Efficient Unknown Tag Detection in Large-Scale RFID Systems with Unreliable Channels

Wei Gong, *Member, IEEE,* Jiangchuan Liu, *Fellow, IEEE,* Zhe Yang, *Member, IEEE*

***Abstract*—One of the most important applications of Radio Frequency Identification (RFID) technology is to detect unknown tags brought by new tagged items, misplacement, or counterfeit tags. While unknown tag identification is able to pinpoint all the unknown tags, probabilistic unknown tag detection is preferred in large-scale RFID systems that need to be frequently checked up, e.g., real-time inventory monitoring. Nevertheless, most of the previous solutions are neither efficient nor reliable. The communication efficiency of former schemes is not well optimized due to the transmission of unhelpful data. Further, they do not consider characteristics of unreliable wireless channels in RFID systems. In this paper, we propose a fast and reliable method for probabilistic unknown tag detection, White Paper (WP) protocol. The key novelty of WP is to build a new data structure of composite message that consists of all the informative data from several independent detection synopses; thus it excludes useless data from communication. Further, we employ packet loss differentiation and adaptive channel hopping techniques to combat unreliable backscatter channels. We implement a prototype system using USRP software-defined radio and WISP tags to show the feasibility of this design. We also conduct extensive simulations and comparisons to show that WP outperforms previous methods. Compared to state-of-the-art protocols, WP achieves more than 2x performance gain in terms of time-efficiency when all the channels are assumed free of errors and the number of tags is 10,000, and achieves up to 12x success probability gain when the burstiness is more than 80%.**

***Index Terms*—RFID, large-scale systems, unknown tag detection, unreliable channels**

I. Introduction

OVER the past decade, Radio Frequency Identification (RFID) technology has witnessed an unprecedented growth in practical applications. It has several distinct advantages. First, RFID tags are so small that they can be embedded in almost everything to give a unique ID. Second, the low price of tags makes large-scale use possible for almost anything that costs more than $1. Third, tags are able to be read wirelessly, from a few inches to several feet. Fourth, it supports parallel processing that can operate thousands of tags at a time. In contrast, other methods, e.g., barcode, can only deal with objects sequentially.

This paper focuses on the fundamental problem of detecting unknown tags in large-scale RFID systems. Accurate and fast unknown tag detection is very important to many applications [1], [2], [3], [4], [5]. For example, in RFID-enabled inventory control, it needs to detect unknown-tag events due to new commodities moved in or item misplacement [3]. When processing a large number of tagged items at a mail service center, unknown tag detection can help efficiently verify a batch of tags [4]. Moreover, unknown tag detection is needed as a filter module in unknown tag identification [6] and missing tag detection [5].

One of the closest problems to unknown tag detection (UTD) is unknown tag identification (UTI). Despite their similarity, both two problems are different and have their own characteristics. The UTI can exactly pinpoint all the unknown tags in a batch, whereas the UTD is able to discover the unknown tag event with desired accuracy. Specifically, the UTD has several salient features as follows. First, it can provide tradeoffs between time efficiency and result accuracy while the UTI cannot. This advantage is much needed when the stringent time is required in some cases. For example, if the to-be-tested tag set is of size 10,000, a standard identification would cost dozens of minutes [4], and the state-of-the-art UTI takes tens of seconds [6]. In contrast, the advanced UTI only needs a couple of seconds or even less [1]. Such significant time differences make trading result accuracy for better time efficiency become practical and necessary in time-stringent and even real-time RFID applications. As an illustration, for cross-border cargo inspection where per-item processing is barely possible due to the continuously large volume [7], sampling (usually below 10%), which sacrifices accuracy for time efficiency, is the de facto standard all over the world [8]. Second, the UTD is an essential complement to the UTI rather than an alternative for identification purposes. In case of frequent and long-term monitoring in RFID-enabled warehouses, a fast UTD is ideal as a pre-processing module of the UTI. If the result of the UTD is positive, the UTI can be invoked for further identification. Otherwise, the UTI can remain untriggered, saving time and energy. Therefore, a more advanced UTD would see further improved time efficiency of the UTI. Third, the UTD naturally fits for applications where privacy is concerned as there is no exposure of identification information involved. The UTI, on the contrary, experiences performance degradation because extra time and computation are needed for the transmission of cipher keys and decryption [9]. To summarize, the UTD can benefit a range of RFID applications where unknown tags are present [1], [3], [4], [5], including the UTI, and is best for scenarios where time-accuracy tradeoffs and privacy are concerned.

Despite the importance of the UTI, its efficiency is still challenged by the following two major factors.

Communication Overhead: Although several probabilistic

W. Gong and J. Liu are with the School of Computing Science, Simon Fraser University, Canada. {gongweig, jcliu}@sfu.ca. Z. Yang is with the School of Computer Science, Northwestern Polytechnical University, China, and the School of Computing Science, Simon Fraser University, Canada, zyang@nwpu.edu.cn.

unknown tag detection schemes have been proposed to find unknown tags in a batch, [1], [4], [10], we observe that the efficiency of existing methods has yet to be well optimized due to the transmission of unhelpful data for detection. For instance, collision slots that contribute nothing in the detection are still included in communication messages [1], [4], [10], losing great opportunities of transmission optimization.

Unreliable Channels: Most of existing schemes assume perfect channel conditions. Therefore the performance of those schemes would degrade significantly with unreliable wireless channels [1], [4]. Although some error-resistant schemes are proposed for large-scale RFID system management, they are based on generic models [11][12][13] and fail to exploit characteristics of backscatter channels. For instance, former schemes are unable to distinguish channel fading losses and collision losses, which are quite common in RFID systems because tags cannot sense each other. Moreover, they do not take channel hopping into account, which is mandatory in RFID standards across the world.

In this paper, we propose a fast and reliable protocol for probabilistic unknown tag detection, named White Paper (WP), *where the communication message is composed of (almost) all zero slots*. There are two major contributions of this design. First, we introduce a new compact detection synopsis and tune its parameter for optimal detection efficiency, i.e., we minimize the failure probability of detection for a given frame length. Meanwhile, based on all the informative slots of detection synopses we construct a novel data structure of composite message to significantly reduce transmission overhead. Hence, we are able to optimize detection and communication efficiency at the same time, resulting in high detection efficiency with minimal transmission overhead. Furthermore, various fundamental energy-time tradeoffs in probabilistic unknown tag detection are also achieved in our analytical framework. Second, an error-resistant scheme is designed to combat unreliable and dynamic backscatter channels. In particular, it consists of two major components: a collision-aware channel estimation module that can decompose the measured packet loss rate into the channel fading loss rate and the collision loss rate; and an adaptive channel hopping module that can measure link burstiness and then leverage channel hopping to avoid massive bursty errors.

We demonstrate the effectiveness of the proposed protocol through a prototype system using USRP software defined radio [14] and WISP tags [15]. Comparisons are done with extensive simulations and trace-driven studies to examine the performance in large-scale settings. We also conduct extensive simulations and comparisons to show that WP outperforms previous methods. Compared to state-of-the-art protocols, WP achieves more than 2x performance gain in terms of time-efficiency when all the channels are assumed free of errors and the number of tags is 10,000, and achieves up to 12x success probability gain where the burstiness is more than 80%.

II. Preliminaries

A. Problem Formulation

Suppose we have a known tag set $\mathcal{S} = \{x_1, x_2, x_3, ...\}$, and a to-be-tested tag set $\mathcal{T} = \{y_1, y_2, y_3, ...\}$. The cardinalities of $\mathcal{S}$ and $\mathcal{T}$ are N and n, respectively. While N is a priori, n is not. The goal of probabilistic unknown tag detection is to find whether there is any unknown tag in $\mathcal{T}$ with the knowledge of $\mathcal{S}$. In practice, the two basic requirements for probabilistic unknown tag detection are i) if all the tags in $\mathcal{T}$ are known, the detection result should be negative for sure, ii) if there is at least one unknown tag in $\mathcal{T}$, the detection result should be declared as positive with high probability, i.e., a little detection failure is allowed. Towards this end, we define two parameters: ε, the detection failure probability, and m, the tolerable maximum number of unknown tags in a batch. Thus an (ε, m) detection scheme should be able to detect an unknown-tag event with probability at least $1 - \varepsilon$ if the number of unknown tags in $\mathcal{T}$ is greater than or equal to m. Intuitively, we want ε to get closer to 0 and m to get closer to 1. Although we do not assume any relationship between N and n, it is worth noting that the unknown tag detection problem becomes even more challenging when $N \gg n$. Our scheme cannot adapt to cases where no failure probability is allowed, but setting ε to any arbitrarily extreme values that are suitable for realworld applications, e.g., 0.001%, can be an alternative when other exact methods, like the UTI, are not available.

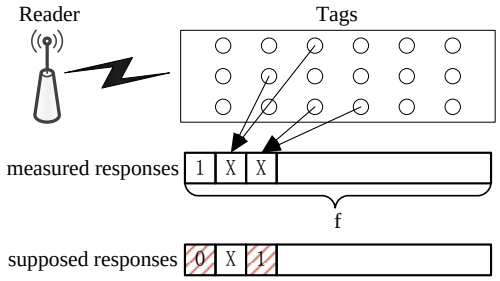


Fig. 1. The ALOHA model and basic detection scheme. 0, 1, and X mean zero, only one, and more than one responses in the slot. An unknown tag event is detected when there are more responses of a slot in measured responses than that of the corresponding slot in supposed responses, e.g., for the first slot, 1 (from measured responses) $>$ 0 (from the supposed responses) means some unknown tag is detected.

B. System Model

A typical RFID system consists of three parts: tags, a reader[1], and a back-end server. Tags may either be read-only, having assigned unique identification information, or may be read/write, where additional data can be stored into the memory on board by the user. The back-end server usually stores all the tags' information and performs various management operations. Generally, we assume that the reader is securely connected to the back-end server through a high-speed channel. Therefore, we denote the reader and back-end server by the reader for simplicity, if not specified.

We assume that the communication between the reader and tags follows the ALOHA model, which is widely used in EPC Global C1G2 standard [16] and many other RFID protocols [4], [17], [18]. As shown in Figure 1, the reader first broadcasts

[1]For multiple readers, we can treat them as a unified virtual reader.

to tags a probing message, which contains the frame size f and the random seed value r. When each tag has received this probing message, it uses preloaded hash functions H to compute its own reply slot number as $sn = H(f, r, ID)$, where ID is the unique identification information. Afterward, the reader issues a slot-start command to all the tags. Then each tag checks whether its supposed reply slot number is equal to the current slot number. If so, it responds instantly. Otherwise, the current slot number increases by 1. According to the number of responses in a single slot, we classify slots into three types: a *zero slot* means no response is in that slot; a *singleton slot* denotes that only one tag's reply is in the slot; a *collision slot* means there are at least two tags' responses in the slot. We also use *non-zero slot* to denote both singleton slot and collision slot. For unknown tag detection, when the ALOHA frame completes, the reader is able to compose the measured responses of size f. Meanwhile, since the back-end server contains all the information of tags (hash functions and unique ID), the reader can virtually construct the supposed responses as if all the tags in $\mathcal{S}$ are present. Therefore, the server can perform the detection by comparing the measured responses to the supposed responses slot by slot. There are two conditions where the server declares there indeed exist unknown tag(s): i) a supposed zero slot turns out to be a singleton or collision slot, e.g., the first position of the supposed responses in Figure 1; ii) a supposed singleton slot turns out to be a collision slot, e.g., the third position in the supposed responses. In summary, *only unknown tags would cause this "add-up" effect on responses*. We use responses and synopsis/synopses interchangeably in this paper, because the use of "responses" is from the perspective of communication and the use of "synopsis/synopses" is from the perspective of data structure.

According to the Philips Semiconductors implementation of C1G2 [19], if we need to distinguish a zero slot from a non-zero slot, the tag only needs to transmit a short response that costs 0.4 ms, denoted as T_s; if we want to distinguish a zero slot from a singleton slot and a collision slot, a long response that is 0.8 ms is required, denoted as T_l. Moreover, if a slot is used to transmit the ID (typically 96 bits) of the tag, it costs 2.4 ms, denoted as T_{tag}. We prefer to use short responses than long responses in terms of time efficiency. In other words, we only distinguish zero slots from non-zero slots in our scheme. We also employ the participation probability for each tag in a frame, denoted as p. For example, if $p = 0.25$, it means this tag would engage in this frame with 25% probability.

III. Fast Probabilistic Unknown Tag Detection

In this section, we first present our motivation and then describe our basic idea for fast unknown tag detection. Later we consolidate it with detailed communication protocols for both the reader and tag sides. Corresponding analysis and practical design issues are discussed as well. Not that perfect channels are assumed in this section. We will then expand the new protocol to work under unreliable channels in the next section.

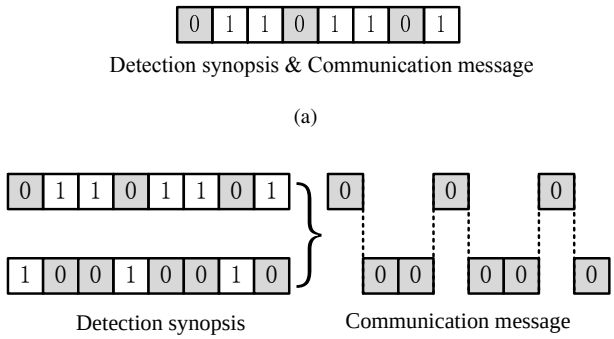


Fig. 2. (a) The detection synopsis and communication message are the same; (b) The detection synopsis and communication message are separated (0 denotes zero slot and 1 denotes non-zero slot).

A. Motivation

Here, we examine previous work by using SEBA (SEBA-2) as a case study [4], in which it only distinguishes zero slots from non-zero slots. Although the following analysis is based on SEBA, it can also apply to most existing schemes. As aforementioned, we know that zero slots are important in unknown tag detection. If an unknown tag responds in a supposed zero slot, it would cause the actual slot to be a non-zero slot, indicating that an unknown-tag event is detected. Thus, we argue that the transmission of non-zero slots in supposed responses is a waste of time because those slots contribute nothing for detection. Actually the amount of this waste is significant: about 50% of the total communication time is wasted in SEBA as later elaborated in section III-C. As shown in Figure 2a, the grey-colored zero slots stand for useful transmission.

Therefore, it motivates us to treat communication messages and detection synopses differently by transmitting useful slots (zero slots) as many as possible, greatly improving detection and communication efficiency. Let's see an example. Suppose that the detection success probability of a SEBA synopsis is 80% and we are going to detect a single unknown tag. As shown in Figure 2b, if we have 2 independent SEBA synopses in advance, then the detection success probability of using 2 synopses together is $1 - (1 - 80\%)^2 = 96\%$. If we have even more independent synopses, e.g., l, it is easy to see that $1 - (1 - 80\%)^k$ could fast approach to 100% as l increases. Later we will show how to achieve efficient communication by encoding several independent synopses into a composite message. We also consider multiple responses from each tag in a single frame, which can largely improve the detection efficiency of a synopsis.

B. Basic Ideas

The basic idea of our scheme is to build a composite message data structure that consists of all the informative data from several independent detection synopses, excluding the useless data from communication.

Based on the observation that non-zero slots contributes little for detection, our scheme aims to improve communication

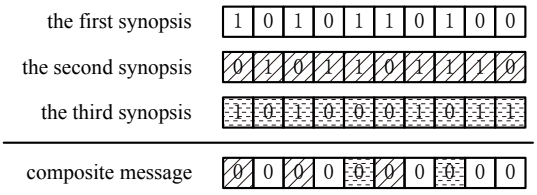


Fig. 3. The construction of a sample composite message. A composite message that is full of zeros can be built through combining all informative slots (zero slots) across different synopses.

efficiency by changing non-zero slots into zero slots. Let us see a toy example in Figure 3, where 0 denotes a zero slot and 1 denotes a non-zero slot. Assume that we have three detection synopses. First, we construct the composite message based on the first synopsis, in which all the zero slots are kept in the composite message. Then by scanning the second synopsis, the composite message continues to 'absorb' useful (zero) slots. The rule of absorption is that the slot of current synopsis is zero and the corresponding slots in all the previous synopses are non-zero. For the third synopsis, we also apply this rule for combination. Finally, we get a composite message full of zero slots, nowhere for unknown tags to hide: *it is easy to spot stains (unexpected responses caused by unknown tags) on a white paper (a composite message).*

C. White Paper Protocol

In this part, we turn our basic idea into detailed protocols for both the reader and tag. As we know that, the construction of a composite message relies on several virtual synopses [2]. Hence, each slot in a composite message is actually an *index* of synopses, indicating which synopsis this slot comes from. We use an index vector to denote a sequence of such indexes.

The protocol consists of four major steps: index vector generation, index vector transmission, response measurement, and unknown tag detection, as shown in Algorithm 1 and 2.

Phase one - index vector generation: As shown in Figure 1, the supposed responses can be virtually generated since the reader knows all the information of tags in $\mathcal{S}$. The key difference is that, in WP each tag replies at k different slots based on k independent hash functions in a frame. The rule of constructing the composite message is first come first serve, i.e., it sequentially selects zero slots from supposed synopses one by one. Note that as we change non-zero slots into zero slots in our best efforts, if there are still some non-zero slots after combing all the synopses, we just keep such remaining non-zero slots from the first synopsis.

Phase two - index vector transmission: As there are l supposed responses, the size of each element in an index vector is $\lceil \log(l+1) \rceil$-bit. For instance, if $l = 3$, then each element is 2-bit long. The whole index vector may not be able to fit into a single transmission if the size of synopses is large. Hence, we can divide it into pieces and each piece includes $\lfloor \frac{96}{\lceil \log l+1 \rceil} \rfloor$ indexes. Using this division, the reader starts the frame and transmits all the pieces of an Index Vector (IV) using T_{tag} slots. At the same time, when the tag receives the frame start command, it will expect an IV piece by piece.

[2]A virtual synopsis is a synopsis generated on the server, not a real measured synopsis.

Algorithm 1 The WP protocol for tags

1: Receive the frame start command and the index vector IV.
2: Receive the frame size f, the participation probability p, and the random seeds $s_0, s_1, ..., s_{l-1}$.
3: Choose to participate in this frame or sleep based on the probability p.
4: If not participate, sleep until another frame starts.
5: Compute reply slot numbers $sn[i][j] = H(f, ID, s_i, k)$ where $(0 \leq i \leq l-1, 0 \leq j \leq k-1)$
6: Initialize the current slot number $csn \leftarrow 0$ and current random seed index $ci \leftarrow 0$.
7: **while** TRUE **do**
8: wait-for-slot-start().
9: $ci \leftarrow IV[csn]$.
10: **for** $i = 0$ to $k-1$ **do**
11: **if** $csn == sn[ci][i]$ **then**
12: Respond instantly and break.
13: $csn \leftarrow csn + 1$.

Phase three - response measurement: The reader continues to broadcast several parameters to tags, including f, the frame size, p, the participation probability, and $s_0, s_1, ..., s_{l-1}$, the random seeds. Upon receiving those parameters, the tag first decides whether to participate in this frame according to p. If it does not participate in this frame, it will sleep until another frame starts. If it chooses to join in, it needs to compute reply slot numbers using the hash function H. The tag generates k supposed reply slot numbers based on the different random seeds. In each time slot, the reader issues a slot start command and waits for responses. At the tag side, if one of supposed reply slot numbers for random seed indexes is equal to the current slot number, the tag responds instantly. Otherwise, it keeps silent. When f time slots are finished, the reader obtains the Measured Responses (MR).

Phase four - unknown detection: The detection process for the reader is relatively easy. First, the reader compares the Measured Responses (MR) to the composite Supposed Responses (SR) slot by slot. If any one slot in the measured responses is non-zero and its corresponding slot in the supposed responses is zero, the reader shall report a positive result, indicating there exist unknown tags in the batch. Otherwise, all the tags in the batch ($\mathcal{T}$) are deemed known since the result is negative. Note that this result may contain false negatives (detection failure), but no false positives.

D. Protocol Analysis

Now, we seek to optimize detection and communication efficiency at the same time.

Detection Efficiency Optimization: From the supposed synopsis generation process that is in the phase one of WP, we know that each tag selects k slots in a frame. Besides, each tag chooses to participate in the frame based on the probability p.

Algorithm 2 The WP protocol for the reader

1: //Phase one - index vector construction.
2: Generate l random seeds $s_0, s_1, ..., s_{l-1}$ and corresponding supposed responses $SR_0, SR_1, ..., SR_{l-1}$.
3: Initialize the index vector $IV[i] \leftarrow 0 (0 \leq i \leq f-1)$.
4: Initialize the combined supposed responses $SR \leftarrow SR_0$.
5: **for** $i = 1$ to $l-1$ **do**
6: **for** $j = 0$ to $f-1$ **do**
7: **if** $IV[j] == 0$ and $SR_i[j] == 0$ **then**
8: $IV[j] \leftarrow i$, $SR[j] \leftarrow 0$.
9: //Phase two - index vector transmission.
10: Divide the IV into pieces and each piece contains $\lfloor \frac{96}{\lceil \log l+1 \rceil} \rfloor$ indexes.
11: Issue a frame start command and transmit the IV piece by piece.
12: //Phase three - response measurement
13: Broadcast the frame size f, the participation probability p, and the random seeds $s_0, s_1, ..., s_{l-1}$.
14: Initialize the measured responses $MR[i] \leftarrow 0 (0 \leq i \leq f-1)$.
15: **for** $i = 0$ to $f-1$ **do**
16: Issue slot-start command.
17: wait-for-tags-response().
18: **if** there is any response in this slot **then**
19: $MR[i] \leftarrow 1$.
20: //Phase four - unknown detection
21: **for** $i = 0$ to $f-1$ **do**
22: **if** $MR[j] == 1$ and $SR[j] == 0$ **then**
23: Report a positive result and return.
24: Report a negative result and return.

Hence, the probability p_0 that one slot in a supposed synopsis is still zero after N tags' responses is

$$p_0 = (1 - p\frac{1}{f})^{kN} \approx e^{-\frac{pkN}{f}}. \quad (1)$$

Meanwhile, for an unknown tag, if all the k slots it chooses are non-zero, it would be hidden in this synopsis. We can calculate this hidden probability, p_h, as

$$p_h = 1 - p + p(1 - p_0)^k \approx 1 - p + p(1 - e^{-\frac{pkN}{f}})^k. \quad (2)$$

In order to maximize the detection efficiency of a synopsis, we needs to minimize the above hidden probability with respect to k given the fixed frame length f. To do so, we first rewrite $(1 - e^{-\frac{pkN}{f}})^k = e^{k\ln(1-e^{-\frac{pkN}{f}})} = e^q$ where $q = k\ln(1 - e^{-\frac{pkN}{f}})$. It is easy to see that minimizing p_h is equal to minimizing q, thus we can obtain its partial derivative as

$$\frac{dq}{dk} = \ln(1 - e^{-\frac{pkN}{f}}) + \frac{kNe^{-\frac{pkN}{f}}}{f(1 - e^{-\frac{pkN}{f}})}. \quad (3)$$

If let this derivative to be 0, we get when $k = \frac{f}{pN}\ln 2$, p_h achieves its global minimum $1 - p + p(\frac{1}{2})^k$ [3].

[3]Using the second derivative test, we know it is a minimum instead of a maximum, since its second derivative value at point $k = \frac{f}{pN}\ln 2$ is greater than 0.

Let us see an illustrative example showing the impact of different k on the hidden probability (detection failure probability). In Figure 4a, our settings are: the size of $\mathcal{T}$ is 1000; the tolerable number of unknown tags is only 1; and the participation probability is 1. By varying the size of frame, we compare different $k(1, ..., 5)$ to the theoretical optimal $k(\frac{f}{N}\ln 2)$. We observe that SEBA ($k = 1$) achieves optimal when hidden probability is above 0.382, which is quite inefficient for detection. In particular, when frame size is 5000, $k = 3$ achieves optimal hidden probability 0.09, whereas its of SEBA is only 0.18. Note that for SEBA, when p_h achieves its minimum, $p_0 \approx e^{-\frac{kN}{f}} = \frac{1}{2}$. It means that about 50% of the total slots in a frame are non-zero slots, which are of no use in unknown tag detection, leading to unnecessary and wasteful transmission. This further makes necessary the idea of constructing composite messages to improve communication efficiency.

Communication Efficiency Optimization: As aforementioned, we know that the composite message is constructed using l supposed synopses. In order to achieve the user-specified requirements for (ε, m), we should obtain the hidden probability of m unknown tags in a composite message, denoted by α_h. To do so, we first calculate the hidden probability of a single unknown tag in a composite message, denoted as β_h. It is obvious that $\alpha_h = \beta_h^m$. Virtually we can divide a final composite message into l layers, each of which only contains the zero slots from i-th ($0 \leq i \leq l-1$) synopsis. Let w_i be the probability of an original slot in the i-th synopsis to be chosen into the i-th layer of the composite message, and γ_i be the hidden probability of the i-th layer of the composite message. Iteratively, according to the criteria that a slot is chosen into the composite message only if it is zero in i-th layer and all of corresponding positions in former layers are non-zero slots, we can have

$$w_i = (1-p_0)^i p_0, \gamma_i = 1-p+p(1-w_i)^k, (0 \leq i \leq l-1). \quad (4)$$

After l iterations, we have

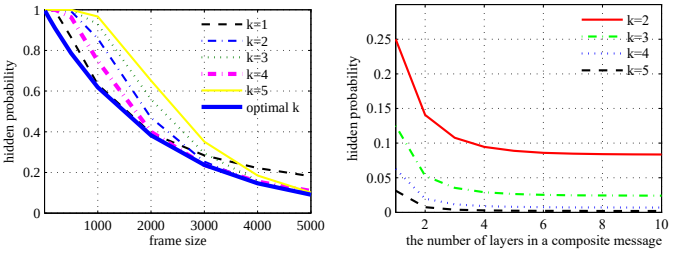
$$\beta_h = \prod_{i=0}^{l-1} \gamma_i. \quad (5)$$

It is easy check that when $i = 0$, the results $w_0 = p_0$ and $\beta_h = \gamma_0$, which are consistent with the former analysis. Therefore, in order to fulfill the requirements of (ε, m), the following equation should be satisfied

$$\varepsilon \geq \alpha_h = \beta_h^m = (\prod_{i=0}^{l-1}(1 - p + p(1 - w_i)^k))^m. \quad (6)$$

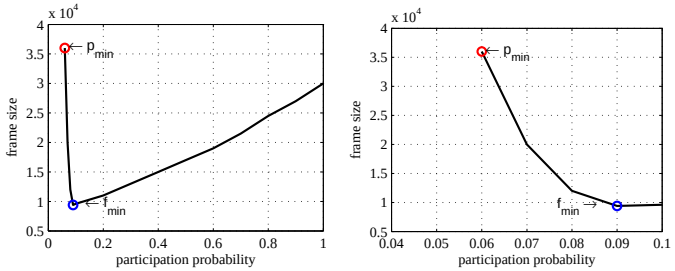
We observe that it follows the law of diminishing marginal returns regarding the number of layers (l) in the composite message. As shown in Figure 4b, we set $m = 1, p = 1$ and $\frac{f}{N} = \frac{k}{\ln 2}$. By varying l, we observe similar trends for $k \in [2, 5]$. Note that the case where l is above 5 is not shown, because the return of increasing l is less than 0.001, which is negligible. Therefore in the following we use $l = 5$ as default unless otherwise specified[4].

[4]For system that may require an extremely low error on the detection failure probability, e.g., 10^{-5}, a larger l should be employed.



(a) (b)

Fig. 4. (a) Different frame size vs hidden probability when $N = 1000, m = 1, p = 1$; (b) Different number of layers in a composite message vs hidden probability when $m = 1, p = 1$, and $\frac{f}{N} = \frac{k}{\ln 2}$.



(a) (b)

Fig. 5. (a) Participation probability vs frame size when $N = 100,000, k = 1, m = 50$, and $\epsilon = 0.05$. (b) Zoom view of (a) for p$\in [p_{min}, \mathcal{F}(f_{min})]$

E. Energy and Time Tradeoffs

The energy cost of tags is another important issue we should carefully cope with. For example, in a large RFID-enabled warehouse, active tags are usually used to label commodities. Since active tags are battery-powered, recharging batteries for thousands of tags is really a heavy work process, and even in some cases the tags are not easily accessible, e.g., tagged commodities may be intensively piled. Here, we mainly focus on the energy consumption caused by wireless transmission, we use the participation probability of tags in a frame, p, to depict the energy consumption of the tags in the detection. The smaller p is, the fewer tags to transmit responses and thus the less energy consumed. As most of the time is spent on the frame, we use the frame size f to represent the time cost of WP. Therefore, we strive to achieve energy-time tradeoffs in probabilistic unknown tag detection. One typical problem is how to minimize the communication time under predefined energy-constraints. The other problem is how to minimize the energy consumption in a limited period of time. In both cases, predefined ε and m requirements should be satisfied at the same time.

Intuitively, one may want both f and p to be as small as possible. However, their choices must satisfy equation 6, which means we cannot minimize both of them at the same time, providing opportunities to make energy-time tradeoffs. Without loss of generality, we can define two functions

$$\mathcal{F}(f) = p, \quad \mathcal{G}(p) = \mathcal{F}^{-1}(p) = f. \tag{7}$$

That is to say, given system parameters (N, k, l) and user-specified parameters (ε, m), $\mathcal{G}(p) = f$ can find the minimum f that satisfies $\varepsilon \geq \alpha_h$. $\mathcal{F}$ is the inverse function of $\mathcal{G}$.

If we set $N = 100,000, k = 1, m = 50$, and $\epsilon = 0.05$, we can plot the curve of $\mathcal{G}$ with varying p, as in Figure 5. This energy-time curve measures energy cost as $n * p$ tags participating in the frame. $\mathcal{G}(p)$ denotes the corresponding optimal frame size. The two distinct points of p_{min} and f_{min} need some explanations.

Finding Minimum p: It is obvious that the participation probability p cannot be arbitrarily small, since $\varepsilon \geq \alpha_h$ may not hold when p it too small. Therefore, there is a minimum participation probability p_{min} that satisfies the user-specified ε. Using $\mathcal{G}$, it is easy to obtain the p_{min} through a binary search. According to the settings in Figure 5, p_{min} is found to be 0.06.

Finding Minimum f: Similarly, it is easy to derive that between p_{min} to 1, there must be a minimum frame size f_{min} that makes $\varepsilon \geq \alpha_h$ hold. By a binary search, we can find f_{min} as 9460 under the settings in Figure 5. Note that for different parameter settings, the curve of $\mathcal{G}$ may be different but the process of finding p_{min} and f_{min} are the same.

Constrained Optimization Problems: There are two closely related constrained optimization problems: energy-constrained least time problem and time-constrained least energy problem. The energy-constrained least time problem always takes the maximum number of tags to participate in the frame, n_u, as input. Therefore, we just set the $\frac{n_u}{N}$ as the maximum participation probability, then a binary search in the range $[p_{min}, \max(\frac{n_u}{N}, \mathcal{F}(f_{min}))]$ would give the optimal. On the other hand, the time-constrained least energy problem often takes the upper bound of frame size f_u as a constraint. After carefully reviewing the energy-time curve in Figure 5, we observe that all the solutions should be in the range $[p_{min}, \mathcal{F}(f_{min})]$. Since if we choose $p > \mathcal{F}(f_{min})$, both the energy cost and time cost are increased. Thus, we set $f = f_u$ and do a binary search in the range $[p_{min}, \mathcal{F}(f_{min})]$ to find the optimal p.

IV. Detection with Unreliable Channels

Most existing detection schemes assume perfect channel conditions, which would make their performance severely degrade with unreliable wireless channels. For instance, the response signal brought by noise could lead to false detection in SEBA [4]. SBF [1] also heavily depends on channel conditions, the detection accuracy would dramatically decrease even if the error rates of tag-to-reader links might be small. Some error-resistant schemes for large-scale RFID system management have been proposed using generic models, e.g., random error model and burst error model [11][12][13]. Those schemes, however, have two major limitations. First, they are unable to distinguish channel fading losses and collision losses. Without diagnosing the causes of packet losses, the (channel) error-resistant scheme that is meant to respond to varying channel fading rather than collisions, becomes ineffective and inefficient. Second, they do not consider channel hopping, which is mandatory in the RFID standards across the world. For example, the FCC RFID regulations dictate that the

average time of occupancy at any frequency must not be larger than 0.4 seconds within any 10-second period if the bandwidth of the hopping channel is larger than 250 kHz. In China the maximum dwell time on a channel is set to 2 seconds.

To address the above issues, we propose an error-resistant scheme to enhance WP, called EWP. EWP consists of two major components: a collision-aware channel estimation module and an adaptive channel hopping module.

Collision-aware Channel Estimation: In RFID systems, the measured packet error rate usually includes both channel fading errors and collision errors. Apparently it would be biased if the measured packet error rate is deemed as the channel fading error rate. So, we seek to approximate the channel fading error rate from the measured packet error rate by estimating collision errors. Generally, the measured error rate can be decomposed as follows.

$$1 - p_m = (1 - p_f)(1 - p_c), \tag{8}$$

where p_m is the measured error rate, p_f is the channel fading error rate, and p_c is the collision error rate. Specifically, p_m, the measured error rate, is obtained using the query-based lightweight and fast probing scheme in [20]. There are a couple of advantages of this lightweight probing method. First, it uses the simplest encoding and slowest baud rate to avoid channel errors being masked by more complex encodings and faster baudrates. Moreover, the reading range of the slowest baudrate is larger than other baudrates and thus we can probe as many tags as possible. In addition, the query-based scheme can be applied to both software defined readers and commercial off-the-shelf readers. For more details, please refer to [20].

Since p_c is unknown in Equation 8, we seek to approximate it using the slotted-Aloha model specified by the C1G2 standard [16]. In the meantime, the ratio of the number of singleton slots to the number of tags in $\mathcal{T}$ can be estimated as

$$\frac{\theta}{n} = (1 - \frac{1}{f})^{n-1}, \tag{9}$$

where θ is the number of singleton slots in a frame. Intuitively, the collision error rate should be the complement of $\frac{\theta}{n}$, i.e.,

$$p_c = 1 - \frac{\theta}{n}. \tag{10}$$

Note that although n is unknown, we can leverage well-researched existing cardinality estimation algorithms to obtain a good estimate, e.g., [11], [21]. Therefore, the channel fading error rate, p_f, can be deduced by combining equation 8, 9, and 10.

Since WP does not rely on any monotonic feature of response signals, given the probability of a non-zero slot being correctly identified, $(1 - p_f)$, the hidden probability of i-th layer of a composite message with unreliable channels becomes $\gamma_i' = 1-p+p(1-w_i(1-p_f))^k$. Hence, an (ε_{EWP}, m) detection scheme under unreliable channels should satisfy

$$\varepsilon_{EWP} \geq (\prod_{i=0}^{l-1}(1 - p + p(1 - w_i(1 - p_f))^k))^m. \tag{11}$$

For possible errors on reader-to-tag links, we adopt a common method that uses CRC checksum [12], [13]. The C1G2 protocol specifies that tags should support 16-bit CRC computation as well. For instance, in WP the index vector might be divided into several segments for transmission. In each transmission, the reader needs to attach a 16-bit CRC checksum at the end of the message. When a tag receives the message, it will check the CRC checksum first. If it is correct, it would proceed since the content is not corrupted. Otherwise, the tag won't join the following executions until it receives the next correct message. This would definitely bring additional communication overhead. It, however, ensures we achieve the desired detection accuracy with error-prone wireless channels.

Adaptive Channel Hopping: Like many existing schemes in RFID management [12], [13], the above error-resistant scheme alone still suffers from burstiness errors, which are known for affecting wireless protocol performance seriously [22], [23]. Yet, we observe an opportunity to leverage channel hopping mechanisms in RFID standards to cope with link burstiness. First, we use a well-known metric, β-factor, to quantify burstiness errors, which is shown effective in predicting wireless protocol performance [24]. Then our channel hopping scheme uses the extent of burstiness (β value) to decide whether it needs to quickly switch to another channel.

According to [24], the β-factor quantifies burstiness in terms of the distance of an empirical link to the distance of an independent link with the same packet reception ratio (which is the complement of packet error rate). We give a few brief descriptions here, and readers are referred to [24] for more details. Specifically, the burstiness metric β is defined by

$$\beta = \frac{\mathbf{KW}(I) - \mathbf{KW}(E)}{\mathbf{KW}(I)}, \tag{12}$$

where $\mathbf{KW}()$ denotes the Kantorovich-Wasserstein distance [25] from the ideal bursty link, E is the Conditional Packet Delivery Function (CPDF) of the empirical link in test, and I is the CPDF of an independent link with the same packet reception ratio.

Putting all the above together, we come to design our adaptive channel hopping scheme that works with burstiness and channel correlation. There are two viable schemes. For scheme A, we choose 25 out of 50 channels as a candidate channel set (CCS) based on thresholds, (e.g., $p_f = 0.5$). If there are more than 25 channels, top 25 are picked in the descending order of burstiness. If not, we randomly choose additional ones to make 25 channels complete. Then, let the reader hop randomly throughout the CCS. The maximum dwell time is set to 400 ms. In scheme B, the reader hops to the next channel randomly among channels that are χ (e.g., 4) channels away from the last channel, otherwise the process is exactly the same as the standard channel hopping with 50 channels. While scheme B could avoid jumping to neighbor channels that may be interfered by the same source, we decide to use scheme A as it values consistently good channels and also implicitly filters out neighbour channels that are correlated and bad. Therefore, our channel hopping is fully compatible with standards and makes the use of best channels available.

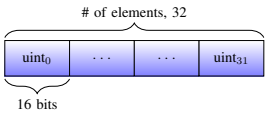


Fig. 6. Pre-stored *Random Number Table* in WISP 5. It consists of 32 unsigned integers, each of which is 16-bit long.

V. Implementation

Our prototype of WP is based on USRP software defined radio and programmable WISP tags.

Setup: We implement the Software-Defined RFID reader (SDReader) using a USRP N210 and the Gen 2 RFID Tools [14]. This SDReader works in 900MHz band based on an RFX900 daughterboard which is connected to Alien circular polarized antennas. Then we connect the SDReader to a laptop via the built-in Ethernet port on the USRP N210. The operating system is Ubuntu 14.04.2 LTS (32-bit).

The tag implementation is based on WISP hardware[15]. The WISP tag is equipped with an ultra-low power MSP430 micro-controller which is able to do basic computations. Since the C1G2 protocol is partially built in WISP 4.1 firmware, we just need to extend it with the functions of WP.

Channel implementation: We implement the channel estimation as in [20] and the adaptive channel hopping scheme where most hopping settings are adopted from the commercial Impinj reader. Our reader hops across the UHF RFID band, 902-928 MHz with hopping occurring between 902.75-927.25 MHz in 500 KHz steps. Based on burstiness and loss rate thresholds, 25 channels are selected out of all available 50 channels. The 20 dB bandwidth of each channel is limited to 500 kHz and the dwell time on each channel is set to 400 ms in a 10-second interval.

Hash implementation: As the power harvested from RF signals is always limited, not much energy could be spared for hash calculation. Sometimes, inappropriate designed computation tasks can even deplete onboard energy and make it into sleep mode (when the voltage is below 1.5 v). Following the common design principle of the WISP platform that trades memory for energy and time, we pre-store hash values onto WISP tags in each round. Actually, this way of slot-count generation has already been implemented in WISP by default. For WISP 4.1, the slot-count and RN16 are preloaded from EPC (unique ID). For WISP 5, it even includes a *Random Number Table* (RNT) that is initialized by data from the onboard temperature sensor, as shown in Figure 6. The WISP5 tag uses the RN16 from the last round as the seed with "Modulo 32" to obtain an index, which can further get an uint of RNT as an RN16. Then the slot-count is calculated from this RN16 and the Q mask [15]. Therefore, we pre-compute the required hash values beforehand and write them into EPC (WISP 4.1) or RNT (WISP 5). If a hash value takes 6 bits, then a EPC and a RNT can host 16 and 85 hash values, respectively, which are enough for microbenchmark. If more hash values are needed, the non-volatile memory on the MSP430 microcontroller can also be used, which typically is 8 KB flash.

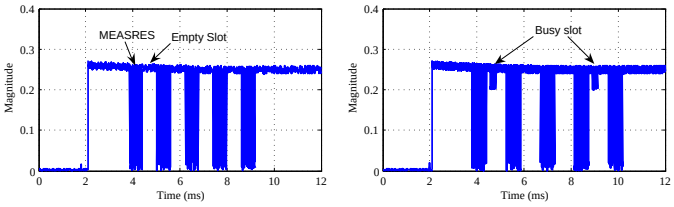


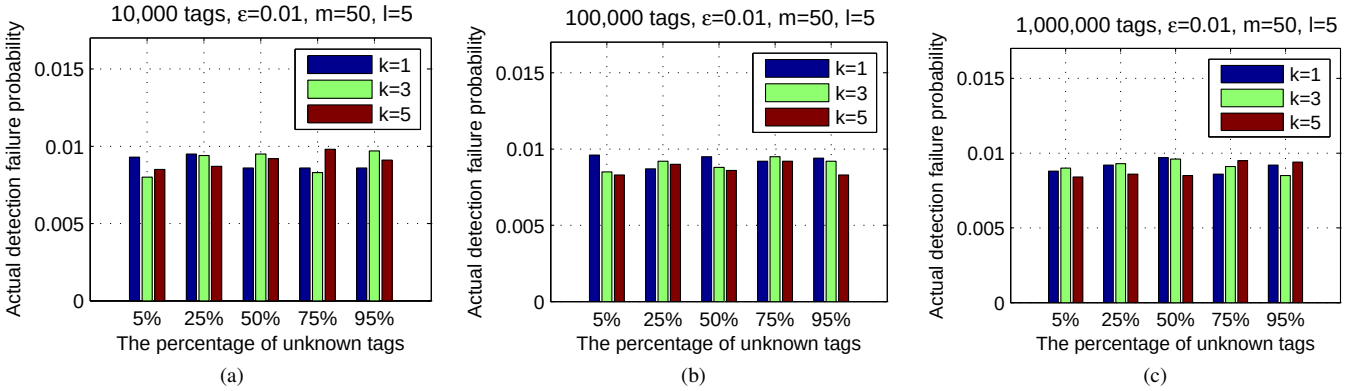
Fig. 7. (a) The communication between the reader and 3 known WISP tags. Each slot after the **MEASRES** command is empty. (b) The communication between the reader and 5 WISP tags (3 known tags and 2 unknown tags). Busy slots are found after the first and the fourth command, respectively, which means unknown tags are detected.

Protocol Implementation: In line with the reader-initiated approach of C1G2, we add two commands into the command set: **TRANSIV** that is used to transmit an index vector, and **MEASRES** that can start the slot and measure the responses from tags. Since the major procedures are already described in section III, we just detail the core part in the phase three here. To measure the responses from tags, the reader sends out a **MEASRES** command along with other parameters, e.g., the participation probability p and random seeds. When the WISP tag has received the **MEASRES** command, it starts computing the reply slot numbers $sn[i][j] = H(f, ID, s_i, k)$. If any of $sn[i][j]$ is equal to the current slot number csn, the tag responds instantly. Otherwise keep silent. To respond, the WISP tag just simply transmits a single tone at 250kHz, which has proven enough for robust detection [11].

Detecting Unknown Tags: We prototype an unknown tag detection system which includes 5 WISP tags and 1 SDReader. Among 5 WISP tags, three of them are known and the other two are unknown. The communication is shown in Figure 7a where all 3 known tags are present. All the slots after **MEASRES** command are empty since there is no unknown tag. Then we put 2 unknown tags into the field. The responses measured are shown in Figure 7b. We find two short responses after the first and the fourth **MEASRES** command, indicating an unknown-tag event being detected.

VI. Evaluation

Although our WP prototype works well in real-time, we turn to large-scale simulations and trace-based experiments for more detailed examinations and comparisons with state-of-the-art methods. There are two reasons for this. First, the large-scale field experiment is still hard for the USRP and WISP platform in terms of programming, debugging, and testing [11]. For example, the operating range of the SDReader is quite limited since the power output is only 200mW for RFX900 daughterboards. Second, we would like to compare with prior schemes in various settings, e.g., the size of frames and unknown ratios. In this section, we evaluate the performance of WP and compare it with the state-of-the-art unknown detection schemes: SEBA [4], SEBA+ [10], and SBF [1].



(a) (b) (c)

Fig. 8. When $\varepsilon = 0.01, m = 50, l = 5$, the percentage of unknowns vs actual detection failure probability. (a) 10,000 tags; (b) 100,000 tags; (c) 1,000,000 tags.

TABLE I
THE RELATIVE ENERGY AND TIME COST OF WP($k = 7, l = 5$) WITH $\mathcal{F}^{-1}(f_{min})$ AND p_{min} TO SEBA, WHEN $\varepsilon = 0.01$ AND $N = 100,000$.

	$\mathcal{F}^{-1}(f_{min})$		p_{min}	
	relative energy cost	relative time cost	relative energy cost	relative time cost
m=10	10.2%	9.8%	9.6%	98.4%
m=50	6.3%	4.3%	5.9%	146.7%
m=100	4.2%	2.1%	3.8%	342.9%

A. Simulation Setup

Our simulation parameters are set according to the Philips I-Code system [26], in which $T_s = 0.4$ ms and $T_{tag} = 2.4$ ms, including the waiting time. The T_s slot is used to transmit tag responses. The broadcast data, including the random seeds, the frame size, the participation probability, and the index vector, are transmitted using multiple T_{tag} slots. The time cost of both downlink and uplink is measured as

$$timecost = |\frac{sizeof(broadcastdata)}{96}|T_{tag} + fT_s.$$

The energy cost is depicted by the participation probability p and the number of tags participated in the frame together, which is np.

B. WP Investigation

Detection Failure Probability: First, we examine the actual detection failure probability of WP, which is an important metric in our scheme. We fix $p = 1, \epsilon = 0.01, m = 50$, and $l = 5$. As shown in Figure 8a, by varying the percentage of unknown tags from 5% to 95%, the actual detection failure probabilities are always below the predefined $\epsilon = 0.01$ for different k. This result shows that WP can effectively detect the unknown-tag event with the desired requirements. Similar results can be found in both Figure 8b and 8c. Those two subfigures further suggest that our WP is able to detect unknown tags in different sizes from 10,000 to 1,000,000.

Energy-time Tradeoffs: The quantified results are given in Table I, when $k = 7, l = 5, \varepsilon = 0.01, N = 100,000$. We show the results at two critical points f_{min} and p_{min}. All those data shows that WP is indeed an efficient probabilistic unknown tag detection protocol in terms of energy cost and time cost, compared to SEBA. In particular, when $m = 100$, the energy cost of WP is only 3.8% of SEBA at p_{min} that achieves least energy cost, and the time cost of WP is just 2.1% of SEBA at f_{min} that is the point of least time cost.

C. Comparison under ideal channels

Here, we compare the performance of WP with SEBA [4], SEBA+ [10], and SBF (SBF-UDP) [1], under different number of tags and the tolerable minimum number of unknown tags. We set $N = 100,000, \varepsilon = 0.01$, $m = 10, 20, 50$, and n ranging from 10,000 to 100,000.

First, we examine the energy cost of the four schemes in terms of number of participated tags. As shown in Figure 9a, WP always has the smallest energy cost among all the protocols. In particular, SEBA and SEBA+ are the same worst (overlapping) due to no energy conservation strategy built-in, i.e., the participation probability is always 1. While SBF employs a sampling probability scheme and is better than SEBA and SEBA+, it is not as good as WP since it does not eliminate the wasteful transmission completely. Similar trends can be seen in Figure 9b and 9c as well.

Then, we study the time cost of different methods. As shown in Figure 10, WP significantly outperforms all the prior schemes in terms of time cost. Specifically, when $N = 100,000, n = 80,000, \varepsilon = 0.01, m = 10$, WP is as much as 9.9x, 9x, and 2.1x faster than SEBA, SEBA+, and SBF, respectively, as shown in Figure 10a. This advantage mainly comes from the compact composite message design that has no wasteful information involved in the communication. Note that in Figure 10b and 10c, the plots for SEBA and SEBA+ are out of range of the vertical axes.

D. Comparison under unreliable channels

Next, we are going to do comparisons with unreliable channel conditions. The comparisons are done with the traces collected using a commercial reader, Impinj speedway R420 and 50 Alien Squiggle tags. The main reason of doing this with commercial devices is that the maximum output power

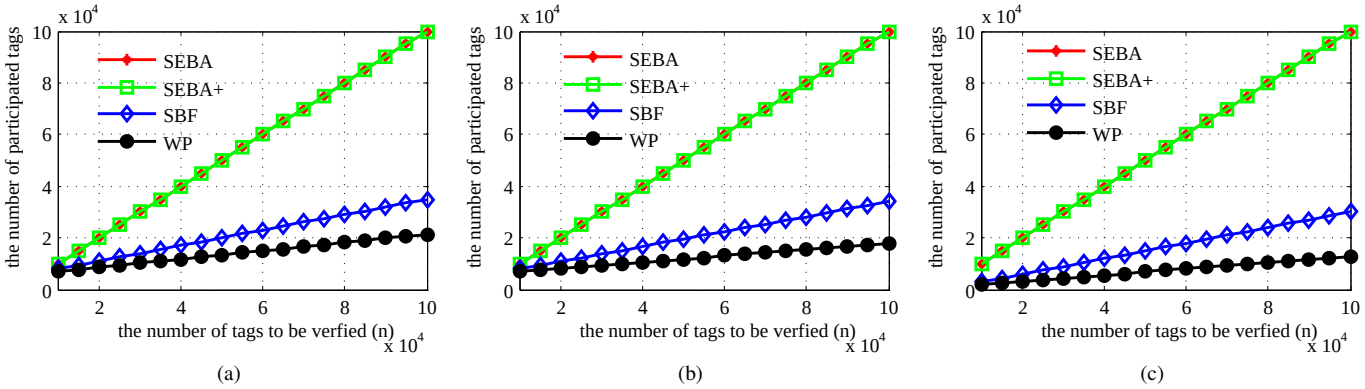


Fig. 9. When $N = 100,000, \varepsilon = 0.01$, the number of tags to be verified VS the number of participated tags. (a) $m = 10$; (b) $m = 20$; (c) $m = 50$.

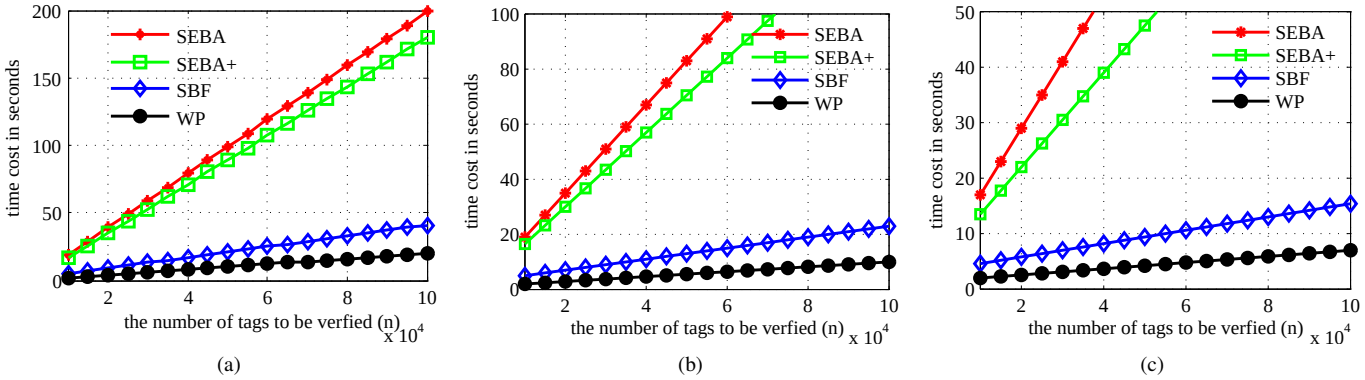


Fig. 10. When $N = 100,000, \varepsilon = 0.01$, the number of tags to be verified VS time cost. (a) $m = 10$; (b) $m = 20$; (c) $m = 50$.

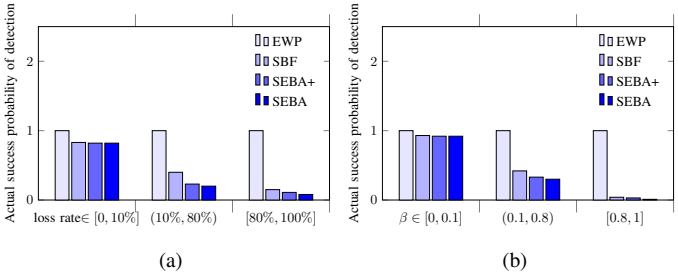


Fig. 11. Actual success probability of detection with different channel conditions. (a) Under different loss rate scenarios, EWP consistently meets the predefined requirement $\varepsilon = 0.01$ while all the others schemes fail; (b) the same observation can be made under different β value cases.

of a USRP RFX900 daughterboard is only 23 dBm, which is far less than that of the Impinj reader, 32.5 dBm. Such limited power make it difficult to observe the characteristics of realworld channels, e.g., path loss and multipath interference, because the maximum reading range is about 2 meters. In the experiments, tags are randomly placed at different distances and orientations to the reader's antennas, creating diverse channel conditions.

We compare EWP with SBF, SEBA+, and SEBA. There is standard channel hopping (each channel is used at the maximal dwell time) for SBF, SEBA+, and SEBA while EWP includes our adaptive channel hopping. Also only EWP has channel estimation whereas other schemes do not. These are because EWP is the first UTD scheme that considers the burstiness, channel estimation, and channel hopping for dynamic channels. We set $N = 1,000, n = 50, m = 10$, and $\varepsilon = 0.01$. First, we classify the links into three groups based on channel fading loss rates: good links with loss rates in $[0, 10\%]$, intermediate links with loss rates in $(10\%, 80\%)$, and poor links with loss rates in $[80\%, 100\%]$. Without loss of generality, we assume the quality of reader-to-tag links is the same as its of tag-to-reader links. For the SBF implementation, we choose its noise-resistant version SBF-UTD and set its multiple copies parameter $c = 2$ according to their report [1]. As shown in Figure 11a, we observe that the results of SEBA and SEBA+ are dramatically biased from the predefined threshold $\varepsilon = 0.01$. While SBF is relatively more robust than SEBA and SEBA+ due to its countermeasures for link errors, it is still unable to deliver the desirable reliability. The reason is that the countermeasures of SBF mainly focus on reader-to-tag links. In contrast, EWP maintains the high success probability of detection due to its error-resistant design on both reader-to-tag and tag-to-reader links and the collision-aware channel estimation scheme. Next, we classify the links into three groups based on β values: nearly independent links with β in $[0, 0.1]$, intermediate bursty links with β in $(0.1, 0.8)$, and highly bursty links with β in $[0.8, 1]$. In this test, the EWP adopts our adaptive channel hopping scheme using the thresholds $p_f = 0.5$ and $\beta = 0.8$. The results are shown

in Figure 11b. As expected, the other three methods fail to meet the goal in all the different β value cases, although the performance with nearly independent links is better than that with intermediate and highly bursty links. On the contrary, EWP's actual success probability of detection can meet the goal all the time because it employs the adaptive channel hopping scheme to counter the effects of bursty and high loss rate links. Particularly, when $\beta \in [0.8, 1]$, the actual success probability of SBF, SEBA+, SEBA are only 15.2%, 11.1%, and 8.2%, respectively, failing to meet the required success probability 99%, while its of EWP is 99.3%.

VII. Related Work

The first probabilistic unknown tag detection scheme, SEBA, is proposed in [4]. In SEBA, the reader first builds a supposed echo sketch in the back-end server, then compares it with the measured echo sketch to detect unknown tags. Later SEBA+ [10] is introduced to improve the performance of SEBA based on the bloom filter. By further exploring the characteristics of bloom filter, Liu et al. [1] combine the standard bloom filter and a sampling process to propose the Sampling Bloom Filter (SBF) for fast unknown tag detection. Although those probabilistic schemes can effectively pinpoint unknown tag events, they still suffer from inefficient communication due to the wasteful transmission of unhelpful data.

Several unknown tag identification schemes are proposed to exactly find all the unknown tags in a batch [6]. When applied in unknown tag detection applications, those schemes cost much more time and energy than probabilistic detection methods [1]. There are also a number of probabilistic solutions for many other RFID problems. Probabilistic estimation schemes are proposed to acquire the approximate size of tags in interested regions [11], [17], [21], [27], [28], [29]. But those methods only count the number of tags and so are unable to distinguish unknown tags from known ones. Several exact identification and probabilistic detection of missing tags are introduced in [18], [30]. Nevertheless, missing tag problems always assume all the information about the to-be-tested tags are known in advance, which is hard to meet in the unknown tag detection. Furthermore, they can only find missing tags, but not unknown tags.

Countermeasures for unreliable channels in RFID management are examined in [1], [11], [12], [13]. While [1] mainly deals with unreliable reader-to-tag links, [11], [12], [13] do not calibrate measured packet losses due to collisions that are quite common in RFID systems. More importantly, all of these schemes do not handle bursty errors very well. Even though [12], [13] have discussed bursty errors, their solutions basically rely on more retransmission. In contrast, our solution not only differentiates the channel fading losses from collision losses, and also leverages an adaptive channel hopping mechanism to avoid highly bursty channels.

VIII. Discussions

Error Model: Although our model incorporates both channel fading error and collision error that is not considered in previous unknown tag detection, it can be made more comprehensive by including more practical factors, such as hardware imperfection, decoding error, and interference dynamics. Hardware imperfection includes mutual coupling effects due to closely located tags or reader antennas and phase offsets. Coupling effects can be modeled using mutual impedance based on the antenna shape and material [31]. Environmental interferences mainly involve human mobility and tag motion. Usually the impact of tag motion is more serious than its of human mobility, because it changes tag orientation and position. Declutter techniques could help differentiate wanted signals from a number of reflected ones and further eliminate the negative impact brought by mobility [32].

Adaptation to Other Schemes: Although the error-resistant techniques in this paper are designed to work with our protocol WP, some general principles are helpful to its competitors, e.g., SBF, SEBA, and many other RFID applications, including missing-tag detection and tag searching. For instance, SEBA [4] would definitely benefit from the adaptive channel hopping to combat bursty errors because it is totally unaware of dynamic channel conditions. THP [19] could also benefit from our loss-rate calibration model to improve its estimates under collisions. Nevertheless, the implementable adaptation with specific existing protocols [4], [19] requires working closely with different protocol parameters, such as communication rounds and collision numbers. We hope this work could inspire more community interest to design various robust solutions along this line and even more general error-resistant techniques.

IX. Conclusion

In this paper, we have proposed a fast and reliable probabilistic unknown tag detection scheme. At its core, we observed that much data in prior communication messages was of no help in detecting unknown tags. Thus, we have proposed a compact message design that included only informative data, excluding all the unhelpful data from communication. In addition, various energy-time tradeoffs have also been achieved in our analytic framework. An extended protocol dealing with unreliable channels has been considered using novel packet loss differentiation and adaptive channel hopping techniques. Our analysis and experiments have showed that the proposed protocols can significantly outperform previous methods in terms of time efficiency, energy efficiency, and robustness. In the future, we are going to investigate applications of the highly efficient detection schemes in this work, e.g., efficient unknown tag identification solution, tag search and missing-tag identification in the presence of unknown tags.

Acknowledgments

This work was supported by an Industrial Canada Technology Demonstration Program, an NSERC Discovery Grant, an NSERC E.W.R. Steacie Memorial Fellowship, a Mitacs Accelerate Internship, and the Project NSFC under Grant 61402372.

REFERENCES

[1] X. Liu, H. Qi, K. Li, I. Stojmenovic, A. X. Liu, Y. Shen, W. Qu, and W. Xue. Sampling Bloom Filter-Based Detection of Unknown RFID Tags. *IEEE Transactions on Communications*, 63(4):1432–1442, 2015.

[2] J. Han, C. Qian, P. Yang, D. Ma, Z. Jiang, W. Xi, and J. Zhao. GenePrint: Generic and Accurate Physical-Layer Identification for UHF RFID Tags. *IEEE/ACM Transactions on Networking*, 24(2):846–858, 2016.

[3] Q. Xiao, B. Xiao, S. Chen, and J. Chen. Collision-aware churn estimation in large-scale dynamic rfid systems. *IEEE/ACM Transactions on Networking*, 25(1):392–405, Feb 2017.

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

[5] S. Muhammad and A. X. Liu. Fast and Reliable Detection and Identification of Missing RFID Tags in the Wild. *IEEE/ACM Transactions on Networking*, PP(99):1–1, 2016.

[6] X. Liu, K. Li, G. Min, K. Lin, B. Xiao, Y. Shen, and W. Qu. Efficient Unknown Tag Identification Protocols in Large-Scale RFID Systems. *IEEE Transactions on Parallel and Distributed Systems*, 25(12):3145–3155, 2014.

[7] Third business mission focuses on cargo-tracking technology. http://www.winnipegfreepress.com/business/centreport-heading-back-to-china-147708545.html.

[8] S. Chunlin. Sino-u.s.cooperation on marine transportation security: progress and problems. 2010.

[9] OpenBeacon. *http://www.openbeacon.org/*.

[10] G. Bianchi. Revisiting an RFID Identification-Free Batch Authentication Approach. *Communications Letters, IEEE*, 15(6):632–634, 2011.

[11] Y. Zheng and M. Li. Towards More Efficient Cardinality Estimation for Large-Scale RFID Systems. *IEEE/ACM Transactions on Networking*, 22(6):1886–1896, Dec 2014.

[12] W. Luo, S. Chen, Y. Qiao, and T. Li. Missing-Tag Detection and Energy-Time Tradeoff in Large-Scale RFID Systems With Unreliable Channels. *IEEE/ACM Transactions on Networking*, 22(4):1079–1091, 2014.

[13] Mei Chen, Wan Luo, Zhen Mo, Shigang Chen, and Yi Fang. An Efficient Tag Search Protocol in Large-Scale RFID Systems With Noisy Channel. *IEEE/ACM Transactions on Networking*, 24(2):703–716, 2016.

[14] Gen 2 RFID Tools. *https://moo.cmcl.cs.cmu.edu/trac/cgran/wiki/Gen2*.

[15] WISP Platform. *http://wisp.wikispaces.com/WISPFirmware*.

[16] EPCglobal Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz-960MHz. 2008.

[17] M. Kodialam and T. Nandagopal. Fast and Reliable Estimation Schemes in RFID Systems. In *Proc. of ACM MOBICOM*, 2006.

[18] C. Tan, B. Sheng, and Q. Li. How to Monitor for Missing RFID tags. In *Proc. of IEEE ICDCS*, 2008.

[19] T. Li, S. Chen, and Y. Ling. Efficient Protocols for Identifying the Missing Tags in a Large RFID System. *IEEE/ACM Transactions on Networking*, 21(6):1974–1987, 2013.

[20] P. Zhang, J. Gummeson, and D. Ganesan. BLINK: A High Throughput Link Layer for Backscatter Communication. In *Proc. of ACM MobiSys*, 2012.

[21] Z. Zhou, B. Chen, and H. Yu. Understanding RFID Counting Protocols. *IEEE/ACM Transactions on Networking*, 24(1):312–327, 2016.

[22] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level Measurements from an 802.11b Mesh Network. In *Proc. of ACM SIGCOMM*, 2004.

[23] A. Köpke, A. Willig, and H. Karl. Chaotic Maps as Parsimonious Bit Error Models of Wireless Channels. In *Proc. of IEEE INFOCOM*, 2003.

[24] Kannan Srinivasan, Maria A Kazandjieva, Saatvik Agarwal, and Philip Levis. The β-factor: Measuring Wireless Link Burstiness. In *Proc. of ACM SenSys*, 2008.

[25] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Proc. of IEEE ICCV*, 1998.

[26] Philips Semiconductors. I-CODE smart label RFID tags.

[27] W. Gong, I. Stojmenovic, A. Nayak, K. Liu, and H. Liu. Fast and Scalable Counterfeits Estimation for Large-Scale RFID Systems. *IEEE/ACM Transactions on Networking*, 24(2):1052–1064, 2016.

[28] W. Gong, J. Liu, K. Liu, and Y. Liu. Toward more rigorous and practical cardinality estimation for large-scale rfid systems. *IEEE/ACM Transactions on Networking*, PP(99):1–12, 2016.

[29] W. Gong, H. Liu, L. Chen, K. Liu, and Y. Liu. Fast Composite Counting in RFID Systems. *IEEE/ACM Transactions on Networking*, PP(99):1–1, 2015.

[30] Y. Zheng and M. Li. P-MTI: Physical-Layer Missing Tag Identification via Compressive Sensing. *IEEE/ACM Transactions on Networking*, 23(4):1356–1366, 2015.

[31] H. S. Lui, H. T. Hui, and M. S. Leong. A note on the mutual-coupling problems in transmitting and receiving antenna arrays. *IEEE Antennas and Propagation Magazine*, 51(5):171–176, 2009.

[32] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti. WiDeo: Fine Grained Device-Free Motion Tracing using RF Backscatter. In *Proc. of USENIX NSDI*, 2015.

Wei Gong (M'14) received the B.S. degree from the Department of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2003 and the M.S. and Ph.D. degrees in School of Software and Department of Computer Science and Technology from Tsinghua University, Beijing, China, in 2007 and 2012, respectively. His research interests include RFID applications, wireless networks, and mobile computing.

Jiangchuan Liu (S'01-M'03-SM'08-F'17) received B.Eng. (Cum Laude) from Tsinghua University, Beijing, China, in 1999, and Ph.D. from The Hong Kong University of Science and Technology in 2003. He is currently a Full Professor (with University Professorship) in the School of Computing Science at Simon Fraser University, British Columbia, Canada. He is an IEEE Fellow and an NSERC E.W.R. Steacie Memorial Fellow.

He is a Steering Committee Member of IEEE Transactions on Mobile Computing, and Associate Editor of IEEE/ACM Transactions on Networking, IEEE Transactions on Big Data, and IEEE Transactions on Multimedia. He is a co-recipient of the Test of Time Paper Award of IEEE INFOCOM (2015), ACM TOMCCAP Nicolas D. Georganas Best Paper Award (2013), and ACM Multimedia Best Paper Award (2012).

Zhe Yang (S'08-M'13) received the B.S. degree in information engineering from Xian Jiaotong University, Xian, China, in 2005, and the Ph.D. degree in electrical and computer engineering from the University of Victoria, Victoria, BC, Canada,in 2013. He is currently an Associate Professor with the School of Computer Science, Northwestern Polytechnical University, Xian. His current research interests include cross-layer design, scheduling and resources allocation for wireless networks, and synchronization.