IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 28, 2018, accepted September 23, 2018, date of publication October 5, 2018, date of current version October 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2873669

On Using Sampling Bloom Filter for Unknown Tag Identification in Large-Scale RFID Systems

HONGLONG CHEN[1], (Member, IEEE), LU LIU[1], RONGJIE CHE[2], KAI LIN[1], XIN AI[1], AND YANJUN LI[3]

[1]College of Information and Control Engineering, China University of Petroleum, Qingdao 257000, China
[2]Sinopec Petroleum Engineering Corporation, Dongying 257026, China
[3]School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Honglong Chen (honglongchen1984@gmail.com)

This work was supported in part by NSFC under Grants 61772551, 61772472, 61872385, and 21606255, in part by the Shandong Provincial Key Program of Research and Development under Grants 2018GGX101035 and 2018GGX101025, in part by the Fundamental Research Funds for the Central Universities under Grant 18CX07003A and Grant 18CX02133A, in part by the Natural Science Foundation of Zhejiang Province under Grant LY17F020020, and in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2016BQ16.

ABSTRACT Radio Frequency Identification (RFID) is one of the key technologies for Internet of things. In RFID systems, the tags without registering in advance are called unknown tags, which usually appear in the scenarios, where the tag-attached objects are moved into or misplaced in the reader's interrogating area. Consequently, unknown tag identification is significant for RFID-based applications, which is the concentration of this paper. We first propose a basic efficient unknown tag identification protocol based on sampling Bloom filter called UTI-SBF, which consists of known tag deactivation phase and unknown tag identification phase. The idea behind the UTI-SBF protocol is to deactivate the known tags to counteract their interference on the unknown tag identification. Then, we propose an enhanced protocol called EUTI-SBF, which eliminates the non-homogeneous slots based on the UTI-SBF protocol to improve the time efficiency. The parameters of the two protocols are theoretically analyzed to maximize the efficiency. We conduct extensive simulations to evaluate the proposed UTI-SBF and EUTI-SBF protocols and the simulation results illustrate that the UTI-SBF and EUTI-SBF protocols outperform the BUIP protocol. In particular, the EUTI-SBF protocol only consumes about 70% of deactivation time compared with the BUIP protocol in the known tag deactivation phase.

INDEX TERMS Known tag deactivation, radio frequency identification (RFID) systems, time efficiency, unknown tag identification.

I. INTRODUCTION

Radio frequency identification (RFID), which has greatly promoted the development of Internet of things (IoTs) [1], [29], [30], [33], [39]–[41], is widely used in various industrial fields [9], [34], [43] after considering the security and privacy issues [28], [31], [32], [38]. RFID systems are usually composed of electronic tags, readers, and back-end server [12]. Electronic tags are used to identify objects, while the reader has a powerful capacity of computation and storage, which can communicate with the tags via radio frequency signal [22] without line-of-sight. The back-end server is the control center that stores and identifies information collected by the reader [11]. The specific characteristics of RFID, including the low cost of tags [21], non-line-of-sight reader-tag communications [22], and so on, lead to the wide applications of RFID systems. In most of the RFID applications, the tags are attached to different objects, which can then be efficiently managed by interrogating the tags, conducted by the reader. Typically, RFID tags are divided into three categories: passive tags, active tags and semi-active tags. Passive tag is the most widely used one, which has no internal power supply and can be driven by electromagnetic waves sent by the reader, resulting in a relatively short communication range (generally 3 to 5 meters). The active tag has an internal power supply, which is more expensive than the passive tag and has a longer communication range. For the semi-active tag, it has an internal power supply for the information processing, while its communication is still driven by the reader's electromagnetic waves.

In practical applications, RFID technology is widely deployed in many areas, such as warehouse management [25], intelligent transportation [10], supply chain management [18], [19], and localization [6], etc. In these applications, such as warehouse management, the RFID system makes it easier for the staff to count the number of items, i.e., cardinality estimation of tags [16], [19], finding lost items, i.e., missing tag detection and identification [8], [13], [23], and discovering new-coming items, i.e., unknown tag identification [14], [15]. Solving these problems efficiently helps to minimize the economic losses, i.e., increasing the time efficiency and reducing the labor cost.

In RFID systems, the tags without registering in advance are called unknown tags, which usually appear in the scenarios, where the tag-attached objects are moved into or misplaced in the reader's interrogating area. In this paper, we focus on investigating the problem of identifying the unknown tags efficiently and completely for large-scale RFID systems. For this purpose, there are two challenges need to be solved. The first one is the presence of known tags, which will involve in the unknown tag identification procedure that may reduce the communication efficiency between the reader and unknown tags, which is based on the Framed Slotted Aloha (FSA) protocol [37], [42]. As the known tags will also participate in the unknown tag identification and respond to the reader, resulting in more collision slots in the slotted frame, which will reduce the slot utilization and time efficiency.

Another challenge is how to improve time efficiency. In this paper, we have made the following three efforts to improve the efficiency of unknown tag identification: 1) We adopt the sampling Bloom filter to increase the ratio of singleton slots, which can increase the slot utilization and time efficiency; 2) We build the mathematical model between the efficiency and the related parameters, after which the parameters are set to theoretically minimize the known tag deactivation time; 3) Last but not the least, the tags, which have been identified, will no longer participate in the unknown tag identification procedure, resulting in less interference and higher efficiency.

The main contributions of this paper are the following:

- We propose a basic efficient unknown tag identification protocol based on sampling Bloom filter called UTI-SBF, which consists of known tag deactivation phase and unknown tag identification phase;
- We propose an enhanced protocol called EUTI-SBF, which eliminates the non-homogeneous slots based on the UTI-SBF protocol to further improve the time efficiency;
- The parameters of the proposed UTI-SBF and EUTI-SBF protocols are theoretically analyzed to maximize the identification efficiency;
- We conduct extensive simulations to evaluate the proposed UTI-SBF and EUTI-SBF protocols and the simulation results illustrate that the UTI-SBF and EUTI-SBF protocols outperform the BUIP protocol.

The rest of the paper is shown as follows. Section II reviews the related work. Section III presents the problem statement including the mathematical model and problem formulation. In Section IV, we propose the UTI-SBF protocol with theoretical analysis. In Section V, an enhanced protocol called EUTI-SBF is proposed. In Section VI, we evaluate the performance of the proposed UTI-SBF and EUTI-SBF protocols. Finally, this paper is concluded in Section VII.

II. RELATED WORK

As one of the key technologies in the IoTs, RFID is widely studied and lots of works [9], [22], [34], [43] have been proposed. Generally, the missing tag detection [4], [5], missing tag identification [7], [8], unknown tag detection [15] and identification [13] can be summarized as anomaly detection in RFID systems. In this section, we will discuss on the Bloom filter based anomaly detection and unknown tag detection and identification.

A. APPLICATION OF BLOOM FILTER IN ANOMALY DETECTION

Standard Bloom Filter [3] is a data structure that efficiently represents collection affiliation, which is used the h Hash functions to map the elements in the set to a vector. Firstly, all bits in the vector are initialized to zero. When inserting an element into a Bloom filter by h Hash functions, the corresponding bit in the vector is changed from '0' to '1'. To determine whether an element is in the set, it is only necessary to determine whether the corresponding h bits in the vector are '1s', otherwise the element is out of the set. But Bloom filter has a drawback, that is, false positive, which means that even if the corresponding bits of the element are all '1s', the element may not belong to the set.

Anomaly detection using Bloom filter: In [36], Bloom filter is used to first deactivate the unexpected tags to reduce their interference to the missing tag detection, after which whether the expected tags are missing or not can be detected. In [15], an unknown tag detection protocol called SBF-UDP is proposed, which uses a sampling Bloom filter (an extension of the standard Bloom filter) to detect the existence of unknown tags. SBF-UDP achieves better performance than that using the standard Bloom filter, since the sampling probability can be optimally set to maximize the efficiency.

B. UNKNOWN TAG DETECTION AND IDENTIFICATION

Unknown tag detection is to detect the unknown-tag event in the RFID systems. SBF-UDP [15] is proposed to detect the unknown tags, which adopts the sampling Bloom filter that each Hash function corresponds to a sampling probability. A disadvantage of SBF-UDP protocol is that it has the false positive characteristic, i.e., the unknown-tag event can not be deterministically detected. While the *unknown tag identification* is to identify the unknown tags' IDs. BUIP protocol [17] is proposed to identify the unknown tags, which deactivates the known tags to reduce the interference, and can achieve complete identification. However, there is still some potential to improve the efficiency during the known tag deactivation phase. SEBA [35] adopts the Framed Slotted Aloha (FSA) [37], [42] to determine whether unknown tags exist based on the unexpected response received by the reader. SEBA+ [2] is proposed to use the standard Bloom filter to discover an unknown tag when an expected empty slot becomes a non-empty slot. SEBA and SEBA+ are the most basic unknown tag detection protocols. The CU protocol [24] collects unknown tag IDs by continuously scanning and taking advantage of the information which is gathered by previous scanning operations. However the CU protocol is a probabilistic protocol that cannot guarantee the complete identification of all the unknown tags. FUTI and IFUTI [13] protocols are proposed to identify the unknown tags. The FUTI protocol uses the EDFSA protocol to collect the unknown tags' IDs, resulting in lower efficiency, especially when unknown tags' density is relatively large. While the IFUTI protocol does not deactivate the known tags, which will interfere with the process of collecting IDs of unknown tags, thus reduces the time efficiency.

III. PROBLEM STATEMENT

A. SYSTEM MODEL

In large-scale RFID systems, multiple readers are usually deployed, in which there are two kinds of collisions [26], [27]: 1) Reader-reader collision; 2) Reader-tag collision. Therefore, in a multi-reader RFID system, we need to schedule the readers' working order to avoid these collisions. So that two readers with overlapping areas will not work at the same time. The existing reader scheduling protocols [27] can be adopted to avoid collisions, which is out of the scope of this paper. Our proposed protocols are suitable for multiple reader environment as well as a single reader system. For ease of analysis, we use a single reader system as an example. We assume that the RFID system consists of a back-end server, a reader, k known tags and u unknown tags, all the unknown tags and known tags are within the reader's interrogating area. Each tag has a unique ID and is loaded with multiple unified Hash functions. For convenience of description, we refer to both the reader and back-end server as reader in this paper. In this paper, we adopt the sampling Bloom filter, in which each element is pseudo-randomly mapped to the filter with h Hash functions, each of which is with a sampling probability P.

B. UNKNOWN TAG IDENTIFICATION PROBLEM

In this paper, we concentrate on solving the problem of identifying all the unknown tags in large-scale RFID systems and improving the time efficiency. In an RFID system, there are k known tags and u unknown tags. We assume that there are no missing tag, all known tags' IDs are pre-stored in the reader. However, the reader does not know the IDs and the number of the unknown tags. Our goal is to identify all unknown tags as quickly as possible.

The reader communicates with the tags based on the Framed Slotted Aloha (FSA) protocol [37], [42].

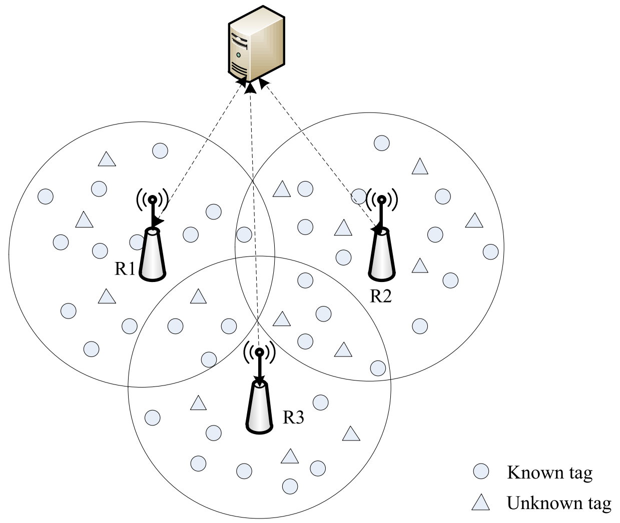


FIGURE 1. Illustration of a multi-reader RFID system with known and unknown tags.

According to the FSA protocol, the communications between the reader and the tags are divided into a serial of frames. In each frame, the reader first broadcasts a query command. After receiving the query command, each tag randomly selects a slot to reply by calculating $H(ID, R) \bmod f$. After the reader executes a frame completely, the slots can be divided into three different types, depending on the number of responding tags in each slot: 1) empty slot: no response; 2) singleton slot: only one response; 3) collision slot: two or more responses. Only the tags replying in a singleton slot can be correctly recognized by the reader. The slots can also be classified into three types according to their lengths: 1) tag slot, in which the tag sends a 96 bit information, denoted as t_{tag}; 2) long-response slot, in which the tag sends 8 to 10 bits of information, denoted as t_l; 3) short-response slot, in which the tag sends one bit of information, denoted as t_s. We set $t_{tag} = 2.4\ ms$, $t_l = 0.8\ ms$ and $t_s = 0.4\ ms$, respectively according to some previous works [4], [13], [15]. In particular, in this paper the long-response slot will be used for the tags to respond to the reader, since the reader needs to distinguish whether the slot is singleton or collision.

IV. UTI-SBF PROTOCOL

To identify the unknown tags for the RFID system, we propose a basic efficient unknown tag identification protocol based on sampling Bloom filter called UTI-SBF, which consists of two phases: known tag deactivation and unknown tag identification. In this section, we will first give a description of the UTI-SBF protocol. Afterwards, we will discuss the parameter optimization to minimize the execution time.

A. DESCRIPTION OF UTI-SBF PROTOCOL

The UTI-SBF protocol consists of two phases: 1) known tag deactivation phase; 2) unknown tag identification phase.

In this paper, we focus on the known tag deactivation phase, in which it requires to estimate the number of unknown tags. Here we will adopt the unknown tag cardinality estimation method in [8]. In the second phase, the existing protocol TIP [20] will be adopted to identify the unknown tags.

The known tag deactivation phase consists of multiple rounds. At the beginning of the each round, the reader first broadcasts the parameters (f_i, R_i, P_i, h) as the query command to the tags, where f_i is the actual execution frame length, R_i denotes the random seed, h represents the number of Hash functions, and P_i is the sampling probability ($0 < P_i < 1$). Note that the sampling probability indicates the probability that a tag replies to the reader in its selected slot, which can be achieved by introducing the "virtual" frame [22]. After receiving the parameters, each tag determines the slot indexes to respond by calculating $H_1(ID, R_i) \bmod f_i$, $H_2(ID, R_i) \bmod f_i$, $\cdots$, $H_h(ID, R_i) \bmod f_i$. Since the IDs of the known tags have been stored in the reader beforehand, the reader can estimate the state of each slot based on the known tag IDs, which can be saved as an f_i-bit expected vector, denoted as EV_i. This process is named expected frame. Each bit in EV_i is set as '1' if the associated slot is expected to be homogeneous and '0' otherwise. Here the homogeneous slot represents a slot which is selected by only one tag's one or multiple sampled Hash mappings.[1] Thus, a singleton slot is also a homogeneous slot, but not vice versa. After that the reader executes the frame, and each tag will reply a long response in each of its selected slots with probability P_i. The reader will detect the actual state (empty, singleton or collision) of each slot. When the frame is completely executed, the reader can generate an f_i-bit actual vector denoted as AV_i according to the actual state of each slot. If the slot is singleton, the corresponding bit in AV_i is marked as '1', and in other cases it will be marked as '0'. Note that the EV_i and AV_i can be used to estimate the known tag cardinality, and the reader can record the detailed information (empty, singleton or collision) of each slot in both the expected and executed frames.

The states of each corresponding slot in the expected frame and executed frame may be different due to the presence of unknown tags. It can be specifically divided into the following scenarios:

- The state of the corresponding slot is homogeneous in the expected frame, then the following two scenarios are considered.
  1) The corresponding slot in the executed frame is singleton, indicating that only one known tag is mapping to this slot. The reader can determine that the tag is a known tag, which will be deactivated in both the two phases;
  2) The corresponding slot in the executed frame is collision, indicating that this slot is also selected by at least one unknown tag. There will be no operation for the reader since it can not differentiate the known tag with unknown tag.

[1] Note that a tag has multiple Hash mappings.

- The state of the corresponding slot is empty in the expected frame, then the following two scenarios are considered.
  1) The corresponding slot in the executed frame is singleton or collision. It indicates that one or more unknown tags are mapping to this slot. Then the reader can determine that the tags are unknown tags, which will be deactivated in the current phase and participate in the unknown tag identification phase;
  2) The corresponding slot in the executed frame is empty. There will be no operation for the reader since no tag replies in this slot.
- The state of the corresponding slot is collision and non-homogeneous in the expected frame, then there will be no operation for the reader since at least two known tags select this slot and the reader can not determine whether this slot is selected by unknown tags.

The known tag deactivation phase consists of multiple rounds, at the end of which all the known tags will be deactivated. The reader then begins the unknown tag identification phase by adopting TIP [20], in which all the unknown tags, including the deactivated ones in the first phase, will participate.

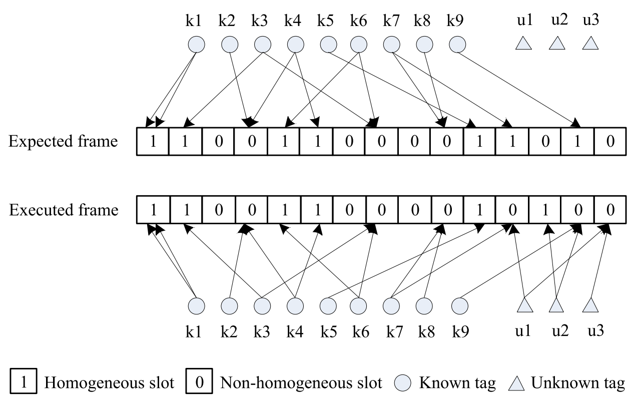


FIGURE 2. Illustration of the known tag deactivation in Phase I of the UTI-SBF protocol.

Fig. 2 illustrates an example of UTI-SBF in the known tag deactivation phase. We suppose that $f_i = 15$, $P_i = 0.75$, and $h = 2$. As shown in Fig. 2, the bits 1, 2, 5, 6 and 11 in the EV_i are set as '1s' and the corresponding bits in the AV_i are also '1s'. Note that the corresponding slots are homogeneous, and only one known tag is mapping to each of the corresponding slots. Thus these known tags can be deactivated. The 12-th and 14-th bits in EV_i are set as '1s', and the corresponding bits in AV_i are '0s'. It indicates that there are unknown tags mapping to the corresponding slots, making the singleton slots turn out to be collision. Therefore, the reader does nothing for them. The 13-th bit in EV_i is '0', and the corresponding bit in AV_i is '1', indicating that only one unknown tag is mapping to this slot. Then the reader can deactivate this unknown tag in the current phase, which will only participate in the unknown tag identification phase. The 15-th slot in the expected frame is empty, which turns out to be collision in the executed frame. It indicates that at least two unknown tags are mapping to this slot, which will be deactivated in the current phase. The 4-th, 8-th and 10-th bits are '0s' in both the EV_i and AV_i, thus there will be no operation for the tags mapping to these collision slots.

B. PARAMETER OPTIMIZATION

To maximize the time efficiency of the UTI-SBF protocol, we need to optimize the parameters f_i and P_i. In this section, we will discuss the parameter settings of f_i and P_i.

The average time of deactivating a known tag in the i-th round, denoted as C_i can be calculated as:

$$C_i = \frac{T_i}{D_i}, \tag{1}$$

where T_i represents the total execution time of the i-th round, and D_i denotes the expected number of deactivated known tags in the i-th round. Then T_i can be calculated as:

$$T_i = t_l \cdot f_i + t_{tag}, \tag{2}$$

where $t_l \cdot f_i$ indicates the time that the tags respond to the reader and t_{tag} is the time for the reader to broadcast the parameters.

Moreover, D_i can be estimated as follows:

$$D_i = k_i \cdot Pd_i, \tag{3}$$

where k_i denotes the number of undeactivated known tags before round i, and Pd_i is the probability for a known tag to be deactivated in round i. Before calculating Pd_i, we first discuss the probability Pu_i, which indicates the probability that a known tag can not be deactivated by one of its h Hash mappings in round i. Thus, there are two parts for Pu_i, the first one is the probability that the Hash mapping is not sampled, and the other one is the probability that the Hash mapping is sampled but the corresponding slot is non-empty and non-homogeneous. Consequently, we can get:

$$Pu_i = P_i \cdot \left(1 - \left((1-P_i) + P_i \cdot (1 - \frac{1}{f_i})\right)^{(k_i+u_i-1)h}\right) + (1-P_i). \tag{4}$$

Thus, the probability Pd_i can be obtained as follows:

$$Pd_i = 1 - Pu_i^h = 1 - \left(P_i \cdot \left(1 - \left((1-P_i) + P_i \cdot (1 - \frac{1}{f_i})\right)^{(k_i+u_i-1)\cdot h}\right) + (1-P_i)\right)^h. \tag{5}$$

Then we can get D_i as:

$$\begin{aligned} D_i &= k_i \cdot Pd_i = k_i \cdot \left(1 - \left(P_i \cdot \left(1 - \left((1-P_i) + P_i \cdot (1 - \frac{1}{f_i})\right)^{(k_i+u_i-1)\cdot h}\right) + (1-P_i)\right)^h\right) \\ &= k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}}\right)^h\right). \end{aligned} \tag{6}$$

Finally, C_i can be obtained as follows:

$$C_i = \frac{T_i}{D_i} = \frac{t_l \cdot f_i + t_{tag}}{k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}}\right)^h\right)}. \tag{7}$$

In order to minimize the deactivation time, the partial derivatives of C_i with respect to f_i and P_i can be calculated as follows:

$$\frac{\partial C_i}{\partial f_i} = \frac{t_l \cdot \left(1-\alpha^h\right) - \left(t_l \cdot f_i + t_{tag}\right) \cdot h \cdot P_i \cdot \alpha^{h-1} \cdot \beta}{k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}}\right)^h\right)^2}, \tag{8}$$

where $\alpha = 1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}}$, $\beta = e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}} \cdot \frac{P_i \cdot h \cdot \left(k_i+u_i-1\right)}{f_i^2}$.

$$\frac{\partial C_i}{\partial P_i} = \frac{\left(t_l \cdot f_i + t_{tag}\right) \cdot h \cdot \alpha^{h-1} \cdot \sigma}{k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}}\right)^h\right)^2}. \tag{9}$$

where $\sigma = e^{-\frac{P_i \cdot h \cdot (k_i+u_i-1)}{f_i}} \cdot \left(\frac{P_i \cdot h \cdot \left(k_i+u_i-1\right)}{f_i} - 1\right)$.

Let Eq. (8) and Eq. (9) equal 0, and we can numerically obtain the optimal frame size f_i^* and sampling probability P_i^* at the beginning of round i.

V. EUTI-SBF PROTOCOL

In the UTI-SBF protocol, only the homogeneous slots are used to deactivate the known tags, in which the reader only needs to detect the states of the slots in the execute frame, which are homogeneous in the expected frame. However, the proportion of homogeneous slots in each frame is relatively small, indicating that lots of slots in the frame are not utilized. In this section, we propose an enhanced protocol called EUTI-SBF that can improve the slot utilization by eliminating non-homogeneous slots.

A. OVERVIEW OF EUTI-SBF

EUTI-SBF protocol also consists of two phases: known tag deactivation phase and unknown tag identification phase. we focus on the known tag deactivation phase, which consists of multiple rounds. At the beginning of the i-th round, the reader broadcasts parameters (f_i, R_i, P_i, h) as the query command to the tags, where f_i is the expected frame length, R_i is the random seed, h represents the number of Hash functions and P_i is the sampling probability ($0 < P_i < 1$) of each Hash mapping. When receiving the parameters, each tag determines its slot indexes by calculating $H_1(ID, R_i)$ mod f_i, $H_2(ID, R_i)$ mod f_i, $\cdots$, $H_h(ID, R_i)$ mod f_i, each of which is sampled with probability P_i. Since the IDs of the known tags have been stored in the reader beforehand, the reader can estimate the state of each slot, which can be saved as an f_i-bit expected vector, denoted as EV_i. Each bit in EV_i is set as '1' if the associated slot is expected to be homogeneous, otherwise, it is set as '0'. The reader then divides EV_i into $\lceil f_i/96 \rceil$

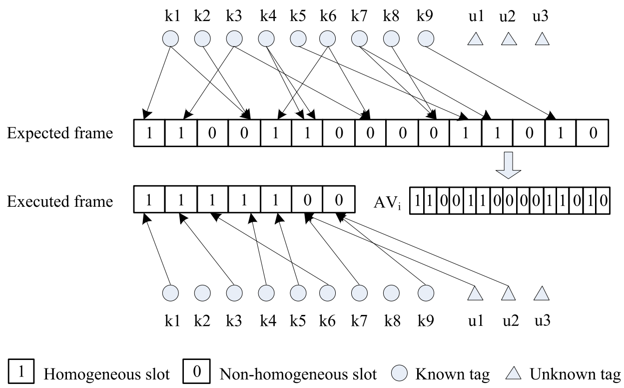


FIGURE 3. **Illustration of the known tag deactivation in Phase I of the EUTI-SBF protocol.**

segments and then broadcasts them sequentially to the tags. When receiving the segments, each tag can check the value of each of its associated bits in V_c according to the expected slot indexes. Only if the corresponding bit is '1' in EV_i, indicating that the slot is homogeneous, the tag will transmit a long response in the executed frame. In the executed frame, the non-homogeneous slots will be eliminated and each tag can update the reply slot index in the executed frame by calculating the number of '1s' prior to the corresponding bit in EV_i. For instance, as shown in Fig. 3, the 5-th slot is homogeneous, which is selected by known tag k_6 in the expected frame, and there are two '1s' prior to corresponding bit in EV_i. Therefore, k_6 will reply in the 3-th slot in the executed frame and two non-homogeneous slots (the 3-th and 4-th slots) in the expected frame will be eliminated.

In each slot of the executed frame, the reader will detect the slot states (empty, singleton or collision). After that an f_i-bit actual vector denoted as AV_i according to the actual state of each slot. If the slot is singleton, the corresponding bit in AV_i is set as '1', and in other cases it will be set as '0'. When the frame is completely executed, the reader detects the status of each slot. There are two scenarios to be discussed:

- The slot is singleton in the executed frame, indicating that only one known tag is mapping to this slot. The reader can determine that the tag is a known tag, which will be deactivated in both of the two phases;
- The slot is collision in the executed frame, indicating that this slot is also selected by at least one unknown tag. There will be no operation, since the reader can not differentiate the known tag with the unknown tag.

The reader repeats the above process until all the known tags are deactivated. Then the reader begins the unknown tag identification phase by adopting TIP protocol [20].

Fig. 3 illustrates an example of the known tag deactivation phase in the EUTI-SBF protocol. We assume the number of the Hash functions $h = 2$. As shown in Fig. 3, slots 1, 2, 5, 6, 11, 12 and 14 in the expected frame are homogeneous, the corresponding tags will respond in the executed frame. For example, the 5-th slot in the expected frame is homogeneous and there are two homogeneous slots prior to it, thus k_6 will respond in the third slot in the executed frame. Consequently, k_6 will be deactivated in both the two phases. In this example, there are totally 8 non-homogeneous slots in the expected frame, which are eliminated in the executed frame, resulting in greatly improving the efficiency.

B. PARAMETER OPTIMIZATION

Similarly, we will discuss the optimal parameter settings of f_i and P_i in this section.

The average time to deactivate a known tag in the i-th round can be obtained as follows:

$$C_i = \frac{T_i}{D_i}, \tag{10}$$

where T_i represents the total execution time of the i-th round, and D_i denotes the expected number of deactivated known tags in the i-th round. Similar to Eq. (6) in the UTI-SBF protocol, D_i can be calculated as:

$$D_i = k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}}\right)^h\right), \tag{11}$$

where k_i denotes the number of active known tags before round i, P_i denotes the sampling probability of each Hash mapping in round i, h represents the number of Hash functions, and u denotes the number of unknown tags.

Moreover, T_i can be calculated as follows:

$$T_i = Nh_i \cdot t_l + \left(\left\lceil \frac{f_i}{96} \right\rceil + 1\right) \cdot t_{tag}, \tag{12}$$

where Nh_i represents the expected number of homogeneous slots in round i. In Eq. (12), $Nh_i \cdot t_l$ refers to the time that the tags respond to the reader in the executed frame, and $\left(\left\lceil \frac{f_i}{96} \right\rceil + 1\right) \cdot t_{tag}$ is the time for the reader to broadcast the parameters and the segments of AV_i. In order to obtain Nh_i, we first discuss the probability Ph_i, which indicates the probability that a slot is homogeneous in round i. Ph_i can be estimated as follows:

$$\begin{aligned} Ph_i &= \binom{1}{k_i} \cdot \left(1 - \left(P_i \cdot (1 - \frac{1}{f_i}) + (1 - P_i)\right)^h\right) \\ &\quad \cdot \left((1 - P_i) + P_i \cdot (1 - \frac{1}{f_i})\right)^{(k_i - 1) \cdot h} \\ &= k_i \cdot \left(1 - e^{-\frac{P_i \cdot h}{f_i}}\right) \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}}. \end{aligned} \tag{13}$$

Then we can get Nh_i as:

$$Nh_i = f_i \cdot Ph_i = f_i \cdot k_i \cdot \left(1 - e^{-\frac{P_i \cdot h}{f_i}}\right) \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}}. \tag{14}$$

Finally, C_i can be obtained as follows:

$$\begin{aligned} C_i &= \frac{T_i}{D_i} \\ &= \frac{f_i \cdot k_i \cdot \left(1 - e^{-\frac{P_i \cdot h}{f_i}}\right) \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}} \cdot t_l + \left(\left\lceil \frac{f_i}{96} \right\rceil + 1\right) \cdot t_{tag}}{k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}}\right)^h\right)}. \end{aligned} \tag{15}$$

In order to minimize the deactivation time, the partial derivatives of C_i with respect to f_i and P_i can be calculated as follows:

$$\frac{\partial C_i}{\partial f_i} = \frac{\gamma_1 - \gamma_2}{\left(k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u_i - 1)}{f_i}}\right)^h\right)\right)^2}, \tag{16}$$

$$\frac{\partial C_i}{\partial P_i} = \frac{\gamma_3 - \gamma_4}{\left(k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u_i - 1)}{f_i}}\right)^h\right)\right)^2}, \tag{17}$$

where $\gamma_1 = \left(t_l \cdot k_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}} \cdot \left(\left(\left(1 - e^{-\frac{P_i \cdot h}{f_i}}\right) \cdot \left(1 + \frac{P_i \cdot h \cdot (k_i - 1)}{f_i}\right) - \frac{P_i \cdot h}{f_i} \cdot e^{-\frac{P_i \cdot h}{f_i}}\right) + \left\lceil \frac{t_{tag}}{96} \right\rceil\right)\right) \cdot \delta$, $\gamma_2 = \lambda \cdot k_i \cdot h \cdot P_i \cdot \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}}\right)^{h-1} \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}} \cdot \frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i^2}$, $\gamma_3 = f_i \cdot t_l \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}} \cdot \left(\frac{h}{f_i} \cdot e^{-\frac{P_i \cdot h}{f_i}} - \frac{h \cdot (k_i - 1)}{f_i} \cdot \left(1 - e^{-\frac{P_i \cdot h}{f_i}}\right)\right) \cdot \delta$, $\gamma_4 = \lambda \cdot h \cdot \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}}\right)^{h-1} \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}} \cdot \left(1 - \frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}\right)$, $\delta = k_i \cdot \left(1 - \left(1 - P_i \cdot e^{-\frac{P_i \cdot h \cdot (k_i + u - 1)}{f_i}}\right)^h\right)$, $\lambda = f_i \cdot t_l \cdot k_i \cdot (1 - e^{-\frac{P_i \cdot h}{f_i}}) \cdot e^{-\frac{P_i \cdot h \cdot (k_i - 1)}{f_i}} + \left(\left\lceil \frac{f_i}{96} \right\rceil + 1\right) \cdot t_{tag}$.

Let Eq. (16) and Eq. (17) equal 0, and we can numerically obtain the optimal frame size f_i^* and sampling probability P_i^* at the beginning of round i.

VI. PERFORMANCE EVALUATION

In this section, we evaluate the performance UTI-SBF and EUTI-SBF protocols in MATLAB, and compare them with the BUIP protocol [17]. We first describe the simulation settings. After that, we conduct performance comparisons to illustrate the effectiveness of our proposed UTI-SBF and EUTI-SBF protocols.

A. SIMULATION SETTINGS

We adopt the execution time of the protocols as performance metric. In the simulations, we consider the single reader scenario. However the protocols we proposed can also be applied to the multi-reader scenario. And we consider the error-free communication channel between the reader and the tags. We set the time for each tag or reader to transmit the tag ID or 96-bit segment as $t_{tag} = 2.4\ ms$, and it takes $t_l = 0.8\ ms$ for each tag to reply a long response to reader. Each result is obtained by averaging 100 simulations. We mainly consider two parameters, i.e., the number of known tags k and the number of unknown tags u, and study their effects on the protocols' execution time. Finally, we use the TIP protocol [20] to collect unknown tag IDs in the unknown tag identification phase.

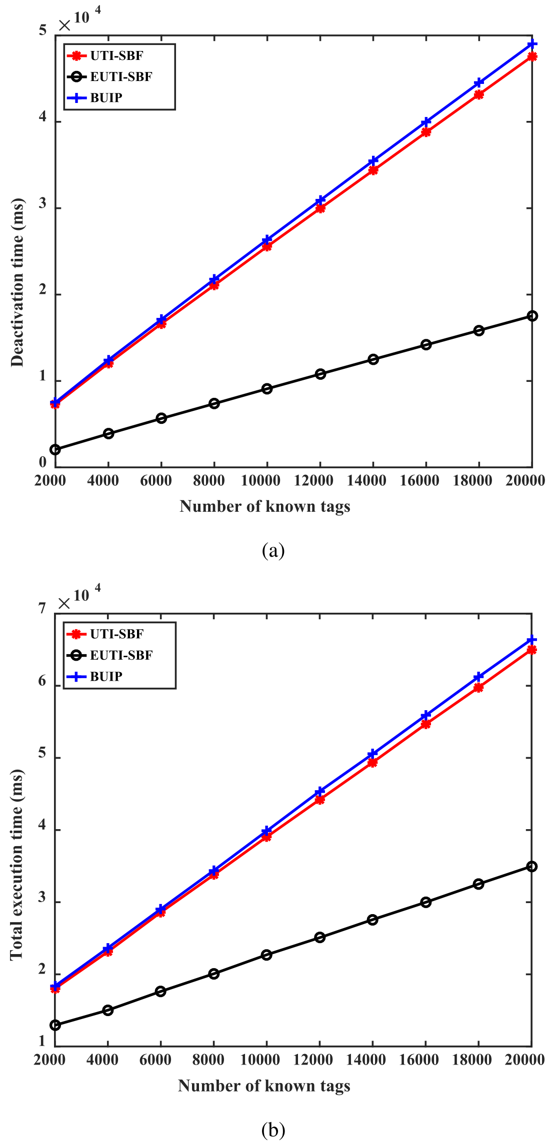
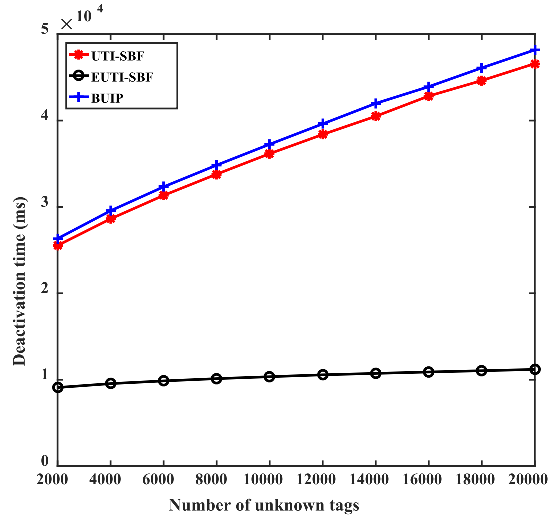


FIGURE 4. Effects of number of known tags on the three protocols. (a) Effects of k on known tag deactivation time. (b) Effects of k on total execution time.

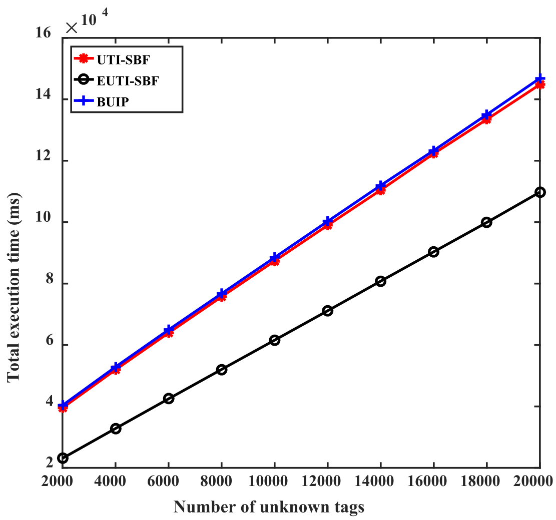
B. SIMULATION RESULTS

1) EFFECTS OF NUMBER OF KNOWN TAGS

Fig. 4 shows the effects of the number of known tags on the deactivation phase and total execution time of our proposed UTI-SBF and EUTI-SBF protocols and the BUIP protocol. We set $u = 2000$, and k gradually increases from 2000 to 20000. Fig. 4 shows that as the number of known tags increases, the deactivation time and total execution time of all the three protocols also increase. Fig. 4(a) shows the trend of the deactivation time as the number of known tags varies. It illustrates that the deactivation time of the UTI-SBF protocol is a bit less than that of the BUIP protocol, while the EUTI-SBF protocol consumes about 30% of deactivation time of the BUIP protocol. Fig. 4(b) shows the trend of the total execution time as the number of known tags varies, which is similar to that of Fig. 4(a). In addition, the increasing rate of the deactivation time and total execution time of the proposed EUTI-SBF protocol is much less than that of the UTI-SBF and BUIP protocols as the number of known tags increases. Fig. 4 illustrates that the proposed UTI-SBF and EUTI-SBF protocols outperform the BUIP protocol under different numbers of known tags.



(a)



(b)

FIGURE 5. Effects of number of unknown tags on the three protocols. (a) Effects of *u* on known tag deactivation time. (b) Effects of *u* on total execution time.

Fig. 5 shows the effects of the number of unknown tags on the deactivation time and total execution time. We set $k = 10000$, and u gradually increases from 2000 to 20000. Fig. 5(a) shows that as the number of unknown tags increases, the deactivation time of the UTI-SBF protocol and BUIP protocol increases as well. However, the increase of the number of unknown tags has very limited effect on the deactivation time of the EUTI-SBF protocol since the non-homogeneous slots are eliminated. Also, the UTI-SBF protocol consumes less deactivation time than that of the BUIP protocol. As shown in Fig. 5(b), when the number of unknown tags increases, the total execution time of the three protocols also increases. However, the total execution time of the EUTI-SBF protocol increases relatively slowly compared with the UTI-SBF and BUIP protocols. Fig. 5 illustrates that the proposed UTI-SBF and EUTI-SBF protocols outperform the BUIP protocol under different numbers of unknown tags.

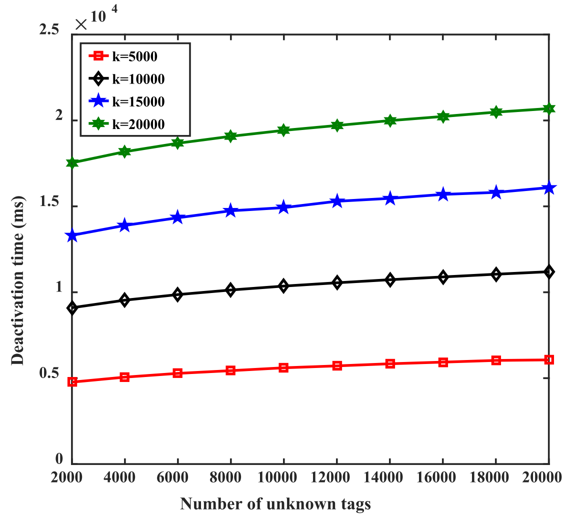


FIGURE 6. Effects of number of unknown tags on the deactivation time of the EUTI-SBF protocol with different number of known tags.

Fig. 6 shows the effects of the number of unknown tags on deactivation time of the EUTI-SBF protocol with different numbers of known tags. We set $k = 5000$, $k = 10000$, $k = 15000$, $k = 20000$ respectively, and u gradually increases from 2000 to 20000. As shown in Fig. 6, the deactivation time in the EUTI-SBF protocol is mainly affected by the number of known tags. However, as the number of unknown tags increases, interference to the known tag deactivation caused by the unknown tags will strengthen, thus the deactivation time will increase slightly.

VII. CONCLUSIONS

In this paper, we investigated the problem of identifying the unknown tags for large-scale RFID systems. We first proposed a basic efficient unknown tag identification protocol based on sampling Bloom filter called UTI-SBF, which consists of known tag deactivation phase and unknown tag identification phase. Then we proposed an enhanced protocol called EUTI-SBF to improve the time efficiency. The parameters of the two protocols were theoretically analyzed to maximize the efficiency. We conducted extensive simulations to illustrate the effectiveness of the proposed UTI-SBF and EUTI-SBF protocols.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[2] G. Bianchi, "Revisiting an RFID identification-free batch authentication approach," *IEEE Commun. Lett.*, vol. 15, no. 6, pp. 632–634, Jun. 2011.

[3] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Commun. ACM*, vol. 13, no. 7, pp. 422–426, Jul. 1970.

[4] H. Chen, G. Ma, Z. Wang, Q. Wang, and J. Yu, "MAC: Missing tag iceberg queries for multi-category RFID systems," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2018.2863726.

[5] H. Chen, G. Ma, Z. Wang, F. Xia, and J. Yu, "Probabilistic detection of missing tags for anonymous multicategory RFID systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11295–11305, Dec. 2017.

[6] H. Chen, G. Ma, Z. Wang, J. Yu, L. Shi, and X. Jiang, "Efficient 3-dimensional localization for RFID systems using jumping probe," *Pervas. Mobile Comput.*, vol. 41, pp. 300–318, Oct. 2017.

[7] H. Chen, Z. Wang, F. Xia, Y. Li, and L. Shi, "Efficiently and completely identifying missing key tags for anonymous RFID systems," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2915–2926, Aug. 2018.

[8] H. Chen, G. Xue, and Z. Wang, "Efficient and reliable missing tag identification for large-scale RFID systems with unknown tags," *IEEE Internet Things J.*, vol. 4, no. 3, pp. 736–748, Jun. 2017.

[9] M. Chen, W. Luo, Z. Mo, S. Chen, and Y. Fang, "An efficient tag search protocol in large-scale RFID systems with noisy channel," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 703–716, Apr. 2016.

[10] S. Chen, M. Zhang, and B. Xiao, "Efficient information collection protocols for sensor-augmented RFID networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3101–3109.

[11] Y. Fu, Z. Qian, G. Ji, X. Gao, and Q. Zhu, "Fast unknown tag identification in large-scale RFID systems," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Oct. 2017, pp. 1–6.

[12] C.-H. Hsu, H.-C. Chao, and J. H. Park, "Threshold jumping and wrap-around scan techniques toward efficient tag identification in high density RFID systems," *Inf. Syst. Frontiers*, vol. 13, no. 4, pp. 471–480, 2011.

[13] X. Liu, K. Li, G. Min, Y. Shen, A. X. Liu, and W. Qu, "Completely pinpointing the missing RFID tags in a time-efficient way," *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 87–96, Jan. 2015.

[14] X. Liu *et al.*, "Efficient unknown tag identification protocols in large-scale RFID systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3145–3155, Dec. 2014.

[15] X. Liu *et al.*, "Sampling Bloom filter-based detection of unknown RFID tags," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1432–1442, Apr. 2015.

[16] X. Liu *et al.*, "RFID cardinality estimation with blocker tags," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 1679–1687.

[17] X. Liu, B. Xiao, S. Zhang, and K. Bu, "Unknown tag identification in large RFID systems: An efficient and complete solution," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1775–1788, Jun. 2015.

[18] Y. Qi, Q. Yao, Y. Chen, and X. Zhong, "Study on RFID authentication protocol theory," *China Commun.*, vol. 8, pp. 65–71, Jan. 2011.

[19] C. Qian, H. Ngan, Y. Liu, and L. M. Ni, "Cardinality estimation for large-scale RFID systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1441–1454, Sep. 2011.

[20] Y. Qian, Z. He, and D. Zhang, "TIP: Time-efficient identification protocol for unknown RFID tags using Bloom filters," in *Proc. IEEE Int. Conf. Parallel Distrib. Syst.*, Dec. 2016, pp. 151–158.

[21] M. Roberti, "A 5-cent breakthrough," *RFID J.*, vol. 5, no. 6, p. 1, 2006.

[22] M. Shahzad and A. X. Liu, "Every bit counts: Fast and scalable RFID estimation," in *Proc. ACM MOBICOM*, 2012, pp. 365–376.

[23] M. Shahzad and A. X. Liu, "Fast and reliable detection and identification of missing RFID tags in the wild," *IEEE/ACM Trans. Netw.*, vol. 24, no. 6, pp. 3770–3784, Dec. 2016.

[24] B. Sheng, Q. Li, and W. Mao, "Efficient continuous scanning in RFID systems," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[25] C. C. Tan, B. Sheng, and Q. Li, "Efficient techniques for monitoring missing RFID tags," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 1882–1889, Jun. 2010.

[26] S. Tang, C. Wang, X.-Y. Li, and C. Jiang, "Reader activation scheduling in multi-reader RFID systems: A study of general case," in *Proc. IEEE Int. Parallel Distrib. Process. Symp.*, May 2011, pp. 1147–1155.

[27] S.-J. Tang, J. Yuan, X.-Y. Li, G. Chen, Y. Liu, and J. Zhao, "RASPberry: A stable reader activation scheduling protocol in multi-reader RFID systems," in *Proc. IEEE Int. Conf. Netw. Protocols*, Oct. 2009, pp. 304–313.

[28] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, "Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 4, pp. 591–606, Jul./Aug. 2018.

[29] Z. Wang, Q. Cao, H. Qi, H. Chen, and Q. Wang, "Cost-effective barrier coverage formation in heterogeneous wireless sensor networks," *Ad Hoc Netw.*, vol. 64, no. 9, pp. 65–79, 2017.

[30] Z. Wang, H. Chen, Q. Cao, H. Qi, Z. Wang, and Q. Wang, "Achieving location error tolerant barrier coverage for wireless sensor networks," *Comput. Netw.*, vol. 112, no. 1, pp. 314–328, 2017.

[31] Z. Wang *et al.*, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2018.2861393.

[32] Z. Wang *et al.*, "Privacy-preserving crowd-sourced statistical data publishing with an untrusted server," *IEEE Trans. Mobile Comput.*, 2018, doi: 10.1109/TMC.2018.2861765.

[33] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018.

[34] L. Yang, J. Cao, W. Zhu, and S. Tang, "Accurate and efficient object tracking based on passive RFID," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2188–2200, Nov. 2015.

[35] L. Yang, J. Han, Y. Qi, and Y. Liu, "Identification-free batch authentication for RFID tags," in *Proc. 18th IEEE Int. Conf. Netw. Protocols*, Oct. 2010, pp. 154–163.

[36] J. Yu, L. Chen, R. Zhang, and K. Wang, "Finding needles in a haystack: Missing tag detection in large RFID systems," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2036–2047, May 2017.

[37] A. Zanella, "Adaptive batch resolution algorithm with deferred feedback for wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3528–3539, Oct. 2012.

[38] H. Zhang, Y. Qi, H. Zhou, J. Zhang, and J. Sun, "Testing and defending methods against DoS attack in state estimation," *Asian J. Control*, vol. 19, no. 4, pp. 1295–1305, 2017.

[39] H. Zhou, M. Ruan, C. Zhu, V. C. M. Leung, S. Xu, and C. Huang, "A time-ordered aggregation model-based centrality metric for mobile social networks," *IEEE Access*, vol. 6, pp. 25588–25599, 2018.

[40] H. Zhou, H. Wang, X. Li, and V. Leung, "A survey on mobile data offloading technologies," *IEEE Access.*, vol. 6, no. 1, pp. 5101–5111, Jan. 2018.

[41] C. Zhu, V. C. M. Leung, J. J. P. C. Rodrigues, L. Shu, L. Wang, and H. Zhou, "Social sensor cloud: Framework, greenness, issues, and outlook," *IEEE Netw.*, vol. 32, no. 5, pp. 100–105, Sep./Oct. 2018.

[42] L. Zhu and T.-S. P. Yum, "Optimal framed aloha based anti-collision algorithms for RFID systems," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3583–3592, Dec. 2010.

[43] W. Zhu, J. Cao, Y. Xu, L. Yang, and J. Kong, "Fault-tolerant RFID reader localization based on passive RFID tags," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 2065–2076, Aug. 2014.

HONGLONG CHEN (M'15) received the M.E. degree in control theory and control engineering from Zhejiang University, China, in 2008, and the Ph.D. degree in computer science from The Hong Kong Polytechnic University, Hong Kong, in 2012. He was a Post-Doctoral Researcher with the School of CIDSE, Arizona State University, from 2015 to 2016. He is currently an Associate Professor with the College of Information and Control Engineering, China University of Petroleum, China. He has published over 50 research papers in prestigious journals and conferences including IEEE TVT, IEEE IoT-J, IEEE TMC, IEEE TETC, IEEE INFOCOM, IEEE ICPP, IEEE ICDCS, and so on. His current research interests are in the areas of RFID and Internet of things. He is a member of ACM. He is an Associate Editor of the IEEE Access journal.

LU LIU received the B.E. degree in automation from Qufu Normal University, China, in 2015. She is currently pursuing the master's degree in control engineering with the College of Information and Control Engineering, China University of Petroleum, China. Her research interest is in the field of radio frequency identification.

RONGJIE CHE received the B.E. degree in automation from the China University of Petroleum, China, in 2006, and the M.E. degree in control theory and control engineering from the China University of Petroleum, China, in 2009. He is currently a Senior Engineer with Sinopec Petroleum Engineering Corporation, Dongying, China. His research interests include Internet of Things.

KAI LIN received the B.E. degree in automation from Tianjin Polytechnic University, China, in 2015. He is currently pursuing the master's degree in control engineering with the College of Information and Control Engineering, China University of Petroleum, China. His current research interests include RFID system and Internet of things.

XIN AI received the B.E. degree in automation from Qufu Normal University, China, in 2017. She is currently pursuing the master's degree in control theory and control engineering with the College of Information and Control Engineering, China University of Petroleum, China. Her research interest is in the field of RFID security.

YANJUN LI received the B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively, and the Ph.D. degree from Nancy University, Villers-les-Nancy, France, in 2010. She is currently an Associate Professor with the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. Her research area includes energy harvesting sensor networks and Internet of things.