Differential Estimation in Dynamic RFID Systems

Qingjun Xiao, Bin Xiao
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
Email: {csqjxiao, csbxiao}@comp.polyu.edu.hk

Shigang Chen
Department of Computer & Info of Science & Engineering
University of Florida
Gainesville, Florida 32611
Email: sgchen@cise.ufl.edu

***Abstract*—Efficient estimation of tag population in RFID systems has many important applications. In this paper, we present a new problem called *differential cardinality estimation*, which tracks the population changes in a dynamic RFID system where tags are frequently moved in and out. In particular, we want to provide quick estimation on (1) the number of new tags that are moved in and (2) the number of old tags that are moved out, between any two consecutive scans of the system. We show that the traditional cardinality estimators cannot be applied here, and the tag identification protocols are too expensive if the estimation needs to be performed frequently in order to support real-time monitoring. This paper presents the first efficient solution for the problem of differential cardinality estimation. The solution is based on a novel differential estimation framework, and is named *zero differential estimator*. We show that this estimator can be configured to meet any pre-set accuracy requirement, with a probabilistic error bound that can be made arbitrarily small.**

I. INTRODUCTION

In the past decades, RFID technology has been applied to a wide range of applications, including warehouse management, logistic control, and smart shelf [1], [2]. In RFID systems, a canonical research problem is *cardinality estimation* which figures out the number of RFID tags within the radio range of an RFID reader [3], [4]. Such knowledge can be used for inventory monitoring, and also for accelerating the speed of tag IDs collection by setting the optimal ALOHA frame size.

This paper considers a new problem named *differential cardinality estimation*. This problem differs from traditional cardinality estimation by focusing on dynamic population. It estimates (1) the number of missing tags that have left the RF range of a reader and (2) the number of new tags that have entered this range. Specifically speaking, in our scheme, an RFID reader scans its surrounding periodically. It would find some tags that were present in the previous scan and disappear in the current scan, and find some other tags that were previously unknown and appear in the current scan. We call the former kind of tags *missing tags*, and the latter *new tags*. Our purpose is to quickly estimate their numbers.

Keeping track of dynamic tag populations can benefit many industries for efficient and accurate object management. Just imagine in a large industrial warehouse, the products are attached with RFID tags and are moved in and out frequently. If the "number of new tags" is made available to the manager in real time, he knows the quantity of inbound goods during any time interval. Moreover, if this quantity is not as expected (e.g., inconsistent with the reported number at the checkin counter), then there must exist some management faults, such as stock misplacement or vendor fraud. Meanwhile, if the "number of missing tags" is made available in real time, it can help reveal stolen stocks, which is another critical application.

However, the differential cardinality estimation problem cannot be solved by the traditional estimation algorithms [3], [5]. For example, a traditional algorithm may tell you that there are 5,000 tags in the system at time 1 and there are 7,500 tags at time 2. But it cannot tell how many new tags are moved into the system and how many existing ones are moved out between time 1 and time 2. It could be that there is no missing tag while 2,500 new tags have been moved in, it could be that there are 5,000 missing tags while 7,500 new tags have been moved in, or it could be any of the numerous possible combinations of missing and new tags that yield the same result of 7,500 tags observed at time 2.

A straightforward solution for this problem is to collect all tag IDs in both the previous and current scans through a tag identification protocol [6], [7], [8], [9]. The missing or new tags can be easily identified by comparing two tag populations. There are other protocols [10], [11] that are designed to identify the IDs of missing/new tags more efficiently. However, collecting individual IDs from tags is a far more expensive operation than estimating the numbers of missing/new tags. This is particularly true when thousands or more tags are involved and if we want to track the numbers of missing/new tags in real time, which means we need to perform differential cardinality estimation frequently. In many applications, users may only need to know the numbers of missing/new tags, and do not need their exact IDs. Therefore, ID collection can be both unnecessary and time-wasting.

In this paper, we propose the first efficient solution that is specially designed for the problem of differential cardinality estimation. Our solution requires each tag to send just 1 bit to indicate its presence. In comparison, the best tag identification protocol requires each tag to transmit its 96-bit ID for 2.72 times on average in order to resolve collisions. In each scan of our solution, tags randomly pick slots in the time frame to transmit their 1-bit responses. The RFID reader monitors the state of every slot in the frame. There are two types of states: empty/non-empty, corresponding to the cases where zero/non-zero tags transmit in the slot. To solve the problem of differential cardinality estimation, we jointly consider two consecutive scans and pair up the corresponding slots in the

978-1-4673-5946-7/13/$31.00

two time frames. Each pair of slots have four possible state combinations: {empty, non-empty} / {empty, non-empty}. We count the number of pairs that fall in each of the four state combinations. Based on these numbers, we construct the *Zero Differential Estimator* (ZDE) to estimate the numbers of missing/new tags. We show that this estimator can be configured to meet any pre-set accuracy requirement, with an error bound that can be made arbitrarily small.

II. Related Work

In RFID research community, there are two canonical problems: tag identification problem and cardinality estimation problem. The *identification* problem is to allow RFID reader to collect tag IDs within its radio range [6], [7], [8], [9]. The *cardinality estimation* problem is to estimate the number of tags in an RFID system [3], [4]. This problem has its practical use, because the tag identification protocols based on framed ALOHA can achieve best performance when the frame size (f) is equal to the number of tags (n) in the system. The RFID-related research also stretched in several other directions, such as faster cardinality estimation [5], missing tag identification [12], [13], [14], and continuous monitoring [10], [11]. However, there is no previous work that gives an efficient solution as how to address the differential cardinality estimation problem.

III. PROBLEM DEFINITION

In this section, we formulate the problem of *differential cardinality estimation*. It assumes the tag population covered by an RFID reader to be a dynamic process, and estimates the population differences at two adjacent time points.

A. System Model

The group of RFID tags in the radio range of an RFID reader can change over time, as the tags enter or leave the range. Such a dynamic group is called a *dynamic tag population* which can be modeled as a time series: T_0, T_1, $\ldots$, T_t, T_{t+1}. Here, $0, \ldots, t$ is the discrete time, and symbol T_t denotes the tag population that can be inventoried by RFID reader at time t. This time series is illustrated in Fig. 1 where each ellipse represents a tag population at a particular time point.

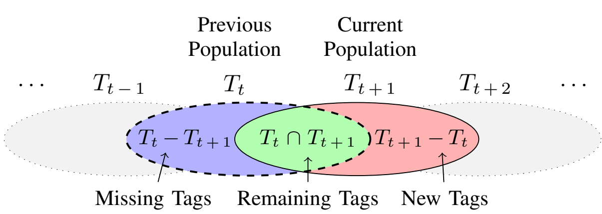


Fig. 1. Dynamic process of a tag population monitored by an RFID reader.

We focus on the tag populations at two adjacent time points, i.e., the previous population T_t and the current population T_{t+1}. In Fig. 1, T_t is depicted as a dashed ellipse, and T_{t+1} is shown as a solid ellipse. We are interested in the intersections and discrepancies of the two populations to show differential cardinality. Thus, totally three kinds of tags can be found:

- *missing tags* which are denoted by $T_t - T_{t+1}$ and depicted as a blue region in Fig. 1 (i.e., found in previous population but disappear in current population),
- *remaining tags* which are denoted by $T_t \cap T_{t+1}$ and are depicted as a green region in Fig. 1 (i.e., exist in both previous and current populations),
- *new tags* which are denoted by $T_{t+1} - T_t$ and are depicted as a red region in Fig. 1 (i.e., previously unknown but appear in the current population).

B. Problem Definition

Tag Number Estimation. Our research problem is to design an algorithm that can estimate the numbers of the three kinds of tags (i.e., missing/remaining/new tags). We use the following symbols to denote their numbers.

- n_- is the number of missing tags: $n_- = |T_t - T_{t+1}|$.
- n_0 is the number of remaining tags: $n_0 = |T_t \cap T_{t+1}|$.
- n_+ is the number of new tags: $n_+ = |T_{t+1} - T_t|$.
- n_u is the total number of tags: $n_u = |T_t \cup T_{t+1}|$.

Here, n_u means the total number of tags in current population and previous population. Thus, we have $n_u = n_- + n_0 + n_+$.

Keeping track of number of missing tags n_- and the number of new tags n_+ can benefit many real-world industries. For example, in a warehouse, if the manager knows the number of "missing tags" in real time, he can arrange a replenishment of inventory immediately when such number exceeds a threshold. If the manager knows the number of "new tags", he is able to quickly identify the occurrence of stock misplacement.

We formalize our problem of *differential cardinality estimation* in Definition 1. Here, we give out the metrics to quantify cardinality estimation errors, and identify the need to bound such errors to meet the pre-set accuracy requirement.

Definition 1 (Differential Estimation Problem): The purpose is to estimate the numbers of missing/remaining/new tags. We denote the ground truth by n_-, n_0, n_+, and denote the estimated values by $\hat{n}_-$, $\hat{n}_0$, $\hat{n}_+$ with carets over the letters. The estimation errors are calculated as $\frac{\hat{n}_- - n_-}{n_-}$, $\frac{\hat{n}_0 - n_0}{n_0}$, $\frac{\hat{n}_+ - n_+}{n_+}$, because we normalize the estimation errors by dividing them by ground truth.

To meet the pre-set accuracy requirement α (e.g., 95% or 99%, etc.), the normalized estimation errors should be contained within a confidence interval $(\alpha - 1,\ 1 - \alpha)$, i.e.,

$$\frac{\hat{n}_* - n_*}{n_*} \in (\alpha - 1,\ 1 - \alpha), \quad \text{where } n_* \text{ means } n_-,\ n_0 \text{ or } n_+.$$

The probability for this range to contain the estimation error is the confidence level β (e.g., β can be 68%, 95%, 99.7%, etc).

Tag Ratio. The accuracy of counting missing/remaining/new tags can be affected by their ratios in union population $T_t \cup T_{t+1}$. Thus, we define their ratios as follows.

- γ_- is the ratio of missing tags, with $\gamma_- = n_- / n_u$.
- γ_0 is the ratio of remaining tags, with $\gamma_0 = n_0 / n_u$.
- γ_+ is the ratio of new tags, with $\gamma_+ = n_+ / n_u$.

The sum of the three ratios is equal to one. Thus, using only two ratios γ_- and γ_+, we can fully specify the relation between previous tag population T_t and current tag population T_{t+1}. For example, when γ_- and γ_+ are equal to 0.15 and 0.2 respectively, there are 15% missing tags, 65% remaining tags and 20% new tags in the union population $T_t \cup T_{t+1}$.

IV. DIFFERENTIAL ESTIMATION FRAMEWORK

In this section, we describe a differential estimation framework which collects the input raw data to feed into the differential cardinality estimation algorithm.

Estimation Framework. At each time t, the RFID reader uses an ALOHA frame with f slots to interacts with its surrounding tags. To start the frame, the reader broadcasts a query message containing frame size f. Then, each tag that hears this query will choose a time slot to respond to the reader. Such choice is made using a uniform hash function $h_f(id, r)$, where id is the tag identifier, f is frame size, and r is a random seed used for the whole frame. So the initial query message needs to contain both parameters f and r.

We focus on two ALOHA frames at time t and $t+1$, which are depicted in Fig. 2. The frame at time $t+1$ is the "current frame" where the tags in set T_{t+1} send their responses to the reader, e.g., $T_{t+1} = \{3, 4, 5, \dots, 10\}$ in Fig. 2. The frame at time t is the "previous frame" which contains the responses of previous tag population T_t, e.g., $T_t = \{0, 1, 2, \dots, 7\}$ in Fig. 2.

We assume that, if a tag replies in both frames, it will choose the same slot index. For example, in Fig. 2, tag 5 chooses the same slot 6 in both frames. This assumption is realized by letting the two frames use the same frame size f and random seed r. Thus, in Fig. 2, you can see both frames configure the frame size f to 10 (with slot indices range from 0 to 9).

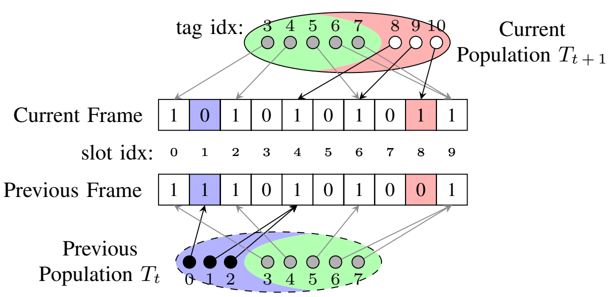


Fig. 2. Slot states in previous frame and current frame.

In a frame, the RFID reader can detect the empty/non-empty state of each time slot, which means zero/non-zero tag responses in that slot. We label the two kinds of slot states by digits 0 and 1 respectively in Fig. 2. Note that, to differentiate the two states, we need each tag to respond with only 1 bit for presence indication, rather than 96 bits for tag ID collection. This can save over 90% energy (for active tags) and time.

For each frame, the RFID reader can scan only the preceding f' ($\leq f$) time slots to collect the needed raw data. When the f' value is smaller than frame size f, the reader can prematurely yield the current frame without transmitting the subsequent slots (see UHF EPC C1G2 standard [1] for *NAK* command which forces all the tags into *arbitrate* state). The purpose of killing the current frame is to save time and energy when the estimation accuracy already satisfies user's need.

Estimator Inputs. To acquire raw data for differential estimation, we will observe the preceding f' slots in previous frame and in current frame, which is described in details as follows.

We use symbol x_i to denote the state of slot i ($0 \leq i < f'$) in previous frame, and symbol y_i to indicate the state of slot i in current frame. Since we jointly consider slot states x_i and y_i, we can define four kinds of slot pairs: {empty, non-empty} / {empty, non-empty}. For example, an empty/empty slot pair is empty in both previous and current frames; an empty/non-empty slot pair is empty in previous frame but becomes non-empty currently. The RFID reader, by scanning the preceding f' time slots in both frames, can obtain the slot number $N_{x,y}$ which means the number of slots that are previously in x state and currently in y state ($0 \leq x, y \leq 1$). There are totally four such slot numbers corresponding to the four kinds of slot pairs.

In summary, we use the slot numbers $N_{x,y}$ ($0 \leq x, y \leq 1$) as algorithm inputs to conduct differential estimation. For the ease of presentation, we define four other slot number variables: $N_{*,y} = \sum_{x=0}^{1} N_{x,y}$ and $N_{x,*} = \sum_{y=0}^{1} N_{x,y}$. Therefore, $N_{*,0}$ is the number of empty slots in current frame, and $N_{0,*}$ is the number of empty slots in previous frame.

V. ZERO DIFFERENTIAL ESTIMATOR (ZDE)

This section presents the zero differential estimation (ZDE) algorithm which utilizes the slot numbers $N_{0,*}$, $N_{*,0}$ and $N_{0,0}$, i.e., the numbers of slot pairs with at least one empty slot.

Motivation. The previous work tells us that, from the number of empty slots in an ALOHA frame, we can estimate the number of tags that responded in this frame [3]. However, it does not help to give differential estimates (i.e., the values of n_-, n_0 and n_+), because from the number of empty slots in current frame (or previous frame), we can only estimate the tag number $n_0 + n_+$ (or $n_- + n_0$). We are unable to extract n_-, n_+, n_0 from these two tag numbers. Therefore, we propose to use empty/empty slot number $N_{0,0}$ to derive the total tag number n_u from which we can infer the differential estimates.

Estimator Details. We describe zero differential estimator (ZDE) as follows. We firstly analyze the expected values of slot numbers $N_{*,0}$, $N_{0,*}$ and $N_{0,0}$ which are under observation.

Theorem 1 (Expected Values of Empty Slot Numbers): In the current frame, the expected number of empty slots is

$$E(N_{*,0}) = f' (1-p)^{n_u - n_-},$$

where f' is the number of slots scanned by RFID reader, p is the reciprocal of frame size f, and $n_u - n_-$ is the number of tags in current frame. Similarly, in the previous frame, the expected number of empty slots is $E(N_{0,*})$ shown as follows.

$$E(N_{0,*}) = f' (1-p)^{n_u - n_+}$$

For the number of slots $N_{0,0}$ that are empty in both frames, its expected value is $E(N_{0,0})$.

$$E(N_{0,0}) = f' (1-p)^{n_u}$$

For the equations in Theorem 1, we can substitute the expected values by the observations ($N_{*,0}$, $N_{0,*}$, $N_{0,0}$), in order to get estimates for the unknown variables (n_-, n_+, n_u). This is because the observations ($\frac{N_{*,0}}{f'}$, $\frac{N_{0,*}}{f'}$, $\frac{N_{0,0}}{f'}$) follows Gaussian distributions. Their variances are proportional to $\frac{1}{f'}$ and thus are close to zero when f' is large (see Appendix A).

By solving the equations in Theorem 1, we can get the estimators in Definition 2. For example, by solving the first and third equations, we have $n_u - n_- = \log_{1-p} \frac{N_{*,0}}{f'}$ and $n_u = \log_{1-p} \frac{N_{0,0}}{f'}$. Then, we can estimate the number of missing tags as $\hat{n}_- = \log_{1-p} \frac{N_{0,0}}{f'} - \log_{1-p} \frac{N_{*,0}}{f'} = \log_{1-p} \frac{N_{0,0}}{N_{*,0}}$.

Definition 2 (Zero Differential Estimator): By scanning the leading f' slots in adjacent ALOHA frames and counting the three slot numbers $N_{0,0}$, $N_{0,*}$ and $N_{*,0}$, the RFID reader can estimate the tag numbers n_-, n_0 and n_+ as follows.

$$\hat{n}_- = \log_{1-p} \frac{N_{0,0}}{N_{*,0}}, \quad (1)$$

$$\hat{n}_+ = \log_{1-p} \frac{N_{0,0}}{N_{0,*}}, \quad (2)$$

$$\hat{n}_0 = \log_{1-p} \frac{N_{0,*}\, N_{*,0}}{f'\, N_{0,0}}. \quad (3)$$

Accuracy Analysis. We analyze the means and variances of the ZDE estimators presented in Definition 2. Here, we show only the variances of missing tag estimation and new tag estimation. To save space, we do not give variance analysis for the remaining tag variable, which can be derived similarly.

Theorem 2: When frame size f and tag number n_u are large enough, the error of ZDE estimators (i.e., $\hat{n}_- - n_-$ and $\hat{n}_+ - n_+$) approximately follows Guassian distributions of zero mean. Their variances are as follows:

$$Var(\hat{n}_-) \approx g\big(E(N_{*,0}),\, n_-\big)$$
$$Var(\hat{n}_+) \approx g\big(E(N_{0,*}),\, n_+\big)$$

where $g(N, n) = \frac{1}{p\,N}\frac{1}{p}\big[e^{p\,n} - (1 + p\,N \cdot p\,n)\big]$.

If β is the confidence level, the confidence interval for estimating missing tag number is between $n_- \pm Z_{\frac{\beta}{2}} \cdot \sqrt{Var(n_-)}$, where $Z_{\frac{\beta}{2}}$ is quantile function of standard Gaussian distribution. Thus, the accuracy α of missing tag estimation is $1 - Z_{\frac{\beta}{2}} \cdot \frac{\sqrt{Var(\hat{n}_-)}}{n_-}$ (for new tags $1 - Z_{\frac{\beta}{2}} \cdot \frac{\sqrt{Var(\hat{n}_+)}}{n_+}$ similarly).

Proof: The proof can be found in Appendix A. ■

From Theorem 2, we know the estimation error of missing tag number is determined by two factors. One is $p\,n_- = \frac{n_-}{f}$, i.e., the density of missing tags in previous frame. The other is $p\,E(N_{*,0}) = \frac{E(N_{*,0})}{f}$, i.e., the expected ratio of empty slots in current frame. These empty slots are important for missing tag estimation, because if they were meanwhile non-empty in previous frame, then the RFID reader can detect the presence of missing tags. For example, in Fig. 2, slot 1 is empty in current frame and non-empty in previous frame. The RFID reader can tell some tags must be missing, i.e., tag with ID 0.

The similar conclusion can be drawn for new tag estimation, whose accuracy is determined by two factors: (1) $p\,n_+$ or the density of new tags in current frame, and (2) $p\,E(N_{0,*})$ or the expected ratio of empty slots in previous frame. For these empty slots, an example of new tag detection can be found in Fig. 2, i.e., slot 8 which was empty in previous frame but become non-empty in current frame. This is a strong signal for the presence of new tags.

The variance $Var(\hat{n}_-)$ in Theorem 2 can be further interpreted as a function of only one parameter - union load factor $p\,n_u$. This is because its two deciding factors can be converted as function of $p\,n_u$, when the tag ratios γ_- and γ_+ are fixed and the ratio of observed slots pf' is known.

$$p\,n_- = p\,n_u \cdot \gamma_-$$
$$p\,E(N_{*,0}) = pf' \cdot (1-p)^{n_u - n_-} \approx pf' \cdot e^{p\,n_u(1-\gamma_-)}$$

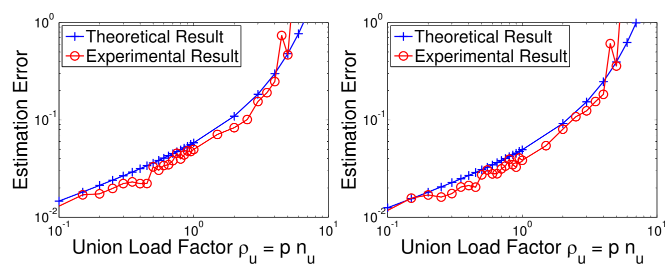
Similarly, the variance $Var(\hat{n}_+)$ can be reduced to a function of union load factor $p\,n_u$, which is plotted in Fig. 3.

VI. SIMULATION

In this section, we use simulations to evaluate the accuracy of the proposed differential cardinality estimator. Our simulation results will also analyze the impacts of two adjustable parameters, i.e., union load factor and number of observed slots.

Simulation Settings. Our experiments firstly simulate the distribution of tags in previous frame and current frame, when given the numbers of missing/remaining/new tags (i.e., $n_- = 450$, $n_0 = 1950$, $n_+ = 600$, $n_u = 3000$). The node distribution in frames can be controlled by one parameter, i.e., union load factor $\rho_u = \frac{n_u}{f}$. Then, the RFID reader, by observing the preceding f' slots, can estimate the tag numbers n_- and n_+ by ZDE algorithm. We will investigate the impacts of two parameters ρ_u and f' on estimation error. Here, f' has been normalized as $\frac{f'}{f} = pf'$. For each experiment, we will execute it one hundred times and calculate the average estimation error.

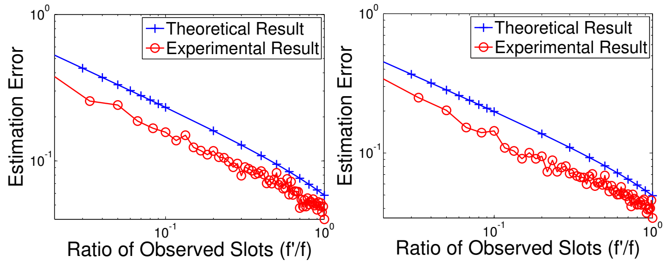
Simulation Results. We plot simulation results in Fig. 3 and Fig. 4. The former varies the union load factor ρ_u to evaluate its impact on estimation error, and meanwhile fixes the number of observed slots f' to be equal to frame size f. In contrast, the latter varies f' to evaluate its impact, and fixed ρ_u to one.



(a) Missing tag estimation error. (b) New tag estimation error.

Fig. 3. Impact of union load factor ρ_u, with observed number of slots $f' = f$.

Firstly, these two figures show that the trends for missing tag estimation and new tag estimation are similar, because they are symmetric process. Secondly, the two figures show that the trend obtained from theoretical derivation is consistent with experimental results. The theoretical estimation error is calculated by $\frac{\sqrt{Var(\hat{n}_-)}}{n_-}$ and $\frac{\sqrt{Var(\hat{n}_+)}}{n_+}$ (see Theorem 2 for variance formulas). Thirdly, the figures show that, to reduce



(a) Missing tag estimation error. (b) New tag estimation error.

Fig. 4. Impact of observed number of slots f', with union load factor $\rho_u = 1$.

estimation error, we should reduce the union load factor ρ_u, or increase the ratio of observed slots f'/f. For example, if we want to achieve over 99% accuracy (i.e., 1% expected error), then Fig. 3 shows that we can configure observed slots f' to frame size f, and keep union load factor ρ_u small than 0.1.

VII. Conclusion

In this paper, we presented the *differential estimation* problem to estimate the numbers of missing/new tags, without the need to collect tag IDs. This is a new problem of how to capture the dynamics in large-scale RFID tag populations. To solve this problem, we have proposed the ZDE estimator that observes empty slots in previous ALOHA frame and current ALOHA frame. We have thoroughly investigated the statistical properties of ZDE estimator, by both theoretical analysis and simulation results. The results show that ZDE estimator performs well in the cases with small load factor.

Acknowledgment

This work was supported in part by the National Science Foundation under grant CNS-1115548.

References

[1] "EPC™ radio-frequency identity protocols class-1 generation-2 UHF RFID protocol for communications at 860MHz-960MHz v1.2.0," 2008.
[2] "ISO/IEC 18000 information technology - RFID for item management - Part 6: Parameters for air interface communications at 860MHz-960MHz," 2004.
[3] M. Kodialam and T. Nandagopal, "Fast and reliable estimation schemes in RFID systems," in *Proc. of ACM MobiCom*, 2006, pp. 322–333.
[4] C. Qian, H. Ngan, and Y. Liu, "Cardinality estimation for large-scale RFID systems," in *Proc. of IEEE PerCom*, 2008.
[5] C. Qian, Y. Liu, H. Ngan, and L. M. Ni, "ASAP: Scalable identification and counting for contactless RFID systems," in *Proc. of IEEE ICDCS*, 2010, pp. 52–61.
[6] J. Myung and W. Lee, "Adaptive splitting protocols for RFID tag collision arbitration," in *Proc. of ACM MOBIHOC*, 2006.
[7] H. Vogt, "Efficient object identification with passive RFID tags," in *Proc. of IEEE PERCOM*, 2002, pp. 98–113.
[8] B. Zhen, M. Kobayashi, and M. Shimizu, "Framed ALOHA for multiple RFID objects identification," *IEICE Trans. on Communications*, 2005.
[9] S.-R. Lee, S.-D. Joo, and C.-W. Lee, "An enhanced dynamic framed slotted ALOHA algorithm for RFID tag identification," in *Proc. of IEEE MOBIQUITOUS*, 2005, pp. 166–174.
[10] B. Sheng, Q. Li, and W. Mao, "Efficient continuous scanning in RFID systems," in *Proc. of IEEE INFOCOM*, 2010.
[11] Q. Xiao, K. Bu, and B. Xiao, "Efficient monitoring of dynamic tag populations in RFID systems," in *Proc. of IEEE/IFIP EUC (Best Paper Award)*, 2011.
[12] T. Li, S. Chen, and Y. Ling, "Identifying the missing tags in a large RFID system," in *Proc. of ACM MOBIHOC*, 2010, pp. 1–10.
[13] C. C. Tan, B. Sheng, and Q. Li, "How to monitor for missing RFID tags," in *Proc. of IEEE ICDCS*, 2008, pp. 295–302.
[14] K. Bu, B. Xiao, Q. Xiao, and S. Chen, "Efficient misplaced-tag pinpointing in large RFID systems," in *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 23, no. 11, 2012, pp. 2094–2106.
[15] W. Feller, *An introduction to probability theory and its applications*. Wiley, 1968.

Appendix A
Proof of ZDE Estimator Variance

For ZDE estimator, we use the following asymptotic results in deriving the mean and the variance of its missing tag estimate $\hat{n}_- = \log_{1-p} \frac{N_{0,0}}{N_{*,0}}$. The statistical properties of $\hat{n}_+$ can be analyzed similarly.

Firstly, we assume that $N_{*,0}$ is fixed, and analyze the conditional probability $N_{0,0} \,|\, N_{*,0}$. The physical meaning is, among the empty slots in current frame, the number of slots that do not contain missing tags. This random variable approximately follows a normal distribution when $N_{*,0}$ and n_- are large enough. See Feller [15] for a proof of the normality.

$$N_{0,0} \,|\, N_{*,0} \quad \sim \quad \mathcal{N}\big(\, \mu_0(N_{*,0},\, n_-), \quad \sigma_0(N_{*,0},\, n_-)\,\big)$$

Secondly, we know the mean value $\mu_0(N_{*,0},\, n_-)$ is equal to $N_{*,0}\,(1-p)^{n_-}$, because of physical meaning of $N_{0,0} \,|\, N_{*,0}$ mentioned before. This expected value tells us the estimate $\hat{n}_- = \log_{1-p} \frac{N_{0,0}}{N_{*,0}}$ is approximately an unbiased estimator. This is because we know that, no matter whatever fixed value $N_{*,0}$ is, $\frac{N_{0,0}}{N_{*,0}} - (1-p)^{n_-}$ follows a normal distribution with zero mean,. Thus, $\log_{1-p}(\frac{N_{0,0}}{N_{*,0}}) - \log_{1-p}(1-p)^{n_-} = \hat{n}_- - n_-$ is approximately with zero mean (see Theorem 6 of [3]).

Thirdly, we derive the variance $\sigma_0(N_{*,0},\, n_-)$ as follows. Let $N = N_{*,0}$ and $n = n_-$, we have

$$\sigma_0^2(N,\, n) \quad = \quad E\left[(\textstyle\sum_{j=1}^{N} X_j)^2\right] \,-\, \left(E\left[\textstyle\sum_{j=1}^{N} X_j\right]\right)^2$$

where X_j is the event that j^{th} empty slot in the current frame contains zero missing tags. Note that for $i \neq j$, $E[X_i X_j] = (1-2p)^n$. Plugging this result in the expression for the variance, we get

$$\sigma_0^2(N,\, n) \quad = \quad N(N-1)(1-2p)^n + N(1-p)^n - N^2(1-p)^{2n}$$

According to Taylor expansion, we have $(1-p)^{2n} = (1-2p+p^2)^n = (1-2p)^n + np^2(1-2p)^{n-1} + O(p^4)$. Thus, we can approximate $\sigma_0^2(N,\, n)$ as follows.

$$\begin{aligned} \sigma_0^2(N,\, n) \quad &\approx \quad N\left[(1-p)^n - (1-p)^{2n} - {\scriptstyle (N-1)np^2(1-2p)^{n-1}}\right] \\ &\approx \quad N\,(1-p)^{2n}\left[(1-p)^{-n} - 1 - p\,N \cdot p\,n\right] \end{aligned}$$

Fourthly, we can derive the variance of $\log_{1-p} N_{0,0} \,|\, N_{*,0}$ as follows, since logarithm is a monotonic continuous function:

$$Var(\log_{1-p} N_{0,0} \,|\, N_{*,0}) \quad = \quad \frac{\sigma_0^2(N,n)}{[\mu_0'(N,n)]^2}$$

where $N = N_{*,0}$ and $n = n_-$. Then, we have

$$\begin{aligned} Var(\log_{1-p} N_{0,0} \,|\, N_{*,0}) \quad &= \quad \frac{1}{N\ln^2(1-p)}\left[{\scriptstyle (1-p)^{-n} - 1 - p\,N \cdot p\,n}\right] \\ &\approx \quad \frac{1}{Np^2}\left[e^{p\,n} - (1 + p\,N \cdot p\,n)\right] \end{aligned}$$

By substituting the temporally fixed variable $N_{*,0}$ (or variable N) with the its expected value, we can get the approximation of $Var(\log_{1-p} \frac{N_{0,0}}{N_{*,0}})$ shown in Theorem 2.