

# A Fast Approach to Unknown Tag Identification in Large Scale RFID Systems

Xiulong Liu\*, Keqiu Li\*, Yanming Shen\*, Geyong Min<sup>§</sup>, Bin Xiao<sup>†</sup>, Wenyu Qu<sup>‡</sup> and Hongjuan Li\*

\*School of Computer Science and Technology, Dalian University of Technology, China

<sup>§</sup>School of Computing, Informatics and Media, University of Bradford, U.K.

<sup>†</sup>Department of Computing, The Hong Kong Polytechnic University, Hong Kong

<sup>‡</sup>School of Information Science and Technology, Dalian Maritime University, China

Corresponding author: Keqiu Li (keqiu@dlut.edu.cn)

**Abstract**—Radio Frequency Identification (RFID) technology has been widely applied in many scenarios such as inventory control, supply chain management due to its superior properties including fast identification and relatively long interrogating range over barcode systems. It is critical to efficiently identify the unknown tags because these tags can appear when new tagged objects are moved in or wrongly placed. The state-of-the-art Basic Unknown tag Identification Protocol-with Collision-Fresh slot paring (BUIP-CF) protocol can first deactivate all the known tags and then collect all the unknown tags. However, BUIP-CF protocol investigates an ALOHA-like technique and causes too many tag responses, which results in low efficiency. This paper proposes a Fast Unknown tag Identification (FUI) protocol which investigates an indicator vector to label the unknown tags with a given accuracy and removes the time-consuming tag responses in the deactivation phase. FUI also adopts the classical Enhanced Dynamic Framed Slotted ALOHA (EDFSA) protocol to collect the labeled unknown tags. We then investigate the optimal parameter settings to maximize the performance of the proposed FUI protocol. Extensive simulation experiments are conducted to evaluate the performance of the proposed FUI protocol and the experimental results show that it considerably outperforms the state-of-the-art protocol.

## I. INTRODUCTION

Compared with conventional barcode systems, Radio Frequency Identification (RFID) technology possesses several attractive properties, such as remote and multiple access, simple computational ability, non-sight limitation, and relatively large rewritable memory. Owing to these superior features, RFID devices have been widely deployed in many application scenarios such as localization [1]–[4], supply chain management [5]–[9], and inventory control [10]–[15]. In these applications, an RFID system usually consists of a single (multiple) reader(s) and a large number of tags. RFID tags are labeled in designated objects where each tag has a small size of memory to store its unique ID and some other information (e.g., product price). The RFID tags are generally categorized as either active tags or passive tags. The passive tag has no battery supply and transmits data by reflecting the received signals from the reader. The passive tag is only activated when it is within the electromagnetic field of a reader. The active tag is powered by an internal battery. An RFID reader is a device used to read (write)

data from (to) the RFID tags and usually can communicate with the central computer through a high-rate network link. The reader manages or monitors the RFID tags within its interrogating range as well as the corresponding items via wireless channels.

The important *RFID tag monitoring* issues are generally classified into two types of research work: *Missing Tag Identification* and *Unknown Tag Identification*. This study investigates the problem of identifying the unknown RFID tags efficiently, which is of great importance in many scenarios, such as the inventory control in a large warehouse, retailing store, shipside, or the baggage management in an airport. In these scenarios, some unregistered items may be mixed with the registered items due to management fault. Obviously, manually checking the misplaced items in a certain region is seriously laborious and of low time-efficiency. If each item is attached with an RFID tag, a reader then automatically identifies the unknown tags so as to find out the corresponding unregistered items. This type of fully automated checking procedure that can be efficiently performed is greatly desirable.

Few research work studies this problem of identifying the unknown RFID tags. To the best of our knowledge, the state-of-the-art scheme is the Basic Unknown tag Identification Protocol-with Collision-Fresh slot paring (BUIP-CF) protocol which is proposed by Liu *et al* in Ref. [16]. BUIP-CF protocol consists of two phases: (i) *known tags deactivation* phase, in which an ALOHA-like technique is used to deactivate all the known tags and label all the unknown tags. And all the known tags are deactivated and do not participate in the next phase; (ii) *unknown tags collection* phase, in which BUIP-CF employs the classical Enhanced Dynamic Framed Slotted ALOHA (EDFSA) protocol [17] to collect all the labeled tags (i.e., unknowns tags). The *highlight* of BUIP-CF is the *known tag deactivation* tactic employed in the first phase, whereas, its *known tags deactivation phase* involves too many tag responses inherent from the classical *ALOHA* communication mechanism that is time-consuming and leads to its low time-efficiency.

To this end, a Fast Unknown tag Identification (FUI) protocol is proposed in this paper. The FUI also consists of two phases: *Unknown Tags Filtering* phase and *Unknown*

*Tags Collecting* phase. In the first phase, FUI employs an indicator vector to filter and label a fraction of unknown tags, and this *filtering process* is repeated for multiple rounds to achieve the desired accuracy. Only the labeled tags (unknown tags) participate in the *Unknown Tags Collecting* phase. In this paper, we further investigate the minimum round count required to guarantee the desired accuracy. In the first phase (i.e., tag separation phase), unlike BUIP-CF protocol, the FUI technique does not involve any tag responses so as to achieve better time-efficiency. In the second phase, the same with BUIP-CF, the proposed FUI protocol also employs the classical EDFSA technique to collect the labeled unknown RFID tags. We then present how to optimize the parameters of the FUI protocol, specifically, the optimal indicator length and the minimum round count. Sufficient simulations are conducted and the results show that the proposed FUI protocol considerably outperforms the state-of-the-art BUIP-CF protocol. The major contributions of this paper can be summarized as follows:

- 1) This paper proposes a Fast Unknown tag Identification (FUI) protocol, in which an indicator vector tactic is investigated to filter out the unknown tags without the time-consuming tag responses inherent from the ALOHA communication mechanism.
- 2) This paper investigates the optimal parameter settings, specifically, the optimal indicator vector length and the minimum round count, of the proposed FUI protocol in order to maximize its performance.
- 3) The extensive simulations are conducted to evaluate the performance of the proposed FUI protocol and the results show its superiority over the state-of-the-art scheme.

The rest of this paper is organized as follows. Section II describes the problem to be addressed in this paper and presents the system model. We propose the FUI protocol and present the related proofs in Section III. In Section IV, extensive simulation experiments are conducted to evaluate the performance of the FUI protocol. This paper is concluded in Section V.

## II. PROBLEM DESCRIPTION AND SYSTEM MODEL

We consider a large-scale RFID system as illustrated in Fig. 1, which consists of a single reader,  $N$  known tags and  $M$  unknown tags, where IDs of  $N$  known tags are stored in the central database but the IDs of unknown tags and the number  $M$  are not available in the database due to a management fault (e.g., misplacing items or moving the unregistered items in). Following the prior RFID related papers [8] [16] [18], we consider a single reader in this paper and assume the reader has adequate power to communicate with all the RFID tags via a *error-free* wireless channel. The known tag set is denoted as  $T_{known}$ , i.e.,  $T_{known} = \{t_1, t_2, \dots, t_i, \dots, t_N\}$ ; and the unknown tag set is denoted as  $T_{unknown}$ , i.e.,  $T_{unknown} = \{tu_1, tu_2, \dots, tu_j, \dots, tu_M\}$ . Each tag has a

unique ID and is equipped with the same uniform Hash generator  $H(\cdot)$ . The reader has access to a database that stores the IDs of all  $N$  known tags. This can be implemented easily if the IDs are stored into (deleted from) the database when the objects are moved into (out of) the system. The RFID reader can communicate with the RFID tags through a one-hop low-rate wireless network link. In contrast, the reader and the central computer can communicate through a single-hop or multiple-hop wireless or wired link. The tags can be active RFID tags or passive RFID tags, or a mixture of them. The problem that this paper investigates is to identify the unknown tags with a desired accuracy in a fast way.

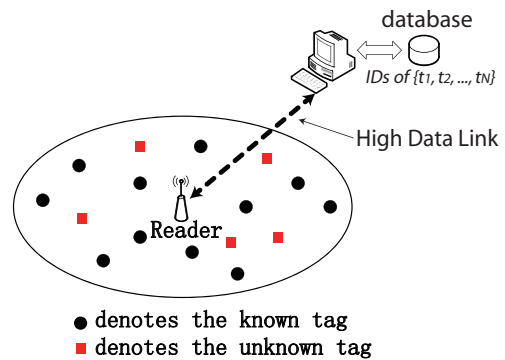


Figure 1. The System Model.

Communications between the reader and tags are in a time-slotted way. The reader synchronizes the slot clocks of all tags by a control signal. In Ref. [19], Li *et al* presented a method of classifying the time slots based on their length: *tag slots*, *long-response slots* and *short-response slots*. The length of a tag slot is denoted as  $t_{tag}$ , which allows the transmission of a tag ID (96 bits), either from the reader to the tags or from a tag to the reader. The length of a long-response slot is denoted as  $t_{long}$ , which can afford transmitting a long response carrying 10 bits information. The length of a short-response slot is denoted as  $t_{short}$ , which allows the transmission of a short response carrying only one bit information. According to the specification of the Philips I-Code system [20], the rate from a tag to a reader is  $53\text{ Kb/s}$  and the rate from a reader to a tag is  $26.5\text{ Kb/s}$ . And any two consecutive transmissions (from a tag to a reader or vice versa) are separated by a waiting time of  $302\mu\text{s}$ . Hence, the length  $t_{tag}$  of *tag slot* is set to  $2.4\text{ms}$  for transmission of a tag ID (96 bits) from a tag to a reader or vice versa; the length  $t_{long}$  and  $t_{short}$  are set to  $0.8\text{ms}$  and  $0.4\text{ms}$ , respectively. The used notations are summarized in Table I.

## III. THE PROPOSED FAST UNKNOWN TAG IDENTIFICATION PROTOCOL

In this section, we first present and analyze the deficiency of two straightforward techniques and the state-of-the-art

Table I  
NOTATIONS.

Symbols	Descriptions
$N$	The number of known tags
$M$	The number of unknown tags
$T_{known}$	The known tag set
$T_{unknown}$	The unknown tag set
$V$	The indicator vector
$v$	The length of the indicator vector $V$
$R$	The random seed that is fresh in each round
$H(\cdot)$	The Hash generator with a uniform random distribution
$e$	The natural constant
$\alpha$	The identification accuracy
$c$	The minimum round count that satisfies the given identification accuracy

scheme for unknown tag identification, which inspires the motivation and ideas of the proposed FUI protocol. We then detail the protocol design of the proposed FUI and investigate how to optimize the parameter settings for achieving its best performance.

#### A. Motivation

The unknown tag identification problem is to collect the IDs of unknown tags. Note that the IDs of known tags are already in the database and we are not interested in collecting their IDs.

The straightforward technique is, *Step 1*: collecting IDs of all tags (both known tags and unknown tags) by employing the *Tag Collection* protocol, for instance, the classical and well-known EDFSA algorithm proposed in Ref. [17]. *Step 2*: comparing the collected results with the IDs of all known tags, then we can get the IDs of unknown tags. The *time complexity* of this technique is  $O((N + M) \cdot e \cdot t_{tag})$  but far from the efficiency of only collecting the IDs of unknown tags, which is just  $O(M \cdot e \cdot t_{tag})$ . The basic reason to this deficiency is re-collecting a lot of redundant known tag IDs. Hence, the challenge is how to *separate* the unknown tags and the known tags, then only collect the IDs of unknown tags.

The simplest tactic for separating the known tags and the unknown tags is a polling method. Specifically, the reader polls the known tags by their IDs one by one, and any tag that finds its ID information in the polling request will keep silent in the following *Tag Collection* phase. Obviously, after this tag polling phase, known tags are filtered out and only unknown tags are active to be collected. The *time complexity* of this method is  $O(N \cdot t_{tag} + M \cdot e \cdot t_{tag})$ , which is better than the above *Collecting-Comparing* technique but still far from the objective  $O(M \cdot e \cdot t_{tag})$ . The fundamental reason is: transmitting IDs of known tags in the *Polling-Separating* phase is severely time-consuming. Hence, the new challenge is how to separate the unknown tags and the known tags without transmitting tag IDs.

The state-of-the-art BUIP-CF protocol investigated by Liu *et al* in Ref. [16] proposed an ALOHA-like technique

to completely deactivate all the known tags and label all the unknown tags. BUIP-CF employs the classical EDFSA algorithm to collect the remaining labeled tags (i.e., the unknown tags). The tactic investigated to deactivate known tags and label unknown tags is a variant of the classical ALOHA communication mechanism, which is the highlight and innovation of BUIP-CF protocol. The *time complexity* of BUIP-CF is about  $O(N \cdot e \cdot t_{long} + M \cdot e \cdot t_{tag})$ . Although known tag deactivation phase does not include the transmission of IDs, it involves too many tag responses inherent from ALOHA technique, which results in the low time-efficiency of BUIP-CF.

To overcome the disadvantages of the above tactics, the FUI protocol is proposed in this paper, aiming at removing both the transmission of IDs and tag responses in the unknown tag filtering phase. This paper mainly focuses on how to efficiently separate the unknown tags from the known tags, and the corresponding tactic is the highlight of our study.

#### B. Protocol Design

The proposed FUI protocol consists of two phases: *Unknown Tags Filtering* phase and *Unknown Tags Collecting* phase. In the *Unknown Tags Filtering* phase, the FUI protocol repeats for many rounds. As illustrated in Fig. 2, in an arbitrary round, the proposed FUI protocol employs an indicator vector  $V$  to help filter out some unknown tags and label them. Specifically, based on the indicator vector  $V$ , some unknown tags can be labeled. The more rounds are repeated, the more unknown tags are expected to be filtered out and labeled. In the following subsections, we further investigate how to optimize the parameters including the optimal length  $v$  of indicator vector  $V$  and the minimum round count  $c$  to maximize the performance of the proposed FUI protocol and guarantee the desired accuracy  $\alpha$ . The labeled RFID tags are to be collected by employing the classical EDFSA technique. In what follows, we present the details of the two phases.

1) *Unknown Tags Filtering phase*: The reader gets full knowledge of the known tag IDs stored in the central database. As illustrated in Fig. 2 (a), the reader maps the IDs of all known tags to an indicator vector  $V$  with  $v$  bits. Specifically, an arbitrary known tag  $t_i$  with the  $ID_i$  is mapped to the  $\gamma_i^{th}$  bit in the vector  $V$  based on a uniform Hash function  $H(\cdot)$ , where  $\gamma_i = H(ID_i, R) \bmod v$  ( $R$  is a random seed number that is fresh in each round). And tag  $t_i$  refers to the  $\gamma_i^{th}$  bit as its *representative* bit. If one or more known RFID tags are mapped to the  $\gamma_i^{th}$  bit in the indicator vector  $V$ , this bit is set to '1'. In contrary, if none of the known RFID tags is mapped to the  $\gamma_i^{th}$  bit, this bit is set to '0'. Then the reader broadcasts the used parameters including random seed  $R$ , the vector length  $v$  as well as the indicator vector  $V$  to all the RFID tags (both known tags and unknown tags). Obviously, the locations

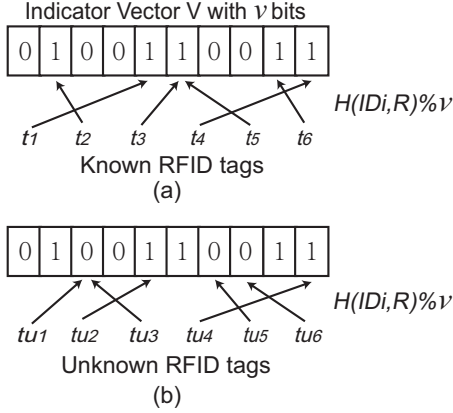


Figure 2. The basic tactic that FUI employs to filter out unknown RFID tags. (a) Mapping all the known tag IDs to the bits in the indicator vector  $V$ ; (b) Based on  $V$ , some unknown RFID tags (e.g.,  $tu_1$ ,  $tu_3$ ,  $tu_5$  and  $tu_6$ ) can be filtered out and labeled in this round.

of ‘1s’ in the indicator vector  $V$  indicate the mapping distribution of known RFID tags. Each tag (including known tags and unknown tags) employs the Hash function  $H(\cdot)$ , the received random seed  $R$  and the length  $v$  of indicator vector  $V$  to calculate its *representative* bit, specifically, using  $H(ID_i, R) \bmod v$  as the index of its *representative* bit. It is easy to get the following conclusions: (i) the representative bits of the known RFID tags in the indicator vector  $V$  are ‘1s’; (ii) the representative bits of some unknown RFID tags are ‘1s’, when they select the bits that one or more known RFID tags are mapped to; (iii) the representative bits of some unknown RFID tags are ‘0s’, when they select the bits that none of the known RFID tags are mapped to. During the process of receiving the indicator vector  $V$ , each tag checks its representative bit in the vector  $V$ . For an arbitrary tag, if it finds that its representative bit turns out to be ‘0’, it asserts that it is an unknown tag and labels itself. In a round, some unknown RFID tags can be labeled. As illustrated in Fig. 2(b), the unknown tags  $tu_1$ ,  $tu_3$ ,  $tu_5$  and  $tu_6$  can be filtered out and labeled, because their representative bits are ‘0s’. The more rounds are repeated, the more unknown tags are expected to be filtered out and labeled. This filtering phase is repeated until the desired accuracy  $\alpha$  is achieved.

2) *Unknown Tags Collecting phase*: Only the labeled RFID tags (i.e., unknown tags) participate in this ID collecting phase. The same as BUIP-CF protocol, we also employ the classical and well-known *Tag Collecting* technique to collect the labeled unknown tags.

Note that the indicator vector  $V$  in the *Unknown Tags Filtering* phase may be too long to be transmitted in a *tag slot* (i.e.,  $t_{tag}$ ). To address this problem, the long indicator vector  $V$  is divided into multiple segments of 96-bits to be sequentially transmitted in multiple *tag slots*.

The *Unknown Tags Filtering* phase is of highlight in this study, so we mainly focus on optimizing the parameters

involved in the first phase, specifically, the optimal length  $v$  of indicator vector  $V$  and the minimum round count  $c$ .

### C. Determining the Optimal Length $v$ of Indicator Vector

In an arbitrary round  $k$  of the *Unknown Tags Filtering* phase, a certain unknown RFID tag can be labeled when it selects a bit in  $V$  that no known RFID tag is mapped to. Since a uniform Hash function is used, the probability that a tag is mapped to a certain bit is  $\frac{1}{v}$ . The probability that a certain unknown RFID tag can be filtered out and labeled in the  $k^{th}$  round is denoted as  $P_k$ , which can be given by

$$P_k = \binom{v_k}{1} \cdot \frac{1}{v_k} \cdot \left(1 - \frac{1}{v_k}\right)^N \approx e^{-\frac{N}{v_k}}, \quad (1)$$

where  $N$  is the number of unknown RFID tags and  $v_k$  is the length of indicator vector  $V$  in the  $k^{th}$  round. Given that  $v_k$  is normally very large,  $P_k$  can be simplified to  $e^{-\frac{N}{v_k}}$  in Eq. 1, where  $e$  is the *natural constant*. And Eq. (1) indicates that an arbitrary unknown RFID tag can be filtered out with a probability of  $P_k$ . We denote the number of unknown tags as  $M$ . Each of the  $M$  unknown tags is expected to be labeled with the probability  $P_k$  in this round. Hence, the number of unknown tags that are labeled in this round is expected to be  $M \cdot P_k$ , i.e.,  $M \cdot e^{-\frac{N}{v_k}}$ . Then we discuss the *time cost* of this round. To begin with, it takes the proposed FUI protocol one *tag slot* (i.e.,  $t_{tag}$ ) to transmit the random seed  $R$  and the length  $v$  of the indicator vector  $V$ . Furthermore, the time for transmitting the indicator vector  $V$  with length of  $v_k$  is  $\lceil \frac{v_k}{96} \rceil \cdot t_{tag}$ , where 96 is the length of a segment of  $V$ . Therefore, the *time cost* of this round is  $t_{tag} + \lceil \frac{v_k}{96} \rceil \cdot t_{tag}$ . In this round, the average time  $T_{avg}$  for labeling an unknown tag can be given by

$$T_{avg} = \frac{t_{tag} + \lceil \frac{v_k}{96} \rceil \cdot t_{tag}}{M \cdot e^{-\frac{N}{v_k}}} \approx \frac{\frac{v_k}{96} \cdot t_{tag}}{M \cdot e^{-\frac{N}{v_k}}} \quad (2)$$

Note that, since the indicator vector  $V$  is normally very long (i.e., the vector length  $v_k$  is very large),  $t_{tag} + \lceil \frac{v_k}{96} \rceil \cdot t_{tag}$  is simplified to  $\frac{v_k}{96} \cdot t_{tag}$  in Eq. (2). Obviously, we need to minimize  $T_{avg}$  thereby getting the best time-efficiency. Hence, we get its derivative as follows:

$$\begin{aligned} & \frac{d(T_{avg})}{d(v_k)} \\ &= \frac{M \cdot e^{-\frac{N}{v_k}} \cdot \frac{t_{tag}}{96} \cdot \left(1 - \frac{N}{v_k}\right)}{\left(M \cdot e^{-\frac{N}{v_k}}\right)^2} \end{aligned} \quad (3)$$

By setting the derivative in Eq. (3) to *zero*, we obtain  $v_k = N$ . When  $v_k > N$ , this derivative is larger than 0; when  $v_k < N$ , it is smaller than 0. Clearly,  $T_{avg}$  achieves the minimum value when indicator vector length  $v_k$  is set to  $N$ . Therefore, we should set  $v = N$  in each round to achieve the best time-efficiency.

### D. Determining the Minimum Round Count $c$

The unknown tags filtering process is repeated for multiple rounds and the round count is denoted as  $c$ . After  $c$

rounds, the probability that a certain unknown tag can be labeled *at least once* is denoted as  $P$ , which can be given by

$$P = 1 - \prod_{k=1}^c (1 - P_k) \quad (4)$$

Since the indicator vector length  $v_k$  is optimized to  $N$  in each round, then  $P_k$  becomes a constant  $\frac{1}{e}$  and we have:

$$\begin{aligned} P &= 1 - \prod_{k=1}^c (1 - P_k) \\ &= 1 - \left(1 - \frac{1}{e}\right)^c \end{aligned} \quad (5)$$

According to Eq. (5), each unknown tag has the independent probability  $P$  to be labeled at least once. So the ratio of unknown tags that can be labeled is also expected to be  $P$ . If we desire to identify the unknown RFID tags with an accuracy at least  $\alpha$ , we should guarantee that the expected proportion  $P$  of unknown tags that can be labeled is not less than  $\alpha$ , then we have:

$$P = 1 - \left(1 - \frac{1}{e}\right)^c \geq \alpha \quad (6)$$

By solving the above inequality, we get

$$c \geq \log_{1-\frac{1}{e}}(1 - \alpha) \quad (7)$$

That is, to identify the unknown RFID tags with an accuracy  $\alpha$ , the minimum round count  $c$  should be at least  $\lceil \log_{1-\frac{1}{e}}(1 - \alpha) \rceil$ . In the previous subsection, the theoretical analysis shows that the optimal indicator vector length  $v$  is equal to the number  $N$  of known RFID tags. Hence, the execution time of each round is  $t_{tag} + \lceil \frac{v}{96} \cdot t_{tag} \rceil$ . In this subsection, we have investigated the minimum round count  $c$ . Hence, in order to identify the unknown tags with a desired accuracy  $\alpha$ , the total execution time of *Unknown Tags Filtering* phase can be given as

$$T_{Phase1} = (t_{tag} + \lceil \frac{N}{96} \rceil \cdot t_{tag}) \cdot \lceil \log_{1-\frac{1}{e}}(1 - \alpha) \rceil \quad (8)$$

It is significant to find that the execution time of *Unknown Tags Filtering* phase only depends on  $N$  and  $\alpha$ , and is independent on  $M$ , where  $N$  and  $M$  indicate the number of known tags and the number of unknown tags, respectively, and  $\alpha$  is the desired identification accuracy.

#### IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed FUI protocol via simulators which are developed in C++. According to the specification of Philips I-Code system [20],  $t_{tag}$  and  $t_{long}$  are set to  $2.4ms$  and  $0.8ms$  respectively, the detailed analysis has been given in the Section-III. For fair comparison, we adopt the same parameter settings in the proposed FUI and other compared protocols. The same with Ref. [16], we also consider a single reader in the simulations and ignore the channel errors. And each simulation is conducted for 100 times and we get the average value.

##### A. Tag Separation Time

The innovation of both the proposed FUI and BUIP-CF in Ref. [16] is how to efficiently separate the unknown tags and the known tags. Specifically, the BUIP-CF in Ref. [16] investigates an ALOHA-like tactic to deactivate all the known RFID tags and label all the unknown tags. The proposed FUI protocol investigates an indicator vector  $V$  to help to filter out the unknown tags and label them with a desired accuracy  $\alpha$ . Hence, in this subsection, extensive simulations are conducted to evaluate the efficiency of *Unknown Tag Filtering* technique of the proposed FUI protocol (i.e., the efficiency of the first phase). Actually, in Ref. [16], Liu *et al* proposed a series of protocols to identify the unknown RFID tags, including Basic Unknown tag Identification Protocol (BUIP), BUIP with Collision-Empty Slot Pairing (BUIP-CE) and BUIP with Collision-Fresh Slot Pairing (BUIP-CF). Hence, in this subsection, we evaluate the proposed FUI protocol by comparing it with BUIP, BUIP-CE and BUIP-CF. We set the identification accuracy  $\alpha = 99\%$  in the simulations.

1) *M changes* ( $N = 10,000$ ): Fig. 3 (a) illustrates the simulation results with the fixed known tag number  $N$  (keeping 10,000) and the varying unknown tag number  $M$  (varying from 1,000 to 10,000). The simulation results in Fig. 3 (a) manifest that the *Tag Separation Time* of BUIP, BUIP-CE and BUIP-CF increases when the unknown tag number  $M$  increases. Whereas, the *Tag Separation Time* of the proposed FUI technique is independent on unknown tag number  $M$  and almost a constant. The proposed FUI protocol outperforms the the protocols in Ref. [16] by dramatically reducing the *Tag Separation Time*. For instance, when  $M = 5,000$ , the *Tag Separation Time* of BUIP, BUIP-CE and BUIP-CF is  $31.0s$ ,  $28.6s$ , and  $22.7s$ , respectively. The proposed FUI protocol requires only  $2.8s$ , representing 91.0%, 90.2%, and 87.7% reduction of the *Tag Separation Time*, when compared to the BUIP, BUIP-CE and BUIP-CF. Moreover, the more unknown tags exist, the better the FUI technique performs than these three protocols. For example, when  $M = 10,000$ , the *Tag Separation Time* of BUIP, BUIP-CE and BUIP-CF is  $37.1s$ ,  $35.3s$ , and  $28.4s$ , respectively. The proposed FUI protocol still requires only  $2.8s$ , which represents 92.5%, 92.1%, and 90.1% reduction of *Tag Separation Time*.

2) *N changes* ( $M = 2,000$ ): Fig. 3 (b) illustrates the simulation results when keeping the number  $M$  of unknown tags fixed to 2,000 and varying the number  $N$  of known tags from 1,000 to 10,000. As  $N$  increases, the protocols, BUIP, BUIP-CE and BUIP-CF, experience a sharp rise, whereas, the proposed FUI protocol enjoys a steady rise. The simulation results illustrated in Fig. 3 (b) demonstrate that the proposed FUI protocol outperforms BUIP, BUIP-CE and BUIP-CF in terms of *Tag Separation Time*. For instance, when  $N = 8,000$ , the *Tag Separation Time* of BUIP, BUIP-

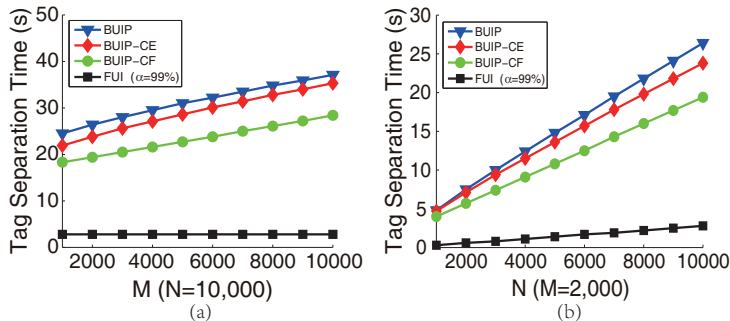


Figure 3. Evaluating the *Tag Separation Time* of different protocols. (a) the number  $M$  of unknown RFID tags changes from 1000 to 10,000 while keeping the number  $N$  of known RFID tags unchanged; (b) the number  $N$  of known RFID tags changes from 1000 to 10,000 while keeping the number  $M$  of unknown RFID tags unchanged.

CE and BUIP-CF is 21.8s, 19.8s, and 16s, respectively. The proposed FUI protocol requires only 2.2s, representing 89.9%, 88.9%, and 86.3% reduction of the *Tag Separation Time*, when compared to the BUIP, BUIP-CE and BUIP-CF.

The main reason for the superiority of the proposed FUI protocol is that the three protocols in Ref. [16] employ ALOHA-like communication mechanism which begets too many time-consuming tag responses in the first phase. However the proposed FUI protocol investigates an efficient indicator vector to label unknown RFID tags thereby removing the time-consuming tag responses.

### B. Identifying Accuracy

In the *Unknown Tags Filtering* phase of the proposed FUI protocol, the tag filtering process is repeated for multiple rounds to achieve the desired accuracy  $\alpha$ . The minimum *Round Count*  $c$  has been investigated in Section IV-D. Specifically, to achieve the identification accuracy  $\alpha$ , the round count  $c$  should be at least  $\lceil \log_{1-e^{-1}}(1-\alpha) \rceil$ . In this subsection, extensive simulations are conducted to investigate the identifying accuracy of the proposed FUI protocol.

As illustrated in Fig. 4, the minimum round count  $c$  varies from 3 to 12. For each certain round count, 100 independent simulations are conducted. This figure shows the *Maximum Accuracy*, *Minimum Accuracy* and *Average Accuracy* for each round count. For instance, when the minimum round count  $c$  is set to 6, among the 100 simulation results, the *Maximum Accuracy* that FUI achieves is 95.1%; the *Minimum Accuracy* that FUI achieves is 92.3%; the *Average Accuracy* of 100 simulation results that FUI achieves is 93.6%. And the more rounds are repeated, the better identification accuracy the proposed FUI achieves. For example, when the minimum round count  $c$  is set to 12, among the 100 simulation results, the *Maximum Accuracy* that FUI achieves is 99.9%; the *Minimum Accuracy* that FUI achieves is 99.3%; the *Average Accuracy* of 100 simulation results that FUI achieves is 99.6%. It is important to find

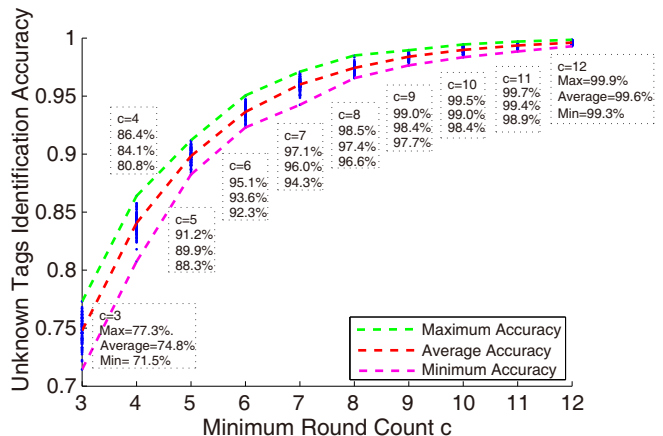


Figure 4. Identification Accuracy of the proposed FUI protocol with respect to the minimum round count  $c$ .

that the three types of identification accuracy (*Maximum Accuracy*, *Minimum Accuracy* and *Average Accuracy*) are of convergence with *Minimum Round Count*  $c$  increasing.

Recall that the round count  $c$  should be at least  $\lceil \log_{1-e^{-1}}(1-\alpha) \rceil$  for achieving the desired identification accuracy  $\alpha$ . For example, if the desired accuracy  $\alpha$  is set to 95%, by calculating  $\lceil \log_{1-e^{-1}}(1-0.95) \rceil$ , we can get that the corresponding minimum round count  $c$  is 7. According to the simulation results in Fig. 4, when the minimum round count  $c$  is 7, the maximum and the average identification accuracy are 97.1% and 96.0% respectively, which can guarantee the desired accuracy 95%. However, the minimum identifying accuracy is 94.3% that is a bit less than the desired accuracy  $\alpha = 95\%$ . The reason for this deviation between the simulation and theoretical results is due to the variance inherent from the probabilistic methods. Hence, we can combine the theoretical analysis and the simulation results in Fig. 4 to determine the appropriate round count for achieving the specified identification accuracy  $\alpha$ .

### C. Total Execution Time

In this subsection, we evaluate the total execution time of the proposed FUI protocol. Besides the state-of-the-art protocols proposed in Ref. [16], the continuous scanning protocol proposed in Ref. [12] can also identify the unknown tags. Hence, in this section, we compare the proposed FUI protocol with the continuous scanning protocol CU and the state-of-the-art protocols investigated in Ref. [16]. In the simulations, we set the identification accuracy  $\alpha$  to 99%. and fix the known tag number  $N = 10,000$  and vary the unknown tag number  $M$  from 1000 to 10,000. Extensive simulation results, as illustrated in Table II, demonstrate that the proposed FUI tactic outperforms the CU (99%), BUIP, BUIP-CE and BUIP-CF protocols. For instance, when  $M = 5000$ , the total execution time of CU (99%), BUIP, BUIP-CE, and BUIP-CF is 93.5s, 67.0s, 64.6s, and 58.7s,



Table II  
TOTAL EXECUTION TIME OF PROTOCOLS WHEN M CHANGES ( $N = 10,000$ ).

Alg. Name	Total Execution Time (s)									
	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
CU (99%)	50.5	56.9	68.8	80.7	93.5	106.4	118.5	131.9	137.3	150.1
BUIP	31.7	40.8	49.6	58.3	67.0	75.4	83.9	92.4	100.7	109.1
BUIP-CE	29.1	38.2	47.0	55.9	64.6	73.3	81.8	90.4	98.8	107.3
BUIP-CF	25.5	33.8	42.1	50.4	58.7	67.0	75.4	83.7	92.0	100.4
FUI ( $\alpha = 99\%$ )	9.9	17.1	24.2	31.3	38.4	45.6	52.7	59.8	67.0	74.1

respectively. Furthermore, the proposed FUI protocol requires only 38.4s, representing 58.9%, 42.7%, 40.6%, and 34.6% reduction of the total execution time.

## V. CONCLUSION

This paper studies the important problem of identifying the unknown tags in large-scale RFID systems. Specifically, we aim at identifying a fraction of the unknown tags with a desired accuracy  $\alpha$ . The solutions to this problem are desirable in many valuable applications, such as identifying the unregistered items due to a management fault. This paper investigates an indicator vector to efficiently filter out and label the unknown tags and proposes a Fast Unknown tag Identification (FUI) protocol, in which the time-consuming tag responses inherent from the traditional ALOHA communication mechanism is avoided. Then the classical EDFSA protocol is employed to collect these labeled unknown tags. More importantly, this paper theoretically presents how to choose the optimal parameters of the FUI protocol, i.e., optimizing the indicator vector length  $v$  and the minimum round count  $c$ . Furthermore, extensive simulation experiments have been conducted to evaluate the performance of the FUI protocol and the results demonstrate that the proposed FUI considerably outperforms the state-of-the-art protocols.

## ACKNOWLEDGMENT

This work was supported by NSFC (Grant No.s 60973117, 61173160, 61173162, and 60903154), New Century Excellent Talents in University (NCET) of Ministry of Education of China, and the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61225010).

## REFERENCES

- [1] C. C. Tan, B. Sheng, and Q. Li, "Landmarc: Indoor location sensing using active rfid," *Proc. of IEEE PerCom*, 2003.
- [2] C. Wang, H. Wu, and N.-F. Tzeng, "RFID-Based 3-D Positioning Schemes," *Proc. of IEEE INFOCOM*, 2007.
- [3] A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra, "Tuning to Your Position: FM radio based Indoor Localization with Spontaneous Recalibration," *Proc. of IEEE PerCom*, 2010.
- [4] A. Nemmaluri, M. D. Corner, and P. Shenoy, "Sherlock: Automatically Locating Objects for Humans," *Proc. of ACM MobiSys*, 2008.
- [5] L. Xie, B. Sheng, C. C. Tan, H. Han, Q. Li, and D. Chen, "Efficient Tag Identification in Mobile RFID Systems," *Proc. of IEEE INFOCOM*, 2010.
- [6] B. Sheng, C. C. Tan, Q. Li, and W. Mao, "Finding Popular Categories for RFID Tags," *Proc. of ACM MobiHoc*, 2008.
- [7] C. Lee and C. Chung, "Efficient Storage Scheme and Query Processing for Supply Chain Management using RFID," *Proc. of ACM SIGMOD*, 2008.
- [8] S. Chen, M. Zhang, and B. Xiao, "Efficient Information Collection Protocols for Sensor-augmented RFID Networks," *Proc. of IEEE INFOCOM*, 2011.
- [9] R. Zhang, Y. Liu, Y. Zhang, and J. Sun, "Fast Identification of the Missing Tags in a Large RFID System," *Proc. of IEEE SECON*, 2011.
- [10] M. Kodialam, T. Nandagopal, and W. C. Lau, "Anonymous Tracking using RFID tags," *Proc. of IEEE INFOCOM*, 2007.
- [11] C. C. Tan, B. Sheng, and Q. Li, "Efficient Techniques for Monitoring Missing RFID Tags," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1882–1889, 2010.
- [12] B. Sheng, Q. Li, and W. Mao, "Efficient Continuous Scanning in RFID systems," *Proc. of IEEE INFOCOM*, 2010.
- [13] K. Bu, B. Xiao, Q. Xiao, and S. Chen, "Efficient Pinpointing of Misplaced Tags in large RFID Systems," *Proc. of IEEE SECON*, 2011.
- [14] W. Luo, S. Chen, T. Li, and S. Chen, "Efficient Missing Tag Detection in RFID Systems," *Proc. of IEEE INFOCOM*, 2011.
- [15] K. Bu, B. Xiao, Q. Xiao, and S. Chen, "Efficient misplaced-tag pinpointing in large RFID systems," *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [16] X. Liu, S. Zhang, K. Bu, and B. Xiao, "Complete and Fast Unknown Tag Identification in Large RFID Systems," *Proc. of IEEE MASS*, 2012.
- [17] S. Lee, S. Joo, and C. Lee, "An Enhanced Dynamic Framed Slotted ALOHA Algorithm for RFID Tag Identification," *Proc. of IEEE MobiQuitous*, 2005.
- [18] Y. Qiao, S. Chen, T. Li, and S. Chen, "Energy-efficient Polling Protocols in RFID Systems," *Proc. of ACM MobiHoc*, 2011.
- [19] T. Li, S. Chen, and Y. Ling, "Identifying the Missing Tags in a Large RFID System," *Proc. of ACM MobiHoc*, 2010.
- [20] P. Semiconductors, "I-CODE Smart Label RFID Tags," [http://www.nxp.com/acrobat\\_download/other/identification/SL092030.pdf](http://www.nxp.com/acrobat_download/other/identification/SL092030.pdf), Jan 2004.