Anonymous Temporal-Spatial Joint Estimation at Category Level over Multiple Tag Sets

Youlin Zhang* Shigang Chen* You Zhou* Yuguang Fang†
*Department of Computer & Information Science & Engineering,
†Department of Electrical & Computer Engineering,
University of Florida, Gainesville, FL 32611, USA,
Email: {youlin, sgchen, youzhou}@cise.ufl.edu fang@ece.ufl.edu

***Abstract*—Radio-frequency identification (RFID) technologies have been widely used in inventory management, object tracking and supply chain management. One of the fundamental system functions is called cardinality estimation, which is to estimate the number of tags in a covered area. We extend the research of this function in two directions. First, we perform joint cardinality estimation among tags that appear at different geographical locations and at different times. Moreover, we collect category-level information, which is more significant in practical scenarios where we need to monitor the tagged objects of many different types. Second, we require anonymity in the process of information gathering in order to preserve the privacy of the tagged objects. These capabilities will enable new applications such as tracking how products are moved in a large, distributed supply network. We propose a novel protocol design to meet the requirements of anonymous category-level joint estimation over multiple tag sets. We formally analyze the performance of our estimator and determine the optimal system parameters. Extensive simulations show that the proposed protocol can efficiently obtain accurate category-level estimation, while preserving tags' anonymity.**

I. Introduction

Recent research on RFID technologies has established new system functions that support novel applications in inventory management, logistics, product tracking, and supply chain management [3], [5], [6], [9], where tags are attached to objects in an area of surveillance, and an RFID reader is deployed with one or multiple antennas placed at chosen locations to monitor the set of tagged objects.

Cardinality Estimation: One of the fundamental system functions is called *cardinality estimation*, which is to estimate the number of tags in a surveillance area. This function has wide applications in warehouse management such as detecting management errors, theft, and vendor fraud [13]. It is also useful for other applications (e.g., transferring commercial goods at a port) that only require a reader to estimate the number of tagged objects, without the need of accessing the tag IDs for the purpose of keeping the anonymity of customer products. Numerous solutions [5]–[7], [11], [18], [20], [21] have been proposed, and they take much less time than the traditional approach of identifying all tag IDs and then counting the number. For warehouse applications, time efficiency is important for minimizing disruptions to normal operations in a busy environment [10]. One limitation of the aforementioned work is that they only consider cardinality estimation of *a single tag set* [5], [6], [10], [11], [18], [20].

Multiple Tag Sets: Consider a large, distributed supply-chain network, where products are tagged and shipped through the network from location to location. Clearly, the set of tags at the storage facility of any location will change as products are moved in and out over time. We consider each tag set as a spatial-temporal function with respect to location and time, representing the set of products at a given location and a given time.

Suppose an RFID reader (possibly with multiple antennas) is installed at every location to take a snapshot of the local tag set periodically after each time interval (e.g., a day), where we define snapshot as a data structure that anonymously encodes the information of a tag set, without carrying any tag ID. It will be very useful to develop a system function that performs joint estimation over snapshots from different locations or from the same location but at different times. For example, given a snapshot from location A on date 1 and another snapshot from location B on date 2, if the function can estimate the number of common tags in the two tag sets, we will know how many products are shipped from A to B between the dates. By monitoring such pairs of snapshots on other dates, we will gain a good picture about how products are moved between these two locations over time. We can generalize this function to three or more snapshots: Given an arbitrary number of snapshots from different locations at chosen times, the joint-estimation function estimates the number of common tags that show up in all tag sets that the snapshots represent. This function allows us to learn information about the volume of products moving along a chain of locations during the times under consideration. When we apply the function to different location chains, we will gather a detailed picture about how products are moved in the whole network.

There is recent work studying joint estimation of *two tag sets*, i.e., estimating the cardinality of the intersection of the two sets [15], [17]. Their solutions cannot be easily extended to joint estimation over *an arbitrary number of tag sets*. This more difficult problem is solved by [8], [16]. One practically important limitation is that all the above work [8], [15]–[17] can only tell us the total volume of all products moving from location to location, but cannot tell how each type of products flows through the network.

Multiple Categories: In the example of supply-chain network, there may be numerous different types of products.

978-1-5386-4128-6/18/$31.00 ©2018 IEEE

Practically, it is much more useful to know how each type of products flow through the network than the total number of products shipped from one location to another. To support product types, we put tags into categories, one category for each product type, with all tags in the same category sharing a common category ID. Given an arbitrary number of snapshots from different locations at chosen times, the problem of joint estimation at category level is to estimate the number of *common tags in each category*, which show up in all the snapshots; recall that each snapshot records a tag set. That is, we want to anonymously estimate the cardinality of the intersection over multiple tag sets for each category.

We mentioned earlier the protocols [8], [16] that estimate the cardinality of the intersection of multiple tag sets. One may suggest that we can apply them repetitively, one category at a time, to obtain category-level information. This can be done by the reader announcing one category ID each time so that only tags with that category ID will respond. Such an approach however breaks the anonymity of category IDs (which may be more important than individual object IDs because they reveal the product types). More importantly, the approach is inefficient as we will demonstrate in this paper.

There is very limited prior work that supports tag categories. Related is the work that classifies the categories in a single tag set [10], [12]. The only work that performs category-level joint estimation [4] can deal with only two tag sets, and its analytical framework cannot be easily extended to more sets. This paper proposes a new protocol for anonymous category-level joint estimation over multiple tag sets. To keep the anonymity of tags, we thoroughly mix the information from tags of all categories in each snapshot, preventing unauthorized readers to guess the tag/category IDs even if they are able to eavesdrop on the communication. To perform joint estimation, we combine the snapshots and remove the noise (due to mixing) in the combination to obtain the cardinality of the intersection of multiple tag sets for each category. There are two kinds of noise: inter-set noise and inter-category noise. We use statistical methods to estimate and then remove the noises.

The main contributions of our protocol are summarized as follows: First, we extend the traditional RFID estimation problem to more practical scenarios for estimating the category-level information over multiple tag sets at different locations and/or different times. Not only is the problem more challenging, but the proposed solution allows us to learn the spatial-temporal dynamics among these tag sets and their associated objects.

Second, we enforce anonymity in the proposed multi-set category-level tag estimation protocol.

Third, we formally analyze the performance of our protocol. Through statistical analysis, we show that our estimator for category-level information is asymptotically unbiased and can be made to meet any preset requirement on estimation accuracy. Our simulations show that the proposed protocol can efficiently obtain accurate category-level estimation, while preserving tags' anonymity.

II. System Model and Problem Definition

A. System Model

Consider a distributed RFID system, where all the objects are classified into m different categories with a set M of *category IDs*, $\{cid_1, cid_2, ..., cid_m\}$. Each object is attached with a tag and can be uniquely identified by a *tag* ID id, which contains a category ID cid and an object ID tid, with the former specifying which category the object belongs to and the latter being unique in the same category. The length of a tag ID is typically 96 bits or 128 bits long. Using 16 bits for category IDs can support 65,536 different categories, which is sufficient for a large RFID-assisting supply-chain network with tens of thousands of suppliers.

Tags are powered by the RF waves from the antennas of a reader, and communicate with the reader by backscattering and modulating the received signals, using a frame-slotted ALOHA protocol. The reader initiates communication by broadcasting a request and records a snapshot of the tag set based on information sent back from the tags. We will define the snapshot structure later.

Suppose unauthorized adversaries may plant readers at chosen locations to eavesdrop on the communication between tags and readers, from which they try to infer private information such as tag IDs and category IDs about the products. We assume that the adversaries do not have prior knowledge of the tag IDs or category IDs in the system. We want to prevent them from knowing the IDs.

B. Problem Definition

Consider k snapshots, denoted as $B_1, B_2, ..., B_k$, which are captured at different locations or at the same location but different times. The tag sets that they encode are $T_1, T_2, ..., T_k$, respectively. Let C_i^{cid} be the subset of tags in T_i that belong to a category $cid \in M$. Clearly, $T_i = C_i^{cid_1} \cup C_i^{cid_2} ... \cup C_i^{cid_m}$. We will study the joint property of the k tag sets for each category in M.

Let $C_*^{cid} = C_1^{cid} \cap C_2^{cid} ... \cap C_k^{cid}$ and $n_*^{cid} = |C_*^{cid}|$, where the subscript $*$ means the *common tags* among the k subsets. Because all operations are applied to each category independently and separately, in the sequel we will leave out the superscript cid in operation description to simplify the notations. We abbreviate C_i^{cid}, C_*^{cid} and n_*^{cid} as C_i, C_* and n_* respectively.

The problem of category-level joint estimation over multiple tag sets is to estimate n_* as $\hat{n}_*$ for each category with the following accuracy requirement:

$$Prob\{|\hat{n}_* - n_*| \leq e\} \geq \alpha. \quad (1)$$

where e is an absolute error bound and α is a probability, which is referred as *confidence level*. For example, if $\alpha = 95\%$ and $e = 50$, the estimation requires that the probability for the estimation error $|\hat{n}_* - n_*|$ to be bounded by 50 is at least 95%.

The prior work [5]–[7], [20], [21] on cardinality estimation of a *single* tag set adopts a relative error model:

$$Prob\{|\hat{n}_* - n_*| \leq \varepsilon n_*\} \geq \alpha. \quad (2)$$

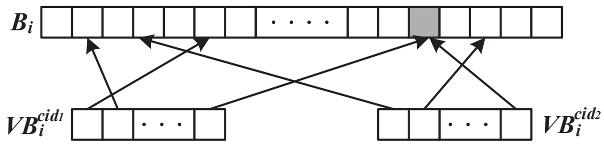


Fig. 1: An illustration of drawing bits randomly from a bitmap B_i to construct virtual bitmaps $VB_i^{cid_1}$ and $VB_i^{cid_2}$. The bit in grey is shared by both virtual bitmaps.

where ε is the relative error requirement. However, for joint estimation over multiple sets, the execution time is inversely related to the Jaccard similarity, $J = \frac{n_*}{n}$, where n_* is the number of common tags in the tag sets and n is the total number of tags in all sets. For example, the time complexity of [8] is $\Theta(\frac{1}{\varepsilon^2 J} \ln \frac{1}{1-\alpha})$, under the relative error model. While n is typically big for a large RFID system, the value of n_* can be large or small, even down to zero, causing the term $\frac{1}{J}$ to approach infinity. That is the reason why the more recent work of [16] advocates the absolute error model (1), which we adopt in this paper. For the prior work [5]–[7], [20], [21] on a single set, their Jaccard similarity is one since $n = n^*$. Therefore, the relative error model is fine.

C. Performance Metrics

We use three metrics for performance evaluation.

1) Execution Time: Since RFID systems operate in low-rate communication channels, time efficiency is an important performance metric, especially when the number of tags is very large. Therefore, it is desirable to design a protocol that can reduce execution time as much as possible.

2) Estimation Accuracy: The accuracy requirement is specified in (1).

3) Anonymity: We use two probability values to quantify the preserved anonymity of tag IDs and category IDs respectively. More specifically, p_{id} (p_{cid}) is the probability that any tag (category) ID is not revealed by an adversary that eavesdrops on all wireless communications. In practice, we want to make p_{id} and p_{cid} as close to 1 as possible.

III. Anonymous Temporal-Spatial Joint Estimation at Category Level over Multiple Tag Sets

In this section, we present our new protocol for anonymous temporal-spatial Joint Estimation at Category level over Multiple tag sets (JECM). The protocol consists of two phases: an online encoding phase where snapshots are taken, each encoding a tag set at a certain location and a certain time in the system, and an offline analysis phase in which all snapshots are brought to a server where joint-estimation queries are made. We adopt an asymmetric design that pushes most complexity to the central server, while keeping the tag operation simple. The only thing that a tag needs to do is to make a single transmission in response to a reader's request during online encoding.

A. Structure of Snapshot

Consider an arbitrary tag set T_i at a certain location and a certain time in a large distributed RFID system. To take a snapshot, the reader broadcasts a request, which is followed by a slotted ALOHA frame. Upon receipt of the request, each tag pseudo-randomly selects a slot in the frame and sends back a signal response in the slot. The reader monitors the status of each slot, which is referred to either as an *empty slot* where no tag responds or as a *busy slot* where one or more tags respond. The reader converts the time frame into a bitmap B_i, zero for each empty slot and one for each busy slot. We use B_i as the snapshot of T_i.

Encoding category-level information is tricky. Establishing one bitmap for each category is problematic. As discussed in the introduction, this requires the reader to broadcast one request per category, carrying a category ID to ask only tags in the category to respond, which breaks anonymity. Moreover, this approach takes long execution time; see Section IV. Our idea is to mix information from all categories in the same bitmap to improve anonymity and time efficiency as no category ID will be transmitted and it takes just one request-response round to build the snapshot for all categories. To do so, we must introduce additional structure to the snapshot B_i. For each category cid, we pseudo-randomly select a certain number l of bits from B_i to encode tags of that category. Logically, these bits form a *virtual bitmap* VB_i^{cid}. Fig. 1 illustrates the idea of drawing bits from B_i to form two virtual bitmaps, $VB_i^{cid_1}$ and $VB_i^{cid_2}$, for two categories cid_1 and cid_2, respectively. Different categories may share bits in B_i due to random bit selection, which brings two benefits: First, information from different categories is mixed, which is good for anonymity. Second, it improves time efficiency. The value of l has to be set reasonably large so that there are sufficient bits to encode large categories. If separate bits were designated for different categories, many bits for small categories may be left unused. Thanks to random sharing, in our design, the unused bits for small categories can be picked up by other categoies, which reduces the total number of bits (time slots) needed.

The challenge is how to take a snapshot with embedded category structure and how to perform accurate joint estimation when category level information is mixed.

B. Online Encoding

Denote the jth bit in the bitmap as $B_i[j]$, $0 \le j \le f-1$. Consider an arbitrary category cid, whose jth bit is denoted as $VB_i^{cid}[j]$, $0 \le j \le l-1$. Since our discussion is involved only a single category ID, we will leave out the superscript cid. The selection of $VB_i[j]$ from B_i is formally defined as

$$VB_i[j] \equiv B_i[H_j(cid)]. \tag{3}$$

where $H_j()$ is a hash function. Instead of requiring l different hash functions, we implement them based on a common master hash function $H(...)$ and l different random seeds, $r_0, r_1, ..., r_{l-1}$,

$$H_j(cid) = H(cid \oplus r_j), 0 \le j \le l-1. \tag{4}$$

For online encoding, the reader broadcasts a request, including the frame size f and l random seeds. The request is

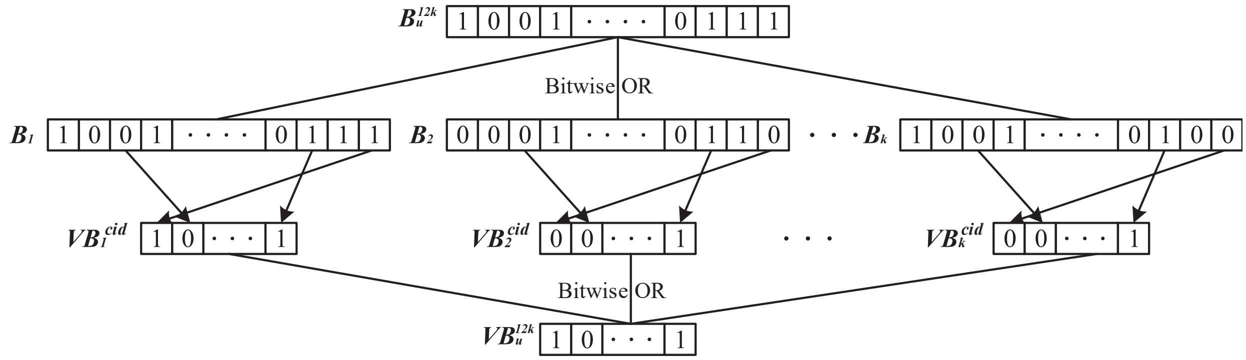


Fig. 2: An illustration of how to construct B_u^{12k} and VB_u^{12k} for subsets $\{B_1, B_2, B_k\}$ and $\{VB_1, VB_2, VB_k\}$, respectively.

followed by a time frame of f slots. Consider an arbitrary tag. Without losing generality, suppose the tag belongs to category cid. It should be encoded by one of the bits in the category's virtual bitmap VB_i. The tag uses another hash function $h(tid) \in [0, l-1]$ to choose a bit pseudo-randomly from VB_i, where $h()$ may also be implemented using the master hash function with a pre-defined seed. By (3), the bit $VB_i[h(tid)]$ is actually the $H_{h(tid)}(cid)$th bit in B_i, which corresponds to the $H_{h(tid)}(cid)$th slot in the time frame. Hence, the tag will choose that time slot to transmit. Once the reader finds the $H_{h(tid)}(cid)$th slot is busy, it sets

$$B_i[H_{h(tid)}(cid)] = 1. \tag{5}$$

C. Offline Category-level joint estimation over multiple Tag Sets

With online encoding, snapshots of different tag sets are sent to a central server. Consider joint estimation over an arbitrary set of k snapshots, $\{B_1, B_2, ..., B_k\}$. There are $2^k - 1$ subsets of $\{B_1, B_2, ..., B_k\}$, excluding the empty subset. Consider an arbitrary subset $\{B_{c_1}, B_{c_2}, ..., B_{c_i}\}$, where $1 \le i \le k$ and $1 \le c_1 < c_2 < ... < c_i \le k$. The server combines the bitmaps in the subset by bitwise OR. Namely, the combined bitmap $B_u^{c_1 c_2 ... c_i}$ is defined as

$$B_u^{c_1 c_2 ... c_i} = B_{c_1} \vee B_{c_2} \vee ... \vee B_{c_i}. \tag{6}$$

where $\vee$ is the bitwise-OR operation. As a result, the information from tags in different bitmaps is combined. For example, B_u^{12} is the combined bitmap of B_1 and B_2. The combined bitmaps will be used later in our estimation.

The server retrieves per-category information from the k tag sets by reconstructing the bitmap VB_i (short for VB_i^{cid}) as follows:

$$VB_i[j] \equiv B_i[H_j(cid)], \; 1 \le i \le k, 0 \le j \le l-1. \tag{7}$$

Again, there are $2^k - 1$ subsets of $\{VB_1, VB_2, ..., VB_k\}$, excluding the empty subset. Consider an arbitrary subset $\{VB_{c_1}, VB_{c_2}, ..., VB_{c_i}\}$, where $1 \le i \le k$ and $1 \le c_1 < c_2 < ... < c_i \le k$. The server constructs the combined virtual bitmap $VB_u^{c_1 c_2 ... c_i}$ as

$$VB_u^{c_1 c_2 ... c_i} = VB_{c_1} \vee VB_{c_2} \vee ... \vee VB_{c_i}. \tag{8}$$

Fig. 2 shows an example of how to construct B_u^{12k} and VB_u^{12k} for subsets $\{B_1, B_2, B_k\}$ and $\{VB_1, VB_2, VB_k\}$, respectively.

For each snapshot, virtual bitmaps for all categories share the bits in the same underlying bitmap B_i, $1 \le i \le k$. A bit '1' in VB_i may not be set by tags belonging to category cid, but instead be set by tags of other categories, resulting inter-category noise in virtual bitmaps. When bitmaps and virtual bitmaps are combined, a bit in $B_u^{c_1 c_2 ... c_i}$ or $VB_u^{c_1 c_2 ... c_i}$ may be set to one by tags from different sets, which introduces inter-set noise. In deriving our estimation formula, we must consider the inter-set and inter-category noises caused by bit sharing.

Below we give the formula that the server uses to estimate the number n_* of common tags in all k tag sets for category cid. The actual derivation comes next. Let $\hat{n}_*$ be the estimator. We have

$$\hat{n}_* = \frac{\sum_{i=1}^{k} [(-1)^{i+1} \sum_{1 \le c_1 < ... < c_i \le k} (\ln V_{c_1 c_2 ... c_i} - \ln U_{c_1 c_2 ... c_i})]}{\ln(1 - \frac{1}{l}) - \ln(1 - \frac{1}{f})}. \tag{9}$$

where $V_{c_1 c_2 ... c_i}$ is the fraction of bits in $VB_u^{c_1 c_2 ... c_i}$ that are zeros, and $U_{c_1 c_2 ... c_i}$ is the fraction of bits in $B_u^{c_1 c_2 ... c_i}$ that are zeros.

D. Derivation of $\hat{n}_*$

We use probabilistic methods to analyze the expected fraction of zero bits in all the bitmaps and virtual bitmaps we obtain and derive the estimator $\hat{n}_*$ for n_*.

First, we want to prove the following theorem.

Theorem 1: For an arbitrary bit b in an f-bit bitmap, a tag t has a probability $\frac{1}{f}$ to set it as one.

Proof: Assume we have an f-bit bitmap B and an l-bit virtual bitmap VB, which belongs to tags of category cid. Let random variable X be the number of bits that are set by the virtual bitmap VB in B. Since a tag sets a bit in B in a uniformly random way, this problem can be cast into bins and balls problem. We have

$$P(X=x)=\frac{\binom{f}{x}\times x!\times S(l,x)}{f^l}. \tag{10}$$

where $S(l,x)$ is the stirling number [1] and $S(l,x) = \frac{1}{x!}\sum_{i=0}^{x}(-1)^i\binom{x}{i}(x-i)^l$. $S(l,x)$ computes the number of partitions to place a set of l tags into x bits in the bitmap. For each bit b in the bitmap, it has a probability of $1-\binom{f-1}{x}/\binom{f}{x}$ to be one of these x bit. Besides, each tag t has a probability $\frac{1}{x}$ to be mapped to these x bits. Thus, we obtain the probability for a bit b to be set by a tag t as:

$$\begin{aligned}P_b&=\sum_{x=1}^{l}\frac{\binom{f}{x}\times x!\times S(l,x)}{f^l}\times(1-\frac{\binom{f-1}{x}}{\binom{f}{x}})\times\frac{1}{x}\\&=\frac{1}{f^{l+1}}\times\sum_{x=1}^{l}[\binom{f}{x}\times x!\times S(l,x)]=\frac{1}{f}.\end{aligned} \tag{11}$$

■

Before we continue our derivation of $\hat{n}_*$, we want to first present the basis for our estimator. In set theory, the cardinality of the intersection of k tag sets can be derived based on the well-known *principle of inclusion and exclusion*:

$$\begin{aligned}|C_1\cap C_2...\cap C_k| =& \sum_{1\le c_1\le k}|C_{c_1}| - \sum_{1\le c_1<c_2\le k}|C_{c_1}\cup C_{c_2}| + ...\\&+(-1)^{i+1}\sum_{1\le c_1...<c_i\le k}|C_{c_1}\cup C_{c_2}...\cup C_{c_i}|\\&+(-1)^{k+1}|C_1\cup C_2\cup ...\cup C_k|.\end{aligned} \tag{12}$$

where $C_{c_1}\cup C_{c_2}...\cup C_{c_i}$ is the union of i tag sets and $|C_1\cap C_2...\cap C_k|$ is the category-level joint information of k sets we want to estimate. From equation (12), we can observe that in order to obtain the joint information of k tag sets, we need to compute the cardinality of all 2^k-1 combined sets first. JECM uses the 2^k-1 bitmaps obtained in Section III-C to estimate the cardinality of each combined set and finally obtains category-level joint information of k tag sets by using the principle of inclusion and exclusion. Next we will show how we derive the cardinality of each combined set.

Let us first start with estimating the cardinality of C_i in one tag set T_i with bitmaps B_i and VB_i. In the sequel, we will leave out index i for B_i, VB_i, T_i and C_i, since we now focus on just one set. Remember we also leave out the superscript cid in C_i as mentioned in Section II-B. We denote t as the number of tags in T (all tags of all categories) and n as the number of tags in C (tags belonging to category cid).

For a bitmap B, we denote U as the fraction of zero bits in B. Let $\mathcal{A}_j$ be the event that the jth $0\le j\le f-1$ bit in B remains 0 after online encoding, and 1_{A_j} be the corresponding indicator random variable, that is,

$$1_{\mathcal{A}_j}=\begin{cases}1, & \text{if } B[j]=0,\\0, & \text{if } B[j]=1.\end{cases}$$

Using Theorem 1, we have $P(\mathcal{A}_j)=(1-\frac{1}{f})^t$. Moreover, U is the fraction of zero bits in B and we have $U=\frac{\sum_{j=0}^{f-1}1_{\mathcal{A}_j}}{f}$. Thus,

$$\begin{aligned}E(U)&=\frac{1}{f}\sum_{j=0}^{f-1}E(1_{\mathcal{A}_j})\\&=\frac{1}{f}\sum_{j=0}^{f-1}[1\times P(\mathcal{A}_j)+0\times(1-P(\mathcal{A}_j)]\\&=(1-\frac{1}{f})^t.\end{aligned} \tag{13}$$

Now we will move on to investigate the properties of a virtual bitmap VB. We denote V as the fraction of zero bits in VB. Let $\mathcal{B}_j$ be the event that the jth $0\le j\le l-1$ bit in VB remains 0 after online encoding, and $1_{\mathcal{B}_j}$ be the corresponding indicator random variable. Similarly,

$$1_{\mathcal{B}_j}=\begin{cases}1, & \text{if } VB[j]=0,\\0, & \text{if } VB[j]=1.\end{cases}$$

In this condition, in order to make a bit in VB remain 0, neither the tags in category cid nor the tags belong to other categories can set $VB[j]$. The probability for a tag in category cid not to set $VB[j]$ is $(1-\frac{1}{l})^n$ and the probability for a tag belonging to other categories not to set $VB[j]$ is $(1-\frac{1}{f})^{t-n}$. Thus, we have $P(\mathcal{B}_j)=(1-\frac{1}{l})^n(1-\frac{1}{f})^{t-n}$ and the expected value of V can be derived as:

$$\begin{aligned}E(V)&=\frac{1}{l}\sum_{j=0}^{l-1}E(1_{\mathcal{B}_j})\\&=\frac{1}{l}\sum_{j=0}^{l-1}[1\times P(\mathcal{B}_j)+0\times(1-P(\mathcal{B}_j)]\\&=(1-\frac{1}{l})^n(1-\frac{1}{f})^{t-n}.\end{aligned} \tag{14}$$

Combining (13) with (14), we have

$$E(V)=(1-\frac{1}{l})^n(1-\frac{1}{f})^{-n}E(U). \tag{15}$$

Substituting $E(U)$ and $E(V)$ with U and V respectively, and taking a logarithm on both sides, we derive an estimator for n as:

$$\hat{n}=\frac{\ln V-\ln U}{\ln(1-\frac{1}{l})-\ln(1-\frac{1}{f})}. \tag{16}$$

Recall that we leave out the tag index i in all these formulas. In this way, we can obtain the category-level cardinality information in each tag set.

Now let us investigate the properties of combined bitmaps $B_u^{c_1c_2...c_i}$ and $VB_u^{c_1c_2...c_i}$, $1\le i\le k$.

Let $\mathcal{C}_j$ be the event that jth bit in $B_u^{c_1c_2\cdots c_i}$ remains zero after online encoding and $U_{c_1c_2...c_i}$ be the fraction of zeros in $B_u^{c_1c_2\cdots c_i}$. We denote $1_{\mathcal{C}_j}$ as the indicator random variable of $\mathcal{C}_j$. Since $B_u^{c_1c_2\cdots c_i}$ is the combination of i tag sets $c_1, c_2, ...,$ and c_i, z will remain zero if and only if z is not chosen by any tag in these i sets, that is,

$$P(\mathcal{C}_j) = (1-\frac{1}{f})^{t_{c_1c_2\cdots c_i}}. \quad (17)$$

where $t_{c_1c_2...c_i}$ is the number of tags in all these i tag sets and $t_{c_1c_2...c_i} = |T_{c_1} \cup T_{c_2} ... \cup T_{c_i}|$. Therefore, the expected value of $U_{c_1c_2...c_i}$ can be derived as:

$$\begin{aligned} E(U_{c_1c_2...c_i}) &= \frac{1}{f}\sum_{j=0}^{f-1} E(1_{\mathcal{C}_j}) \\ &= \frac{1}{f}\sum_{j=0}^{f-1}[1\times P(\mathcal{C}_j) + 0\times(1-P(\mathcal{C}_j)] \\ &= (1-\frac{1}{f})^{t_{c_1c_2\cdots c_i}}. \end{aligned} \quad (18)$$

For a combined virtual bitmap $VB_u^{c_1c_2\cdots c_i}$, let $\mathcal{D}_j$ be the event that the jth bit in $VB_u^{c_1c_2\cdots c_i}$, and $1_{\mathcal{D}_j}$ be the corresponding indicator random variable. In this situation, $1_{\mathcal{D}_j}$ will be true only if the following two conditions are satisfied:

1) The jth bit is not chosen by any tag in $C_{c_1} \cup C_{c_2} ... \cup C_{c_i}$.
2) The jth bit is not chosen by any tag in $(T_{c_1} \cup T_{c_2} ... \cup T_{c_i}) - (C_{c_1} \cup C_{c_2} ... \cup C_{c_i})$.

For the first condition, the probability q_1 for it to be satisfied is

$$q_1 = (1-\frac{1}{l})^{n_{c_1c_2\cdots c_i}}. \quad (19)$$

where $n_{c_1c_2...c_i}$ is the number of tags belonging to category cid in all k sets and $n_{c_1c_2...c_i} = |C_{c_1} \cup C_{c_2} ... \cup C_{c_i}|$. Similarly, the probability q_2 for the second condition to be satisfied can be calculated as

$$q_2 = (1-\frac{1}{f})^{t_{c_1c_2\cdots c_i} - n_{c_1c_2\cdots c_i}}. \quad (20)$$

Combining (19) and (20), we have

$$\begin{aligned} P(\mathcal{D}_j) &= q1 \times q2 \\ &= (1-\frac{1}{l})^{n_{c_1c_2\cdots c_i}}(1-\frac{1}{f})^{t_{c_1c_2\cdots c_i} - n_{c_1c_2\cdots c_i}}. \end{aligned} \quad (21)$$

Let $V_{c_1c_2...c_i}$ be the fraction of zeros in $VB_u^{c_1c_2\cdots c_i}$ and the expected value can be derived as

$$\begin{aligned} E(V_{c_1c_2...c_i}) &= \frac{1}{l}\sum_{j=0}^{l-1} E(1_{\mathcal{D}_j}) \\ &= (1-\frac{1}{l})^{n_{c_1c_2\cdots c_i}}(1-\frac{1}{f})^{t_{c_1c_2\cdots c_i} - n_{c_1c_2\cdots c_i}}. \end{aligned} \quad (22)$$

Apply (18) to (22)

$$E(V_{c_1c_2...c_i}) = (1-\frac{1}{l})^{n_{c_1c_2\cdots c_i}}(1-\frac{1}{f})^{-n_{c_1c_2\cdots c_i}} E(U_{c_1c_2...c_i}). \quad (23)$$

Substitute $E(U_{c_1c_2...c_i})$, $E(V_{c_1c_2...c_i})$ with the observed value $U_{c_1c_2...c_i}$, $V_{c_1c_2...c_i}$ respectively, take a logarithm on both sides, and the estimator for $n_{c_1c_2...c_i}$ can be derived as:

$$\hat{n}_{c_1c_2...c_i} = \frac{\ln V_{c_1c_2...c_i} - \ln U_{c_1c_2...c_i}}{\ln(1-\frac{1}{l}) - \ln(1-\frac{1}{f})}. \quad (24)$$

Combining (12), (16) and (24), we have our estimator $\hat{n}_*$ as

$$\hat{n}_* = \frac{\sum_{i=1}^{k}[(-1)^{i+1}\sum_{1\le c_1<...<c_i\le k}(\ln V_{c_1c_2...c_i} - \ln U_{c_1c_2...c_i})]}{\ln(1-\frac{1}{l}) - \ln(1-\frac{1}{f})}. \quad (25)$$

E. Mean and variance of $\hat{n}_*$

Now, we analyze the statistical properties, mean and variance of $\hat{n}_*$.

In order to derive the mean and variance of $\hat{n}_*$, we need to first derive the mean and variance of $-\ln U_{c_1c_2...c_i}$ and $-\ln V_{c_1c_2...c_i}$. Let $\hat{u}_{c_1c_2...c_i} = -\ln U_{c_1c_2...c_i}$ and $\hat{v}_{c_1c_2...c_i} = -\ln V_{c_1c_2...c_i}$.

In [14], K. Whang *et al.* use Taylor expansion to estimate the fraction of zeros in a bitmap and obtain the following results:

$$E(\hat{u}_{c_1c_2...c_i}) = \frac{1}{f}(t_{c_1c_2...c_i} + \frac{e^{\omega}-\omega-1}{2}), \quad (26)$$

$$Var(\hat{u}_{c_1c_2...c_i}) = \frac{1}{f}(e^{\omega}-\omega-1). \quad (27)$$

where $t_{c_1c_2...c_i}$ is the number of tags in all these i tag sets and $\omega = \frac{t_{c_1c_2...c_i}}{f}$. Usually the frame size f is chosen such that ω is very small and $(e^{\omega}-\omega-1)$ is negligible when compared to $t_{c_1c_2...c_i}$. In this case, we will have $E(\hat{u}_{c_1c_2...c_i}) \simeq \frac{t_{c_1c_2...c_i}}{f}$ and the standard derivation, which is the root of $Var(\hat{u}_{c_1c_2...c_i})$ will also be insignificant compared to the mean.

Next we derive the mean and variance of $\hat{v}_{c_1c_2...c_i}$. In [19], M. Yoon *et al.* use Taylor expansion and statistical methods to estimate the fraction of zeros in a virtual bitmap and gives the results as follows:

$$E(\hat{v}_{c_1c_2...c_i}) = \alpha + \frac{e^{\alpha}-\omega'-1}{2l}, \quad (28)$$

$$Var(\hat{v}_{c_1c_2...c_i}) = \frac{1}{l}(e^{\alpha}-\omega'-1). \quad (29)$$

where $\alpha = \frac{t_{c_1c_2...c_i}-n_{c_1c_2...c_i}}{f} + \frac{n_{c_1c_2...c_i}}{l}$, $\omega' = \frac{n_{c_1c_2...c_i}}{l}$, and $n_{c_1c_2...c_i}$ is the number of tags belonging to category cid in all k sets. Similarly, if l is large enough, we can obtain $E(\hat{v}_{c_1c_2...c_i}) \simeq \alpha$.

Combining (26) and (28), we have:

$$\begin{aligned} E(\ln V_{c_1c_2...c_i} - \ln U_{c_1c_2...c_i}) &= E(\hat{u}_{c_1c_2...c_i}) - E(\hat{v}_{c_1c_2...c_i}) \\ &\simeq \frac{t_{c_1c_2...c_i}}{f} - \alpha \\ &= \frac{n_{c_1c_2...c_i}}{f} - \frac{n_{c_1c_2...c_i}}{l}. \end{aligned} \quad (30)$$

Thus, the mean of $\hat{n}_*$ can be derived as:

$$E(\hat{n}_*) = \frac{\sum_{i=1}^{k}[(-1)^{i+1}\sum_{1\le c_1<\ldots<c_i\le k}E(\ln V_{c_1\ldots c_i} - \ln U_{c_1\ldots c_i})]}{\ln(1-\frac{1}{l}) - \ln(1-\frac{1}{f})}$$
$$\simeq \frac{\sum_{i=1}^{k}[(-1)^{i+1}\sum_{1\le c_1<\ldots<c_i\le k}E(\ln V_{c_1\ldots c_i} - \ln U_{c_1\ldots c_i})]}{\frac{1}{f}-\frac{1}{l}}$$
$$\simeq \sum_{i=1}^{k}[(-1)^{i+1}\sum_{1\le c_1<\ldots<c_i\le k} n_{c_1c_2\ldots c_i}]$$
$$= n_*. \tag{31}$$

Similarly, the variance of $\hat{n}_*$ can be calculated as:

$$Var(\hat{n}_*) = \frac{Var\sum_{i=1}^{k}[(-1)^{i+1}\sum_{1\le c_1<\ldots<c_i\le k}(\hat{u}_{c_1\ldots c_i} - \hat{v}_{c_1\ldots c_i})]}{r^2}. \tag{32}$$

where r is a constant and $r = \ln(1-\frac{1}{l}) - \ln(1-\frac{1}{f})$.

In order to derive $Var(\hat{n}_*)$, we need to calculate $Var(\hat{u}_{c_1c_2\ldots c_i})$, $Var(\hat{v}_{c_1c_2\ldots c_i})$ and the covariance of $\hat{u}_{c_1c_2\ldots c_i}$ and $\hat{v}_{c_1c_2\ldots c_i}$. The covariance can also be derived using Taylor expansion, which is similar to the process in [14], [19]. As [19] shows, the covariance of $\hat{u}_1$ and $\hat{u}_{12}$ can be approximated as:

$$\begin{aligned} Cov(\hat{u}_1, \hat{u}_{12}) &= E(\hat{u}_1\hat{u}_{12}) - E(\hat{u}_1)E(\hat{u}_{12}) \\ &= -E(\hat{u}_1)E(\hat{u}_{12}) - \ln E(\hat{u}_1)\ln E(\hat{u}_{12}) \\ &+ \ln E(U_1)E(\hat{u}_{12}) + E(\hat{u}_1)\ln E(U_{12}). \end{aligned} \tag{33}$$

Substituting the formula of $E(U_1)$, $E(U_{12})$, $E(\hat{u}_1)$ and $E(\hat{u}_{12})$, which we have obtained already, we can obtain the covariance. With the covariance of $\hat{u}_{c_1c_2\ldots c_i}$ and $\hat{v}_{c_1c_2\ldots c_i}$, $Var(\hat{u}_{c_1c_2\ldots c_i})$ and $Var(\hat{v}_{c_1c_2\ldots c_i})$, we can calculate the $Var(\hat{n}_*)$ by expanding (32).

F. Analysis of Anonymity

In this section, we analyze the preserved anonymity of a tag while executing our protocol. Let l_{id} and l_{cid} be the length of tag IDs and category IDs (in binary), respectively. Since the unauthorized adversary does not have any prior knowledge of the tag IDs or category IDs in our system, it can only infer one tag ID or category ID for each slot it eavesdrops on. Therefore, the anonymity of our protocol can be characterized by the probability that the adversary identifies the correct tag ID or category ID.

1) Anonymity of Category IDs: For an l_{cid}-bit category ID, there are $2^{l_{cid}}$ possible category IDs. Each category is assigned into an l-bit virtual bitmap drawn from an f-bit bitmap. As a result, each bit in the bitmap will correspond to an average of $\frac{l\cdot 2^{l_{cid}}}{f}$ categories. Since an adversary does not have any prior information about any categories that are mapped to the same slot, the probability for the adversary to infer the correct category ID of a tag is $\frac{1}{\frac{l\cdot 2^{l_{cid}}}{f}} = \frac{f}{l\cdot 2^{l_{cid}}}$. Therefore, the anonymity of a category ID for JECM, namely p_{cid}, is $p_{cid} = 1 - \frac{f}{l\cdot 2^{l_{cid}}}$.

Protocol	p_{cid}	p_{id}
CCF	0	$1-\frac{f}{2^{l_{id}-l_{cid}}}$
MJREP	0	$1-\frac{f}{2^{l_{id}-l_{cid}}}$
JECM	$1-\frac{f}{l\cdot 2^{l_{cid}}}$	$1-\frac{f}{2^{l_{id}}}$

TABLE I: Preserved anonymity of different protocols.

2) Anonymity of Tag IDs: For an l_{id}-bit long tag ID, the bits that are available for an object ID tid are $(l_{id} - l_{cid})$-long. As a result, there are $2^{(l_{id}-l_{cid})}$ possible object IDs per category. Meanwhile, each tag belonging to the same category is randomly assigned to a slot in an l-bit virtual bitmap. Hence, the average number of tags that are mapped to one slot in the same virtual bitmap is $\frac{2^{l_{id}-l_{cid}}}{l}$. According to Section III-F1, the adversary has a probability of $\frac{f}{l\cdot 2^{l_{cid}}}$ to infer the correct category id of a tag. Thus, the probability for the adversary to infer the correct tag ID is $\frac{f}{l\cdot 2^{l_{cid}}} \times \frac{l}{2^{l_{id}-l_{cid}}} = \frac{f}{2^{l_{id}}}$. As a result, the anonymity of a tag ID is given as $p_{id} = 1 - \frac{f}{2^{l_{id}}}$.

Table I shows the preserved anonymity of different protocols when performing category-level joint estimation of multiple tag sets. We can observe that only our JECM protocol can preserve both category and tag ID anonymity, while CCF and MJREP cannot preserve category ID anonymity. In terms of tag ID anonymity, JECM is the highest among three protocols when frame sizes are the same among them.

G. Parameter Setting

In order to reduce the execution time of our protocol, we optimize the parameters f and l in JECM protocol under the accuracy constraints given in (1). In Section III-D, we prove that $\hat{n}_*$ is asymptotically unbiased and they approximate Gaussian distributions. For a Gaussian distribution with $E(\hat{n}_*) \simeq \hat{n}_*$, equation (1) can be translated to

$$Var(\hat{n}_*) \le (e/Z_\delta)^2. \tag{34}$$

where Z_δ is $1-\frac{\delta}{2}$ percentile for standard Gaussian distribution and $\delta = 1-\alpha$. Therefore, we first set f and l such that (34) is satisfied. Then we will decrease f and l empirically to minimize the execution time. The process is terminated until (34) is not satisfied and we pick the last pair (f, l) as the optimal value.

IV. Simulation Results

A. Simulation Settings

We evaluate the performance of JECM by simulations. There is no prior work on estimating category-level joint information over an arbitrary number of tag sets. The most related work to our problem is CCF [8] and MJREP [16], but their protocols were designed to perform joint estimations at set (not category) level. As discussed earlier, we can adapt CCF and MJREP to perform estimation on one category at a time: The reader chooses a category ID cid to broadcast in a request. Only if a tag's category ID matches cid, the tag will participate in the execution of CCF (or MJREP). In this way, we can repeat the protocol to estimate one category at a time.

We use the performance metrics in Section II-C for evaluation. We will first compare the execution times of JECM,

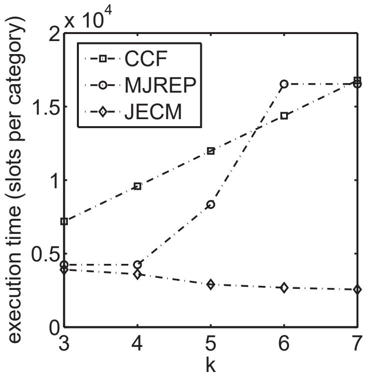


Fig. 3: Execution time comparison with respect to number of tag sets, subject to the same accuracy requirement with $\alpha = 95\%$ and $e = 50$.

CCF and MJREP, subject to the same accuracy requirement. Execution time is measured as the number of time slots each protocol needs to perform category-level joint estimation. We will then evaluate how well the proposed JECM can achieve a given accuracy requirement. Finally, we will investigate anonymity of JECM, CCF and MJREP.

The system model is a distributed RFID system of k locations, with an accuracy requirement of $e = 50$ and $\alpha = 95\%$. At each location, a reader periodically takes a snapshot of the local tag set. We set the number m of categories in each set to be 500 and the number of tags in each category to be 1000. We let the number n_* of common tags follow a zipf distribution [2] in $[10, 1000]$ and vary k from 3 to 7. We set $l_{cid} = 16$ out of $l_{id} = 96$.

We set the parameters for JECM based on Section III-G, and we set those of CCF and MJREP by exactly following [8], [16]. Specifically, for CCF, the length value is $\lceil \log(k * 1000) \rceil$ and the number of synopses is $\Theta(\frac{1}{\varepsilon^2 J} \ln \frac{1}{1-\alpha})$; for MJREP, f is optimized as is described in [16].

B. Execution Time

The first set of simulations evaluate the average protocol execution time. We apply the same accuracy requirement to CCF, MJREP and the proposed JECM, with $e = 50$ and $\alpha = 95\%$. Fig. 3 compares their execution times. The x-axis is the number k of tag sets, and the y-axis is the average number of slots needed per category by each protocol. When $k = 3$, MJREP and JECM have comparable time costs, while CCF takes longer. As k increases, the execution time of JECM decreases, while those of CCF and MJREP increase. For JECM, a larger number of tag sets provide more opportunity to filter out non-common tags during the inclusion/exclusion set joint process, which means a smaller time frame can be used to meet a certain accuracy requirement, resulting in smaller execution time. For CCF and MJREP, by doing one category at a time, the small number of common tags will take a larger time frame to separate them out from other tags, which is not a problem for JECM that records all categories together, ensuring a larger number of common tags. The curve of MJREP takes the non-smooth shape because the time frame for each category is set to a power of 2, with a large discrete jump between different settings. For a specific comparison, for joint estimation over 5 tag sets, JECM needs 2,902 slots per category, while CCF and MJREP need 11,983 and 8,192 slots respectively.

k	3	4	5	6	7
p_{cid}	98.90%	98.28%	98.99%	98.72%	99.02%

TABLE II: Preserved cid anonymity of JECM under given simulation settings.

C. Estimation Accuracy

The second set of simulations evaluate the accuracy of JECM. We vary the number k of tag sets from 3 to 5 and set the system parameters based on Section IV-A. Fig. 4 shows the results from joint estimation over 500 categories. Each point in the plot represents one category, where the x coordinate is the number n_* of common tags and the y coordinate is the estimated value $\hat{n}_*$. The equality line, $y = x$, is drawn for reference: the closer a point is to the equality line, the more accurate the estimation result is. We can observe from the figure that most estimation results are clustered around the equality line, demonstrating good accuracy of our protocol under different numbers of tag sets. Fig. 5 shows the cumulative distribution function (CDF) of estimation errors. The x coordinate is the estimation error and the y coordinate is the probability for the estimation error to fall below this range. The red dotted line is the preset error bound e. For k = 3, 4 and 5, the probabilities for estimation error being bounded by 50 are 0.954, 0.952 and 0.964 respectively, which confirms that our parameter setting in Section IV-A can indeed meet the pre-defined accuracy requirement of $e = 50$ and $\alpha = 95\%$ in all simulation cases.

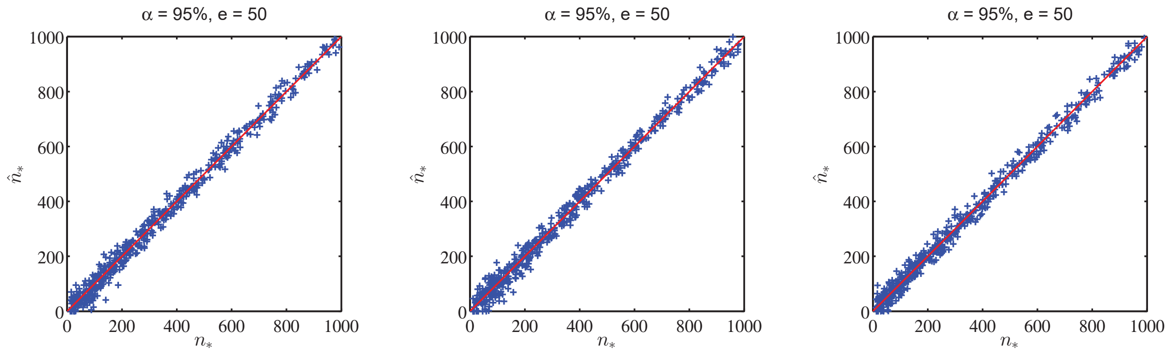
D. Anonymity

The third set of simulations study anonymity of the three protocols. Recall that in Table I, $p_{id} \approx 1$ for all these three protocols when $l_{id} = 96$ and $l_{cid} = 16$. So we only study the p_{cid} of these three protocols.

Table II shows the preserved anonymity of JECM when the number k of tag sets varies from 3 to 7. The first row shows the value of k and the second row presents the corresponding preserved anonymity of category ID cid, namely p_{cid}. The table shows that the p_{cid} values are close to 1 in all simulations of JECM, which means the probability for an unauthorized adversary to reveal any category ID is very low. (The slight variance among the p_{cid} values is due to the randomness in simulations.) For CCM and MJREP, since the reader will broadcast category IDs one at a time, an unauthorized adversary can easily acquire these IDs by eavesdropping the communication channel, making $p_{cid} = 0$ for both protocols.

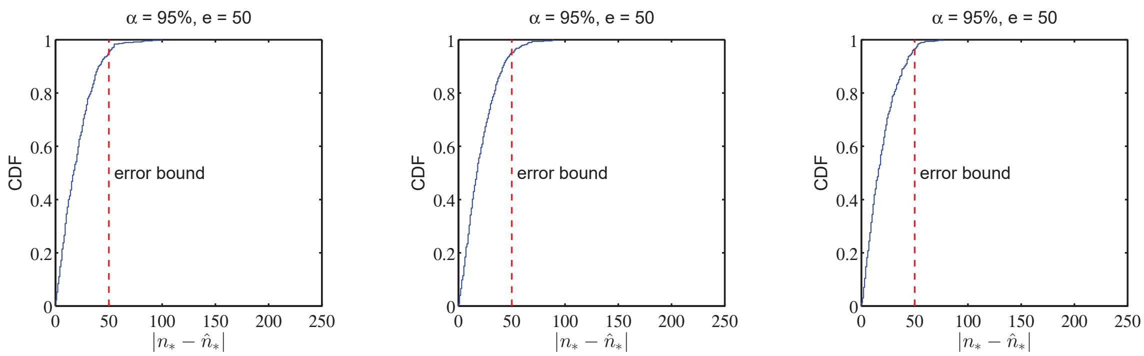
V. Conclusion

This paper studies a new problem of anonymous category-level joint estimation over multiple tag sets in RFID systems: for any category in a large RFID system, we want to anonymously estimate the cardinality of the intersection among



(a) $k = 3$ (b) $k = 4$ (c) $k = 5$

Fig. 4: Estimation results for $k = 3, 4, 5$ sets with 500 categories.



(a) $k = 3$ (b) $k = 4$ (c) $k = 5$

Fig. 5: CDF of estimation errors for $k = 3, 4, 5$ sets with 500 categories.

multiple tag sets. We design a protocol called JECM based on temporal or spatial snapshots. We derive an estimator, perform statistical analysis on it, and provide formulas for optimizing system parameters. Through extensive simulations, we evaluate the performance of our protocol and demonstrate that our protocol outperforms the prior art in time cost reduction and anonymity preservation.

VI. Acknowledgements

This work was in part supported by the National Science Foundation under grants CNS-1409797, CNS-1718708 and National Science Foundation of China under grant 61472385.

References

[1] Stirling Numbers of the Second Kind. *https://en.wikipedia.org/wiki/Stirling_numbers_of_the_second_kind*.

[2] Zipf's Law. *https://en.wikipedia.org/wiki/Zipf's_law*.

[3] M. Chen, S. Chen, Y. Zhou, and Y. Zhang. Identifying State-free Networked Tags. *IEEE/ACM Transactions on Networking*, 2017.

[4] M. Chen, J. Liu, S. Chen, and Q. Xiao. Efficient Anonymous Category-level Joint Tag Estimation. *Proc. of IEEE ICNP*, November 2016.

[5] H. Han, B. Sheng, C. Tan, et al. Counting RFID Tags Efficiently and Anonymously. *Proc. of IEEE INFOCOM*, 2010.

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

[7] T. Li, S. Wu, S. Chen, and M. Yang. Energy Efficient Algorithms for the RFID EStiimation Problem. *Proc. of IEEE INFOCOM*, 2010.

[8] H. Liu, W. Gong, L. Chen, et al. Generic Composite Counting in RFID Systems. *Proc. of IEEE ICDCS*, 2014.

[9] J. Liu, Y. Zhang, M. Chen, S. Chen, and L. Chen. Collision-resistant Communication Model for Stateless Networked Tags, poster paper. *Proc. of IEEE ICNP*, 2016.

[10] W. Luo, Y. Qiao, S. Chen, and M. Chen. An Efficient Protocol for RFID Multigroup Threshold-based Classification Based on Sampling and Logical Bitmap. *IEEE/ACM Transactions on Networking*, 24(1), 2016.

[11] M. Shahzad and A. X. Liu. Every Bit Counts - Fast and Scalable RFID Estimation. *Proc. of ACM Mobicom*, 2012.

[12] B. Sheng, C. C. Tan, Q. Li, and W. Mao. Finding popular categories for RFID tags. *Proc. ACM MobiHoc*, May 2008.

[13] C. Tan, B. Sheng, and Q. Li. How to monitor for missing RFID tags. *Proc. of ICDCS*, 2008.

[14] K. Y. Whang, B. T. Vander-Zanden, and H. M. Taylor. A Linear-time Probabilistic Counting Algorithm for Database Applications. *ACM TODS*, 22(9), 1990.

[15] Q. Xiao, M. Chen, S. Chen, and Y. Zhou. Temporally or spatially dispersed joint RFID estimation using snapshots of variable lengths. *Proc. of ACM MobiHoc*, 2015.

[16] Q. Xiao, S. Chen, and M. Chen. Joint Property Estimation for Multiple RFID Tag Sets Using Snapshots of Variable Lengths. *Proc. of ACM MobiHoc*, 2016.

[17] Q. Xiao, B. Xiao, and M. Chen. Differential Estimation in Dynamic RFID Systems, Proc. of IEEE INFOCOM,. *Proc. of IEEE INFOCOM*, 2013.

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

[19] M. Yoon, T. Li, S. Chen, and J. Peir. Fit a Compact Spread Estimator in Small High-Speed Memory. *IEEE Transactions on Networking*, 19(5), 2011.

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

[21] Y. Zheng, M. Li, and C. Qian. PET: Probabilistic Estimating Tree for Large-Scale RFID Estimation. *Proc. of IEEE ICDCS*, 2011.