Efficient Missing Tag Identification in Large High-Dynamic RFID Systems

Xinning Chen[1], Xuan Liu[1](✉), Ying Xu[1], Jiangjin Yin[1], and Shigeng Zhang[2]

[1] Hunan University, Changsha 410082, China
{chenxinning,xuan_liu,xuyinghnu,jiangjinyin}@hnu.edu.cn
[2] Central South University, Changsha 410083, China
sgzhang@csu.edu.cn

Abstract. Missing tag detection is an important for many radio frequency identification (RFID) systems. Most existing methods can only work in relatively static systems, where there are no unknown tags entering the system. For highly dynamic RFID systems where tags move into and out from the system frequently, it is challenging to identify missing tags because of the interference from unknown tags. In this paper, we propose a new time efficient protocol called HDMI to identify missing tags in highly dynamic RFID systems. Our idea is to combine the index of the replying slot and the bits replied by the tag to efficiently filter out unknown tags and identify missing tags simultaneously. We theoretically analyze how to set optimal parameters (e.g., frame length and bit number replied by tags) to minimize the execution time while ensuring the recognition accuracy of missing tags. Extensive experimental results show that HDMI identify missing tags with a high accuracy rate, and achieving higher efficiency than state-of-art solutions.

Keywords: Radio frequency identification · Missing tag identification · Dynamic system · Time efficiency

1 Introduction

Radio frequency identification (RFID) has been extensively applied in warehouse management [16], inventory control [10] and supply chain management [3]. An RFID system usually consists of one or more readers and thousands of tags. The tags are small and inexpensive and can be attached to almost every object with a unique ID to identify the item. The reader is deployed with one or several antennas to scan tags in the monitoring area via wireless communications. With RFID technology, the tags can be monitored in nonline-of-sight manner and several tags can be read simultaneously.

Missing tag detection and identification plays an important role in the RFID applications. For example, consider a large storehouse in which there are tens of thousands of goods labeled by RFID tags. We might want to know whether some goods are lost and which good are lost because of thief or delivery. Thus how

H. Jiang et al. (Eds.): ICECI 2020, LNICST 368, pp. 141–156, 2021.
https://doi.org/10.1007/978-3-030-73429-9_9

to monitor these tagged items by identifying the missing tags is a challenging issue. A simple solution is to query all tags' IDs one by one and verify which one is missing. This method, while can find all the missing tags, is time-consuming because most tags are actually not missing. Another method is to broadcast tag ID one by one and identify missing tags by checking whether there are replies from tags matching the broadcasted IDs. This method is not suitable for large RFID systems that contains a large number of tags.

Existing solutions to missing tag identification always make each tag randomly select a slot to return a 1-bit response which shows the tag's existence in each frame. When the reader receives data transmitted from the slot selected by a tag, the reader knows that a tag is present and confirms that it is not missing. Otherwise, if the reader does not receive data as expected, the tag mapping to this slot is considered missing. However, existing missing tag identification protocols assume there are no unknown tags when performing missing tag identification. In practice, tags might enter the system and leave the system frequently. The newly entered tags will interfere with the identification process of missing tags and thus will decrease accuracy of missing tag identification.

Recently there are few works considering the scenario where unknown tags and missing tags co-exist in the RFID system [1,14]. They generally consist of two phases. In the first phase, the reader deactivates the unknown tags. In the second phase, most existing protocols on missing tag identification can be used to solve the remaining issues. However, these protocols are time inefficient because they handle unknown tags and missing tags separately. Moreover, the interference between unknown tags and missing tags degrades accuracy in both phases.

In this paper, we propose a protocol to identify missing tags in a highly dynamic RFID system in which unknown tags exist. The protocol combines the tag's replying slot location in the frame along with the bits information to determine whether a tag is missing. The recognition and deactivation of unknown tags can be done simultaneously with the identification of missing tags, without the need of two phases as in previous works [1,14]. The main contributions of this paper are as follows:

- We propose a novel method that can simultaneously recognize unknown tags and missing tags. The idea is to combine both the tag's replying slot location in the frame and the tag's replying bits to increase the ability to distinguish different types of tags. With this method, missing tags are identified with high accuracy and efficiency.
- We carry out theoretical analysis on how to optimize frame length and the reply bits length to minimize the execution time under the condition of ensuring the recognition accuracy of missing tags.
- We conduct extensive simulation experiments to evaluate the performance of the proposed algorithm and make comparison with existing algorithms.

The remainder of this paper is organized as follows. The related work is discussed in Sect. 2. The problem statement is introduced in Sect. 3, and we build the system model. In Sect. 4, we propose our protocol, and then calculate the optimal parameters in theoretical analyses in Sect. 5. Finally, our protocol is simulated to test the performance and summed up in Sect. 6 and Sect. 7 respectively.

2 Related Work

Tag identification, which is one of the most focus aspects in such areas as RFID, is divided into two categories: ALOHA-based protocols [6,7,12] and tree-based protocols [2,4,9]. Most existing missing or unknown tag identification methods are often designed on the basis of ALOHA protocol. In ALOHA protocol, a period of time few greater than sending a tag's echoes is defined as a slot. Several slots are combined into a frame. At the beginning of each frame, the reader first broadcasts the frame size f and a random seed r to all tags in its interrogating. Each tag computes $s = H(ID, r) \bmod f$ to "randomly" select only one slot to respond in a frame after it receives these parameter. If a tag successfully establishes communication with the reader, it will keep silence in the following frame. According to the listening condition of reader, the slots can be divided into three types. If no tag responds in this slot, we call it empty slot. If unique tag responds in this slot and communicates successfully, we call it single slot. If more than one tag reply in this slot, we call it collision slot.

As for the issue of missing tag problem, the study in the missing tag problem can be divided into two directions: missing tag detection and missing tag identification. When we just want to know if something is stolen, it is suitable to adopt the missing tag detection scheme. Luo et al. [13] used multiple random seeds, considering the balance between time-effectiveness and energy consumption, to increase the single-slot probability to detect the existence of missing tag. Yu et al. [15] used Bloom filters which have different lengths, considering the circumstance of multiple-group and multiple-region, to combine the responses from the tags receiving different parameters in each region as a Group filter. Compared with the pre-populated filter, the existence of missing tag can be detect.

If we try not only to judge the appearance of missing tag, but also to pick out which tag is missing, the missing tag identification scheme will go into operation. Li et al. [11] proposed a protocol to make the conflicting tags participate in the reply probabilistically. In this way, the probability of collision slots could be reduced. On the contrary, the chance of single slots increases. Liu et al. [5] applied hash function to adjust some predicted 2-collision and 3-collision slots into single slots in the additional vector, then the tags which are only precomputed on the single slots respond message to reader.

In recent years, most studies focus on solving the topics mentioned above, but these studies have a key limitation that the tag set always contains some unexpected tags whose IDs we never know in reality. In view of the actual situation, few protocol was designed. Shahzad et al. [8] only paid attention to the alteration of the predicted single slots to detect missing tag. Chen et al. [1] proposed a solution including two phases. In the first phase, the reader listened in the predicted empty slots to deactivate the unknown tags. In the second phase, the reader listened in the predicted single slots to identify the missing tag. Similarly, Yu et al. [14] utilized Bloom filter to complete the same tasks in two phases as well. However, with the increasing frequency of in and out warehousing, existing methods are inefficient or even cannot work. The missing tag identification interfered by unknown tags is still a valuable research issue.

3 System Model and Problem Statement

3.1 System Model

We consider an RFID system consisting of three parts: an RFID reader, a backend server, and a large number of RFID tags. The background server is responsible for coordinating the processing and analyzing the information received by the reader. The reader broadcasts the parameters determined by the server to the tags, and transmits the tag replies back to the server. The readers communicate with tags by using the frame slotted ALOHA protocol [12]. Tag might move in and out from the system frequently. We assume that the background server knows all original tags in the system, and will update the tag list after each missing tag operation.

According to the states of tagged items, tags can be classified into three categories: present tag, unknown tag, and missing tag. In the monitoring field, the tag attached to the item that always exists in the system is called *present* tag. The tag attached to the strange item newly added into the system is called *unknown* tag. The tag attached to the item that has been taken out or lost from the system is called *missing* tag.

3.2 Problem Formulation

Definition 1 (The missing tag identification problem). *The current tags set in the system is* N'. *The original known tags set of the system is* N_0. *Referring to the system model mentioned above, there are present tags, missing tags and unknown tags in the RFID system. Given a specified probability* α *from 0 to 1, we need to identify the missing tag at least* $\alpha * N_m$ *under the influence of the appearance of an unknown tag, that is*

$$\frac{N_m - M}{N_m} \leq 1 - \alpha, \tag{1}$$

where N_m is the real number of missing tags in the system which is not known in advance but can be roughly estimated by some estimation algorithms. And M represents the number of missing tags actually identified by the reader. For example, if we set the specified probability α to 0.9, while 1000 tags were really lost in the system, only 100 missing tags are allowed to be misidentified on average.

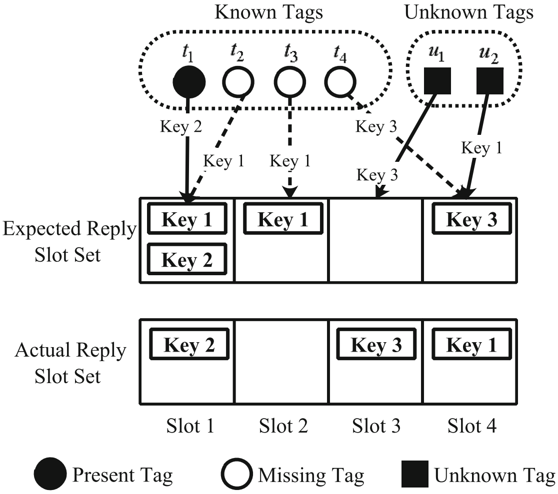


Fig. 1. Illustration of the solution.

4 Protocol Design

4.1 Design Overview

Our proposed protocol HDMI judges the category of a tag by means of slot position and reply Key together. Without divided into two phases, the missing tags can be quickly identified without the interference of unknown tags in a frame. If the tag should reply RN16, a hash function $P = H(id, r) \bmod 2^l$ is used to generate random bits which can be predicted. We define the random bits as the reply Key of the tag, whose total number is marked as L. The figure of random bits is called the reply Key length, which is marked as l. Hence $L = 2^l$. By the replied Key, some expected single slots and some expected collision slots can be able to participate in missing tag identification.

At the beginning, we predict the echoed slot and calculate the reply $Keys$ of the initial known tags. Then the reader starts to query the tags and collects the reply $Keys$ from all interrogated tag. First, some missing tags and unknown tags can be easily distinguish by comparison of pre-single slot and pre-empty slot like the previous works. Then, the remaining unidentified tags can be differentiated by contrasting the slot reply Key set. Finally, all tags are differentiated so that all the missing tags are identified. The details will be given in the next section.

For example, as shown in the Fig. 1, the known tag t_4 should have reply in the fourth slot but it has been lost. Unfortunately, an unknown tag u_2 responds actually in the fourth slot. As we all know, a 1-bit reply can not solve this case, but this situation can be solved easily when each tag has its own Key. It is evident from the Fig. 1 that the known tag t_4 holds the Key 3 while the unknown tag u_2 holds the distinct Key 1. When the reader expects to receive a

Key 3 but actually receives a *Key* 1, it can immediately judge that the actual reply *Key* replied by an unknown tag and the initial known tag t_4 is missing.

This method can also identify missing tags even when multiple tags are expected to reply in a certain slot but only one tag actually participates in the reply. A typical example is given in Fig. 1, the initial known tag t_1, t_2 all choose to reply in the first slot but only the tag t_1 replies in the first slot actually. If all the tags reply a 1-bit, this case can't be found at all. When each tag has a *Key*, all the tags' *Keys* which expected to reply in a certain slot can be known by the initial known tags. After comparing the actual reply *Key* with all the expected *Keys*, the category of the tag can be confirmed. As we can see, the known tag t_1 holds *Key* 2 and the known tag t_2 holds *Key* 1. When the *Key* 2 is actually received, it is compared with the expected *Key* 1 and *Key* 2. We can find that the actual reply *Key* 2 is replied by the known tag t_1 and the known tag t_2 has been lost.

4.2 Detail Description

In this section, we describe the details of the proposed protocol HDMI. And an example is given later.

Firstly, the system backend server generates the required parameters including the random seed r_1 which used to calculate the slot reply position in the frame of a round, and the random seed r_2 which used to calculate reply *Keys* held by the tags, combining the frame size f and the reply *Key* length l. Then the system server predicts the slot positions of all the known tags in a frame based on $S_i = H(id, r_1) \bmod f$, and infers the reply *Key* of each tag from $P_i = H(id, r_2) \bmod 2^l$. Associating with slot position and reply *Key*, the expected reply slot set E of a round is completed. If a expected reply slot selected by no tag, it is recorded as 0 in the expected reply slot set, i.e. $E[i] = \{0\}$.

Secondly, the reader broadcasts the same parameters r_1, r_2, f, l which was used to generate the expected reply slot set E to the known tags. At this moment, only the tags existing the coverage area can receive the parameters broadcast by the reader. With the parameters r_1, r_2, f, l, each tag performs $S_i = H(id, r_1) \bmod f$ as its reply slot position and $P_i = H(id, r_2) \bmod 2^l$ to determine its reply *Key*. Then the reader starts the slot signal at the beginning of each slot, and the tag determines whether replies in this slot according to the value of the counter where its slot position number is stored. If the counter value is not zero, the tag does not reply, storing the value subtracting 1 and waiting for the next slot signal. When the counter value of the tag is zero, the tag replies to the calculated reply *Key*. The reader receives reply *Key* in each slot. According to receive actual reply *Key* in each slot, the real reply slot set R can be formed. When no reply *Key* of the tag is received in a slot, the slot is an empty slot and we sign the reply slot set of this slot as 0, i.e. $R[i] = \{0\}$. When only one reply *Key* is received in a slot which is a single slot, the reply *Key* can be clearly read, i.e. $R[i] = \{Key\ L\}$. When multiple reply *Keys* are received in a slot which becomes a collision time slot, the reader cannot clearly distinguish several reply $Keys$ and we sign the reply slot set of this slot as X, i.e. $R[i] = \{X\}$.

Finally, by comparing every reply slot set of the actual reply slot set $R[i]$ and the expected reply slot set $E[i]$, the missing tags will be identified. The compared results between $E[i]$ and $R[i]$ are listed as follow:

- C1:$E[i] = \{0\}/E[i] = \{Key\ L_i\}/E[i] = \{Key\ 1, ...,$ $Key\ L_i\}, R[i] = \{0\}$: There should be one or more known tags replying in this slot but no reply is received. Most previous works identify the missing tags through this situation, of course, our protocol HDMI can also identify the missing tags in this situation.
- C2:$E[i] = \{Key\ L_i\}/E[i] = \{Key\ 1, ..., Key\ L_i\}$, $R[i] = \{Key\ L_{ri}\}$: In this case, our proposed HDMI can use these slots to identify the missing tags. If $E[i] = \{Key\ L_i\}$, $R[i] = \{Key\ L_{ri}\}$, it means a known tag should reply in this slot and a reply is actually received in this slot. Then we compare the expected Key with the actual Key received by the reader. There will be two possible identification results. One is that the tag replying in this slot is a present tag while the expected $Key\ L_i$ is equal to the actual $Key\ L_{ri}$. The other is that the tag actually replying in this slot is a unknown tag and the tag which holds the expected $Key\ L_i$ in this slot is missing while the expected $Key\ L_i$ is not equal to the actual $Key\ L_{ri}$. If $E[i] = \{Key\ 1, ..., Key\ L_i\}$, $R[i] = \{Key\ L_{ri}\}$, it means more than one known tags should reply in this slot and a reply is actually received in this slot. There will also be two possible identification results. One is that the tag replying in this slot is a present tag while the actual $Key\ L_{ri}$ is equal to one of the expected $Key\ L_i$. The other is that the tag actually replying in this slot is a unknown tag and all the known tags which expected to reply in this slot are missing while there is no expected $Key\ L_i$ equaling to the actual $Key\ L_{ri}$.
- C3:$E[i] = \{Key\ L_i\}/E[i] = \{Key\ 1, ..., Key\ L_i\}$, $R[i] = \{X\}$: There should be one or more known tags replying in this slot but multiple tags actually reply in this slot so that no Key can be read. Therefore, the tags replying in this slot can not be identified and will continue to participate in reply in the next frame.
- C4:$E[i] = \{0\}$, $R[i] = \{Key\ L_{ri}\}/R[i] = \{X\}$: None of the known tags should reply in this slot but some $Keys$ are received in the slot actually. These must be the $Keys$ of the unknown tags. So, in this case, no missing tag will be identified. In contrast, the unknown tags can be easily distinguished in this slot.

5 Parameter Optimization

5.1 Setting the *Key* Length *l*

In this section, we discuss how to set Key length l for each round to achieve the required identification accuracy. For selecting the same slot, each known tag can be completely distinguished only if their $Keys$ are different. Assume that for the ith slot, k tags are expected to choose to reply l-length Key in this slot simultaneously. That being the case, $L = 2^l$ is all the kinds of $Keys$ which a tag may hold. Let $p_{n=k}$ be the probability that a reply Key is different from others. Then

$$p_{n=k} = \frac{L}{L} \cdot \frac{L-1}{L} \cdots \frac{L-k+1}{L} = \frac{A_L^k}{L^k} \tag{2}$$

For the ith slot, where a new tag replies in reality and one of expected tags is missing, the unknown tag can be successfully recognized with the different Key from all the other expected $Keys$ instead of being mistaken for the expected tag that has been lost. As a result, the missing tag can be correctly identified. So the probability p_{uk} that the missing tag can be recognized considering the presence of the unknown is

$$\begin{aligned} p_{uk} &= p_{n=k} \cdot \frac{L-k}{L} = \frac{A_L^k}{L^k} \cdot \frac{L-k}{L} \\ &= \frac{L}{L} \cdot \frac{L-1}{L} \cdots \frac{L-k+1}{L} \cdot \frac{L-k}{L} = \frac{A_L^{k+1}}{L^{k+1}} \end{aligned} \tag{3}$$

which needs to be greater than the specified probability value α, i.e. $p_{uk} \geq \alpha$. With the Stirling Formula $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n = \sqrt{2\pi n}e^{-n} \cdot n^n$, we can get

$$L \geq \frac{(2\ln\alpha - 1)N_0}{2\ln\alpha \cdot f} \tag{4}$$

where $L = 2^l$ is all the possible $Keys$ for a tag in this round. Furthermore, assume that the number of original known tags is N_0 and the frame size is f in this round, there are $k = \frac{N_0}{f}$ excepted tags responding in each slot on average.

5.2 Determining the Optimal Frame Size *f*

In this section, we discuss how to set frame length f for each round to achieve the best efficiency. For the sake of analysis, we assume that the number of present tags is denoted as N_s, with the number of missing tags is N_m and the number of unknown tags is N_u. Also, the number of original known tags is denoted as N_0, the frame size is f and the reply Key size is l, hence the possible $Keys$ are $L = 2^l$.

Firstly, we discuss the execution time of a round. Let $\rho = \frac{N_0}{f}$, the probability of the empty slot p_0, the single slot probability p_1, and the collision slot probability p_2 are respectively given by

$$p_0 = (1 - \frac{1}{f})^{N_0} = e^{-\rho} \tag{5}$$

$$p_1 = \frac{N_0}{f} \cdot (1 - \frac{1}{f})^{N_0 - 1} = \rho \cdot e^{-\rho} \tag{6}$$

$$p_2 = 1 - p_0 - p_1 = 1 - (1 + \rho) \cdot e^{-\rho} \tag{7}$$

When the transmission rate is 62.5 Kbps, referring to the EPC C1G2 specification, the time costs of empty slot, single slot and collision slot are

$$t_0 = 182.5\,\mu\text{s}, t_1 = t_c = (182.5 + 16l)\ \mu\text{s} \tag{8}$$

Therefore, the total execution time of a round t_{total} is:

$$\begin{aligned} t_{total} &= f \cdot p_0 \cdot t_0 + f \cdot p_1 \cdot t_1 + f \cdot p_2 \cdot t_c \\ &= f \cdot \{t_0 \cdot e^{-\rho} + (t_0 + 16l) \cdot \rho \cdot e^{-\rho} \\ &\quad + (t_0 + 16l) \cdot \left[1 - (1 + \rho) \cdot e^{-\rho}\right]\} \\ &= f \cdot \left(t_0 + 16l - 16l \cdot e^{-\rho}\right) \end{aligned} \tag{9}$$

Secondly, we discuss the number of tags that can be identified in a round. According to the protocol, when the actual reply slot is a collision slot, all the corresponding tags in this slot can not be identified. Let event A be k excepted tags choosing a slot to reply their $Keys$ and event B be the number of tags which actually reply their $Keys$ in this slot. Thus the probability that tags in a slot are not identified is

$$\begin{aligned} P\{B \geq 2, A = k\} = &(1 - P\{B = 0|A = k\} \\ &- P\{B = 1|A = k\}) \cdot P\{A = k\} \end{aligned} \tag{10}$$

Let the probability of losing a tag be $P_m^k = \frac{Nm}{N_0}$, k excepted tags in this slot are all missing, hence

$$P\{B = 0|A = k\} = P_m^k \cdot (1 - \frac{1}{f})^{Nu} = (\frac{Nm}{N_0})^k \cdot (1 - \frac{1}{f})^{Nu} \tag{11}$$

Let the probability of present tag in this slot be $P_s^k = \frac{Ns}{N_0}$, According to the protocol described above, when a reply Key is actually received from a tag, either an unknown tag appears with the k missing tags, or only one known tag exists with $k - 1$ missing tags. Hence

$$\begin{aligned} P\{B{=}1|A{=}k\} &= P_m^k \cdot \frac{1}{f}(1{-}\frac{1}{f})^{Nu-1} \cdot C_{Nu}^1 + C_k^{k-1} P_m^{k-1} P_s (1{-}\frac{1}{f})^{Nu} \\ &= (\frac{Nm}{N_0})^k \cdot Nu \cdot \frac{1}{f}(1 - \frac{1}{f})^{Nu-1} + k \cdot (\frac{Nm}{N_0})^{k-1} \cdot \frac{Ns}{N_0} \cdot (1 - \frac{1}{f})^{Nu} \end{aligned} \tag{12}$$

And substituting $P\{A = k\} = C_{N_0}^k \cdot (\frac{1}{f})^k (1 - \frac{1}{f})^{N_0 - k}$ with $Eq.$ (17) and (18) into $Eq.$ (16), the probability that the missing tag can not be identified in this slot p_k is:

$$\begin{aligned} p_k &= P\{B \geq 2, A = k\} = (1 - P\{B = 0|A = k\} \\ &\qquad - P\{B = 1|A = k\}) \cdot P\{A = k\} \\ &= C_{N_0}^k \cdot (\frac{1}{f})^k (1 - \frac{1}{f})^{N_0 - k} \cdot \{1 - (\frac{Nm}{N_0})^k (1 - \frac{1}{f})^{Nu} \cdot [1 + k \cdot \frac{Ns}{Nm} - \frac{Nu}{1 - f}]\} \end{aligned} \tag{13}$$

Since the quantity of tags mapped to this slot k is between 1 and N_0. Therefore, the expected unidentified tags in this slot become $\sum_{k=1}^{N_0} k \cdot p_k$, the expected number of unidentified tags in this round is:

$$\begin{aligned} E(Q) = f \cdot \sum_{k=1}^{N_0} k \cdot p_k = f \cdot \sum_{k=1}^{N_0} k \cdot C_{N_0}^{k} (\frac{1}{f})^k (1 - \frac{1}{f})^{N_0 - k} \\ \cdot \{1 - (\frac{Nm}{N_0})^k (1 - \frac{1}{f})^{Nu} \cdot [1 + k \cdot \frac{Ns}{Nm} - \frac{Nu}{1-f}]\} \end{aligned} \tag{14}$$

Because $f \gg 1$, the Eq. 20 can be turned into

$$\begin{aligned} E(Q) = f \cdot \sum_{k=1}^{N_0} k \cdot p_k = f \cdot \sum_{k=1}^{N_0} k \cdot C_{N_0}^{k} (\frac{1}{f})^k (1 - \frac{1}{f})^{N_0 - k} \\ \cdot \{1 - (\frac{Nm}{N_0})^k (1 - \frac{1}{f})^{Nu} \cdot [1 + k \cdot \frac{Ns}{Nm} + \frac{Nu}{f}]\} \end{aligned} \tag{15}$$

Then the number of identified tags is $N_0 - E(Q)$.

Finally, by the total time and the identified tags number in a round, the time average identifying one tag η is give by:

$$\begin{aligned} \eta &= \frac{t_{total}}{N_0 - E(Q)} = \frac{t_{total}}{N_0 - f \cdot \sum_{k=1}^{N_0} k \cdot p_k} \\ &= \frac{f \cdot (t_0 + 16l - 16l \cdot e^{-\rho})}{N_0 - f \cdot \sum_{k=1}^{N_0} k \cdot p_k} = \frac{t_0 + 16l - 16l \cdot e^{-\rho}}{\rho - \sum_{k=1}^{N_0} k \cdot p_k} \end{aligned} \tag{16}$$

Substituting Eq. (21) into Eq. (22) gives:

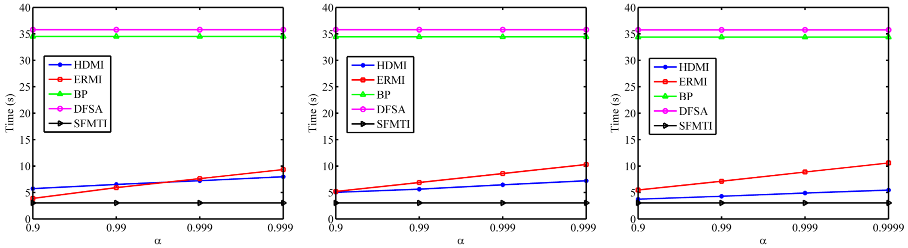
$$\eta = \frac{(t_0 + 16l) \cdot e^{(\frac{Nu+Ns}{N_0} - 2 \cdot \frac{Ns}{N_0^2}) \cdot \rho} - 16l \cdot e^{(\frac{Nu+Ns-N_0}{N_0} - 2 \cdot \frac{Ns}{N_0^2}) \cdot \rho}}{\rho + \frac{Nm}{N_0^2} \left[(Nu+Ns) - \frac{2Ns}{N_0} - \frac{N_s^2}{NmN_0} \right] \cdot \rho^2 - \frac{NuNsNm}{N_0^4} \cdot \rho^3} \tag{17}$$

In order to minimize the execution time to achieve the best efficiency, we want to get the average time of identifying one tag η as short as possible. In addition, the high efficiency not only to be met in the process of the missing tag identification, but it is also necessary to ensure the accuracy of the identification. To achieve the required accuracy, we refer to Eq. (7) and an equation is used for subsequent analysis as follows:

$$2^l = \frac{(2 \ln \alpha - 1) N_0}{2 \ln \alpha \cdot f} \tag{18}$$

Due to $\rho = \frac{N_0}{f}$, the length of reply Key in this round is given by

$$l = \log_2 \left(\frac{(2 \ln \alpha - 1) \cdot \rho}{2 \ln \alpha} \right) \tag{19}$$



(a) (b) (c)

Fig. 2. Impact of the required identification accuracy on the performance of total execution time, where $N_0 = 10000$ and the number of missing tags and unknown tags are set to $(a)\, N_m = N_u = 1000, (b)\, N_m = N_u = 5000, (c)\, N_m = N_u = 9000$.

Therefore, by substituting *Eq.* (25) into *Eq.* (23), a linear equation with one unknown equation for ρ, η can be obtained. Then, we just need to find out the minimal value of η to get the optimal frame size f.

6 Performance Evaluation

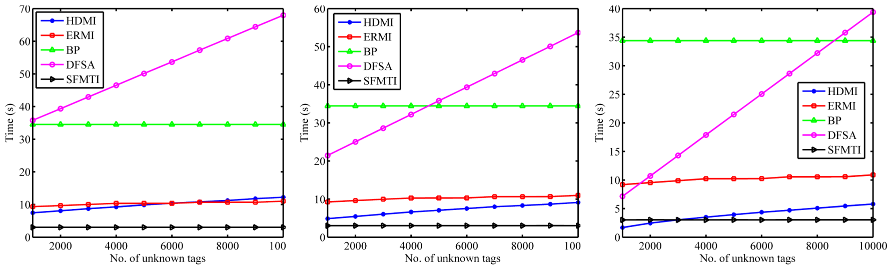
We evaluate the performance of the proposed HDMI algorithm and compare it with state-of-the-art solutions, including ERMI [1] and SFMTI [5]. Meanwhile, we also compare HDMI with two straightforward solutions: the first collects all the tags' ID by using the DFSA protocol and the second broadcasts all tag IDs one by one. We call the former DFSA and the latter one BP.

When evaluating the performance of different algorithms, we consider the following parameters: (1) the missing tag identification accuracy α; (2) the ratio of missing tags and unknown tags in the system; (3) the total number of tags in the system. For each parameter setting, we repeat the simulation experiment 20 times and report the averaged data over the 20 experiments.

6.1 The Impact of Required Identification Accuracy α

In this subsection, we investigate the performance of total execution time under different required identification accuracy α. We set the number of the initially known tags as 10000, i.e. $N_0 = 10000$. We consider three scenarios: (a) the low-dynamic scenario where $N_m = N_u = 1000$; (b) the median-dynamic scenario where $N_m = N_u = 5000$; (c) the highly-dynamic scenario where $N_m = N_u = 9000$. In these tag dynamic case, we obtain the total execution time with different required identification accuracy α, i.e., when $\alpha = 0.9, \alpha = 0.99, \alpha = 0.999, \alpha = 0.9999$ respectively.

Figure 2 shows the identification time in three different scenarios for different α. It can be observed that HDMI consumes less time than all other algorithms except SFMTI when α becomes larger. Actually, HDMI significantly outperforms



(a) (b) (c)

Fig. 3. Impact of the number of unknown tags on the performance of total execution time, where $N_0 = 10000$, $\alpha = 0.9999$ and the number of missing tags are set to $(a)\, N_m = 1000, (b)\, N_m = 5000, (c)\, N_m = 9000$.

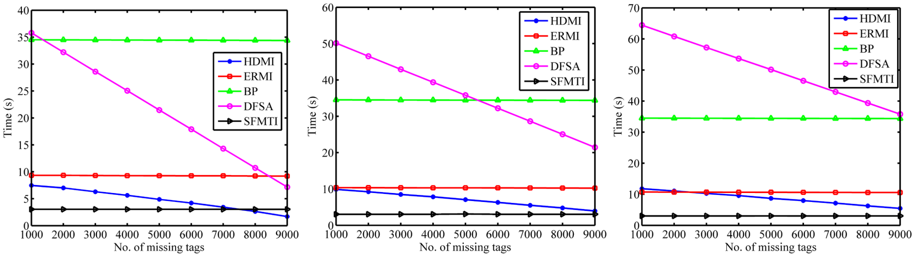
ERMI by reducing total time nearly 50%. SFMTI uses least time and its execution time is not affected by α. However, as to be discussed in the next section, SFMTI cannot satisfy the required identification accuracy. A large portion of missing tags cannot be detected by SFMTI.

6.2 The Impact of the Number of Unknown Tags

In this subsection, we investigate the total execution time under different unknown tag ratios. We set the number of initially known tags to 10000, i.e. $N_0 = 10000$. We also set required identification accuracy to 0.9999, i.e. $\alpha = 0.9999$. First, for the low-dynamic case, we set the missing tag ratios to 10

Figure 3 shows the total execution time in the three different number of missing tags for changing number of unknown tags. In three cases, with the increase of unknown tags, the execution time of most protocols is gradually increasing. Only the total execution time of BP and SFMTI is still unchanged. This is because the execution time of the BP is only related to the number of initial known tags as described in the previous section. The number of initial known tags remains unchanged, so does the execution time. As for the SFMTI protocol, since it does not consider the existence of unknown tags, the change of unknown tags' number has little effect on its execution time. What's more, Fig. 3(c) illustrates that our proposed protocol HDMI is able to better identify the missing tags in the case of higher and higher tag dynamic environment. Compared with the latest method ERMI, our method has shortened the execution time by more than half in this case.

What needs to be explained here is that more and more tags are missing, fewer and fewer tags are actually in the coverage of the reader within the same number of unknown tags. So it takes the less time for DFSA to identify the tags in the current reader range, which is why the execution time of DFSA in the highly-dynamic case is getting shorter and shorter. As can be seen in Fig. 3(c), the execution time is even less than ERMI when the number of unknown tags is 0.



(a) (b) (c)

Fig. 4. Impact of the number of missing tags on the performance of total execution time, where $N_0 = 10000$, $\alpha = 0.9999$ and the number of unknown tags are set to $(a)\, N_u = 1000, (b)\, N_u = 5000, (c)\, N_u = 9000$.

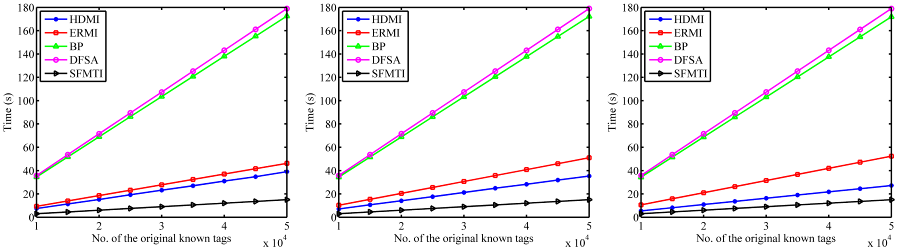
Similarly, the SFMTI silences or identifies tags based on the 1-bit reply from without distinguishing the unknown tags. Many unknown tags are silenced as the identified present tags, which causes the missing tags can not be found out accurately. Thus, its identification accuracy is far from the required identification accuracy α, which will be proved in the following section.

6.3 The Impact of the Number of Missing Tags

In this subsection, we investigate the performance of total execution time under different missing tag ratios. The number of initially known tags is set to 10000 as before, i.e. $N_0 = 10000$. The required identification accuracy is still 0.9999, i.e. $\alpha = 0.9999$. In the three tag dynamic cases, we set the unknown tag ratios to three ratios of the initial known tags, which is 0.1, 0.5, 0.9 respectively, hence $N_u = 1000$, $N_u = 5000$, $N_u = 9000$. Then we vary the number of missing tags N_u from 1000 to 9000 to get the total execution time.

Figure 4 shows the total execution time in the three cases of unknown tags with changing number of missing tags. With the increase of missing tags, the execution time of HDMI, ERMI, DFSA are gradually declining. The DFSA needs to identify all tags in the coverage of the reader currently. The more tags are missing, the fewer tags exist in the area, so the execution time of DFSA varies greatly. Moreover, with the increasing number of missing tags, HDMI can achieve better efficiency than ERMI and DFSA. Even our HDMI's total execution time is less than all the other protocols when N_m is 9000 and N_u is 1000.

What's more, the total execution time of BP and SFMTI are still unchange. The number of initial known tags remains unchanged, so the execution time of the BP remains unchanged. In SFMTI protocol, it identifies the missing tags by the slot position of each known tags, so its execution time almost remains unchanged with the fixed number of initial known tags. Similarly, the SFMTI's identification accuracy does not meet the required identification accuracy α, even if it often has the least time.



(a) (b) (c)

Fig. 5. Impact of the initial known tags N_0 on the performance of total execution time, where $\alpha = 0.9999$ and the number of missing tags and unknown tags are set to $(a)\, N_m = N_u = 0.1 * N_0, (b)\, N_m = N_u = 0.5 * N_0, (c)\, N_m = N_u = 0.9 * N_0$.

6.4 The Impact of the Number of Initial Known Tags

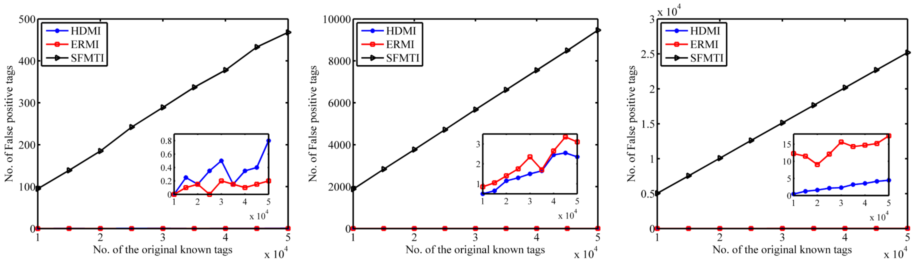
In this subsection, we investigate the performance of total execution time under different number of initial known tags. We set the number of missing tags N_m equal the number of unknown tags N_u, which is $N_m = N_u = 0.1N_0$, $N_m = N_u = 0.5N_0$, $N_m = N_u = 0.9N_0$ respectively. And the number of initially known tags N_0 varies from 10000 to 50000. The required identification accuracy α is 0.9999.

Figure 5 shows the total execution time of all protocols when $N_m = N_u = 0.1N_0$, $N_m = N_u = 0.5N_0$, $N_m = N_u = 0.9N_0$ respectively. With the increasing number of initially known tags, the execution time of all the protocols is gradually increasing. In these case, both the total number of initial tags N_0 and the total number of current tags N' are increasing, so the execution time of BP and DFSA is getting longer. In particular, our HDMI protocol is more efficient than ERMI, BP and DFSA in the low-dynamic case, the higher-dynamic case and the highly-dynamic case. Especially, the efficiency of HDMI has obvious advantages in the highly-dynamic case. Although SFMTI consumes little time, its identification accuracy does not satisfy α,which will be explained in the next section.

6.5 False Positive Tags

In this subsection, we investigate the performance of the number of false positive tags to study the identification accuracy of the protocols. The BP and DFSA can identify the missing tags by the ID, so their identification accuracy can reach 100%. Next, we will not study the accuracy of these two protocols, but discuss the accuracy of the other three protocols.

Figure 6 shows the number of false positive tags which is identified incorrectly when $\alpha = 0.9999$, $N_m = N_u = 0.1N_0$, $N_m = N_u = 0.5N_0$ and $N_m = N_u = 0.9N_0$ respectively. It can be clearly seen that the SFMTI identifies a large number of false positive tags in the three dynamic cases. From 6(a) and (b), the HDMI and ERMI can meet the identification accuracy α in the low-dynamic case and the higher-dynamic case. However, as can be seen in 6(c), the false positive tags



(a) (b) (c)

Fig. 6. Impact of the initial known tags N_0 on the performance of the number of false positive tags, where $\alpha = 0.9999$ and the number of missing tags and unknown tags are set to $(a)\, N_m = N_u = 0.1 * N_0, (b)\, N_m = N_u = 0.5 * N_0, (c)\, N_m = N_u = 0.9 * N_0$.

of ERMI can not meet the requirements in the highly-dynamic case. Only our HDMI can identify the missing tags under the identification accuracy in the highly-dynamic case.

7 Conclusion

In this paper, based on the high-dynamic RFID system, where missing tags, unknown tags, and present tags coexist, we propose a protocol HDMI to identify missing tags with high accuracy and efficiency, which maximizes the use of tag reply slot information without the interference of the unknown tags. The core is identifying a tag by the slot reply position with the reply Key. This paper also discuss the optimal frame length f and reply Key length l. Furthermore, we implemented our HDMI to evaluate its performance. Our proposed protocol HDMI can make a great contribution to the convenience of logistics monitoring in practical high-dynamic RFID scenarios.

Acknowledgements. This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61602167 and 61972140), the Natural Science Foundation of Hunan Province under grant No. 2020JJ3016, and the National Defense Basic Research Plan under Grant JCKY2018110C145.

References

1. Chen, H., Xue, G., Wang, Z.: Efficient and reliable missing tag identification for large-scale RFID systems with unknown tags. IEEE Internet Things J. **PP**(99), 1 (2017)
2. Hou, Y., Zheng, Y.: PHY assisted tree-based RFID identification. In: IEEE Infocom -IEEE Conference on Computer Communications (2017)
3. Jia, L., Xiao, B., Kai, B., Chen, L.: Efficient distributed query processing in large RFID-enabled supply chains. In: Infocom. IEEE (2014)

4. Lei, P., Wu, H.: Smart trend-traversal: a low delay and energy tag arbitration protocol for large RFID systems. In: IEEE INFOCOM (Rio de Janeiro, pp. 2571–2579 (2009)
5. Liu, X., Li, K., Min, G., Shen, Y., Liu, A.X., Qu, W.: Completely pinpointing the missing RFID tags in a time-efficient way. IEEE Trans. Comput. **64**(1), 87–96 (2014)
6. Maguire, Y., Pappu, R.: An optimal Q-algorithm for the ISO 18000-6C RFID protocol. IEEE Trans. Autom. Sci. Eng. **6**(1), 16–24 (2009)
7. Schoute, F.C.: Dynamic frame length aloha. Mob. Commun. **31**(4), 565–568 (1983)
8. Shahzad, M., Liu, A.X.: Expecting the unexpected: fast and reliable detection of missing RFID tags in the wild. In: Computer Communications (2015)
9. Shahzad, M., Liu, A.X.: Probabilistic optimal tree hopping for RFID identification. IEEE/ACM Trans. Netw. **23**(3), 796–809 (2015)
10. Smith, A.D., Smith, A.A., Baker, D.L.: Inventory management shrinkage and employee anti-theft approaches. Int. J. Electron. Finan. **5**(3), 209–234 (2011)
11. Tao, L., Chen, S., Ling, Y.: Identifying the missing tags in a large RFID system. In: Eleventh ACM International Symposium on Mobile Ad Hoc Networking and Computing (2010)
12. Wang, J., Hassanieh, H., Katabi, D., Indyk, P.: Efficient and reliable low-power backscatter networks. In: ACM Sigcomm Conference on Applications (2012)
13. Wen, L., Chen, S., Yan, Q., Tao, L.: Missing-tag detection and energy time tradeoff in large-scale RFID systems with unreliable channels. IEEE/ACM Trans. Netw. **22**(4), 1079–1091 (2014)
14. Yu, J., Lin, C.: Finding needles in a haystack: missing tag detection in large RFID systems. IEEE Trans. Commun. **65**(5), 2036–2047 (2019)
15. Yu, J., Lin, C.: On missing tag detection in multiple-group multiple-region RFID systems. IEEE Trans. Mob. Comput. **PP**(99), 1 (2019)
16. Zheng, Y., Li, M.: Fast tag searching protocol for large-scale RFID systems. IEEE/ACM Trans. Netw. **21**(3), 924–934 (2013)