RFID Estimation with Blocker Tags

Xiulong Liu   Bin Xiao   Keqiu Li   Alex X. Liu   Jie Wu   Xin Xie   Heng Qi

***Abstract*—With the increasing popularization of RFID technology in the retail and logistics industry, RFID privacy concern has attracted much attention because a tag responds to queries from readers no matter they are authorized or not. An effective solution is to use a commercially available blocker tag that behaves as if a set of tags with known blocking IDs are present. However, the use of blocker tags makes the classical RFID estimation problem much more challenging, as some genuine tag IDs are covered by the blocker tag and some are not. In this paper, we propose REB, the first RFID estimation scheme with the presence of blocker tags. REB uses the framed slotted Aloha protocol specified in the EPC C1G2 standard. For each round of the Aloha protocol, REB first executes the protocol on the genuine tags and the blocker tag, and then virtually executes the protocol on the known blocking IDs using the same Aloha protocol parameters. REB conducts statistical inference from the two sets of responses and estimates the number of genuine tags. Rigorous theoretical analysis of parameter settings is proposed to guarantee the required estimation accuracy, meanwhile minimizing the time cost and energy cost of REB. We also reveal a fundamental trade-off between the time cost and energy cost of REB, which can be flexibly adjusted by the users according to the practical requirements. Extensive experimental results reveal that REB significantly outperforms the state-of-the-art identification protocols in terms of both time-efficiency and energy-efficiency.**

***Index Terms*—RFID Estimation, RFID Privacy, Blocker Tags.**

I. Introduction

A. Background & Motivation

RADIO Frequency Identification (RFID) technique has risen to be a revolutionary element in supply chain management and inventory control [2]–[10], as the cost of commercial passive RFID tags is negligible compared with the value of the products to which they are attached (*e.g.*, as low as 5 cents per tag [11]). For example, in Hong Kong International Airport where RFID systems are used to track shipment, the average daily cargo tonnage in May 2010 was 12K tonnes and has been on the rise [12]. As real-time information is made available, the administration and planning processes can be significantly improved. An RFID system typically consists of a reader and a population of tags [13]. A reader has a dedicated power source with significant computing capability. It transmits commands to query a set of tags, and the tags respond over a shared wireless medium. A tag is a microchip with an antenna in a compact package that has limited computing capability and a longer communication range than barcodes. There are two types of tags: *passive tags* that do not have their own power sources and are powered up by harvesting the radio frequency energy from readers, and *active tags* that have their own power sources.

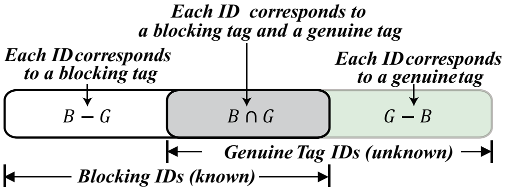

Each ID corresponds to a blocking tag and a genuine tag
Each ID corresponds to a blocking tag
Each ID corresponds to a genuine tag
$B - G$
$B \cap G$
$G - B$
Genuine Tag IDs (unknown)
Blocking IDs (known)

Fig. 1. Three types of IDs in the system containing blocker tags.

An inevitable fact is that the widely used RFID tags impose serious privacy concerns, as when a tag is interrogated by an RFID reader; no matter whether the reader is authorized or not, it blindly responds with its ID and other stored information (such as manufacturer, product type, and price) in a broadcast fashion. For example, a woman may not want her dress sizes and a patient may not want his/her medication, to be publicly known. Some cryptography based authentication protocols have been proposed to circumvent malicious scanning [14]. However, none of these protocols is compliant with the C1G2 standard. Furthermore, these protocols often require computational resources that exceed the capability of commercial C1G2-compliant passive RFID tags. If such additional computational capability is indeed implemented, the cost of such tags will be much more expensive than the C1G2-compliant passive tags. Hence, we use blocker tags, which are easy to deploy, to protect RFID privacy. A *blocker tag* is an RFID device that is preconfigured with a set of known RFID tag IDs, which we call *blocking IDs*. The blocker tag behaves as if all tags with its blocking IDs are present. A blocker tag protects the privacy of the set of genuine tags whose IDs are among the blocking IDs of the blocker tag because any response from a genuine tag is coupled with the simultaneous response from the blocker tag; thus, the two responses collide and attackers cannot obtain private information.

B. Problem Statement

This paper concerns with the problem of RFID estimation with the presence of a blocker tag. Formally, the problem is defined as follows. *Given (1) a set of unknown genuine tags G of unknown size g, (2) a blocker tag with a set of blocking IDs B, which is configured by the system manager, (3) a required*

Xiulong Liu, Keqiu Li, Xin Xie and Heng Qi are with the School of Computer Science and Technology, Dalian University of Technology, China. (e-mail: {xiulongliudut,xiexindut}@gmail.com; {hengqi,keqiu}@dlut.edu.cn)

Bin Xiao is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. (e-mail: csbxiao@comp.polyu.edu.hk)

Alex X. Liu is with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210008, China. (e-mail: alexliu@nju.edu.cn)

Jie Wu is with the Department of Computer and Information Sciences, Temple University, USA. (e-mail: jiewu@temple.edu)

Corresponding Authors: Keqiu Li and Alex X. Liu.

The preliminary version of this paper, titled “RFID Cardinality Estimation with Blocker Tags”, was published in IEEE INFOCOM 2015 [1].

confidence interval $\alpha \in (0, 1]$, and (4) a required reliability $\beta \in [0, 1)$, we want to estimate the number of genuine tags in G, denoted as $\hat{g}$, so that $P\{|\hat{g} - g| \leq \alpha g\} \geq \beta$. Besides time-efficiency, we also take the energy-efficiency into consideration if the battery-powered active RFID tags are used. We assume that the blocker tag is trusted because the manager is in control of the blocker tag. Hence, the blocking ID set B is known by the manager in prior. In contrary, for the genuine tag set G, we know neither its size nor the exact tag IDs in it. As shown in Fig. 1, the sets B and G may overlap.

This problem may arise in many applications. Consider an RFID-enabled logistics center, where each package is affixed with an RFID tag that contains the delivery address and the item information. The information of some packages, *e.g.*, the medicine someone purchased on Amazon, are closely related to the customers' privacy. To protect their personal privacy, we use a blocker tag to prevent the tags from malicious scanning. Meanwhile, the manager may want to use an RFID estimation protocol to monitor the number of packages for the purpose of making an efficient delivery plan. How about turning off the blocker tag and then using prior RFID estimation schemes to estimate the number of genuine tags? Turning off the blocker tag will give attackers a time window to breach privacy, especially for the scenarios in which RFID estimation schemes are being continuously performed for monitoring purposes.

C. Limitations of Prior Art

To the best of our knowledge, this paper is the first to investigate RFID estimation with the presence of a blocker tag. Although some RFID estimation schemes have been proposed [13], [15]–[21], none of them considers the presence of a blocker tag. Furthermore, none of them can be easily adapted to solve our problem. For example, the state-of-the-art Average Run-based Tag estimation (ART) protocol uses the Framed Slotted Aloha communication mechanism specified in the C1G2 standard. The reader queries the tags by initializing a slotted frame, and each tag randomly selects a slot to reply the response. ART leverages the average run length of non-empty slots observed from the time frame to estimate the tag cardinality. Clearly, ART can only tell the tag cardinality of the universal set $U = B \cup G$, which, however, is not what we want. Due to the same reason, all the other existing tag estimation protocols cannot address the new problem of RFID estimation with the presence of a blocker tag.

How about using the tag identification protocols? Generally there are two categories of identification protocols: Aloha-based protocols [22] and Tree-based protocols [23]. It is a well-recognized fact that the identification protocols are slow because their execution time is proportional to the tag population. What is worse, their efficiency will further deteriorate with the presence of a blocker tag. Their basic principles can be found in Section V. Here, we only elaborate why these two types of identification protocols become more inefficient with the presence of a blocker tag. As exemplified in Fig. 2 (a), the responses from two tags with the same ID in $B \cap G$ always collide with each other. Thus, the genuine tag IDs in $B \cap G$ can never be identified. Moreover, the large number of *continuous collisions* also seriously hinders the identification of the tags in $(B - G) \cup (G - B)$.

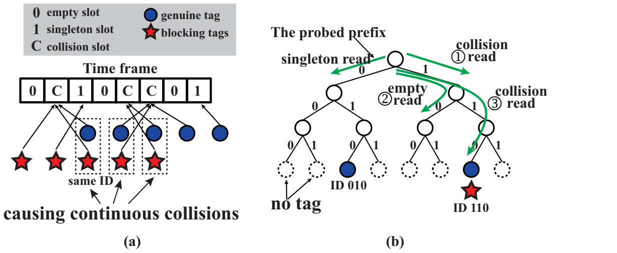


Fig. 2. Exemplify the impact of blocker tag on the tag identification protocols. (a) Aloha-based protocols. (b) Tree-based protocols.

The tree-based identification protocols can identify the IDs in $(B - G) \cup (G - B)$ when a queried prefix is followed by a successful read; and identify the IDs in $B \cap G$ when a prefix whose length is equal to tag ID but still followed by a collision read. Then, we can get the set G, by calculating $\{(B - G) \cup (G - B) - B\} \cup (B \cap G)$. The cardinality g is obtained upon getting G. When there is blocker tag in the system, the tree-based identification becomes slow because the continuous collisions caused by the tags in $B \cap G$ always "lure" the reader to continuously extend the probed prefix by 1 bit each time, until the probed prefix reaches the ID length. As exemplified in Fig. 2 (b), to identify the ID '010', the reader only needs to probe one prefix '0'; However, to identify the ID '110' (an ID in $B \cap G$), the reader needs to successively probe three prefixes: '1'→ '10'→'110' (until the prefix length is equal to ID length). Since the length of prefix '110' is already equal to the length of tag ID, we know the ID '110' is necessary to correspond to two tags: one of them is a blocking tag, and the other one is a genuine tag. We observed from the simulation results that the identification protocols are seriously slow, and their time cost is even hundreds of times longer than REB.

D. Proposed Approach

In this paper, we propose an *RFID Estimation scheme with Blocker tags* (*REB*). The communication protocol used by REB is the standard framed slotted Aloha protocol, in which a reader first broadcasts a value f and a random number R to the tags, where f represents the number of time slots in the forthcoming frame. Then, each tag computes a hash using the random number R and its ID, where the resulting hash value h is within $[0, f - 1]$, and the tag replies during slot h. For each slot, if the reader does not receive any tag response, we represent this slot as 0; if the reader successfully receives a tag response, we represent this slot as 1; if the reader senses the collided tag responses (two or more tags respond simultaneously), and we represent this slot as c. Note that a reader can detect if there is a collision according to the C1G2 standard. Executing this protocol for the blocking IDs (simulated by the blocker tag) and genuine tags, we get a ternary array $\mathbb{BG}[0..f - 1]$ where each bit is 0, 1, or c. As we know the blocking IDs, we can virtually execute the framed slotted Aloha protocol using the same frame size f and random number R for the blocking IDs; thus, we get a ternary array $\mathbb{B}[0..f-1]$ where each bit is 0, 1, or c. Here, if no blocking ID is hashed to this position, it is represented by 0; if only one blocking ID is hashed to this position, it is represented by 1; if two or more blocking IDs are hashed to this position, it is represented by c. From the two arrays $\mathbb{BG}[0..f-1]$ and $\mathbb{B}[0..f-1]$, REB counts two numbers: N_{00}, which is the number of slots i such that both $\mathbb{BG}[i]=0$ and $\mathbb{B}[i]=0$; and N_{11}, which is the number of slots i such that both $\mathbb{BG}[i]=1$ and $\mathbb{B}[i]=1$. REB is based on the key insight that in general the smaller N_{00} is, the larger $|B\cup G|$ is and the larger N_{11} is, the larger $|B-G|$ is. In this paper, we establish a monotonous functional relationship between N_{00} and $|B\cup G|$, and a monotonous functional relationship between N_{11} and $|B-G|$. Thus, from the observed N_{00} and N_{11}, we can estimate $|B\cup G|$ and $|B-G|$. Then, we can calculate the size of G because $|G|=|B\cup G|-|B-G|$.

E. Challenges and Proposed Solutions

The first challenge is to guarantee the required estimation accuracy that is specified by the confidence interval $\alpha\in(0,1]$ and the reliability $\beta\in[0,1)$. The estimator is not precise due to its probabilistic nature. Since a single frame is usually not able to output an accurate estimate, we use the estimate averaged from multiple frames to give a fine-grained estimate. We first theoretically propose the expression of the estimator variance in a single frame. Then, we investigate how many frames are necessary to reduce the estimator variance to a sufficiently small value such that the averaged estimate can satisfy the required $\langle\alpha,\beta\rangle$ accuracy.

The second challenge is to minimize the time cost when passive RFID tags are used, on the premise that the required accuracy is guaranteed. In reality, the frame size f is usually less than 512, for practical reasons [13]. To make REB scalable to the large-scale RFID systems, we use the persistence probability p [13]. The reader initializes a frame with the size of f/p, but sends commands to terminate the frame after the first $f\le 512$ slots. The settings of f and p are important to the performance of REB. Hence, we propose sufficient theoretical analysis to optimize the parameters f and p to minimize the time cost of REB.

The third challenge is to minimize the energy cost when the battery powered active tags are used. We also investigate the optimization of frame size f and persistence probability p to minimize the energy cost of REB. However, we find that the time cost and energy cost of REB cannot be minimized at the same time. When the energy cost is minimized, the execution time of REB can be quite long. Hence, we should jointly consider the time-efficiency and energy-efficiency of REB instead of separately considering them. Finally, we reveal a trade-off between the time cost and energy cost, which can be flexibly adjusted by the protocol parameters.

F. Novelty and Advantage over Prior Art

The key novelty of this paper is in formulating the practically important problem of RFID estimation with the presence of a blocker tag, and taking the first step towards an efficient solution. The key technical depth of this paper is in proposing the unbiased estimator of genuine tags and addressing the three aforementioned technical challenges. The key advantage of REB over prior art is threefold: (1) REB is compliant to the EPC C1G2 RFID standard, and does not require any modifications to off-the-shelf tags, it only needs to be implemented on readers as a software module; (2) Compared with the prior estimation protocols, REB jointly uses the number of persistent empty slots and the number of persistent singleton slots to eliminate the interference from the blocker tag, and thus, can correctly estimate the cardinality of genuine tags; (3) REB significantly outperforms the state-of-the-art identification protocols in terms of both time-efficiency and energy-efficiency. For example, when $|U|=50,000$ and the tag ratio $|B-G|:|B\cap G|:|G-B|=1:1:1$, REB runs 178x faster than EDFSA [22] and 2785x faster than TH [23], meanwhile revealing 281x and 333x improvement over EDFSA and TH in terms of energy-efficiency, respectively.

The rest of this paper is organized as follows. In Section II, we describe the detailed design of REB, and give the functional estimator as well as the minimum frame number that can guarantee the required estimation accuracy. In Section III, we propose rigorous analysis to optimize the involved parameters to minimize the time cost and energy cost of REB, respectively. In Section IV, we conduct extensive simulations to evaluate the performance of REB. We discuss related work in Section V. Finally, we conclude the paper in Section VI.

II. REB Protocol

In this section, we first describe the system model used in this paper. Then, an efficient *RFID Estimation scheme with Blocker tags (REB)* is proposed to estimate the number of genuine tags by jointly using N_{00} and N_{11} observed in a time frame. We explicitly give the functional estimator, and point out that the estimation using a single time frame is hard to be accurate due to probabilistic variance. Hence, we propose to use multiple independent time frames to refine the estimation. This section finally presents rigorous theoretical analysis to investigate how many frames are needed to guarantee the desired estimation accuracy and how to avoid premature termination of REB.

A. System Model

We consider the RFID system containing a single reader, a single blocker tag, and a population of genuine tags. The set of blocking IDs is represented by B, whose cardinality is b. The set of genuine tags is denoted as G, whose cardinality is g. We use U to denote the universal tag set, where $U=B\cup G$, and $|U|=u$. The IDs in $B-G$ do not correspond to any genuine tags, whose cardinality is denoted as b', *i.e.*, $b'=|B-G|$.

The reader communicates with tags (including both genuine tags and virtual ones *simulated* by the blocker tag) under control of the backend server. The communication between the reader and tags are based on a time slotted way. Any two consecutive transmissions (from a tag to a reader or vice versa) are separated by a waiting time $\tau_w=302us$ [13]. According to the specification of the Philips I-Code system [24], the wireless transmission rate from a tag to a reader is $53Kb/s$, that is,

TABLE I
NOTATIONS USED IN THE PAPER

Notations	Descriptions
G / B / U	set of genuine tags; set of blocking IDs; union set.
g / b' / u	$g = \lvert G \rvert$; $b' = \lvert B - G \rvert$; $u = \lvert B \cup G \rvert$.
α / β	required confidence interval; required reliability.
$\hat{g}$	estimate of g.
f / p	frame size; persistence probability.
$E(\cdot)$ / $Var(\cdot)$	expectation; variance.
Z_β	the percentile of β. *e.g.*, $Z_\beta = 1.96$ when $\beta = 95\%$.
p_{00} / p_{11}	probability that a slot pair is $\langle 0, 0\rangle$; probability that a slot pair is $\langle 1, 1\rangle$.
N_{00} / N_{11}	# of the persistent empty slots in a frame; # of the persistent singleton slots in a frame.
$\mathcal{T}$ / $\mathcal{E}$	time cost of REB; energy cost of REB.
ω	energy cost on an active tag for transmitting RN16.

it takes a tag $\tau_t = 18.9us$ to transmit 1 bit. The rate from a reader to a tag is $26.5Kb/s$, that is, transmission of 1 bit to tags requires $\tau_r = 37.7us$. Then, the time of a slot for transmitting m-bit information from a tag to the reader is $\tau_w + m \times \tau_t$; and the time of a slot for transmitting m-bit information from a reader to the tags is $\tau_w + m \times \tau_r$. The main notations used throughout the paper are summarized in Table I.

B. Protocol Description

Our REB uses the standard framed slotted Aloha protocol specified in EPC C1G2 [25] as the MAC layer communication mechanism. The reader initializes a slotted time frame by broadcasting a binary request $\langle R, f\rangle$, where R is a random number and f is the frame size (*i.e.*, the number of slots in the forthcoming frame). Using the received parameters $\langle R, f\rangle$, each tag initializes its slot counter sc by calculating $sc = H(ID, R) \mod f$, and the hashing result follows a uniform distribution within $[0, f - 1]$. In many existing RFID literature [26]–[28], a widely accepted assumption is that a tag is capable of computing a seeded hash function. Moreover, Luo *et al.* have proposed a scheme to implement the seeded hash function in passive RFID tags with simple circuits [29]. When designing our REB protocol, we could obtain the specified hash function from the RFID manufacturer. The reader broadcasts a `QueryRep` command at the end of each slot. Upon receiving `QueryRep`, a tag decrements its slot counter sc by 1. In a slot, a tag will respond to the reader if its slot counter sc becomes 0. According to the occupation status, slots are classified into three types: *empty slot* in which no tag responds; *singleton slot* in which only one tag responds; *collision slot* in which two or more tags respond.

In the following, we present how our REB estimates the number of genuine tags by observing the slots in a frame. Since the backend server gets full knowledge of the simulated blocking IDs, it is able to predict which slots the blocking IDs are "mapped" to. Thus, it is able to construct a virtual ternary array $\mathbb{B}[0..f-1]$. A bit in $\mathbb{B}[0..f-1]$ is set to 0 when no blocking ID is mapped to this slot; 1 when only one blocking ID is mapped to this slot; c when two or more blocking IDs are mapped to this slot (a hashing collision). On the other hand, by observing the frame, the reader could get another array $\mathbb{BG}[0..f-1]$, also consisting of f bits. A bit in $\mathbb{BG}[0..f-1]$ is set to 0 when no tag responds in this slot; 1 when only one tag responds in this slot; c when two or more tags cause a collision in this slot. To distinguish a singleton slot from a collision one, each tag does not need to respond with the whole 96-bit ID. For efficiency, each tag responds with the RN16 (16-bit) [25] that is much shorter than the 96-bit tag ID. Two slots with the same index in $\mathbb{B}[0..f-1]$ and $\mathbb{BG}[0..f-1]$ are called a slot pair. In our scheme, the reader needs to record the numbers of the following two types of slot pairs.

- N_{00} is the number of *persistent empty* slot pairs $\langle 0, 0\rangle$ (*i.e.*, $\mathbb{B}[i] = 0$ *AND* $\mathbb{BG}[i] = 0$, $i \in [0, f-1]$).
- N_{11} is the number of *persistent singleton* slot pairs $\langle 1, 1\rangle$ (*i.e.*, $\mathbb{B}[i] = 1$ *AND* $\mathbb{BG}[i] = 1$, $i \in [0, f-1]$).

REB can estimate the cardinality of genuine tags by jointly using the number of persistent empty slots and that of persistent singleton slots. A persistent empty slot happens only when no ID in $U = B \cup G$ is mapped to this index. Thus, N_{00} reflects the cardinality of U (*i.e.*, u). Later, we will show that a *monotonous* functional relationship can be established between u and N_{00}. REB uses this function to estimate u from N_{00}. Similarly, a persistent singleton slot happens when only one ID in $B - G$ is mapped to this index. Therefore, N_{11} reflects the cardinality $|B - G|$ (*i.e.*, b'). Clearly, if we know u and b', we can get the cardinality g of genuine tags by calculating $g = u - b'$. It may not be sufficient to satisfy the required estimate accuracy by counting the numbers of N_{00} and N_{11} in a *single* frame. Hence, REB executes k independent frames with different random number R, and uses the averaged estimate as the fine-grained result.

Note that, the frame size should be set to no more than 512 in practice [13], [23], [30] (the detailed reasons can be found in [23]). If a large number of tags contend for such a short frame, most slots will become collision slots. To scale to a large tag population, the reader uses a persistence probability $p \in (0, 1]$ to *virtually* extends the frame size f to f/p, but *actually* terminates the frame after the first f slots [13]. Fundamentally, each tag participates in the actual frame of f slots with a probability p.

C. Functional Estimator

In this section, we derive the functional estimator $\hat{g}$ from N_{00} and N_{11} for the REB protocol in one frame. For an arbitrary slot pair, the probability that it is $\langle 0, 0\rangle$, denoted as p_{00}, is given as follows.

$$p_{00} = \left\{1 - \frac{p}{f}\right\}^u \approx e^{-\frac{up}{f}} \tag{1}$$

The approximation in Eq. (1) holds when f/p is relatively large [9], [13], [15]. The number of slot pairs $\langle 0, 0\rangle$, *i.e.*, N_{00}, follows $Bernoulli(f, p_{00})$. The expectation and variance of the variable N_{00} are presented as follows.

$$E(N_{00}) = fp_{00} = fe^{-\frac{up}{f}} \tag{2}$$

$$Var(N_{00}) = fp_{00}\left\{1 - p_{00}\right\} = fe^{-\frac{up}{f}}\left\{1 - e^{-\frac{up}{f}}\right\} \tag{3}$$

Similarly, we use p_{11} to denote the probability that a slot pair is $\langle 1, 1\rangle$, which is given as follows.

$$p_{11} = \binom{b'}{1} \left\{ \frac{p}{f} \right\} \left\{ 1 - \frac{p}{f} \right\}^{u-1} \approx \frac{b'p}{f} e^{-\frac{up}{f}} \tag{4}$$

The number of $\langle 1, 1\rangle$ slot pairs, *i.e.*, N_{11}, also follows $Bernoulli(f, p_{11})$. The expectation and variance of the variable N_{11} are presented as follows.

$$E(N_{11}) = fp_{11} = b'pe^{-\frac{up}{f}} \tag{5}$$

$$Var(N_{11}) = fp_{11} \{1 - p_{11}\} = b'pe^{-\frac{up}{f}} \left\{ 1 - \frac{b'p}{f} e^{-\frac{up}{f}} \right\} \tag{6}$$

According to Eq. (2), u can be expressed as follows.

$$u = -\frac{f}{p} \ln \left\{ \frac{E(N_{00})}{f} \right\} \tag{7}$$

Dividing Eq. (5) by Eq. (2), we have:

$$\frac{E(N_{11})}{E(N_{00})} = \frac{b'p}{f} \Rightarrow b' = \frac{fE(N_{11})}{pE(N_{00})} \tag{8}$$

According to Eqs. (7)(8), g is expressed as follows.

$$g = u - b' = -\frac{f}{p} \ln \left\{ \frac{E(N_{00})}{f} \right\} - \frac{fE(N_{11})}{pE(N_{00})} \tag{9}$$

By substituting N_{00} for $E(N_{00})$ and N_{11} for $E(N_{11})$ in Eq. (9), we get the estimator of g as follows.

$$\hat{g} = -\frac{f}{p} \ln \left\{ \frac{N_{00}}{f} \right\} - \frac{fN_{11}}{pN_{00}} \tag{10}$$

The estimator in Eq. (10) specifies how to use the observed N_{00} and N_{11} to estimate the cardinality g of genuine tags.

D. Variance of Estimator

The proposed estimator has an inherent variance due to the probabilistic nature of REB. The following lemma calculates the expression of the estimator variance.

Lemma 1. *Let f and p be the frame size and the persistence probability, respectively, b' be the size of $B - G$, and u be the size of $B \cup G$. The variance of the estimator is as follows.*

$$Var(\hat{g}) = \frac{1}{fp^2} e^{\frac{up}{f}} \left(b'^2 p^2 + f^2 - b'fp \right) - \frac{f}{p^2} \tag{11}$$

Proof: According to Eq. (10), $\hat{g}$ is a function of N_{00} and N_{11}. Hence, we denote $\hat{g}$ as $\varphi(N_{00}, N_{11})$, that is, $\hat{g} = \varphi(N_{00}, N_{11})$. We present the Taylor's series expansion [31] of function $\varphi(N_{00}, N_{11})$ around (η_0, η_1), where $\eta_0 = E(N_{00})$ and $\eta_1 = E(N_{11})$.

$$\varphi(N_{00}, N_{11}) \approx \varphi(\eta_0, \eta_1) + (N_{00} - \eta_0) \frac{\partial\varphi}{\partial N_{00}} + (N_{11} - \eta_1) \frac{\partial\varphi}{\partial N_{11}} \tag{12}$$

We have the following equation by taking expectation of both sides of Eq. (12).

$$\begin{aligned} &E\{\varphi(N_{00}, N_{11})\} \\ =&\varphi(\eta_0, \eta_1) + \frac{\partial\varphi}{\partial N_{00}} E(N_{00} - \eta_0) + \frac{\partial\varphi}{\partial N_{11}} E(N_{11} - \eta_1) = g \end{aligned} \tag{13}$$

Eq. (13) infers that $\hat{g}$ is an unbiased estimator of g. In what follows, we calculate the variance of $\hat{g}$.

$$\begin{aligned} &Var(\hat{g}) = E\{\hat{g} - E(\hat{g})\}^2 \\ =&E\left\{ (N_{00} - \eta_0) \frac{\partial\varphi}{\partial N_{00}} + (N_{11} - \eta_1) \frac{\partial\varphi}{\partial N_{11}} \right\}^2 \\ =&Var(N_{00}) \left\{ \frac{\partial\varphi}{\partial N_{00}} \right\}^2 + Var(N_{11}) \left\{ \frac{\partial\varphi}{\partial N_{11}} \right\}^2 + \\ &2Cov(N_{00}, N_{11}) \left\{ \frac{\partial\varphi}{\partial N_{00}} \right\} \left\{ \frac{\partial\varphi}{\partial N_{11}} \right\} \end{aligned} \tag{14}$$

As required by Eq. (14), we need to calculate the covariance $Cov(N_{00}, N_{11}) = E(N_{00}N_{11}) - E(N_{00})E(N_{11})$, in which $E(N_{00}N_{11})$ is calculated below.

$$\begin{aligned} &E(N_{00}N_{11}) = \sum_{x=0}^{f} \sum_{y=0}^{f-x} xyP\{N_{00} = x \wedge N_{11} = y\} \\ =&\sum_{x=0}^{f} \sum_{y=0}^{f-x} xy \binom{f}{x} \{p_{00}\}^x \binom{f-x}{y} \{p_{11}\}^y \{1 - p_{00} - p_{11}\}^{f-x-y} \\ =&p_{11} \sum_{x=1}^{f} f\{f - x\} \binom{f-1}{x-1} \{p_{00}\}^x \{1 - p_{00}\}^{f-x-1} \\ =&\frac{p_{00}p_{11}f^2}{1 - p_{00}} \sum_{x=1}^{f} \binom{f-1}{x-1} \{p_{00}\}^{x-1} \{1 - p_{00}\}^{f-x} \\ &- \frac{f\{f-1\}\{p_{00}\}^2 p_{11}}{1 - p_{00}} \sum_{x=2}^{f} \binom{f-2}{x-2} \{p_{00}\}^{x-2} \{1 - p_{00}\}^{f-x} \\ &- \frac{fp_{00}p_{11}}{1 - p_{00}} \sum_{x=1}^{f} \binom{f-1}{x-1} \{p_{00}\}^{x-1} \{1 - p_{00}\}^{f-x} \\ =&\frac{p_{00}p_{11}f^2}{1 - p_{00}} - \frac{f\{f-1\}\{p_{00}\}^2 p_{11}}{1 - p_{00}} - \frac{fp_{00}p_{11}}{1 - p_{00}} \\ =&f\{f - 1\} p_{00}p_{11} \end{aligned} \tag{15}$$

As also required by Eq. (14), we calculate the first-order partial derivatives of $\varphi(N_{00}, N_{11})$ as follows.

$$\begin{aligned} \frac{\partial\varphi}{\partial N_{00}} |_{N_{11}=\eta_1}^{N_{00}=\eta_0} &= e^{\frac{up}{f}} \left(\frac{b'}{f} - \frac{1}{p}\right) \\ \frac{\partial\varphi}{\partial N_{11}} |_{N_{11}=\eta_1}^{N_{00}=\eta_0} &= -\frac{1}{p} e^{\frac{up}{f}} \end{aligned} \tag{16}$$

We have obtained $E(N_{00}N_{11})$ in Eq. (15), $E(N_{00})$ in Eq. (2), and $E(N_{11})$ in Eq. (5). $Cov(N_{00}, N_{11})$ is calculated as below.

$$Cov(N_{00}, N_{11}) = E(N_{00}N_{11}) - E(N_{00})E(N_{11}) = -b'pe^{-\frac{2up}{f}} \tag{17}$$

By combining Eqs. (3) (6) (16) (17) into Eq. (14), we then get the estimator variance shown in Eq. (11). ∎

The simulation results in Fig. 3 demonstrate that $\hat{g}$ in Eq. (10) is an unbiased estimator of the genuine tags. And the simulation results in Fig. 4 reveal that the estimator variance observed from simulations match well with the theoretical value calculated by Eq. (11).

E. Refined Estimation with k Frames

Because of probabilistic variance, the estimate $\hat{g}$ got from a single frame is difficult to meet the predefined accuracy. By the law of large number [32], we issue k independent frames and

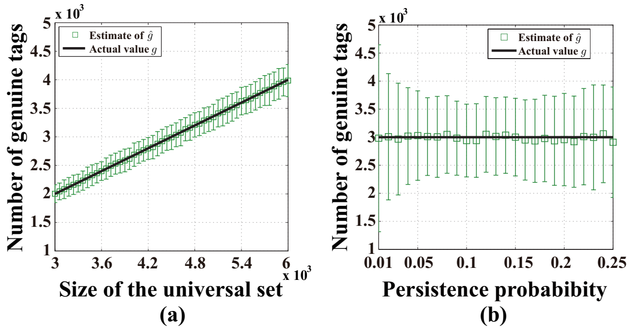


Fig. 3. Comparing the estimates outputted by REB with the actual number of genuine tags. (a) varying $u = |U|$ from 3000 to 6000, where $|B-G|$:$|B \cap G|$:$|G-B|$ = 1:1:1. (b) varying p from 0.01 to 0.25, where $|B-G|$=$|B \cap G|$=$|G-B|$=2000.

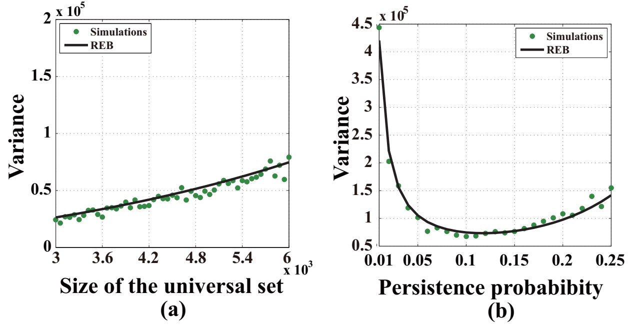


Fig. 4. Comparing the estimator variance observed from simulations with the theoretical value calculated by Eq. (11). (a) varying $u = |U|$ from 3000 to 6000, where $|B-G|$:$|B \cap G|$:$|G-B|$ = 1:1:1. (b) varying p from 0.01 to 0.25, where $|B-G|$=$|B \cap G|$=$|G-B|$=2000.

use the average estimation result $\hat{g}_{\overline{k}} = \frac{1}{k}\sum_{j=1}^{k}\hat{g}_j$ to achieve a more accurate estimate in REB, where $\hat{g}_j$ is the estimate of g derived from the j-th frame. We propose Theorem 1 to give the expression to determine if the frame number is adequate to ensure that REB achieves the required (α, β) accuracy.

Theorem 1. *Let α be the required confidence interval, β be the required reliability, f_j and p_j be the frame size and persistence probability used in the j-th frame, respectively. The average estimate $\hat{g}_{\overline{k}}$ obtained from k frames satisfies the accuracy requirement $P\left\{|\hat{g}_{\overline{k}} - g| \leq \alpha g\right\} \geq \beta$ when the frame number k satisfies the following inequality.*

$$k \geq \frac{Z_\beta}{g\alpha}\sqrt{\sum_{j=1}^{k}\left\{\frac{1}{f_j p_j^2}e^{\frac{up_j}{f_j}}\left(b'^2p_j^2 + f_j^2 - b'f_jp_j\right) - \frac{f_j}{p_j^2}\right\}} \tag{18}$$

Proof: We define $\widehat{g_{\overline{k}}} = \frac{1}{k}\sum_{j=1}^{k}\hat{g}_j$ as the *average* estimate of k successive frames, where $\hat{g}_j$ is the estimate of the j-th frame, $j \in [1, k]$. The reader initializes each frame with a different random seed. Hence, the estimate $\hat{g}_j$ is independent of each other. Thus, we have $E(\widehat{g_{\overline{k}}}) = \frac{1}{k}\sum_{j=1}^{k}E(\hat{g}_j) = g$; and $Var(\widehat{g_{\overline{k}}}) = \frac{1}{k^2}\sum_{j=1}^{k}Var(\hat{g}_j)$. Clearly, the average estimate $\widehat{g_{\overline{k}}}$ still converges to the actual cardinality g. Given a required reliability β, the actual confidence interval is within $[g - Z_\beta\sqrt{Var(\widehat{g_{\overline{k}}})}, g + Z_\beta\sqrt{Var(\widehat{g_{\overline{k}}})}]$, where Z_β is a percentile of β, *e.g.*, if $\beta = 95\%$, Z_β will be 1.96. To guarantee the required confidence α, we should guarantee:

$$\begin{cases} g + Z_\beta\sqrt{Var(\widehat{g_{\overline{k}}})} \leq g + g\alpha \\ g - Z_\beta\sqrt{Var(\widehat{g_{\overline{k}}})} \geq g - g\alpha \end{cases}$$

Substituting $\frac{1}{k^2}\sum_{j=1}^{k}Var(\hat{g}_j)$ for $Var(\widehat{g_{\overline{k}}})$ and solving the above inequalities, we have:

$$k \geq \frac{Z_\beta}{g\alpha}\sqrt{\sum_{j=1}^{k}Var(\hat{g}_j)} \tag{19}$$

According to Eq. (11), we have $Var(\hat{g}_j) = \frac{1}{f_jp_j^2}e^{\frac{up_j}{f_j}}(b'^2p_j^2 + f_j^2 - b'f_jp_j) - \frac{f_j}{p_j^2}$. Substituting it into Eq. (19), we get the inequality in Eq. (18). ■

III. Parameter Optimization

This section proposes rigorous theoretical analysis to optimize the values of frame size f and persistence probability p, to minimize the time cost and energy cost of REB, respectively. Then, we reveal a fundamental trade-off between the time cost and energy cost of REB, which can be flexibly adjusted by the system parameters. Through our analytical framework, we are able to configure the protocol parameters to achieve the desirable performance.

A. Minimizing Time Cost

Let $\mathcal{T}$ represent the total time cost of REB, t_ν represent the time that the reader takes to transmit the parameters for frame initialization, t_μ represent the duration of each slot in the frame, and k represent the frame number required to satisfy the $\langle\alpha, \beta\rangle$ accuracy. Then, we have $\mathcal{T} = k \times (t_\nu + f \times t_\mu)$. We assume the values of f and p are consistently the same across all the frames. Thus, the required frame number k is transformed into:

$$k = \frac{Z_\beta^2}{g^2\alpha^2}\left\{\frac{1}{fp^2}e^{\frac{up}{f}}\left(b'^2p^2 + f^2 - b'fp\right) - \frac{f}{p^2}\right\} \tag{20}$$

Then, the expression of the time cost $\mathcal{T}$ is as follows.

$$\mathcal{T} = \frac{Z_\beta^2(t_\nu + ft_\mu)}{g^2\alpha^2}\left\{\frac{1}{fp^2}e^{\frac{up}{f}}\left(b'^2p^2 + f^2 - b'fp\right) - \frac{f}{p^2}\right\} \tag{21}$$

1) Optimizing p: Since f and p are correlated to minimize the total execution time, we first fix the value of f to get an optimized p. The following theorem calculates the optimal persistence probability p_{op} to minimize the time cost $\mathcal{T}$.

Theorem 2. *Given the sizes of $B \cup G$ and $B - G$ (*i.e.*, u and b'), and the frame size f, if $\frac{\partial\mathcal{T}}{\partial p}|_{(p=1)} \geq 0$, we can solve the equation $e^{\frac{up}{f}}\left\{\frac{ub'^2}{f^2} + \frac{u+b'}{p^2} - \frac{ub'}{fp} - \frac{2f}{p^3}\right\} + \frac{2f}{p^3} = 0$ to obtain the optimal persistence probability p_{op} to minimize the time cost $\mathcal{T}$. On the contrary, if $\frac{\partial\mathcal{T}}{\partial p}|_{(p=1)} < 0$, p_{op} should be 1.*

Proof: We calculate the expression of $\frac{\partial\mathcal{T}}{\partial p}$ as follows.

$$\frac{\partial\mathcal{T}}{\partial p} = \frac{Z_\beta^2(t_\nu + ft_\mu)}{g^2\alpha^2}\left\{e^{\frac{up}{f}}\mathcal{D} + \frac{2f}{p^3}\right\}, \tag{22}$$

where $\mathcal{D} = \frac{ub'^2}{f^2} + \frac{u+b'}{p^2} - \frac{ub'}{fp} - \frac{2f}{p^3}$. First of all, we prove that $\lim\limits_{p \to 0} \frac{\partial \mathcal{T}}{\partial p} < 0$. Then, we prove that the second-order derivative $\frac{\partial^2 \mathcal{T}}{\partial p^2} > 0$, *i.e.*, $\frac{\partial \mathcal{T}}{\partial p}$ is a monotonously increasing function of $p \in (0, 1]$. The corresponding proof can be found in our supplementary file or [33]. When $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} \geq 0$, obviously, there is a value of $p_{op} \in (0, 1]$ that makes $\frac{\partial \mathcal{T}}{\partial p} = 0$. Moreover, we have $\frac{\partial \mathcal{T}}{\partial p} < 0$ when $p < p_{op}$; and $\frac{\partial \mathcal{T}}{\partial p} > 0$ when $p > p_{op}$. Thus, the time cost $\mathcal{T}$ achieves the minimum value when p_{op} is set to the value that satisfies the equation of $\frac{\partial \mathcal{T}}{\partial p} = 0$, by transforming which, we get the equation in theorem statement.

On the other hand, if $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} < 0$, we assert that the first-order derivative $\frac{\partial \mathcal{T}}{\partial p}$ is always less than 0 for any value $p \in (0, 1]$. In this case, $\mathcal{T}$ is a monotonously decreasing function of $p \in (0, 1]$. Therefore, the optimal p_{op} should be set to 1. ■

2) Optimizing f: This section investigates the optimization of f to minimize the time cost $\mathcal{T}$. The following theorem reveals that we should directly set the frame size f to 512 in the large-scale RFID systems where the sizes of $|B \cup G|$ and $|B - G|$ are larger than 512. And in small-scale RFID systems, we will invoke an exhaustive search-based method to find the optimal frame size f. Fortunately, the range of f is an integer and is less than 512. Given a fixed frame size f, we can obtain the corresponding optimal p_{op} according to Theorem 2. By comparing all pairs of $\langle f, p_{op} \rangle$, we can obtain the optimal parameter pair that minimizes the time cost $\mathcal{T}$. Formally, we need to solve the following optimization problem.

$$\begin{aligned} & \text{Minimize } \mathcal{T} = k \times (t_\nu + f \times t_\mu) \\ \text{s.t. } & f \leq 512. \\ & p \text{ is calculated by Theorem 2.} \\ & k = \frac{Z_\beta^2}{g^2\alpha^2} \left\{ \frac{1}{fp^2} e^{\frac{up}{f}} \left(b'^2p^2 + f^2 - b'fp \right) - \frac{f}{p^2} \right\}. \end{aligned} \tag{23}$$

Theorem 3. *Given a large-scale RFID systems where the sizes of $B \cup G$ and $B - G$ are larger than 512 (*i.e.*,$u > b' > 512$), we should set the frame size f_{op} to 512 for achieving the best time-efficiency.*

Proof: Setting $p = 1$ in Eq. (22), we have $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} = \frac{Z_\beta^2(t_\nu + ft_\mu)}{g^2\alpha^2} \left\{ e^{\frac{u}{f}} \mathcal{M} + 2f \right\}$, where $\mathcal{M} = \frac{ub'^2}{f^2} + u + b' - \frac{ub'}{f} - 2f$. We first prove that $\lim\limits_{f \to 0} \{ \frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} \} > 0$ and $\lim\limits_{f \to +\infty} \{ \frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} \} < 0$ (refer to our supplementary file or [33]). Then, we prove that $\frac{\partial^2 \mathcal{T}}{\partial p \partial f}|_{(p=1)} < 0$, which infers that $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1)}$ is a monotonously decreasing function with respect of f (refer to our supplementary file or [33]). Hence, there exists a value of $f_\triangle > 0$ that makes $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1, f=f_\triangle)} = 0$. And $\forall f_* \in (0, f_\triangle]$, we have $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1, f=f_*)} \geq 0$. According to Theorem 2, we could find an optimal persistence probability p_* corresponding to f_*, which satisfies the following equation.

$$e^{\frac{up_*}{f}} \left\{ \frac{ub'^2}{f_*^2} + \frac{u+b'}{p_*^2} - \frac{ub'}{f_*p_*} - \frac{2f_*}{p_*^3} \right\} + \frac{2f}{p_*^3} = 0 \tag{24}$$

According to Eq. (21), we get the time cost $\mathcal{T}_{\langle f_*, p_* \rangle}$ when using frame size f_* and persistence probability p_*:

$$\mathcal{T}_{\langle f_*, p_* \rangle} = t_\nu \times k_{\langle f_*, p_* \rangle} + \frac{Z_\beta^2 t_\mu}{g^2\alpha^2} \times \psi, \tag{25}$$

where $\psi = \frac{1}{p_*^2} e^{\frac{up_*}{f_*}} \left(b'^2p_*^2 + f_*^2 - b'fp_* \right) - \frac{f_*^2}{p_*^2}$. We calculate the first-order derivative $\frac{\partial \psi}{\partial f_*}$ as follows.

$$\frac{\partial \psi}{\partial f_*} = e^{\frac{up_*}{f_*}} \left\{ -\frac{ub'^2p_*}{f_*^2} - \frac{u}{p_*} - \frac{b}{p_*} + \frac{ub'}{f_*} + \frac{2f}{p_*^2} \right\} - \frac{2f_*}{p_*^2} \tag{26}$$

Consider the equation in Eq. (24), we find that $\frac{\partial \psi}{\partial f_*} = 0$. In Theorem 4, we have proved that $(gp\omega)k$ is a decreasing function of f. Hence, k is also a decreasing function of f, we then have $\frac{k_{\langle f_*, p_* \rangle}}{\partial f_*} < 0$. Further, we have $\frac{\partial \mathcal{T}_{\langle f_*, p_* \rangle}}{\partial f_*} = t_\nu \times \frac{k_{\langle f_*, p_* \rangle}}{\partial f_*} + \frac{Z_\beta^2 t_\mu}{g^2\alpha^2} \times \frac{\partial \psi}{\partial f_*} < 0$. That is, the time cost $\mathcal{T}_{\langle f_*, p_* \rangle}$ is a monotonously decreasing function of $f_* \in (0, f_\triangle]$.

Next, we use the method of reductio ad absurdum to prove that $f_\triangle$ is larger than 512 when $u > b' > 512$. If $f_\triangle < 512$ and $u > b' > 512$, we have $\mathcal{M} = \frac{ub'^2}{f^2} + u + b' - \frac{ub'}{f} - 2f > 0$. Then, $\frac{\partial \mathcal{T}}{\partial p}|_{(p=1)} = \frac{Z_\beta^2(t_\nu + ft_\mu)}{g^2\alpha^2} \left\{ e^{\frac{u}{f}} \mathcal{M} + 2f \right\} = 0$ has no solution, which is an absurdum. Hence, when $u > b' > 512$, it is necessary that $f_\triangle > 512$. Recall that $\mathcal{T}_{\langle f_*, p_* \rangle}$ is a monotonously decreasing function of $f_* \in (0, f_\triangle]$. Therefore, we should set f to 512 to minimize $\mathcal{T}$ in this case. ■

B. Minimizing Energy Cost

Energy cost is an important metric when designing an RFID application protocol for the systems where battery-powered active tags are used. This section investigates the impact of f and p on the energy-efficiency of REB. Note that, we only take the energy consumption of genuine tags into consideration because recharging or replacing the batteries of a large number of active tags is seriously laborious. In contrary, the readers and blocker tag devices can be easily recharged because their number is normally limited, hence, we ignore the energy consumption of readers and the blocker tag devices. Following prior RFID literature [34] that also concerns the energy-efficiency, we use the total number of tag transmissions to measure the total energy consumption, because transmitting data consumes much more power than receiving data. In an arbitrary frame, each of the g genuine tags has a probability p of responding RN16 to the reader. Therefore, the total power consumed by the genuine tags across k frames, denoted as $\mathcal{E}$, can be given by $\mathcal{E} = kgp\omega$, where ω represents the energy cost on an active tag for transmitting a response of RN16.

1) Optimizing f: The following theorem infers that we should set $f = 512$ to minimize the energy cost $\mathcal{E}$ of REB.

Theorem 4. *Given any value of the persistence probability* $p \in (0, 1]$, *we should set the frame size f to its maximum value of 512 to minimize the energy cost $\mathcal{E}$.*

Proof: We will prove that the required energy cost $\mathcal{E}$ is a monotonously decreasing function of the frame size f. The sufficient and necessary condition is that the first-order derivative of $\mathcal{E}$ with respect to f is always less than 0. Therefore, we calculate the first-order derivative of $\mathcal{E}$ with respect to f as $\frac{\partial \mathcal{E}}{\partial f} = \frac{Z_\beta^2 \omega \mathcal{X}}{g\alpha^2 p}$, where $\mathcal{X}$ is given by

$$\mathcal{X} = e^{\frac{up}{f}} \left\{ 1 - \frac{up}{f} + \frac{(u-b')b'p^2}{f^2} - \frac{ub'^2p^3}{f^3} \right\} - 1 \tag{27}$$

To prove $\frac{\partial \mathcal{E}}{\partial f} < 0$, we only need to prove $\mathcal{X} < 0$. Next, we will prove $\mathcal{X} < 0$ through two steps. First, we will prove $\mathcal{X}$ is a monotonically increasing function of the frame size f. Second, we will prove $\lim\limits_{f \to +\infty} \mathcal{X}$, *i.e.*, the upper bound on $\mathcal{X}$, is always less than 0. Then, we can know that $\mathcal{X}$ is always less than 0 for any frame size $f \leq 512$.

We calculate the expression of the first-order derivative of $\mathcal{X}$ with respect to f as follows. Additionally, we observe from Eq. (28) that $\frac{\partial \mathcal{X}}{\partial f}$ is always larger than 0. Hence, $\mathcal{X}$ achieves its largest value when $f \to +\infty$.

$$\begin{aligned}\frac{\partial \mathcal{X}}{\partial f} =& e^{\frac{up}{f}} \left\{ \left(\frac{up}{\sqrt{2f^3}} - \frac{\sqrt{2}b'p}{\sqrt{f^3}} \right)^2 + \left(\frac{up}{\sqrt{2f^3}} - \frac{ub'p^2}{\sqrt{f^5}} \right)^2 \right. \\ &\left. + \frac{4ub'^2p^3}{f^4} + \frac{(\sqrt{2}-1)u^2b'p^3}{f^4} \right\}\end{aligned} \tag{28}$$

In the following, we will prove that the *upper bound* on $\mathcal{X}$, *i.e.*, the value of $\mathcal{X}$ when $f \to +\infty$, is always less than 0. Consider the expression of $\mathcal{X}$. For simplicity, we denote $1 - \frac{up}{f} + \frac{(u-b')b'p^2}{f^2} - \frac{ub'^2p^3}{f^3}$ as $\mathcal{Q}$. Thus, $\mathcal{X} = e^{\frac{up}{f}} \mathcal{Q} - 1$. To prove $\mathcal{X} < 0$, we only need to prove $\mathcal{Q} < e^{-\frac{up}{f}}$. Using the Taylor series expansion, we have:

$$\begin{aligned}e^{-\frac{up}{f}} =& 1 - \frac{up}{f} + \frac{(up)^2}{2!f^2} - \frac{(up)^3}{3!f^3} + \frac{(up)^4}{4!f^4} - \frac{(up)^5}{5!f^5} \\ &+ \cdots + \left\{ \frac{(up)^\lambda}{\lambda! f^\lambda} - \frac{(up)^{\lambda+1}}{(\lambda+1)! f^{\lambda+1}} \right\} + \cdots,\end{aligned} \tag{29}$$

where $\lambda = 6, 8, 10, \cdots$. Consider the tail item pair of Eq. (29). Since $\left| \frac{(up)^\lambda}{\lambda! f^\lambda} \right| \Big/ \left| -\frac{(up)^{\lambda+1}}{(\lambda+1)! f^{\lambda+1}} \right| = \frac{(\lambda+1)f}{up} > 1$ when $f \to +\infty$, we have $\left\{ \frac{(up)^\lambda}{\lambda! f^\lambda} - \frac{(up)^{\lambda+1}}{(\lambda+1)! f^{\lambda+1}} \right\} > 0$. Therefore, we have $\mathcal{Y} = 1 - \frac{up}{f} + \frac{(up)^2}{2!f^2} - \frac{(up)^3}{3!f^3} + \frac{(up)^4}{4!f^4} - \frac{(up)^5}{5!f^5} < e^{-\frac{up}{f}}$. To prove $\mathcal{Q} < e^{-\frac{up}{f}}$, we only need to prove $\mathcal{Q} < \mathcal{Y}$. Hence, we calculate the expression of $\mathcal{Y} - \mathcal{Q}$ as follows.

$$\begin{aligned}\mathcal{Y} - \mathcal{Q} =& \frac{u^2p^2}{2f^2} - \frac{ub'p^2}{f^2} + \frac{b'^2p^2}{f^2} + \frac{ub'^2p^3}{f^3} - \frac{u^3p^3}{6f^3} \\ &+ \frac{u^4p^4}{30f^4} + \frac{u^4p^4}{120f^4} - \frac{u^5p^5}{120f^5} \\ =& \left(\frac{up}{2f} - \frac{b'p}{f} \right)^2 + \frac{ub'^2p^3}{f^3} + \left(\frac{up}{2f} - \frac{u^2p^2}{\sqrt{30}f^2} \right)^2 \\ &+ \left(\frac{u^3p^3}{\sqrt{30}f^3} - \frac{u^3p^3}{6f^3} \right) + \frac{u^4p^4}{120f^4} \left(1 - \frac{up}{f} \right)\end{aligned} \tag{30}$$

Since $\frac{up}{f}$ is less than 1 when $f \to +\infty$, the expression of $\mathcal{Y} - \mathcal{Q}$ in Eq. (30) is always larger than 0. Then, we have $\mathcal{Q} < \mathcal{Y} < e^{-\frac{up}{f}}$. Obviously, we have $\mathcal{X} = e^{\frac{up}{f}} \mathcal{Q} - 1 < 0$. Accordingly, the first-order derivative $\frac{\partial \mathcal{E}}{\partial f} = \frac{Z_\beta^2 \omega \mathcal{X}}{g\alpha^2 p} < 0$. Hence, for any fixed value of p, we should set f to its maximum value (*i.e.*, 512 in practice), to minimize the energy cost. ■

2) Optimizing p: The following theorem infers that the energy cost $\mathcal{E} = kgp\omega$ is a monotonously increasing function of the persistence probability p. Hence, we should set the persistence probability $p \in (0, 1]$ as small as possible to decrease the energy cost $\mathcal{E}$.

Theorem 5. *Given any value of the frame size f, the energy cost $\mathcal{E} = kgp\omega$ is a monotonously increasing function with respect to the persistence probability p.*

Proof: We first prove that the second-order derivative $\frac{\partial^2 \mathcal{E}}{\partial p^2}$ is always larger than 0. Then, we can know the first-order derivative $\frac{\partial \mathcal{E}}{\partial p}$ is a monotonously increasing function of p. Thus, $\frac{\partial \mathcal{E}}{\partial p}$ gets close to its lower bound when $p \to 0$. Further, we prove that the lower bound $\lim\limits_{p \to 0} \frac{\partial \mathcal{E}}{\partial p}$ is always larger than 0. Then, we can assert that $\frac{\partial \mathcal{E}}{\partial p}$ is larger than 0 for any value of $p \in (0, 1]$. Specifically, we calculate the second-order derivative of $\mathcal{E}$ as follows.

$$\frac{\partial^2 \mathcal{E}}{\partial p^2} = \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ e^{\frac{up}{f}} \times \mathcal{I} - \frac{2f}{p^3} \right\}, \tag{31}$$

where $\mathcal{I} = \frac{u^2b'^2p}{f^3} + \frac{2ub'^2 - u^2b'}{f^2} + \frac{u^2}{fp} - \frac{2u}{p^2} + \frac{2f}{p^3}$, which can be transformed as follows.

$$\mathcal{I} = \left\{ \frac{ub'\sqrt{p}}{f\sqrt{f}} - \frac{u}{2\sqrt{fp}} \right\}^2 + \frac{2ub'^2}{f^2} + \left\{ \frac{\sqrt{3}u}{2\sqrt{fp}} - \frac{\sqrt{2f}}{p\sqrt{p}} \right\}^2 + \frac{(\sqrt{6}-2)u}{p^2} \tag{32}$$

Using Taylor series expansion, we have $e^{\frac{up}{f}} > 1 + \frac{up}{f} + \frac{u^2p^2}{2f^2} + \frac{u^3p^3}{6f^3} + \frac{u^4p^4}{24f^4}$. Since Eq. (32) reveals that $\mathcal{I} > 0$, we have:

$$\begin{aligned}&\frac{\partial^2 \mathcal{E}}{\partial p^2} = \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ e^{\frac{up}{f}} \times \mathcal{I} - \frac{2f}{p^3} \right\} \\ &> \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ \left(1 + \frac{up}{f} + \frac{u^2p^2}{2f^2} + \frac{u^3p^3}{6f^3} + \frac{u^4p^4}{24f^4} \right) \times \mathcal{I} - \frac{2f}{p^3} \right\} \\ &= \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ \left(\frac{u^3b'p^2\sqrt{p}}{2\sqrt{6f^7}} - \frac{u^3p\sqrt{p}}{4\sqrt{6f^5}} \right)^2 + \frac{u^6p^3}{32f^5} + \frac{u^3}{12f^2} \right. \\ &\quad + \left(\frac{u^2b'p^2\sqrt{u}}{2f^3} - \frac{u^2p\sqrt{u}}{6f^2} \right)^2 + \left(\frac{ub'p\sqrt{2u}}{f^2} - \frac{u\sqrt{u}}{2\sqrt{2}f} \right)^2 \\ &\quad + \left(\frac{b'\sqrt{2u}}{f} - \frac{u\sqrt{u}}{2\sqrt{2}f} \right)^2 + \left(\frac{u^2b'p\sqrt{p}}{2\sqrt{f^5}} - \frac{u^2\sqrt{p}}{2f\sqrt{f}} \right)^2 \\ &\quad \left. + \frac{7u^4b'^2p^3}{12f^5} + \frac{3u^2b'^2p}{f^3} + \frac{u^5p^2}{18f^4} \right\} > 0\end{aligned} \tag{33}$$

Eq. (33) indicates that $\frac{\partial^2 \mathcal{E}}{\partial p^2}$ is always larger than 0, hence, $\frac{\partial \mathcal{E}}{\partial p}$ is a monotonously increasing function of $p \in (0, 1]$. Therefore, $\frac{\partial \mathcal{E}}{\partial p}$ gets close to its lower bound when $p \to 0$.

Next, we prove the lower bound, *i.e.*, $\lim\limits_{p \to 0} \frac{\partial \mathcal{E}}{\partial p}$, is always less than 0. First, we calculate its expression as follows.

$$\frac{\partial \mathcal{E}}{\partial p} = \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ e^{\frac{up}{f}} \left(\frac{ub'^2p}{f^2} + \frac{u}{p} + \frac{b'^2}{f} - \frac{ub'}{f} - \frac{f}{p^2} \right) + \frac{f}{p^2} \right\} \tag{34}$$

For clarity, we denote $\frac{ub'^2p}{f^2} + \frac{u}{p} + \frac{b'^2}{f} - \frac{ub'}{f} - \frac{f}{p^2}$ as $\mathcal{L}$. Then, $\frac{\partial \mathcal{E}}{\partial p} = \frac{Z_\beta^2 \omega}{g\alpha^2} \left\{ e^{\frac{up}{f}} \mathcal{L} + \frac{f}{p^2} \right\}$. To prove $\frac{\partial \mathcal{E}}{\partial p} > 0$ when $p \to 0$, we only need to prove $\mathcal{L} > -\frac{f}{p^2} e^{-\frac{up}{f}}$ when $p \to 0$. Consider the tail pair of items in Eq. (29). Since $\left| \frac{(up)^\lambda}{\lambda! f^\lambda} \right| \Big/ \left| -\frac{(up)^{\lambda+1}}{(\lambda+1)! f^{\lambda+1}} \right| = \frac{(\lambda+1)f}{up} > 1$ when $p \to 0$, we have $\left\{ \frac{(up)^\lambda}{\lambda! f^\lambda} - \frac{(up)^{\lambda+1}}{(\lambda+1)! f^{\lambda+1}} \right\} > 0$. Therefore, we have $\mathcal{Y} = 1 - \frac{up}{f} + \frac{(up)^2}{2!f^2} - \frac{(up)^3}{3!f^3} + \frac{(up)^4}{4!f^4} - \frac{(up)^5}{5!f^5} < e^{-\frac{up}{f}}$. Then, we have $-\frac{f}{p^2}\mathcal{Y} > -\frac{f}{p^2} e^{-\frac{up}{f}}$. To prove $\mathcal{L} > -\frac{f}{p^2} e^{-\frac{up}{f}}$, we only need to prove $\mathcal{L} > -\frac{f}{p^2}\mathcal{Y}$. To this

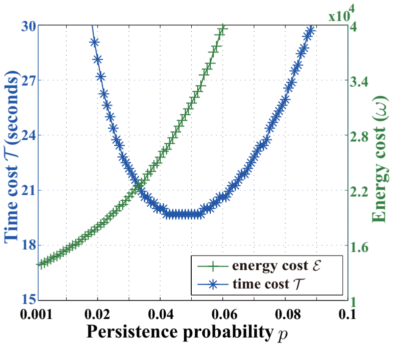


Fig. 5. Trade-off between time cost and energy cost: $b = 10000$, $b' = 5000$, $g = 10000$, $f = 512$, $\langle \alpha, \beta \rangle = \langle 0.02, 98\% \rangle$.

end, we calculate the expression of $\mathcal{L} + \frac{f}{p^2}\mathcal{Y}$ as follows.

$$\begin{aligned}\mathcal{L} + \frac{f}{p^2}\mathcal{Y} =& \frac{ub'^2 p}{f^2} + (\frac{b'}{\sqrt{f}} - \frac{u}{2\sqrt{f}})^2 + (\frac{u}{2\sqrt{f}} - \frac{u^2 p}{\sqrt{30 f^3}})^2 \\ &+ (\frac{u^3 p}{\sqrt{30} f^2} - \frac{u^3 p}{6 f^2}) + \frac{u^4 p^2}{120 f^3}(1 - \frac{up}{f})\end{aligned} \tag{35}$$

In Eq. (35), since $\frac{up}{f} < 1$ when $p \to 0$, we have $\mathcal{L} + \frac{f}{p^2}\mathcal{Y} > 0$. Then, we get $\mathcal{L} > -\frac{f}{p^2}\mathcal{Y} > -\frac{f}{p^2} e^{-\frac{up}{f}}$. Further, we have $\frac{\partial \mathcal{E}}{\partial p} = \frac{Z_\beta^2 \omega}{g \alpha^2} \left\{ e^{\frac{up}{f}} \mathcal{L} + \frac{f}{p^2} \right\} > 0$ when $p \to 0$. Since the lower bound $\lim\limits_{p \to 0} \frac{\partial \mathcal{E}}{\partial p}$ is always larger than 0, we have $\frac{\partial \mathcal{E}}{\partial p}$ is always larger than 0 for any value of $p \in (0, 1]$. That is, the energy cost $\mathcal{E}$ is a monotonously increasing function of p. ■

C. Trade-off between Time Cost and Energy Cost

When the active tags are used, we should jointly take time-efficiency and energy-efficiency into consideration, instead of separately considering these two metrics. Theorems 3 and 4 indicate that we should set the frame size f to 512 for achieving the best time-efficiency and energy-efficiency in large-scale RFID systems. Hence, without otherwise specification, we set the frame size to 512. In terms of persistence probability p, no value of p simultaneously minimizes the time cost and energy cost. Theorem 5 infers that we should set the value of p as small as possible to minimize the energy cost. However, as illustrated in Fig. 5, the execution time of REB will be significantly long when p is too small. Hence, we should make the trade-off between time cost and energy cost by controlling the setting of p, thereby satisfying the twofold requirements on both time-efficiency and energy-efficiency. Specifically, we first leverage Theorem 2 to obtain the optimal persistence probability p_{op} to minimize the time cost. Clearly, both time cost and energy cost increase as p increases within the range of $[p_{op}, 1]$. Hence, we adjust the persistence probability p within the range of $(0, p_{op}]$, while ignoring the range of $p \in (p_{op}, 1]$.

D. Dynamic Parameter Optimization

Recall from the last two sections that we need to know the values of g, u and b' when calculating the optimal frame size f and persistence probability p. However, the actual values of g, u and b' are not previously known, but are precisely what we want to estimate. Hence, we cannot calculate the optimal frame size f and persistence probability p at the start of REB.

For the first frame, we simply set the frame size f to 512 and set the persistence probability to $1/\hat{u}$, where the coarse value of $\hat{u}$ can be obtained by a fast method used in [13], [20], [23]. Specifically, the reader keeps issuing one-slot frames. The persistence probability follows a geometric distribution, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\cdots$, *i.e.*, the persistence probability in the γ-th single-slot frame is $\frac{1}{2^\gamma}$. This process does not terminate until an empty slot appears. Assuming that the ℓ-th slot is the first empty slot, we have a coarse estimation of $\hat{u} = 1.2897 \times 2^\ell$ [17]. Note that, this process takes at most 32 slots in practice.

For the $(x$+$1)$-th frame $(x \geq 1)$, we could use the information observed from previous frames to estimate the values of g, u and b'. Eq. (10) has given the estimator $\hat{g}$ of the genuine tags. According to Eqs. (7)(8), we give the estimators for u and b' as: $\hat{u} = -\frac{f}{p} \ln \left\{ \frac{N_{00}}{f} \right\}$ and $\hat{b'} = \frac{f N_{11}}{p N_{00}}$. Thus, we can use the temporary estimates averaged from the previous x frames, *i.e.*, $\hat{g}_{\overline{x}} = \frac{1}{x} \sum_{j=1}^{x} \hat{g}_j$, $\hat{u}_{\overline{x}} = \frac{1}{x} \sum_{j=1}^{x} \hat{u}_j$, and $\hat{b'}_{\overline{x}} = \frac{1}{x} \sum_{j=1}^{x} \hat{b'}_j$, to calculate the optimal values of f and p to initialize the next frame. We have observed from the simulation results that REB can quickly converge to the near-optimal setting of f and p after a few frames.

E. Avoiding Premature Termination

After each frame, says the x-th frame, REB gets $\hat{u}_{\overline{x}}$, $\hat{b'}_{\overline{x}}$ and $\hat{g}_{\overline{x}}$. However, their estimation is inaccurate due to the probabilistic variance. If we directly use them to calculate the R.H.S. of Eq. (18), the executed frame number x may have a chance to be larger than it, which is not true, and REB will have a premature termination (*i.e.*, the currently achieved accuracy has not met the required one yet). In the following, we present how to avoid the premature termination.

Eq. (11) has given the variance of estimator $\hat{g}$. Using the similar method in Section II-D, we can calculate the variance of the estimators $\hat{u}$ and $\hat{b'}$ as follows.

$$Var(\hat{u}) = \frac{f}{p^2} \left(e^{\frac{up}{f}} - 1 \right), \text{ and } Var(\hat{b'}) = e^{\frac{up}{f}} \left(\frac{b^2}{f} + \frac{b}{p} \right) \tag{36}$$

We use the average estimate $\hat{g}_{\overline{x}}$, $\hat{u}_{\overline{x}}$ and $\hat{b'}_{\overline{x}}$ got from previous x frames, and the frame size f_j as well as the persistence probability p_j used in the j-th frame to calculate the estimation variance in the j-th frame, *i.e.*, $Var(\hat{g}_j)$, $Var(\hat{u}_j)$, and $Var(\hat{b'}_j)$, where $j \in [1, x]$. Then, we could get the variance of the average estimates, *i.e.*, $Var(\hat{g}_{\overline{x}})$=$\frac{1}{x^2} \sum_{j=1}^{x} Var(\hat{g}_j)$, $Var(\hat{u}_{\overline{x}}) = \frac{1}{x^2} \sum_{j=1}^{x} Var(\hat{u}_j)$ and $Var(\hat{b'}_{\overline{x}}) = \frac{1}{x^2} \sum_{j=1}^{x} Var(\hat{b'}_j)$.

To avoid the premature termination, when calculating the R.H.S. of Eq. (18), we use $\hat{u}_\uparrow = \hat{u}_{\overline{x}} + \delta \sqrt{Var(\hat{u}_{\overline{x}})}$ to substitute u, $\hat{b'}_\uparrow = \hat{b'}_{\overline{x}} + \delta \sqrt{Var(\hat{b'}_{\overline{x}})}$ to substitute the *first* b', $\hat{b'}_\downarrow = \hat{b'}_{\overline{x}} - \delta \sqrt{Var(\hat{b'}_{\overline{x}})}$ to substitute the *second* b', $\hat{g}_\uparrow = \hat{g}_{\overline{x}} + \delta \sqrt{Var(\hat{g}_{\overline{x}})}$ to substitute g. The *three-sigma rule* [35] indicates that $\delta = 3$ is large enough. In Section IV, simulation results demonstrate that this tactic can effectively avoid the premature termination.

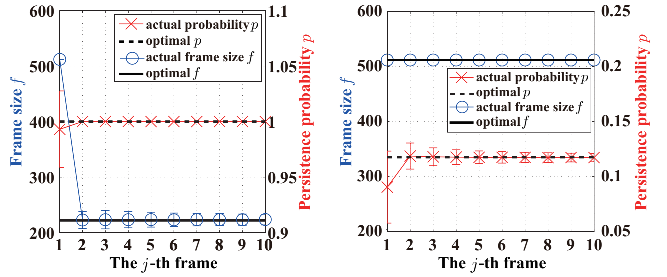


Fig. 6. Verifying the convergence of f and p. $\alpha = 2\%$, $\beta = 98\%$. (a) small-scale RFID system: $|B - G| = 100$, $|B \cap G| = 100$, $|G - B| = 100$; (b) large-scale RFID system: $|B-G| = 2000$, $|B \cap G| = 2000$, $|G-B| = 2000$.

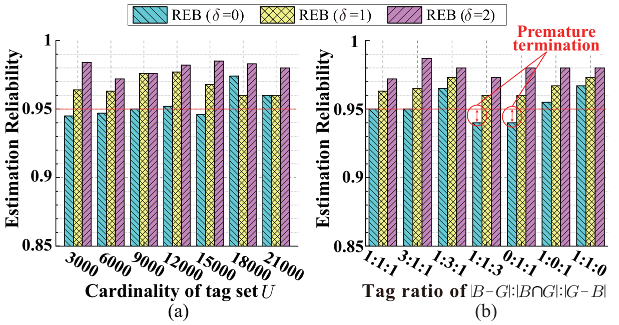


Fig. 7. Evaluating the reliability of REB. $\alpha = 5\%$, $\beta = 95\%$. (a) Tag ratio $|B - G|$:$|B \cap G|$:$|G - B|$ is fixed to $1 : 1 : 1$, and u varies from 3000 to 21000. (b) u is fixed to 9000, and tag ratio varies.

IV. Performance Evaluation

In this section, we conduct extensive simulations to evaluate the performance of REB. Besides REB, we implemented two tag identification protocols including Tree Hopping (TH) protocol [23] and the classical Enhanced Dynamic Framed Slotted ALOHA (EDFSA) protocol [22]. Although no RFID estimation protocol can correctly estimate tag cardinality with the existence of blocker tags, we implemented several representative RFID estimation protocols and conduct comparison side-by-side to investigate the strength and weakness of our REB. The implemented RFID estimation protocols include PZE/PCE [15], FNEB [36], LoF [17], and ART [13]. Following many RFID literature [13], [17], [23], we assume that the communication channel is error-free and a single reader covers all tags. We run each simulation 1000 times and report the average results.

A. Verifying the Convergence of f and p

The frame size f and persistence probability p that are actually picked by REB can approach to the near-optimal values after a few frames. The setting of parameters f and p is important to the performance of REB. To achieve the overall optimal f and p, it is necessary to know the values of u, b' and g before the execution of REB, which, however, are what we want to estimate. As aforementioned in Section III-D, we leverage the estimates of u, b' and g obtained from previous frames to guide the parameter optimization of the next frame. Fig. 6(a) plots the actual values of f and p used by each frame in a *small-scale* RFID system, where the optimal frame size f is 222 and the optimal persistence probability p is 1. On the contrary, Fig. 6(b) plots the actual values of f and p used by each frame in a *large-scale* RFID system, where the optimal frame size f is 512 and the optimal persistence probability p is 0.1177. The simulation results reveal that REB can obtain the near-optimal settings of f and p after just a few frames.

B. Evaluating the Actual Reliability

The δ-sigma method proposed in Section III-E can effectively avoid the premature termination. REB ($\delta = 1$) can always satisfy the required reliability. One of the most important performance metrics for estimation protocols is the actual reliability. In an arbitrary simulation, if the estimate $\hat{g}$ is within $[g(1-\alpha), g(1+\alpha)]$, we refer to it as a successful estimation. We record the *success times* among 1000 independent simulations. The ratio, *i.e.*, *success times*/1000, is treated as the *actual reliability*. Simulation results in Fig. 7 reveal that REB ($\delta = 0$) does not always meet the required reliability (*i.e.*, $\beta = 95\%$). The reason lies in the variances if directly using $\hat{u}_{\overline{k}}$, $\hat{b'}_{\overline{k}}$ and $\hat{g}_{\overline{k}}$ to determine the termination condition. By taking their variances into consideration, the proposed δ-sigma-based termination tactic effectively avoids the premature termination. Simulation results in Fig. 7 reveal that the actual reliability of REB ($\delta = 1$) and REB ($\delta = 2$) is always higher than the required one in various simulation environments.

C. Evaluating the Time-Efficiency

Under the premise that the required estimation reliability is guaranteed, the most important metric is time-efficiency. Recall that no existing estimation protocol can correctly approximate the cardinality of genuine tags in an RFID system with the presence of blocker tags. The only possible solution, to the best of our knowledge, is to perform the comprehensive identification protocols to identify the tags in the system. Hence, we compare REB with two representative identification protocols, *i.e.*, the Tree Hopping (TH) protocol [23] and the Enhanced Dynamic Framed Slotted ALOHA (EDFSA) protocol [22]. TH protocol terminates after it traverses the whole tree. Normally, EDFSA repeats frames round by round until all the tags are identified and keep silent. However, with the presence of blocker tag, EDFSA cannot terminate by itself due to the continuous collisions caused by the tags in $B \cap G$. In frame of EDFSA, only the IDs in $(B - G) \cup (G - B)$ have the chance to be identified. We denote the set of identified IDs as S_{ident}. Since the reader does not know whether all IDs in $(B - G) \cup (G - B)$ are completely identified or even what percentage of them are identified. For the sake of EDFSA, we assume it can "intelligently" terminate once $\{|(B - G) \cup (G - B)| - |S_{ident}|\} < |G| \times \alpha$. In Section IV-B, the simulation results demonstrate that REB ($\delta = 1$) can always satisfy the required estimation accuracy. Hence, we set $\delta = 1$ in REB without otherwise specification, and use REB to denote REB ($\delta = 1$) for simplicity.

1) Impact of Tag Cardinality: REB significantly outperforms the state-of-the-art identification protocols, regardless of the tag population $|U|$. To investigate the impact of tag

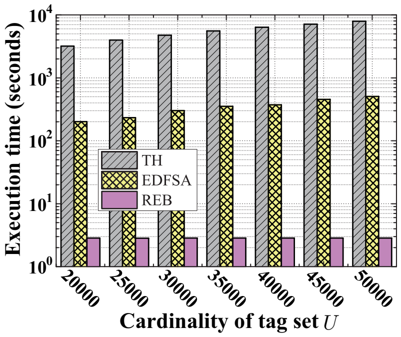


Fig. 8. Evaluating the time-efficiency of protocols with varying u. The tag ratio of $|B-G|$:$|B\cap G|$:$|G-B|$ is fixed to 1:1:1. $\alpha=5\%$, $\beta=95\%$.

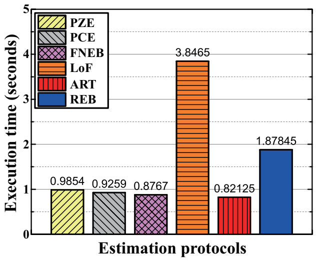


Fig. 10. Evaluating the time-efficiency of RFID estimation protocols without blocker tag. The number of tags is fixed to 10,000. $(\alpha,\beta)=(5\%,95\%)$.

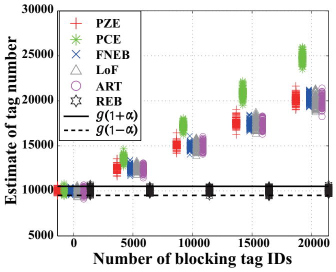
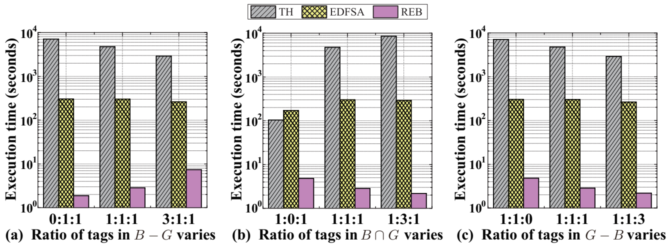


Fig. 11. Evaluating the accuracy of RFID estimation protocols with varying number of blocking tag IDs. $|G|=10,000$, $|B|$ varies from 0 to 20,000. $(\alpha,\beta)=(5\%,95\%)$.



(a) Ratio of tags in $B-G$ varies (b) Ratio of tags in $B\cap G$ varies (c) Ratio of tags in $G-B$ varies

Fig. 9. Evaluating the time-efficiency of protocols with varying tag ratio. u is fixed to 30000, and $\alpha=5\%$, $\beta=95\%$.

cardinality on the execution time of each protocol, we fix the tag ratio $|B-G|$:$|B\cap G|$:$|G-B|$ to 1:1:1, and vary u (indicating the system scale) from 20000 to 50000. The simulation results in Fig. 8 demonstrate that our REB significantly outperforms HT and EDFSA. For example, when $u=50000$, REB runs about 178 times faster than EDFSA, and nearly 2785 times faster than TH. Moreover, the execution time of TH and EDFSA grows linearly as u increases. In contrast, our REB has a stable execution time, which reveals its good scalability against tag cardinality u. The reason for the inefficiency of TH and EDFSA caused by the presence of blocker tag has been explained in Section V.

2) Impact of Tag Ratio: REB significantly outperforms the state-of-the-art identification protocols in terms of time-efficiency, regardless of the tag ratio $|B-G|:|B\cap G|:|G-B|$. The different tag ratio may have a significant impact on the execution time of protocols. Here, we fix $u=30000$, and evaluate the execution time of protocols with a varying tag ratio. The simulation results in Fig. 9 demonstrate that the proposed REB protocol consistently runs hundreds of times faster than the state-of-the-art TH protocol and EDFSA protocol. Moreover, the results in Fig. 9 clearly show the performance trend of the protocols with varying tag ratio.

Besides comparing with the tag identification protocols, we also conduct two set of simulations to compare our REB with the representative RFID estimation protocols including PZE/PCE [15], FNEB [36], LoF [17], and ART [13]. As shown in Fig. 10, we compare REB with prior RFID estimation protocols without blocker tag. The simulation results reveal that our REB is indeed slower than several RFID estimation protocols. For comprehensive comparison, we also conduct simulations to evaluate the estimation accuracy of each protocol with varying number of blocking tag IDs. We observe from Fig. 11 that all protocols can satisfy the required confidence interval $[g(1-\alpha),g(1+\alpha)]$ with a high reliability only without blocker tag (*i.e.*, the number of blocking tag IDs is 0). However, as the number of blocking tag IDs increases, only our REB is able to correctly return the accuracy-guaranteed estimates. Experimental results show that other protocols significantly deviate from the required confidence interval. All in all, our REB is the only protocol that can correctly perform RFID estimation with the presence of blocker tag.

D. Evaluating the Energy-Efficiency

This section evaluates the energy-efficiency of REB when the active RFID tags are used. Both TH and EDFSA only take time-efficiency into consideration. For fair comparison, REB uses the values of f and p that minimizes the time cost.

1) Impact of Tag Cardinality: REB significantly outperforms the state-of-the-art identification protocols in terms of energy-efficiency, regardless of the tag population $|U|$. In this set of simulations, we also fix the tag ratio of $|B-G|:|B\cap G|:|G-B|$ to 1:1:1, and vary the size of universal set U from 20000 to 50000. We make three main observations from the simulation results in Fig. 12. First, the energy cost of TH and EDFSA increases as u (indicating the system scale) increases. The underlying reason is that TH and EDFSA aim at identifying all the tags; their energy cost will be proportional to the number of genuine tags. Since the tag ratio is kept to 1:1:1, the increase of u also indicates an increase of the number of genuine tags. Accordingly, the energy cost will increase. Second, the energy cost of REB is almost independent of the size of U. Third, our REB consistently outperforms TH and EDFSA in terms of energy-efficiency. For example, when $u=50000$, the energy cost of TH and EDFSA is 1.38664×10^6 ω and 1.1686×10^6 ω, respectively. And that of REB is just 4158.47 ω, which represents 333 times and 281 times improvement in terms of energy-efficiency over TH and EDFSA, respectively.

2) Impact of Tag Ratio: REB significantly outperforms the state-of-the-art identification protocols in terms of energy-efficiency, regardless of the tag ratio $|B-G|:|B\cap G|:|G-B|$. In this set of simulations, we fix the value of u and vary the tag ratio of $|B-G|:|B\cap G|:|G-B|$ to investigate the impact of tag ratio on the energy-efficiency of each protocol. The simulation results shown in Fig. 13 reveal

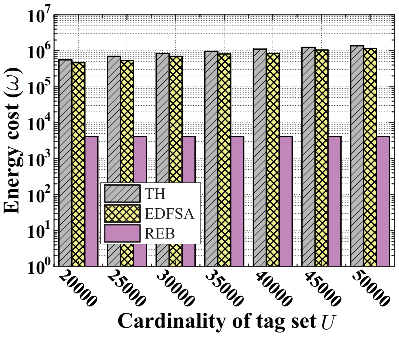
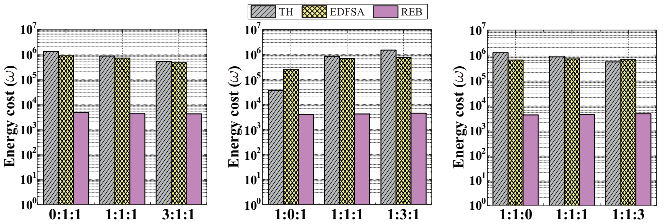


Fig. 12. Evaluating the energy-efficiency of protocols with varying u. The tag ratio of $|B-G|$:$|B \cap G|$:$|G-B|$ is fixed to 1:1:1. $\alpha = 5\%$, $\beta = 95\%$.



(a) Ratio of tags in $B-G$ varies (b) Ratio of tags in $B \cap G$ varies (c) Ratio of tags in $G-B$ varies

Fig. 13. Evaluating the energy-efficiency of protocols with varying tag ratio. u is fixed to 30000, and $\alpha = 5\%$, $\beta = 95\%$.

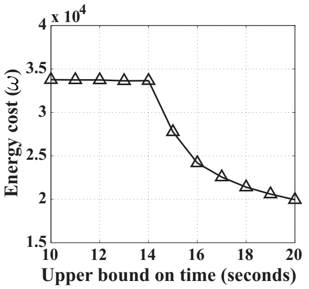


Fig. 14. Time cost of REB vs. given energy threshold. $u = 30000$, $|B-G|$:$|B \cap G|$:$|G-B|$=1:1:3. $\alpha = 2\%$, $\beta = 98\%$.

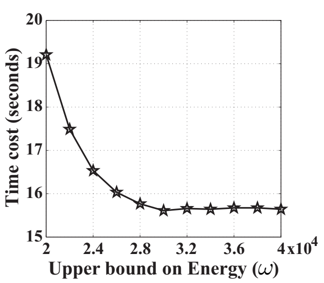


Fig. 15. Energy cost of REB vs. given time threshold. $u = 30000$, $|B-G|$:$|B \cap G|$:$|G-B|$=1:1:3. $\alpha = 2\%$, $\beta = 98\%$.

that our REB consistently outperforms TH and EDFSA under different tag ratios by significantly reducing the energy cost.

E. Performance with Constraints on Time/Energy Cost

In practical applications, we may need to pose constraints on the time cost or energy cost of REB. For example, the trucks carrying tagged items may need to pass through a gate that deploys RFID readers within a short time window; to prolong the lifetime of active tags, we may need to pose a constraint on the energy cost of each execution of REB. Fig. 14 plots the energy cost of REB with a given an upper bound on the time cost. With the simulation settings shown in Fig. 14, the minimum time cost of REB is 15.7s. When the required time threshold is within the range of $[10, 14]$, we have to configure the parameters of REB to minimize its time cost. And thus, the energy cost of REB is stable at a certain level when the time threshold is within $[10, 14]$. On the contrary, when the time threshold is larger than 15.7s, the energy cost of REB decreases as the time threshold increases. The underlying reason is that we can use a relatively small p when time threshold is large, and accordingly, the energy cost of REB decreases. Fig. 15 plots the energy cost with a given constraint on the time cost. As the allowed energy cost increases, the execution time of REB decreases. When the constraint on energy exceeds $3 \times 10^4 \omega$, the constraint on energy will not be the bottleneck, then the time cost of REB remains at the same level.

V. Related Work

In the infant stage of the study of RFID, a great deal of attention was paid to the problem of tag identification that aims to identify the exact tag IDs. Generally, there are two types of identification protocols: Aloha-based protocols [22] and Tree-based protocols [23]. Fundamentally, the Aloha-based protocol is a kind of Time Division Multiple Access (TDMA) mechanism. A tag ID can be successfully identified in a slot when only one tag responds in this slot. As for tree-based protocols, the reader broadcasts a 0/1 string to query the tags. A tag responds with its ID once it finds that the queried string is the prefix of its ID. A reader identifies a tag ID when only one tag responds. The execution time of identification protocols is proportional to the tag population size. As explained in Section I, the performance of tag identification protocols will further deteriorate with the presence of blocker tags.

In recent years, a great effort has been made to study the problem of tag estimation [9], [13], [15]–[20], [28], [30]. The RFID estimation protocols can be used for various purposes. For the first example, when optimizing the performance of an Aloha-based tag identification protocol, we should set the frame size to the number of tags, which, however, is not known in prior. Here, an RFID estimation protocol that can quickly and accurately tell the tag cardinality is desirable. For the second example, consider the stock monitoring in a retailer, we only need to know the tag cardinality instead of the exact tag IDs. Hence, we prefer the fast RFID estimation protocols to the time-consuming tag identification protocols. To the best of our knowledge, the first piece of work concerning with RFID estimation was proposed by Kodialam *et al.* in 2006 [15]. They proposed PZE that uses the number of empty slots in a time frame, and PCE that uses the number of collision slots in a time frame to estimate the number of tags. The underlying reason is that they can establish a monotonous functional relationship between the number of tags and the number of empty slots or collision slots. Then, they can leverage the number of empty slots or collision slots observed from a time frame to perform the estimation of tag cardinality. Kodialam *et al.* have demonstrated that PZE and PCE are complementary to each other well, and combining them gives hybrid estimation protocol called UPE that performs well for a wide range of tag set cardinalities. Qian *et al.* [17] exploited the hashing with geometric distribution to estimate the cardinality of tags, and thus, proposed the Lottery Frame (LoF) scheme. In LoF, each tag has a probability of $\frac{1}{2^i}$ to choose the i^{th} slot to respond to the reader. LoF requires the reader to distinguish empty slots from non-empty slots. An intuitive insight behind LoF's estimator is that the more tags are there in the system, the longer the length of continuous non-empty slots is expected to be. Hence, they can leverage the latter variable observed from a time frame to estimate the number of tags. Shahzad *et al.* proposed the Average Run based Tag estimation (ART) by observing the average length of sequences of consecutive non-empty slots [13]. ART is fast because its estimator has a smaller variance than previous RFID estimation protocols. Zheng *et al.* proposed Zero-One Estimator (ZOE) in which the responses from all the tags aggregate in the single time slot, and thus, allow ZOE to make extensive use of each time slot. Chen *et al.* proposed to put together various estimation schemes as building blocks in a proper manner, and thus achieve a more efficient protocol [18]. Li *et al.* argued that besides time-efficiency, energy-efficiency is also an important issue that must be carefully dealt with when battery-powered active tags are used. They proposed an estimation scheme called Maximum Likelihood Estimator (MLE) to take the energy-efficiency into consideration [20]. For multi-category RFID systems, the literature [21] proposed to estimate tag cardinality in each category with a simultaneous manner; and [4], [9] studied the top-k query problem (*i.e.*, pinpointing the k largest categories). For dynamic RFID systems, literature [7], [28] studied how to quickly estimate the number of absent tags and that of the remaining tags.

VI. Conclusion

This paper formally defines a new problem of RFID estimation with the presence of blocker tags, and makes the first piece of effort that towards providing an efficient solution. The proposed *RFID Estimation scheme with Blocker tags (REB)* is compliant with the commodity EPC C1G2 standard, and does not require any modifications to off-the-shelf RFID tags. REB provides an unbiased functional estimator which can guarantee any degree of estimation accuracy specified by the users. Using REB, a retailer can monitor the product stock in a timely manner; meanwhile, the blocker tags are being used to protect the privacy of some important items. Rigorous analysis is given to optimize the parameters of REB to minimize its time cost and energy cost. A trade-off between the time cost and energy cost can be flexibly controlled to satisfy the practical requirements. Extensive simulation results has revealed the advantages of REB over prior schemes in terms of estimation accuracy, time-efficiency and energy-efficiency.

Acknowledgment

This work is supported by the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61225010); the State Key Program of National Natural Science of China(Grant No. 61432002); NSFC Grant Nos. 61472184, 61321491, 61272417, 61300189, 61370199; Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20130041110019), and the Fundamental Research Funds for the Central Universities (Grant. DUT15QY20). This work is partially supported by Tianjin Key Laboratory of Advanced Networking (TANK), Tianjin 300350, China, the Jiangsu Future Internet Program under Grant Number BY2013095-4-08, and the Jiangsu High-level Innovation and Entrepreneurship (Shuangchuang) Program. Alex X. Liu is also affiliated with the Department of Computer Science and Engineering, Michigan State University, USA. The work of Jie Wu was supported in part by NSF grants CNS 1449860, CNS 1461932, CNS 1460971, CNS 1439672, CNS 1301774, and ECCS 1231461.

References

[1] X. Liu, B. Xiao, K. Li, J. Wu, A. X. Liu, H. Qi, and X. Xie, "RFID Cardinality Estimation with Blocker Tags," *Proc. of IEEE INFOCOM*, 2015.

[2] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-Time Tracking of Mobile RFID Tags to High Precision Using COTS Devices," *Proc. of ACM MobiCom*, 2014.

[3] L. Shangguan, Z. Zhou, X. Zheng, L. Yang, Y. Liu, and J. Han, "ShopMiner: Mining Customer Shopping Behavior in Physical Clothing Stores with Passive RFIDs," *Proc. of ACM SenSys*, 2015.

[4] X. Liu, K. Li, J. Wu, A. X. Liu, X. Xie, C. Zhu, and W. Xue, "TOP-k Queries for Multi-category RFID Systems," *Proc. of IEEE INFOCOM*, 2016.

[5] J. Liu, B. Xiao, S. Chen, F. Zhu, and L. Chen, "Fast RFID grouping protocols," *Proc. of IEEE INFOCOM*, 2015.

[6] L. Shangguan, Z. Yang, A. X. Liu, Z. Zhou, and Y. Liu, "Relative Localization of RFID Tags using Spatial-temporal Phase Profiling," *Proc. of USENIX NSDI*, 2015.

[7] X. Liu, X. Xie, K. Li, B. Xiao, J. Wu, H. Qi, and D. Lu, "Fast Tracking the Population of Key Tags in Large-scale Anonymous RFID Systems," *IEEE/ACM Transactions on Networking*, in press, 2016.

[8] T. Liu, L. Yang, Q. Lin, and Y. Liu, "Anchor-free Backscatter Positioning for RFID Tags with High Accuracy," *Proc. of IEEE INFOCOM*, 2014.

[9] L. Xie, H. Han, Q. Li, J. Wu, and S. Lu, "Efficiently Collecting Histograms Over RFID Tags," *Proc. of IEEE INFOCOM*, 2014.

[10] J. Liu, B. Xiao, K. Bu, and L. Chen, "Efficient Distributed Query Processing in Large RFID-enabled Supply Chains," *Proc. of IEEE INFOCOM*, 2014.

[11] M. Roberti, "A 5-cent Breakthrough," *RFID Journal*, vol. 5, no. 6, 2006.

[12] "http://www.centreforaviation.com/news/share-market/2010/06/17/hong-kong-airport-sets-new-cargo-traffic-record-fedex-sees-surging-asian-exports/page1."

[13] M. Shahzad and A. X. Liu, "Fast and Accurate Estimation of RFID Tags," *IEEE/ACM Transactions on Networking*, vol. 23, no. 1, pp. 241–254, 2015.

[14] M. Chen and S. Chen, "ETAP: Enable Lightweight Anonymous RFID Authentication with O(1) Overhead," *Proc. of IEEE ICNP*, 2015.

[15] M. Kodialam and T. Nandagopal, "Fast and Reliable Estimation Schemes in RFID Systems," *Proc. of ACM Mobicom*, 2006.

[16] M. Kodialam, T. Nandagopal, and W. C. Lau, "Anonymous Tracking using RFID tags," *Proc. of IEEE INFOCOM*, 2007.

[17] C. Qian, H. Ngan, Y. Liu, and L. M. Ni, "Cardinality Estimation for Large-scale RFID Systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1441–1454, 2011.

[18] B. Chen, Z. Zhou, and H. Yu, "Understanding RFID Counting Protocols," *Proc. of ACM MobiCom*, 2013.

[19] Y. Zheng and M. Li, "ZOE: Fast Cardinality Estimation for Large-Scale RFID Systems," *Proc. of IEEE INFOCOM*, 2013.

[20] T. Li, S. Wu, S. Chen, and M. Yang, "Generalized Energy-Efficient Algorithms for the RFID Estimation Problem," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1978–1990, 2012.

[21] X. Liu, K. Li, A. X. Liu, S. Guo, A. L. Wang, X. Xie, and J. Wu, "Multi-category RFID Estimation," *IEEE/ACM Transactions on Networking*, in press, 2016.

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

[23] M. Shahzad and A. X. Liu, "Probabilistic Optimal Tree Hopping for RFID Identification," *IEEE/ACM Transactions on Networking*, vol. 23, no. 3, pp. 796–809, 2015.

[24] P. Semiconductors, "I-CODE Smart Label RFID Tags," *http://www.nxp.com/acrobat_download/other/identification/SL092030.pdf*, 2004.

[25] E. Inc, "Radio-frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 mhz-960 mhz," *EPCGlobal*, Inc, 1.2.0 ed., 2008.

[26] 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 Transactions on Networking*, vol. 24, no. 2, pp. 703–716, 2016.

[27] Y. Zheng and M. Li, “Fast Tag Searching Protocol for Large-Scale RFID Systems,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 3, pp. 924–934, 2013.

[28] Q. Xiao, B. Xiao, and S. Chen, “Differential Estimation in Dynamic RFID Systems,” *Proc. of IEEE INFOCOM*, 2013.

[29] 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.

[30] W. Gong, K. Liu, X. Miao, Q. Ma, Z. Yang, and Y. Liu, “Informative Counting: Fine-grained Batch Authentication for Large-Scale RFID Systems,” *Proc. of ACM MobiHoc*, 2013.

[31] D. E. Smith, “A source book in mathematics,” *Courier Dover Publications*, 2012.

[32] M. Schilling, “Understanding Probability: Chance Rules in Everyday Life,” *The American Statistician*, vol. 60, no. 1, pp. 97–98, 2006.

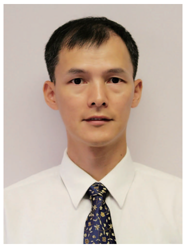
[33] http://pan.baidu.com/s/1i3YgJGh.

[34] Y. Qiao, S. Chen, T. Li, and S. Chen, “Energy-efficient Polling Protocols in RFID Systems,” *Proc. of ACM Mobihoc*, 2011.

[35] S. N. V and D.-B. IV, “Mathematische Statistik in der Technik,” *Deutscher Verl. der Wissenschaften*, 1963.

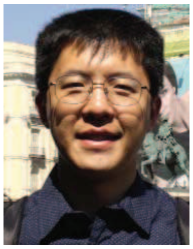
[36] H. Han, B. Sheng, C. C. Tan, Q. Li, W. Mao, and S. Lu, “Counting RFID Tags Efficiently and Anonymously,” *Proc. of IEEE INFOCOM*, 2010.

Xiulong Liu received the B.E. degree from the School of Software Technology, Dalian University of Technology, China, in 2010. Currently, he is a Ph.D. candidate in the School of Computer Science and Technology, Dalian University of Technology, China. He served as a Research Assistant in the Hong Kong Polytechnic University in 2014, and a Visiting Scholar in Temple University in 2015. His research interests include RFID systems and wireless sensor networks.



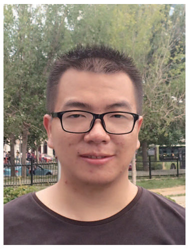
Bin Xiao is currently an Associate Professor in the Department of Computing, The Hong Kong Polytechnic University. Dr. Xiao received the B.Sc and M.Sc degrees in Electronics Engineering from Fudan University, China, and Ph.D. degree in computer science from University of Texas at Dallas, USA. His research interests include distributed wireless systems, network security, and software-defined networks (SDN). Dr. Xiao has published more than 100 technical papers in international top journals and conferences. Currently, he is the associate editor of the Journal of Parallel and Distributed Computing (JPDC) and Security and Communication Networks (SCN). He is the IEEE Senior member, ACM member and the recipient of the best paper award of the international conference IEEE/IFIP EUC-2011.

Keqiu Li received the bachelor’s and master’s degrees from the Department of Applied Mathematics at the Dalian University of Technology in 1994 and 1997, respectively. He received the Ph.D. degree from the Graduate School of Information Science, Japan Advanced Institute of Science and Technology in 2005. He also has two-year postdoctoral experience in the University of Tokyo, Japan. He is currently a professor in the School of Computer Science and Technology, Dalian University of Technology, China. He has published more than 100 technical papers, such as IEEE TPDS, ACM TOIT, and ACM TOMCCAP. He is an Associate Editor of IEEE TPDS and IEEE TC. His research interests include data center networks, cloud computing and wireless networks.



Alex X. Liu received the Ph.D. degree in computer science from the University of Texas at Austin in 2006. He is an Associate Professor with the Department of Computer Science and Engineering, Michigan State University. He is an Associate Editor of IEEE/ACM TRANSACTIONS ON NETWORKING and an Area Editor of Journal of Computer Communications (Elsevier). He received the IEEE & IFIP William C. Carter Award in 2004 and an NSF CAREER award in 2009. He received the Withrow Distinguished Scholar Award in 2011 at Michigan State University. He received Best Paper Awards from ICNP-2012, SRDS-2012, and LISA-2010. His research interests focus on networking and security.

Jie Wu is the Associate Vice Provost for International Affairs at Temple University. He also serves as the Chair and Laura H. Carnell professor in the Department of Computer and Information Sciences. Prior to joining Tempe University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Service Computing and the Journal of Parallel and Distributed Computing. Dr. Wu was general co-chair/chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, and ACM MobiHoc 2014, as well as program co-chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a CCF Distinguished Speaker and a Fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.



Xin Xie received the B.Sc. B.E degree in computer science from Dalian University of Technology, Dalian, China, in 2013. He is currently pursuing the Ph.D in Computer application technology at Dalian University of Technology. His research interests includes RFID technologies and wireless networks. Now, he is serving as the research assistant in Hong Kong Polytechnic University, supervised by Prof. Bin Xiao.

Heng Qi is an associate professor at the School of Computer Science and Technology, Dalian University of Technology, China. He received bachelor’s degree from Hunan University in 2004 and master’s degree from Dalian University of Technology in 2006. Then he received his Ph.D. degree from Dalian University of Technology in 2012. His research interests include computer network, wireless network and multimedia computing.