider the following simple scenario: a set of tags is read by two different readers, call them r_1 and r_2, in two separate reader sessions. The set has N tags and the probability that tag i cannot be read during a session (e.g. because there is an obstacle between the tag and the reader) is p_i. We assume that this probability is the same for all tags $p_1 = p_2 = \ldots = p_N = p$. It is further assumed that the error is *static* and occurs throughout a reader session. The best way to think about it is that, in a given session, the tag has a position that makes it unreadable. Furthermore, from one reader session to another, the position of a given tag is randomized (e.g. due to the change of the handheld position), such that the event of not being readable becomes independent of the previous session and occurs again with probability p. Note that both p and N are not known a priori and need to be estimated. In our approach initially we rely on idealized conditions to make simple relations between parameters. Later we test the algorithms against more realistic cases and show that the estimators are robust when the idealized assumptions are violated. For example, this is the case when the probability that a tag is readable is not independent from the reading outcome in the previous session.

The tag set is first read by r_1 and then read by r_2. The readers are cooperative, in a sense that they share information about the outcome of the reading sessions and thus cooperate towards inferring information about the set of tags. After two reader sessions, we have the following situation: k_1 is the number of tags found in common in both reader sessions, k_{2a} is the number of tags only found in reader session r_1, and k_{2b} is found only in the reader session r_2. This amounts in total to $k_2 = k_{2a} + k_{2b}$, each of them read only once. The number of missed tags is denoted by k_3. The main idea is that, by using the observable values k_1, k_{2a}, k_{2b}, one can estimate k_3, N, and p. The tag readings are independent and the probability that a tag is read by one of the readers is $p(1-p)$, and the probability that a tag is not read in any reader session is $(1-p)^2$. Using these, we can obtain the expected values of k_1 and k_2. Estimates $\hat{p}$ and $\hat{N}$ are obtained by setting k_1 and k_2 equal to their respective expected values,

leading to:

$$k_1 = \hat{N}(1-\hat{p})^2, \qquad k_2 = 2\hat{N}(1-\hat{p})\hat{p}. \tag{1}$$

By solving these two equations, treating the estimates $\hat{N}$ and $\hat{p}$ as unknowns, one can obtain the estimators:

$$\hat{p} = \frac{k_2}{2k_1 + k_2}, \qquad \hat{N} = \frac{k_1 + k_2}{1 - \hat{p}^2}. \tag{2}$$

Using these estimators, one can estimate the probability that there is *at least one missing tag* after the two sessions to be $\hat{P}_M = (1-\hat{p}^2)^{\hat{N}}$. In general, after R reader sessions, the estimates $\hat{p}(R)$ and $\hat{N}(R)$ depend on R and the probability to have at least one missing tag is:

$$\hat{P}_M = (1-\hat{p}^R)^{\hat{N}}, \tag{3}$$

If $\hat{P}_M$ is above the acceptable threshold, another reading session is initiated.

Conceptually, these mechanisms are run at a *reliability layer*, which sits on top of a *arbitration layer* that runs the MAC protocol. The main role of the reliability layer is to run the following sequential decision process: after R reader sessions, use Eq. (3) to estimate the probability that there are missing tags and, if this probability is higher than a predefined threshold, then another session is initiated. Similarly, one can estimate the number of reader sessions R needed for reliable reading given statistics of the setup (second application example).

III. Estimators used at the Reliability Layer

When more than two reader sessions R are performed, the estimates $\hat{p}$ and $\hat{N}$ are not as easily determined as in the case with $R = 2$. This section presents estimators used by the reliability layer, such that the reliability of the estimates increases with the number of reader sessions R. These estimators for p and N are used to calculate $\hat{P}_M$ in Eq. (3). The assumptions under which the estimators are derived are rather idealized: (a) Reader sessions are independent (no correlation between reader sessions); (b) Tags are read independently of each other in one reader session (no correlation between tags); (c) Errors are static and occur due to a tag being in a blind spot; (d) Each tag has an identical error probability p.

Obtaining an estimator that takes into account the outcomes of all $R > 2$ reader sessions is not a trivial task. The estimation procedures are exemplified by a simple setup with $N = 10$ tags and a static error probability $p = 0.1$. The outcomes of R sessions over a set of N tags can be represented in a binary $R \times N$ matrix $\mathbf{S}$. For example, if $R = 4$ the matrix is:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

where $S_{rn} = 1$ if tag n is read in reader session r and $S_{rn} = 0$ otherwise. Note that the number of columns, N, is not a priori known, and that the number of rows R is sequentially increased. The estimation methods require at least two reader sessions, and the reliability layer determines the estimates of P_M using Eq. (3). After a given session

and estimation of P_M, the reliability layer determines if a new session should be started. Clearly, the reliability layer uses the estimated value $\hat{P}_M$ to determine the criterion for stopping the reader sessions. It is desirable that $\hat{P}_M \geq P_M$, where P_M is the true probability of missing a tag found with the not a–priori known values of N and p, since in that case it can be guaranteed that, when the reader sessions are stopped, the probability of missing a tag will be less or equal to the target value.

In principle, it would have been the best to directly estimate the probability of having missed tags P_M. However, such an estimate is difficult to obtain, as this quantity is not straightforwardly related to the number of observed tags. Therefore, we rely on techniques that find estimates of p and N and then determine $\hat{P}_M$. In [11] we have presented two main schemes to obtain $\hat{p}$ and $\hat{N}$. In the *first scheme* p is estimated using the observations and, using $\hat{p}$, obtains $\hat{N}$. Several heuristics for estimating p have been presented in [11]. In the *second scheme*, first the value of $\hat{N}$ is obtained from the data, and $\hat{p}$ is estimated by using the value of $\hat{N}$. For deriving estimator of $\hat{N}$, we rely on the classical result for capture–recapture estimation by Schnabel [14]. The capture–recapture methods are statistical methods that can be used to estimate the size of a population, for example the number of fish in a lake. The simplest estimator is Lincoln–Petersen, which uses only two visits to the lake: in the first visit n_1 fish are captured, marked and released back in the lake; in the second visit, n_2 fish are captured and it is noted that m_2 among them are marked. Then the size of the fish population is estimated as $\frac{n_1 n_2}{m_2}$. In [14], the estimation method is generalized to the case when there are more than two visits to the lake.

To make the analogy, in our case a visit to the lake is a reading session. The tags that are read in the first session correspond to the fish that are marked during the first visit. The tags read in the subsequent sessions correspond to the fish that are re–captured at the corresponding visit. Having established this analogy, we can directly use the Schnabel method in order to estimate $\hat{N}$. For that purpose, let n_i denote the number of tags read in the $i-$th reader session and let m_i denote the number of tags that are re–found in the $i-$th session. Finally, let M_i be the total number of tags found before the $i-$th session. For the example matrix $\mathbf{S}$, $m = [0, 7, 9, 9]$, $M = [0, 8, 10, 10]$. Then using the method from [14], the estimate of N can be found as:

$$\hat{N}_S = \frac{\sum_{i=1}^{R} n_i M_i}{\sum_{i=1}^{R} m_i}.$$

By using $\hat{N}_S$, we can estimate the static error probability p. Note that the probability of error can be estimated for each individual reader session as $1 - \frac{n_i}{\hat{N}_S}$, such that we can obtain an estimate of p as:

$$\hat{p}_S = \frac{1}{R} \sum_{i=1}^{R} \left(1 - \frac{n_i}{\hat{N}_S}\right) \tag{4}$$

We have investigated another type of estimator, which consists of multiple iterative steps. We first estimate $\hat{N}_S$ and $\hat{p}_S$, as explained above. Then $\hat{p}_S$ is used to obtain another estimate of N as follows. If k is the number of distinct tags' ID found in R reader sessions, then the expected value of k is $N(1 - p^R)$. Having k and p_S, one can estimate:

$$\hat{N}_{multi} = \frac{k}{1 - \hat{p}_S^R}.$$

Finally, a new estimate $\hat{p}_{multi}$ for p is obtained by replacing $\hat{N}_S$ in Eq. (4) with $\hat{N}_{multi}$. A schematic representation of the whole procedure is $\mathbf{S} \rightarrow \hat{N}_S \rightarrow \hat{p}_S \rightarrow \hat{N}_{multi} \rightarrow \hat{p}_{multi}$. This estimation algorithm has exhibited robust performance when the initial assumptions are challenged, as the next sections will show.

In Table I we have provided the value of the Schnabel estimator and multi–step estimator for the example with four reader sessions. It can be seen that after $R = 4$ reader sessions, the probability of a missed tag is below 0.3%. Depending on the application, this may, or may not, be sufficient, and additional reader sessions can be carried out. It turns out that the multi–step estimator produces the most reliable estimates. Therefore, in the next section only this estimator is used to carry out the numerical evaluation where experiments are set up to test the reliability of this estimator when the assumptions specified in the construction of the estimator are violated.

IV. Numerical Evaluation

In this section we first show the performance of the estimators in an idealized scenario, in which all assumptions that are used as a basis to derive the estimators are satisfied. After that, we challenge the estimators by evaluating them in scenarios in which some of the assumptions are violated. The results show that the estimation methods are robust and perform well even when the assumptions are not completely satisfied. The most important estimate is that of $\hat{P}_M$, as it shows how many reader sessions are needed to be certain, with predefined probability, that all tags are resolved.

A. Idealized Scenario

In the simple experiment in Fig. 2a with $N = 50$ tags, mean values from 1,000 experiments are shown for ideal conditions, that is uncorrelated reader sessions and static error probability p equal for all tags, where $p = 0.1$ and $p = 0.2$. The curve "true P_M" is obtained by calculating $P_M = (1 - p^R)^N$ with the actual values of p and N, as if they were known a priori. The estimates of P_M are closely following the true value of the probability of having at least one tag missing. For example, let the target reliability for the probability of missing at least one tag be 10^{-3}. Then, if $p = 0.1$, the sequential decision process determines to stop after $R = 5$ reader sessions, and for $p = 0.2$ it is $R = 10$.

B. Estimation with Unequal Error Probability across the Tags

Here the estimators are challenged by carrying out simulations in which the static error probability is not equal for all tags. This models the situation in which the individual probability of tag reading error depends on the position of the tag - e.g. closer/farther from the reader. Thus, we have generated the probability of missing each tag as a Gaussian random variable $p_i \sim \mathcal{N}(p, \sigma^2)$, with averages $p = 0.1$ or $p = 0.2$ and $\sigma = 0.01$. More precisely, p_i is generated by using the normal distribution truncated such that $p_i \in (0, 1)$. As can be seen, also in Fig. 2a, the estimates with randomly generated tag probabilities follow closely those where the errors are not following a random distribution. The figure shows two important things: 1) The standard deviation of the estimates follows the mean values in a way where, by adding a small margin to the number of reader sessions needed, one can

produce estimates with a desired reliability. 2) The estimators are robust in the case of static errors in a sense that regardless of whether all tags have the same probability of error, or the error probabilities are generated by a random distribution, the estimates will follow the mean of that distribution. When p is not constant across the tags, the estimators are approximating the average value of p, which in turn leads to good estimates of P_M, although $\hat{P}_M$ is a function of p_i's and not directly of the average value p.

C. Estimation with Correlated Reader Sessions

One can object the assumption that the reader sessions are independent. For example, a tag that is "stuck" in a bad position during a reader session in the application example with the handheld reader, may stay in a bad position with respect to the other positions of the handheld. In [11] we have developed a model for correlation in order to capture such a dependence on the previous position of the tag. When building the correlation model from [11], we have used the following two principles. First, if during the rth session tag A was read and tag B was not read, then for the session $r+1$ the probability that B is not read is higher than the probability that A is not read. The second principle used is related to the average probability of error: Constrain that the average probability of error for all the tags stays identical. This is also intuitive, as random re–positioning of tags cannot increase or decrease the average probability of tag reading error.

Here we present the results with static errors that are dependent across different reader sessions. We use the correlation model from [11], where a level of correlation, ρ, can be specified in the interval $0 \leq \rho \leq 1$, where $\rho = 0$ corresponds to independent reader sessions and $\rho = 1$ is the case of fully correlated reader sessions. In the evaluation the correlation values $\rho = 0.1$ and $\rho = 0.3$ are used.

The estimate $\hat{P}_M$ when reader sessions are correlated is shown in Fig. 2b, where it can be seen that the correlation affects the estimates. The effect of the correlation is seen in that after R reader sessions the true probability P_M may be higher than the estimated probability $\hat{P}_M$ (e.g. for $p = 0.2$, $\rho = 0.3$ and $R < 6$). To tackle this problem one needs to introduce a margin and thus apply more reader sessions than indicated by the value of $\hat{P}_M$.

D. Estimation with Arbitration Protocols and Dynamic Errors

In practice, the errors are often dynamic (e.g. noise–induced), and each transmission is independently subject to error. Errors can occur both on (a) the *reader–to–tag link*, such that the tag is not initiated to send a response even if it is supposed to, and (b) the *tag–to–reader link*, where the reader either does not detect the tag response (due to low received power) or it receives it with errors.

One essential difference between static and dynamic errors is their relation to the probability of missing tags. If only static errors are present, then the event that the tag experiences static error is equivalent with the event that the tag is missed when the reader session terminates. On the other hand, if the model is dynamic errors, i.e. the probability of error in a single query during arbitration from a reader to tag i back to the reader is q_i, then the probability that a particular tag is missed after the reader session terminates is $p_i(q_i, N)$, an increasing function of q_i and N and, in general, $p_i(q_i, N) \neq q_i$.

In tree–based arbitration protocols [7], the reader identifies a group of tags that should transmit in a given slot based on the outcomes of the previous slots. In determining the group of transmitting tags, the reader probes the population of tags by traversing a binary tree. Fig. 3 depicts an example of a basic variant of the tree protocol, that does not contain additional optimizations. We assume that each tag has an array of bits where each bit is randomly and independently equal to 0 or 1 with probability $\frac{1}{2}$. Initially, in slot s_1, all 8 tags are probed by the reader and they transmit, resulting in collision. In s_2 only the tags with bit–array '0*' (i.e. the bit–array prefix is '0') are probed and enabled to transmit, in s_3 only the tags with bit–array '00*', in slot 4 only the tags with bit–array '01*', etc.

In Fig. 3b we present results of the case, when the multi–step estimator in the reliability module operate by using the basic tree protocol for arbitration, while errors on the communication link are modelled with dynamic errors. The probability of dynamic error is taken from the distribution $q_i \sim \mathcal{N}(\mu_{p=0.1}, \sigma^2)$ and $q_i \sim \mathcal{N}(\mu_{p=0.2}, \sigma^2)$. Here $\mu_{p=0.1} = 0.044$ and $\mu_{p=0.2} = 0.095$ are found using the conversion function $p_i(q_i, N)$ for the binary tree protocol from [15] translating static errors to dynamic errors. As can be seen, this model of dynamic errors does not severely decrease the performance of the estimate. This was expected because of the relation between static and dynamic errors. The estimator shows again robustness, as it implicitly approximates the probability of a missed tag P_M.

V. Conclusion

In this paper we have introduced statistical methods to deal with the problem of missing RFID tags. For that purpose, we have introduced a reliability layer, which ensures that when the reading process is considered over, the probability to still have missing tags, i.e. tags whose ID has not been read at all, is below an acceptable threshold. Conceptually, the reliability layer operates on top of the arbitration (MAC) layer, whose task is to resolve collisions among tags and gather the ID of the tags. In absence of tag reading errors, a single reader session (single execution of the arbitration protocol) is sufficient to gather the ID of all tags. When there are tag reading errors, the reliability layer runs multiple reader sessions and initiates a new session as long as its estimate of the probability of having missed tags is above an acceptable threshold.

We propose several estimators that can be used by the reliability layer to calculate the probability of having a missed tag. These estimators are derived under idealized assumptions. However, they are evaluated with numerical experiments where these idealized assumptions are not met and the estimators exhibit high robustness in the sense that the estimated probability of missing tags is closely approximating the actual probability of missing tags.

There are several interesting directions for future work. First, it is interesting to test the derived estimators in an experimental setup and assess their robustness under error models that are stemming from the actual physical transmission. Another interesting issue is to modify the estimators in order to work with smaller tag sets, as the statistical methods presented here are suitable for relatively large tag sets. In that case it will be crucial to derive new estimators which take into account the prior knowledge on the range of the tag set size.

REFERENCES

[1] R. Angeles, "RFID Technologies: Supply-Chain Applications and Implementation Issues," *Information Systems Management*, vol. 22, no. 1, pp. 51–65, 2005.

[2] A. Juels, D. Molnar, and D. Wagner, "Security and Privacy Issues in E-passports," *1st International Conference on Security and Privacy for Emerging Areas in Communications Networks, SecureComm 2005.*, pp. 74–88, September 2005.

[3] D. Molnar and D. Wagner, "Privacy and security in library RFID: issues, practices, and architectures," in *CCS '04: Proceedings of the 11th ACM conference on Computer and communications security*. Washington DC, USA: ACM, 2004, pp. 210–219.

[4] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. John Wiley & Sons, West Sussex, England, 2003.

[5] *EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz - 960 MHz*, 1st ed., EPCglobal Inc., October 2008.

[6] H. Vogt, "Efficient Object Identification with Passive RFID Tags," in *Pervasive Computing*. Springer-Verlag, Berlin / Heidelberg, Germany, 2002, pp. 98–113.

[7] J. L. Massey, *Collision-Resolution Algorithms and Random-Access Communications*, ser. International Centre for Mechanical Sciences Courses and Lectures. Wien, Austria: Springer-Verlag, 1981, no. 265, pp. 73–137.

[8] P. Popovski, F. H. P. Fitzek, and R. Prasad, "Batch Conflict Resolution Algorithm with Progressively Accurate Multiplicity Estimation," in *DIALM-POMC '04: Proceedings of the 2004 joint workshop on Foundations of mobile computing*. Philadelphia, PA, USA: ACM, 2004, pp. 31–40.

[9] L. W. F. Chaves, E. Buchmann, and K. Böhm, "Tagmark: Reliable Estimations of RFID Tags for Business Processes," in *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2008, pp. 999–1007.

[10] *RFID for logistic applications - Tests results*, EPCglobal France Lab, 2005.

[11] R. M. Jacobsen, K. F. Nielsen, P. Popovski, and T. Larsen, "Reliable Identification of RFID Tags using Multiple Independent Reader Sessions," in *IEEE international conference on RFID*. Orlando, FL, USA: IEEE, 2009.

[12] M. Backes, T. R. Gross, and G. Karjoth, *Tag Identification System*. U. S. Patent Office, August 2008, patent number US 2008/0204243A1.

[13] J. Bart, M. A. Fligner, and W. I. Notz, *Sampling and Statistical Methods for Behavioral Ecologists*. Cambridge University Press, December 1998.

[14] Z. E. Schnabel, "The Estimation of Total Fish Population of a Lake," *The American Mathematical Monthly*, vol. 45, no. 6, pp. 348–352, 1938.

[15] K. F. Nielsen, R. M. Jacobsen, P. Popovski, and T. Larsen, "Fast Capture–Recapture Approach for Mitigating the Problem of Missing RFID Tags," 2009, submitted to IEEE Transactions on Mobile Computing, available at. [Online]. Available: http://kom.aau.dk/~kfyhn/ieee_tmc_v8.pdf

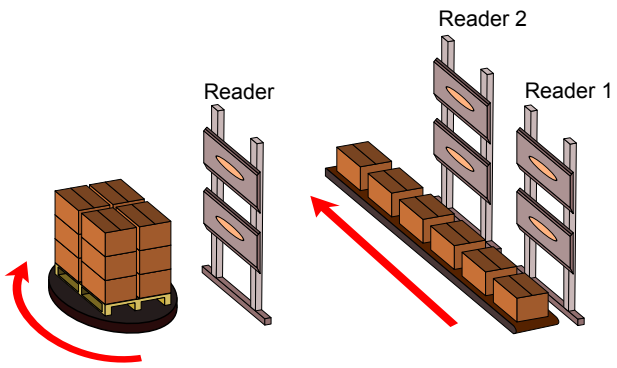
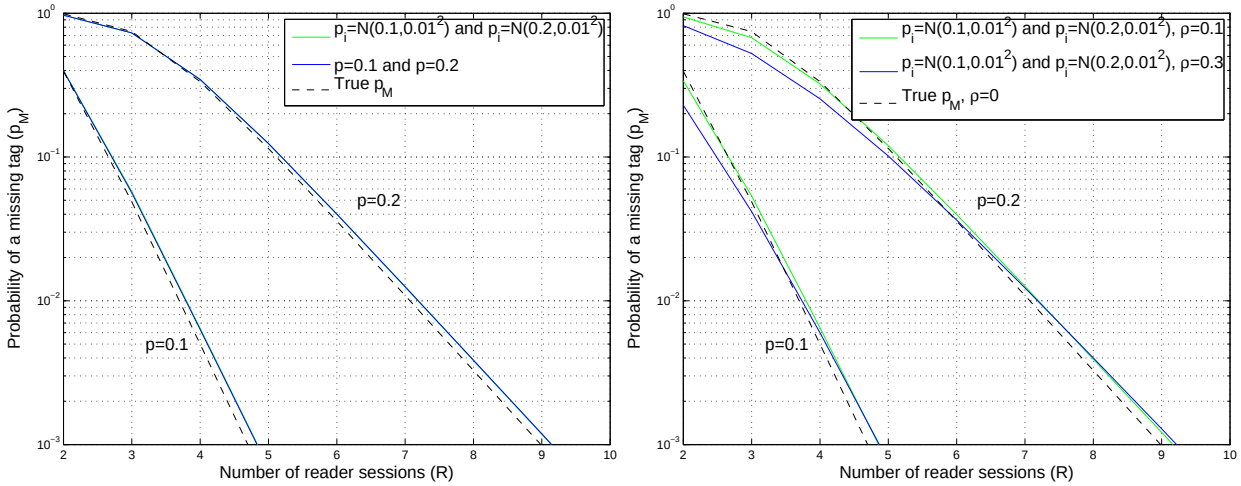
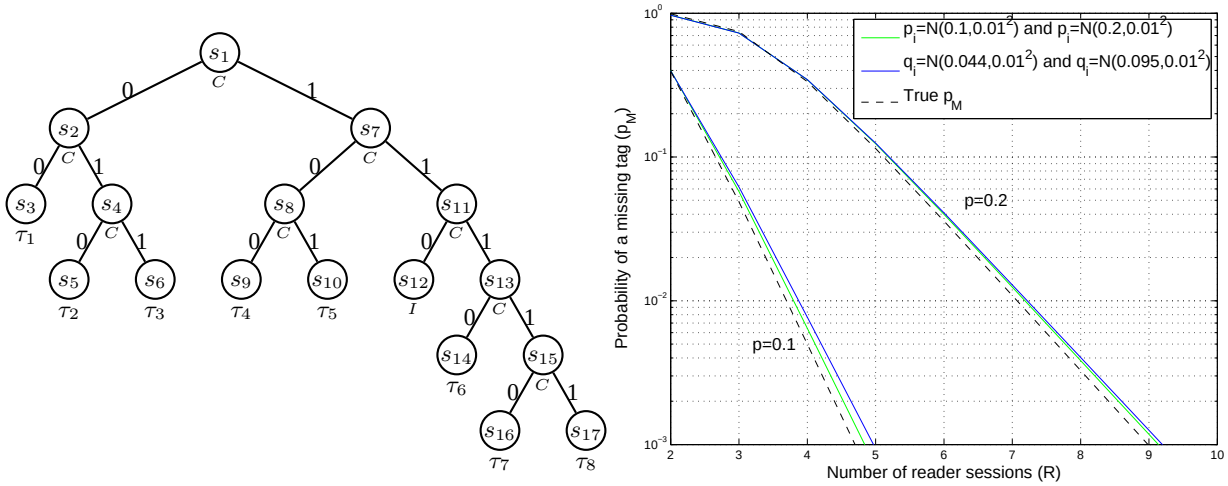


Fig. 1: Situations where multiple reader sessions can used to mitigate the problem of missing tags. The tagged items are in the boxes.



(a) (b)

Fig. 2: Simulated estimates of P_M vs. the number of reader sessions. **(a)** With static errors $p = 0.1$ and $p = 0.2$ (blue lines) and random errors (green lines). In case of random errors, the probability for missing a tag is Gaussian with mean value 0.1 and 0.2, respectively; **(b)** Same parameters, but with correlation between reader sessions, using the correlation model described in Section IV-C.



(a) (b)

Fig. 3: Arbitration with dynamic errors. **(a)** An instance of the binary tree algorithm for $N = 8$. The vertices represent a slot, which state can be Idle (I), Single (S) or Collision (C). For channel state "S", τ_i denotes the resolved tag. **(b)** Comparison of simulated estimates of P_M with the true value of P_M vs. the number of reader sessions for dynamic errors and static errors. Both error probabilities are taken from similar random distributions. As a reference, the true value of P_M is also plotted.

	Schnabel estimator			Multi–step estimator		
R	$\hat{p}$	$\hat{N}$	$\hat{P}_M$	$\hat{p}$	$\hat{N}$	$\hat{P}_M$
2	0.17	10.3	0.270	0.18	10.3	0.276
3	0.14	10.1	0.030	0.14	10.0	0.025
4	0.13	10.1	0.003	0.13	10.0	0.003

TABLE I: Example estimates provided by the two estimation methods.