Time- and Energy-efficient Detection of Unknown Tags in Large-scale RFID Systems

Xiulong Liu*, Heng Qi*, Keqiu Li*§, Yanming Shen*, Alex X. Liu† and Wenyu Qu‡
*School of Computer Science and Technology, Dalian University of Technology, China
†Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, U.S.A.
‡School of Information Science and Technology, Dalian Maritime University, China
§Corresponding Author: (keqiu@dlut.edu.cn)

***Abstract*—Radio Frequency Identification (RFID) technology is widely used in the the retail, warehouse and supply chain management. However, unknown RFID tags appear when the unregistered tagged objects are moved in or tagged objects are misplaced, which leads to huge economic losses (e.g., misplaced chilled food in a warehouse may quickly decay). This paper studies the practically important problem of unknown tag detection. To the best of our knowledge, this is the first piece of work taking both time-efficiency and energy-efficiency into consideration, where the energy-efficiency is very important when the battery-powered active tags are used. This paper proposes two efficient protocols to address the problem of unknown tag detection. Specifically, the Basic Unknown Tag Detection (B-UTD) protocol leverages a cost-effective filter vector to detect the unknown tags, based on which we then propose a Sampling based Unknown Tag Detection (S-UTD) protocol by adopting the well-known sampling idea. We present theoretical analysis to optimize the performance of the proposed protocols. Extensive simulations are conducted to evaluate the performance of the proposed protocols. And the experimental results show that the proposed S-UTD protocol considerably outperforms the most related protocol by reducing more than 90% of the required execution time and energy consumption.**

I. Introduction

Radio Frequency Identification (RFID) technology is widely used in the the localization [1]–[3], warehouse monitoring [4]–[6] and supply chain management [7]–[9] because of its many attractive advantages over the conventional barcode systems, such as remote and multiple access, simple computational ability, and non-sight limitation, etc. However, the *unknown tags* appear when the unregistered tagged objects are moved in a warehouse or the tagged objects are misplaced, which causes the huge economic losses. For example, the chilled food will quickly decay if it is misplaced at the zone with no refrigeration equipment and not discovered in time. And if the unregistered items are mixed with the registered ones, they can never be managed because the back-end server does not have their information (e.g., shelf life). From the view of the reader locating at these zones, both the misplaced tags and the unregistered tags are *unknown tags*. The corresponding problem is the so called Unknown Tag Problem (*UTP*), which can be generally classified into two sub-problems: one is the unknown tag detection problem, which concentrates on how to efficiently detect the unknown tag event with a predefined accuracy; the other one is the unknown tag identification problem, which aims to exactly pinpoint which tags are unknown tags. Intuitively, the unknown tag identification protocols tend to report more information than the unknown tag detection protocols, but consequently consume more time and energy. Therefore, these two types of protocols possess their own superiority and are complementary to each other. Latter, we we will present a proper way to jointly use the detection protocol and the identification protocol.

This paper studies the problem of unknown tag detection. Although it is of great practical importance, to the best of our knowledge, no work has been done on solving the this problem. The most related literature is the latest unknown tag identification protocol—Basic Unknown tag Identification Protocol-with Collision-Fresh slot paring (BUIP-CF) protocol proposed by Liu *et al.* in [10]. It aims to identify the exact unknown tags that appear in a given RFID system. The fundamental principle of BUIP-CF is briefed in the following.

BUIP-CF consists of two phases: (1) *known tags deactivation* phase; and (2) *unknown tags collection* phase. In the first phase, a variant of Framed Slotted ALOHA protocol is used to deactivate all the known tags and label the unknown tags. Specifically, because the reader gets full knowledge of ID information of all known tags, it is able to predict the expected status of each slot if there are no unknown tags: (1) expected empty slots, in which no known tag responds; (2) expected singleton slots, in which one and only one known tag responds; (3) expected collision slots, in which more than one known tags respond. In fact, the unknown tags also participate in this process, which makes it possible that the actual status of a slot different from its expected status. (1) if one and only one tag responds in an expected singleton slot, this tag must be known tag. Then, the reader sends an ACK signal to deactivate it (i.e., telling it to enter the sleep state); (2) if one or more tags respond in an expected empty slot, they must be all unknown tags. then the reader sends a NACK signal to label them (i.e., telling them not to participate in the next round, but still keep active). In a round, some known/unknown tags could be deactivated/labeled, and they will not participate in the following rounds; the other tags will participate in the

978-0-7695-5104-3/13 $26.00 © 2013 IEEE
DOI 10.1109/MASS.2013.75

next round; this process is repeated for multiple rounds until all known tags are deactivated. As a result, the remaining active tags are all unknown ones and will be collected by the classical Enhanced Dynamic Framed Slotted ALOHA (EDFSA) protocol [11] in the next *unknown tags collection* phase.

BUIP-CF is able to identify all the unknown tags but is seriously time-consuming when the number of known tags is very large because it needs to deactivate all the known tags. Moreover, the energy-efficiency is not taken into consideration in [10]; the frequent execution of the BUIP-CF seriously shortens the lifetime of the active tags. In reality, the appearance of unknown tags has two features: (1) *sparse*: the unknown tags do not always appear but only occasionally. For example, given a 100-hour period of time, the unknown tags may only appear at the 10^{th} hour; the other feature is (2) *random*: we cannot predict when the unknown tags appear. As aforementioned, the appearance of unknown tags often leads to serious consequences (e.g., security threats, economic losses, etc.). Although the appearance of the unknown tags is sparse, but because of its random nature, we have to frequently run the BUIP-CF in order to timely discover/identify the unknown tags, e.g., invoking the BUIP-CF protocol every hour. Clearly, among the 100 runs, only the 10^{th} execution identifies the unknown tags, but the other 99 runs are vain, and thus waste a lot of time. Only if a tiny unknown tag detection protocol could judge whether the unknown tags appear in prior. If the detection protocol does not discover any unknown tags, the heavy identification protocol will not be blindly executed. In other words, an unknown tag identification protocol (e.g., BUIP-CF) is invoked only when the detection protocol returns an unknown tag alarm. Then, we only needs to execute 100 times tiny detection protocol and just 1 time heavy identification protocol, instead of 100 times heavy identification protocol. Obviously, it is necessary to jointly using the detection protocol and the identification protocol in practice.

Unfortunately, no literature is found to provide an efficient solution to the problem of unknown tag detection. To fill this gap, we first propose a Basic Unknown Tag Detection (B-UTD) protocol which uses a cost-effective filter vector to detect the unknown tags. To further improve its efficiency, we introduce the well-known sampling idea [12], [13] into B-UTD, and thus propose a Sampling based Unknown Tag Detection (S-UTD) protocol, in which *only* the sampled tags participate in the detection process. This paper also presents the theoretical analysis of system parameters in order to optimize the performance of the proposed protocols. Extensive simulations are conducted to evaluate the performance of the proposed protocols. And the experimental results show that the proposed S-UTD protocol considerably outperforms the most related protocol (i.e., BUIP-CF) by reducing more than 90% of the required execution time and the energy consumption.

The rest of this paper is organized as follows. The previous related work is reviewed in Section II. Section III describes the problem to be addressed in this paper and presents the system model. We propose the B-UTD protocol and the S-UTD protocol in Section IV and Section V, respectively. In Section VI, extensive simulation experiments are conducted to evaluate the performance of the proposed protocols. This paper is concluded in Section VII.

II. Related Work

As aforementioned, RFID is an emerging technology that is widely used in many monitoring applications, where there are two basic RFID research problems soliciting efficient solutions: Missing Tag Problem and Unknown Tag Problem.

In recent years, many efforts have been made to address the missing tag problem. Tan *et al.* proposed the Trust Reader Protocol (TRP) to detect the missing-tag event with a predefined probability α when the number of the missing tags exceeds m [14]. To improve the time-efficiency and energy-efficiency of TRP, Luo *et al.* introduce the sampling idea, and thus propose the Efficient Missing-tag Detection (EMD) protocol, where they use the detection result on the sampled tags to probabilistically reflect the whole intactness [12]. This work inspires us to introduce the sampling idea to efficiently address the problem of unknown tag detection. However, EMD still has a large room to be improved because it contains a large proportion of expected empty/collision slots that cannot be used in missing tag detection. To overcome this deficiency, Luo *et al.* studied a multi-hashing-seed approach to reduce the useless empty slots and collision slots involved in the EMD protocol and thus proposed the Multi-Seed Missing-tag Detection (MSMD) protocol [15]. These protocols (i.e., TRP [14], EMD [12] and MSMD [15]) concentrate on discovering the missing-tag event, instead of pinpointing which tags are missing. The Iterative ID-free Protocol (IIP) proposed in [16] is a variant of the Framed Slotted Aloha protocol, and is able to pinpoint the exact missing tags. The protocols proposed by Zhang *et al.* in [17] accelerate the protocol's execution by leveraging the collaboration of multiple readers.

The unknown tag problem is also practically important in reality, because these unknown tags appear when unregistered tagged objects are moved in or tagged objects are misplaced. Despite of its importance, the unknown tag problem is still under investigated. A straightforward solution is to collect the IDs of all tags including known ones as well as unknown ones [18], [19]. By comparing the collected IDs with the known ID set stored in a database, we are able to recognize the unknown tags. The obvious drawback is that we re-collect a large number of known tag IDs that are already in the database. To overcome this deficiency, Liu *et al.* [10] investigated a series of variant protocols of the Framed Slotted Aloha to separate the known tags and unknown tags, then we can collect the unknown tag IDs only. However, blindly executing the BUIP-CF is of low time-efficiency and energy-efficiency, which has been exemplified in Section I. New efficient unknown tag detection protocols are solicited to jointly tackle the unknown tag problem.

III. PRELIMINARIES

A. System Model and Assumption

We assume a large-scale RFID system that consists of a back-end server, a reader, N known tags, and M unknown tags. Let us denote T_Δ as the known tag set, i.e., $T_\Delta = \{t_1, t_2, \ldots, t_i, \ldots, t_N\}$, and the number N as well as ID information in T_Δ is available in a database of the back-end server. The unknown tag set is denoted as T_Λ, i.e., $T_\Lambda = \{tu_1, tu_2, \ldots, tu_i, \ldots, tu_M\}$, whereas, both the number M and the specific ID information in T_Λ are not known in prior. Each tag has a unique ID and is equipped with the same *uniform* Hash generator $H(\cdot)$. We assume the reader has adequate power to interrogate all the tags including the known ones and the unknown ones. Moreover, the reader communicates with the back-end server via a high-rate network link, and has access to the ID information of all known tags. The communication delay between the reader and back-end server is so small that it is negligible.

B. Slots

The reader communicates with the tags in a time-slotted way, where the slots are synchronized by the '*end slot*' commands broadcasted by the reader. Li *et al.* presented a method of classifying the time slots based on their length in [16], which is widely accepted. Specially, the slots are classified into *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 1 bit information. According to the specification of the Philips I-Code system [20], the wireless transmission rate from a tag to a reader is 53 Kb/s and the rate from a reader to a tag is 26.5 Kb/s. And any two consecutive transmissions (from a tag to a reader or vice versa) are separated by a waiting time of 302 us. Hence, t_{tag} is set to be 2.4 ms for transmission of a tag ID (96 bits) from a tag to a reader or vice versa. Similarly, t_{long} and t_{short} are set to be 0.8 ms and 0.4 ms, respectively.

C. Energy Consumption Model

Because the battery of a reader can be easily recharged or the reader may even use an external power source [15], the energy consumed by the reader is of less concern in this paper. Accordingly, we only care the energy consumption of the battery-powered active tags, particularly, the known tags. During a slot, an active tag has two types of states: *awake* state and *sleep* state [21]. Specifically, a tag needs to be in the awake state (i.e., its CPU operates at full energy and the radio keeps open) for communication. An awake tag may operate one of the three actions during a certain slot: transmitting data to the reader; receiving data from the reader; or just listening the channel for the periodical 'end-slot' commands broadcasted by the reader. Since the radio scanning consumes most of the energy, the above actions of an awake tag almost consume the same amount of energy. Let ω denote the energy consumption of an awake tag during a tag slot. According to the knowledge presented in the previous subsection, the length of a short-response slot is $\frac{1}{6}$ of the tag slot, then the energy consumption of an awake tag during the short-response slot is about $\frac{1}{6}\omega$. Because of the similar reason, the energy consumed in a long-response slot is about $\frac{1}{3}\omega$. In order to conserve battery power, the tag can enter the sleep state, where the CPU works in a low power mode and radio reception is disabled. The ratio of energy consumed between the awake and sleep states is typically on the order of 100 or more, so we neglect the energy consumption of an asleep tag. We use E to denote the energy consumption of N known tags, which is given as $E = \sum_{i=1}^{N}[\eta_{i1} \cdot (\frac{1}{6}\omega) + \eta_{i2} \cdot (\frac{1}{3}\omega) + \eta_{i3} \cdot \omega]$, where η_{i1}, η_{i2} and η_{i3} indicate the number of the short-response slots, the long-response slots and the tag slots that the tag t_i keeps awake for, respectively.

D. Problem Statement

This paper aims to detect the unknown tag event with a predefined accuracy, which is measured by two system parameters: a tolerance threshold m and a confidence level α. The problem addressed in this paper could be formally defined as: we desire to discover the unknown tag event with a probability of at least α when m (or more) unknown tags hide in the RFID systems. For example, $< m = 1, \alpha = 99.9\% >$ means that we want to discover the unknown tag event with a high probability of 99.9% even if there is just one unknown tag. The used notations are summarized in Table I.

E. Performance Metrics

In this paper, we consider two important performance metrics: (1) execution time is the most important performance metric for a unknown tag detection protocol because the shorter the execution time is, the sooner we will discover the unknown tag event and then take timely countermeasures (e.g., replacing the chilled food to the zone equipped with freezers). (2) energy consumption is the second important metric when battery-powered active RFID tags are used to support advanced RFID applications that cover a large area. It is necessary to take the energy-efficiency into consideration so as to prolong the lifetime of active tags. Note that,

Table I
THE USED NOTATIONS.

Symbols	Descriptions
ID	ID information of an RFID tag
N	number of known tags in the system
M	number of unknown tags in the system
m	tolerance threshold
α	detection confidence level
T_Δ	the set of all known tags in the system
T_Λ	the set of all unknown tags in the system
t_{tag}	length of a tag slot
t_{long}	length of a long-response slot
t_{short}	length of a short-response slot
ω	energy consumption of a awake tag during a tag slot
V	filter vector involved in both of the B-UTD and the S-UTD
v	length of the filter vector V
v_{up}	up bound of the filter length
R	random seed which is fresh in every execution
$H(\cdot)$	Hash generator that follows a uniform random distribution
X	a large constant pre-configured in the RFID tags
E_{B-UTD}	energy cost of the B-UTD protocol
T_{B-UTD}	time cost of the B-UTD protocol
E_{S-UTD}	energy cost of the S-UTD protocol
T_{S-UTD}	time cost of the S-UTD protocol
$P(N, M, v)$	detection probability of the proposed B-UTD protocol
$P(N, M, p, v)$	detection probability of the proposed S-UTD protocol
p	sampling probability involved in S-UTD protocol
p_{min}	the minimum sampling probability that can be used by S-UTD
p_t	sampling probability that minimizes the time cost of S-UTD
p_e	sampling probability that minimizes the energy cost of S-UTD

the protocols proposed in this paper are not limited to the active tags although this paper takes both time-efficiency and energy-efficiency into consideration. If only passive tags are used in an application, the time-efficiency becomes the only performance metric.

IV. A BASIC UNKNOWN TAG DETECTION PROTOCOL

In this section, we first present the overview of our method, which is followed by the detailed design of the proposed Basic Unknown Tag Detection (B-UTD) protocol. We then investigate how to configure the system parameters to meet the predefined detection accuracy. Finally, we analyze the performance (i.e., the time cost and the energy cost) of the B-UTD.

A. Principle Overview

A Bloom Filter [22]–[24] is a well-known data structure that probabilistically represents a set of n elements $Y = \{y_1, y_2, \cdots, y_n\}$, which can be used to test set membership. Specially, the Bloom filter compresses this set into a filter vector with w bits by hashing each element in Y into the vector using k hashing functions $h_1, h_2, \cdots, h_k$. A bit in the vector is set to '1' if at least one element is hashed to that index in the vector. When checking whether a given element y belongs to the set Y, we compute $h_1(y), h_2(y), \cdots, h_k(y)$ and assert $y \in S$ if and only if all these k bits are '1s' in the vector; otherwise, $y \notin Y$. Recall the problem of unknown tag detection, essentially, it is equivalent to test if there are new (unknown) tags whose IDs are not in the database. Intuitively, we could compress the IDs of all N known tags into a filter vector, then use the constructed filter to test the identities (known or unknown) of all the tags.

The cost of RFID tags must be remarkably low because of two main reasons: (1) RFID applications are usually large scale, containing a large number (e.g., hundreds of thousands) of tags; (2) the cost of some tagged items are very low. Lower cost implies less hardware [25]. Hence, this paper aims to propose a *lightweight* solution to the problem of unknown tag detection, which is easily accommodated by not only the active tags but also the low-cost passive tags. To reduce the computation overhead on the tag side, we use a special case of Bloom filter, where $k = 1$, i.e., each tag only needs to perform 1 hashing operation when testing its identity. The detailed design of the proposed protocol is presented in the next subsection.

B. Protocol Design

The proposed B-UTD protocol consists of two stages: (1) the *Identity Verification* stage; and (2) the *Unknown Tag Reporting* stage. In the first stage, the reader broadcasts a filter vector V, based on which each tag verifies its identity (a known tag or an unknown tag). In the second stage, the tag that realizes its unknown identity responds an announcement (1-bit information is enough), by which the reader can discover the unknown tag event. If *at least one* unknown tag realizes its unknown identity, the returned detection result is *YES*; otherwise, *NO*. In what follows, we describe the process of these two stages in detail.

1) Identity Verification stage: As aforementioned, the reader has access to a database that stores the ID information of all known tags. As illustrated in Fig. 1, the reader "compresses" (maps) the IDs of N known tags to a filter vector V with v bits, using a uniform hash function $H(\cdot)$. Specifically, an arbitrary known tag t_i with the ID_i is mapped to the l_i^{th} bit in the vector V, where $l_i = H(ID_i, R) \mod v$, and R is a random seed. The l_i^{th} bit is called the *representative* bit of tag t_i. An arbitrary bit in the filter vector V is set to '1', if and only if at leaset one tag is mapped to it; otherwise, it is set to '0'. Then the reader broadcasts the used parameters including the random seed R, the vector length v as well as the filter vector V to all the RFID tags. Obviously, the '1s' in the filter vector V reflect the mapping distribution of known RFID tags. Each tag uses the same hash function $H(\cdot)$, the received random seed R and the filter vector length v to calculate its *representative* bit, specifically, using $H(ID_i, R) \mod v$ as the index of its *representative* bit. During the process of receiving the filter vector V, each tag checks its representative bit in the vector V to judge if its identity is known to the reader or not. Some unknown tags will successfully realize their unknown identity when they find they select '0s'; whereas, the other unknown tags will not realize their unknown identity because they select '1s'. The case that known tags are determined as unknown ones

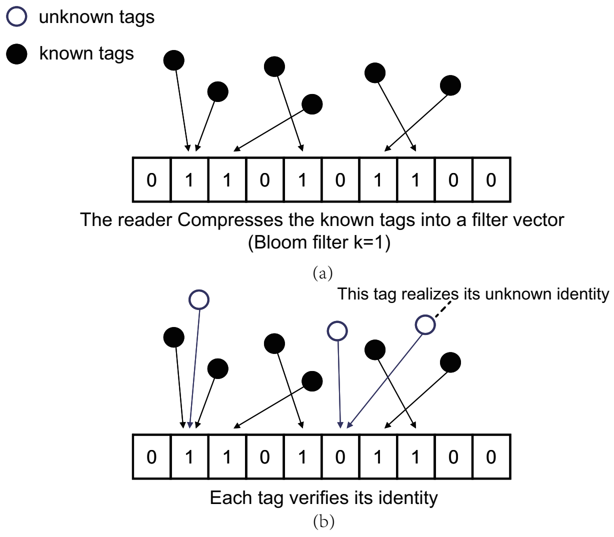

○ unknown tags
● known tags
0 1 1 0 1 0 1 1 0 0
The reader Compresses the known tags into a filter vector (Bloom filter k=1)
(a)
This tag realizes its unknown identity
0 1 1 0 1 0 1 1 0 0
Each tag verifies its identity
(b)

Figure 1. The basic principle of the proposed protocol.

will not occur, because the representative bits of the known RFID tags in the filter vector V are always '1s'.

2) Unknown Tag Reporting stage: The tags realizing their unknown identity frankly announce their unknown identity[1]. If the reader receives any announcement (no matter singleton or collision) after broadcasting the filter vector, it asserts that there exist unknown tags—the unknown tag event is discovered.

As illustrated in Fig. 1, two unknown tags could realize their unknown identity, because they select the '0s'. Intuitively, a large filter length creates more '0s' in V, under the fixed known tag number N, which increases the probability to discover the unknown tags. When M $(\geq m)$ unknown tags exist in an RFID system, it is necessary to investigate the configuration of the filter length v to guarantee the probability of successfully detecting the existing unknown tag event is larger than the required confidence level α. The corresponding theoretical analysis of the system parameter v is presented in the next subsection.

The filter vector V is divided into multiple segments of 96-bits to be sequentially transmitted in multiple *tag slots* [16] when its length is larger than 96 bits (the most bits that can be transmitted in a tag slot).

C. Investigating the Configuration of the Filter Length v

Recall the detection accuracy: *detecting the unknown tag event with more than α probability when there are m or more unknown tags.* We use $P(N, M, v)$ to denote the probability of successfully discovering the unknown tag event when there are N known tags and exactly M unknown tags, and the filter vector length is v. The problem of configuring the parameter v is formulated as follows:

$$\begin{aligned} &Find\ v \\ &s.t.\ \forall M \geq m\ P(N, M, v) \geq \alpha \end{aligned} \tag{1}$$

Theorem 1: Given N, M, and v, we have $P(N, M, v) = 1 - (1 - p_0)^M$, where $p_0 = e^{-\frac{N}{v}}$.

Proof: For a certain one of the M unknown tags, the probability that it can realize its unknown identity is equal to the probability that it selects a '0' (i.e., none of the N known tags is mapped to this bit) in filter vector V. We denote this probability as p_0, it can be given as

$$p_0 = \binom{v}{1} \times \frac{1}{v} \times (1 - \frac{1}{v})^N \approx e^{-\frac{N}{v}}$$

The probability $P(N, M, v)$ that this unknown tag event can be discovered is equal to the probability that *at least one* of the M unknown tags selects '0' in V. Hence, $P(N, M, v)$ can be given as

$$P(N, M, v) = 1 - (1 - p_0)^M,$$

where $p_0 = e^{-\frac{N}{v}}$. ■

Theorem 2: Given N, m and α, if v is set to be larger than $-N/(ln[1 - (1 - \alpha)^{\frac{1}{m}}])$, the detection objective presented in (1) holds.

Proof: Clearly, $P(N, M, v)$ is a monotonically increasing function with respect to M, then we have $P(N, M, v) \geq P(N, m, v)$, $\forall M \geq m$. Hence, for any value of v, if $P(N, m, v) \geq \alpha$, we have $P(N, M, v) \geq P(N, m, v) \geq \alpha$, i.e., the detection objective presented in (1) is satisfied. By solving the inequality $P(N, m, v) \geq \alpha$, we have $v \geq -N/(ln[1 - (1 - \alpha)^{\frac{1}{m}}])$. That is, any value of v larger than $-N/(ln[1 - (1 - \alpha)^{\frac{1}{m}}])$ can satisfy the predefined detection accuracy in (1).

■

D. Analyzing the Performance of B-UTD

In this subsection, we analyze the time-efficiency as well as the energy-efficiency of the proposed B-UTD protocol.

1) Time-efficiency: The time consumed by the computing (both on the reader side and on the tag side) is so minor that it can be neglected when compared with the time consumed by the wireless data transmission. Therefore, we only consider the time consumed by the wireless communication between the reader and the tags.

In the *Identity Verification* stage. One tag slot t_{tag} is adequate for the reader to broadcast the initialized parameters R and v. The time consumed by transmitting the filter vector V is $\lceil \frac{v}{96} \rceil \times t_{tag}$, where t_{tag} can afford transmitting 96 bits data. In the *Unknown Tag Reporting* stage, the reader waits one short-response slot t_{short} for listening the expected announcement from the unknown tags. Hence, the time cost of B-UTD denoted as T_{B-UTD} is given as follows:

$$T_{B-UTD} = t_{tag} + \lceil \frac{v}{96} \rceil \times t_{tag} + t_{short} \tag{2}$$

[1] The assumption that an unknown tag *frankly* announces its unknown identity is reasonable and necessary. Because if the unknown tags intentionally hide and keep silent, no solutions can detect the unknown tag event.

Clearly, when the filter vector length v is minimized to $-N/(ln[1-(1-\alpha)^{\frac{1}{m}}])$, the proposed B-UTD protocol achieves its best time-efficiency.

2) Energy-efficiency: As mentioned before, the long filter vector is divided into segments of 96-bits to be sequentially transmitted. Each of the tags (both known tags and unknown tags) keeps awake before successfully receiving the segment containing its representative bit. For an arbitrary known tag (we do not care the energy consumption of the unknown tags), let the variable I denote the *index* of the segment containing its representative bit. Because each tag employs a *uniform* hash function to select its representative bit, each of the S segments has the same[2] probability $\frac{1}{S}$ to contain the representative bit of a certain tag, where $S = \lceil \frac{v}{96} \rceil$. That is, we have $P(I = i) = \frac{1}{S} \mid i \in [1, S]$. The expectation $E(I)$ of the segment index I can be given as

$$E(I) = \sum_{i=1}^{i=S} [i \times P(I = i)] = \frac{1}{2} \times (S+1) \tag{3}$$

That is, for an arbitrary known tag, it has to keep awake for *one* tag slot to receive the initialized parameters and $\frac{1}{2} \times (S+1)$ (*expectation value*) more tag slots before entering the sleep state. The expected energy consumption of a known tag during one execution of B-UTD is $[1 + \frac{1}{2} \times (S+1)] \times \omega$. For N known tags in total, the energy cost of B-UTD denoted as E_{B-UTD} (excluding the energy consumption of unknown tags) is given as follows:

$$\begin{aligned} E_{B-UTD} &= N \times [1 + \frac{1}{2} \times (S+1)] \times \omega \\ &= N \times [1 + \frac{1}{2} \times (\lceil \frac{v}{96} \rceil + 1)] \times \omega \end{aligned} \tag{4}$$

Clearly, the proposed B-UTD protocol also achieves the best energy-efficiency, when the filter vector length v is minimized.

V. A Sampling based Unknown Tag Detection Protocol

In this section, we first present the motivation of introducing the well-known sampling idea into our protocol. Based on the prior B-UTD protocol, we propose a new Sampling based Unknown Tag Detection (S-UTD) protocol in detail. After that, theoretical analysis of the system parameters is also presented to minimize the execution time and energy consumption of the proposed S-UTD, respectively.

A. Motivation

The well-known sampling idea is widely used in reality, e.g., the number of the products in a factory is so large that checking them all is almost impossible. A common way is to select some samples from all the products, and conducting inspection on the small scale samples only. One of the features of RFID applications is *large scale*. Hence, we could introduce the sampling idea [12], [13] into RFID applications when performing detection on the large number of RFID tags. Intuitively, a more efficient unknown tag detection protocol is achieved if we could discover the unknown tag event by performing detection on *only* sampled tags instead of on all tags.

B. Protocol Design

This section describes the design of the Sampling based Unknown Tag Detection (S-UTD) protocol in detail. A sampling process is added before the execution of the prior B-UTD protocol, and thus we propose the S-UTD protocol, which consists of three stages: the *Sampling stage*, the *Identity Verification stage* and the *Unknown Tag Reporting stage*. In what follows, we mainly describe the details of the new *Sampling stage*.

1) Sampling stage: In this stage, a fraction p of the all the tags (including the known tags as well as the unknown tags) is expected to be selected as samples and the other tags will directly enter the sleep state for conserving energy. The detailed operations in the *Sampling stage* are originated from [12], and are presented in the following.

The reader broadcasts a request $< R_1, x >$, where R_1 is a random number that is fresh in every execution and the integer x is equal to $\lceil p \times X \rceil$, in which p is the sampling probability and X is a sufficiently large constant pre-configured in the tag during the manufacturing process. Using the received random seed and its ID, each tag calculates a hash function $H(ID, R_1) \mod X$ whose result follows a uniform distribution within $[0, X)$. If the hashing result is less than the received parameter x, the tag will participate in the following detection process (i.e., being selected as a sample); otherwise, it will directly enter the sleep state and will not participate in the following detection process. The number of sampled *known* tags is expected to be $N \times p$, where N is the number of known tags and p is the sampling probability.

2) Identity Verification stage: Since all above sampling decisions are made pseudo-randomly depending on the used parameters, hence, the reader can predict all the decisions and exactly know which known tags are sampled. The reader maps the IDs of *sampled* known tags to a filter vector V with v bits. The detailed procedures are the same as these of B-UTD protocol. The reader also sequentially broadcast the parameters R_2, v and the constructed filter vector V, where R_2 is another random number used in this stage. The sampled tags check their representative bits in V to determine if they are unknown to the reader. Specifically, a sampled tag realizes its unknown identity when it finds its representative bit is '0'.

3) Unknown Tag Reporting stage: The unknown tags that are sampled and select '0s' in V will realize and announce their unknown identity. The reader discovers the unknown tag event if it senses a non-empty slot in this reporting stage.

[2] In fact, the probability that the representative bit of a tag lies in the last segment is a little smaller than $\frac{1}{S}$, because the last segment is usually less than 96 bits. But this deviation can be ignored when v is very large.

C. Investigating the Configuration of the Sampling Probability p and the Filter Length v

We use $P(N, M, p, v)$ to denote the probability that the proposed S-UTD protocol can successfully discover the unknown tag event when there are N known tags and exactly M unknown tags; the sampling probability is p; and the filter vector length is v. The problem of configuring the parameters p and v is formulated as follows:

$$\begin{aligned} &Find\ p\ and\ v \\ &s.t.\ \forall M \geq m\ P(N, M, p, v) \geq \alpha \end{aligned} \tag{5}$$

Theorem 3: Given N, M, p, and v, we have $P(N, M, p, v) = 1 - (1 - p_1)^M$, where $p_1 = p \times e^{-\frac{Np}{v}}$.

Proof: For a certain one of the M unknown tags, the probability that it can realize its unknown identity is equal to the probability that it is sampled *and* selects a '0' (i.e., none of the Np sampled known tags is mapped to this bit) in the filter vector V. We denote this probability as p_1, it can be given as

$$\begin{aligned} p_1 &= p \times [\binom{v}{1} \times \frac{1}{v} \times (1 - \frac{1}{v})^{Np}] \\ &\approx p \times e^{-\frac{Np}{v}}, \end{aligned}$$

where p is the sampling probability and v is the filter length.

The probability $P(N, M, p, v)$ that the unknown tag event can be discovered is equal to the probability that *at least one* of the M unknown tags selects '0' in V. Therefore, $P(N, M, p, v)$ can be given as

$$P(N, M, p, v) = 1 - (1 - p_1)^M,$$

where $p_1 = p \times e^{-\frac{Np}{v}}$. ■

Theorem 4: Given N, m, α and p, if v is larger than $-Np/(ln[\frac{1-(1-\alpha)^{\frac{1}{m}}}{p}])$, the detection objective presented in (5) holds.

Proof: Because $P(N, M, p, v)$ is also a monotonically increasing function with respect to M, then we have $P(N, M, p, v) \geq P(N, m, p, v)$, $\forall M \geq m$. Therefore, for any *value pair* $< p, v >$, if $P(N, m, p, v) \geq \alpha$, we have $P(N, M, p, v) \geq P(N, m, p, v) \geq \alpha$, i.e., the detection objective presented in (5) is satisfied. By solving the inequality $P(N, m, p, v) \geq \alpha$, we have $v \geq -Np/(ln[\frac{1-(1-\alpha)^{\frac{1}{m}}}{p}])$. That is, given a sampling probability p, any value of v larger than $-Np/(ln[\frac{1-(1-\alpha)^{\frac{1}{m}}}{p}])$ can satisfy the predefined detection accuracy in (5). ■

Note that, if the sampling probability p is too small, the detection probability $P(N, m, p, v)$ can never be larger than the desired confidence level α for *any* value of filter length v. Such a small sampling probability cannot be used. The smallest sampling probability p_{min} can be calculated by an efficient algorithm Alg. 1. The returned value of p_{min} is within a considerably small error threshold δ from the real minimum value. Because a long filter vector consumes more transmission time, given a fixed sampling probability p, the filter vector length v should be minimized to $-Np/(ln[\frac{1-(1-\alpha)^{\frac{1}{m}}}{p}])$.

Algorithm 1: Finding the Minimum Sampling Probability p_{min}

```
Input  : the number $N$ of known tags, the tolerance threshold $m$ and the
         detection confidence level $\alpha$.
Output : The minimum sampling probability $p_{min}$ that can be used.
$p_{min} = 1$, $\delta = 0.0001$, $v_{up} = 1,000,000$;
while ($p_{min} > 0$) do
    if $p(N, m, p_{min}, v_{up}) \geq \alpha$ then
        $p_{min} = p_{min} - \delta$;
    else
        break;
$p_{min} += \delta$;
return $p_{min}$ ;
```

In the above, we have analyzed how to configure the sampling probability p and the filter length v to meet the predefined detection accuracy.

D. Analyzing the Performance of S-UTD

In this subsection, we analyze the time-efficiency as well as the energy-efficiency of the proposed S-UTD protocol.

1) Time-efficiency: In the *Sampling stage*, *one* tag slot t_{tag} is adequate for the reader to broadcast the sampling parameters R_1 and x. In the *Identity Verification* stage, *one* tag slot t_{tag} is adequate for the reader to broadcast the filtering parameters R_2 and v. The time consumed by transmitting the filter vector V is $\lceil \frac{v}{96} \rceil \times t_{tag}$, where t_{tag} can afford transmitting 96 bits data. In the *Unknown Tag Reporting* stage, the reader waits *one* short-response slot t_{short} for listening the expected announcement from the unknown tags. Hence, the time cost of S-UTD denoted as T_{S-UTD} is given as follows:

$$T_{S-UTD} = t_{tag} + t_{tag} + \lceil \frac{v}{96} \rceil \times t_{tag} + t_{short} \tag{6}$$

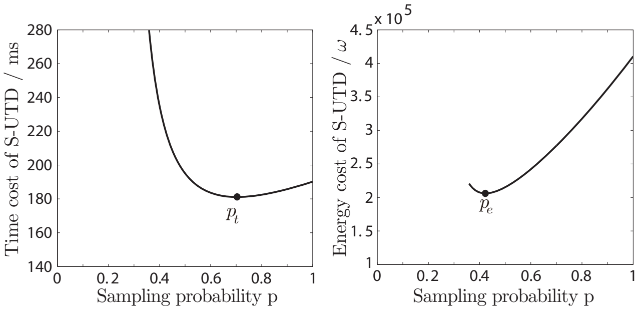
2) Energy-efficiency: In the first stage (i.e., the *Sampling stage*), all the known tags have to keep awake for *one* tag slot to receive the parameters R_1 and x, hence, the corresponding energy consumption of known tags is $N \times \omega$.

Np known tags are expected to keep awake and participate in the second stage (i.e., *Identity Verification stage*). The corresponding energy consumption of the Np awake known tags is $Np \times [1 + \frac{1}{2} \times (\lceil \frac{v}{96} \rceil + 1)] \times \omega$, according to Eq. (4).

We do not care the energy consumption of the unknown tags in the third stage (i.e., the *Unknown Tag Reporting stage*). Hence, the energy cost of the proposed S-UTD denoted as E_{S-UTD} is given as follows:

$$E_{S-UTD} = N \times \omega + Np \times [1 + \frac{1}{2} \times (\lceil \frac{v}{96} \rceil + 1)] \times \omega \tag{7}$$

As illustrated in Fig. 2, there is a sampling probability p_t that minimizes the execution time of S-UTD and a sampling probability p_e that minimizes the energy cost of S-UTD.



(a)

(b)

Figure 2. Evaluating the performance of S-UTD with varying sampling probability p, where the number N of known tags is fixed to 10,000; the tolerance threshold m is pre-configured to 10; and the confidence level α is set to 95%. (a) and (b) show the time cost and the energy cost of S-UTD, respectively.

Note that, p_t and p_e can be easily got through off-line methods, which are not presented in this paper due to the space limitation.

VI. Performance Evaluation

Extensive simulation experiments are conducted to evaluate the performance of the proposed protocols in this section. For fair comparison with the most related literature [10], we simulated the same experimental conditions: (1) considering a single reader in the simulations and assuming it has adequate power to interrogate with a large number of RFID tags; (2) The wireless communication channel between the reader and tags is error-free; (3) the signal interference between the adjacent RFID tags is ignored. And each simulation is conducted for 1000 times and we get the average results.

A. Investigating the Impact of the Tolerance Threshold m and the Confidence Level α

In this subsection, we simulate the proposed protocols under different system parameters m and α: m varies from 20 to 40; under a fixed m, α varies from 90% to 99%; N is fixed to $10,000$. The proposed S-UTD protocol can work at two modes: one is the *time saving mode*, where the sampling probability p is set to p_t as illustrated in Fig. 2 (a), and the other one is the *energy saving mode*, where the sampling probability p is set to p_e as illustrated in Fig. 2 (b). Due to the space limitation, we only configure S-UTD to the *time saving mode* (execution time is the most important performance metric). The simulation results illustrated in Fig. 3 and Fig. 4 show that a small tolerance threshold m (or a large detection confidence level α) will increase the energy cost as well as the time cost of the proposed B-UTD protocol and the S-UTD protocol. That is, if we want to achieve a high detection accuracy, we have to sacrifice the time and energy.

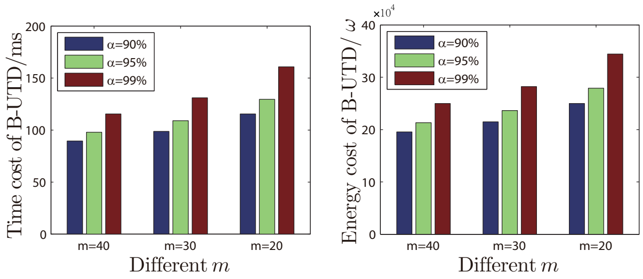


Figure 3. Investigating the impact of the tolerance threshold m and the confidence level α on the B-UTD protocol, where $N = 10,000$.

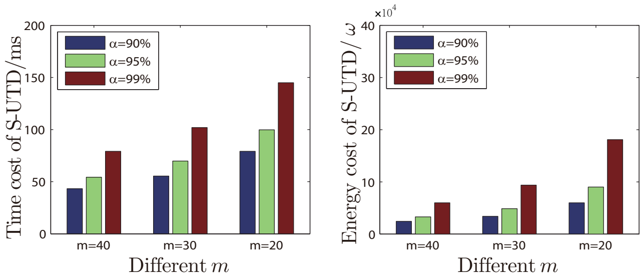


Figure 4. Investigating the impact of the tolerance threshold m and the confidence level α on the S-UTD protocol, where $N = 10,000$.

B. Comparing with the Most Related Protocol

To the best of our knowledge, none of the prior literature concentrates on addressing the problem of unknown tag detection. And the BUIP-CF protocol proposed in [10] is the most related literature, which aims to completely identify all the unknown tags instead of just detecting the unknown tag event. For fair comparison, we only simulate its *known tags deactivation phase*, which is necessary for monitoring the appearance of the unknown tags. Simulation results shown in Table II indicate that the proposed unknown tag detection protocols considerably outperform the BUIP-CF protocol. For example, when $N = 30,000$, $m = 30$, and $\alpha = 95\%$, the execution time of BUIP-CF is 51508.5 ms, and that of proposed S-UTD protocol is just 199.0 ms, representing a reduction of 99.6% in terms of time-efficiency; the energy consumption of BUIP-CF is 2711.4 ω, and that of S-UTD is just 3.5 ω, representing a reduction of 99.9% in terms of energy-efficiency. When monitoring the sparse appearance of unknown tags, periodically executing the the proposed S-UTD protocol instead of BUIP-CF conserves a large amount of time as well as energy, and thus prolongs the lifetime of an active RFID system from several months to several years.

VII. Conclusion

This study has investigated how to detect the unknown RFID tags in a time- and energy-efficient way. The protocols proposed in this paper are complementary to the existing unknown tag identification protocols. These two types of

Table II

Comparing the proposed protocols with the state-of-the-art protocol, where the known tag number N varies from 10,000 to 30,000; m and α are fixed to 30 and 95%, respectively. Time unit: millisecond; Energy unit: ($10^5 \times \omega$)

Alg. Name	B-UTD		S-UTD		BUIP-CF	
	time cost	energy cost	time cost	energy cost	time cost	energy cost
N=10,000	109.0	2.4	69.8	0.5	17233.7	301.6
N=15,000	162.1	5.2	102.1	1.0	25811.6	678.3
N=20,000	215.3	9.2	134.4	1.7	34336.3	1204.9
N=25,000	268.4	14.2	166.7	2.5	42923.2	1883.7
N=30,000	321.5	20.4	199.0	3.5	51508.5	2711.4

protocols should be jointly used to monitor the unknown tags (i.e., the unregistered new items or the misplaced items). A Sampling based Unknown Tag Detection (S-UTD) protocol is proposed in this paper, which adopts the a filter vector and the sampling idea to discover the unknown tag event in an efficient way. We theoretically analyze the system parameters to minimize the execution time as well as energy consumption of the proposed protocols. We conduct extensive simulations to evaluate the performance of the proposed protocols, and the experimental results show that the proposed S-UTD protocol considerably outperforms the most related BUIP-CF protocol by reducing more than 90% of the required energy consumption and execution time.

Acknowledgment

This work was supported by NSFC (grant nos 61173162, 61173160, 61103234, 61272417,61173161,and 61173165), and the National Science Foundation for Distinguished Young Scholars of China (grant no 61225010).

References

[1] C. Wang, H. Wu, and N.-F. Tzeng, "RFID-Based 3-D Positioning Schemes," *Proc. of IEEE INFOCOM*, 2007.

[2] 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.

[3] W. Zhu, J. Cao, Y. Xu, L. Yang, and J. Kong, "Fault-Tolerant RFID Reader Localization Based on Passive RFID Tags," *Proc. of IEEE INFOCOM*, 2012.

[4] C. Qian, Y. Liu, H. Ngan, and L. Ni, "ASAP: scalable identification and counting for contactless RFID systems," *Proc. of IEEE ICDCS*, 2010.

[5] 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.

[6] L. F. Chaves, E. Buchmann, and K. Bohm, "Finding misplaced items in retail by clustering RFID data," *ACM EDBT*, pp. 136–152, 2010.

[7] Y. Zheng and M. Li, "Fast Tag Searching Protocol for Large-Scale RFID," *Proc. of IEEE ICNP*, 2011.

[8] M. Chen, W. Luo, Z. Mo, S. Chen, and Y. Fang, "An Efficient Tag Search Protocol in Large-Scale," *Proc. of IEEE INFOCOM*, 2013.

[9] B. Sheng, C. C. Tan, Q. Li, and W. Mao, "Finding Popular Categories for RFID Tags," *Proc. of ACM MobiHoc*, 2008.

[10] X. Liu, S. Zhang, K. Bu, and B. Xiao, "Complete and Fast Unknown Tag Identification in Large RFID Systems," *Proc. of IEEE MASS*, 2012.

[11] S. Lee, S. Joo, and C. Lee, "An Enhanced Dynamic Framed Slotted ALOHA Algorithm for RFID Tag Identification," *Proc. of IEEE MobiQuitous*, 2005.

[12] W. Luo, S. Chen, T. Li, and S. Chen, "Efficient Missing Tag Detection in RFID Systems," *Proc. of IEEE INFOCOM*, 2011.

[13] L. Yang, J. Han, Y. Qi, and Y. Liu, "Identification-Free Batch Authentication for RFID Tags," *Proc. of IEEE ICNP*, 2010.

[14] 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.

[15] W. Luo, S. Chen, T. Li, and Y. Qiao, "Probabilistic Missing-tag Detection and Energy-Time Tradeoff in Large-scale RFID Systems," *Proc. of ACM MobiHoc*, 2012.

[16] T. Li, S. Chen, and Y. Ling, "Identifying the Missing Tags in a Large RFID System," *Proc. of ACM MobiHoc*, 2010.

[17] 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.

[18] L. G. Roberts, "Aloha Packet System with and without Slots and Capture," *ACM SIGCOMM Computer Communication Review*, vol. 5, no. 2, pp. 28–42, April 1975.

[19] J. I. Capetanakis, "Tree Algorithms for Packet Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 505–515, 1979.

[20] P. Semiconductors, "I-CODE Smart Label RFID Tags," *http://www.nxp.com/acrobat_download/other/identification/SL 092030.pdf*, Jan 2004.

[21] I. Chlamtac, C. Petrioli, and J. Redi, "Energy-Conserving Access Protocols for Identification Networks," *IEEE/ACM Transactions on Networking*, vol. 7, pp. 51–59, 1999.

[22] B. Bloom, "Space/time tradeoffs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, pp. 422–426.

[23] H. Fang, K. Murali, and L. TV, "Building high accuracy bloom filters using partitioned hashing," vol. 35, no. 1, pp. 277–288, 2007.

[24] K. Huang, J. Zhang, D. Zhang, G. Xie, K. Salamatian, A. X. Liu, and W. Li, "A multi-partitioning approach to building fast and accurate counting bloom filters," *Proc. of IEEE IPDPS*, 2013.

[25] L. Kulseng, Z. Yu, Y. Wei, and Y. Guan, "Lightweight mutual authentication and ownership transfer for rfid systems," *Proc. of IEEE INFOCOM*, 2010.