IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 March 2023, accepted 17 April 2023, date of publication 24 April 2023, date of current version 3 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3269571

RESEARCH ARTICLE

Fast Counterfeit RFID Tags Detection With Slot Constraints

BING ZHANG[1], XU HU[2], AND SHAN GAO[1]
[1]Tianjin University of Commerce, Tianjin 300134, China
[2]Tsinghua University, Beijing 100084, China

Corresponding author: Bing Zhang (zhangbing@tju.edu.cn)

ABSTRACT RFID tags are prone to counterfeit attacks in supply chain scenarios. The attacker often uses counterfeit tags to replace stolen tags to avoid being detected by the inventory process. To defend against such attacks, existing hash-based schemes have to know the hash functions embedded in tags, which are usually unavailable in large scale systems. This letter proposes a novel group-based slot constraint (GSC) scheme for lightweight counterfeit tag detection in RFID-enabled supply-chain systems. GSC can be integrated into the identification process by leveraging elaborately designed group hash functions to authenticate tags based on their slot correlation. That is, the tags in the same group can always map to slots with a fixed offset, which offers evidence for identifying counterfeit tags with an abnormal slot index. We also provide a theoretical analysis of the time slots required to achieve the desired accuracy. The simulation results show that GSC provides reliable accuracy without knowing exact hash functions.

INDEX TERMS RFID authentication, counterfeit tag detection, shift-hashing.

I. INTRODUCTION

Radio Frequency Identification (RFID) tags are widely deployed in supply chain systems to support real-time inventory tracking, electronic theft protection and smart business decision-making [1]. Most of the existing RFID systems use tag identity, including EPC ID and TID, to label and identify the tag. These identities help trace the tagged items in the supply chain [2], obtain local real-time inventory [3], and identify theft or counterfeit attacks [4], [5] in the systems. A major challenging issue shared by the above scenarios is *ID leakage*. Since EPC and TID are stored on the open memory bank of the tag, they can be easily obtained by the attacker with a standard Gen2-compliant RFID reader [6]. After that, it is feasible for the attacker to build a counterfeit tag [7] with the same EPC and TID as the valid tag, to cheat and fool the RFID system. The current identity-based system cannot detect these abuses even if valid tags have been replaced by counterfeit tags, thus leading to severe financial losses.

To address this problem, a series of tag authentication schemes have been proposed to achieve effective tag authentication with various additional tag features. They can be classified into three categories: *cryptography approach*, *physical fingerprint approach* and *hash index approach*. Cryptography-based authentication schemes [8] use cryptographic keys and protocols to establish a secure channel between the RFID tag and the reader. This channel ensures that only authorized readers can communicate with the RFID tag and access the information stored on tag. It protects the privacy and integrity of sensitive information stored on RFID tags [9], and prevent unauthorized access to RFID systems. Although the recent advance in elliptic curve cryptography have improve the computation efficiency and robustness of cryptography authentication [10] significantly, such approach is still hard to be deployed on passive tags system with extremely scarce storage and computing resources.

Physical fingerprint-based authentication schemes [4], [5] take advantage of the signal characteristics (RSSI and phase) introduced by the distinct hardware characteristics to authenticate tags. For example, due to their personalized hardware characteristic, different tags may introduce a distinct phase offset to the same reader's query. Also, different tags may have different lifetimes after being charged by wireless signals. However, physical fingerprint approaches are usually suitable for small-scale scenarios with a limited number of tags because it is very hard to find a unique global physical

The associate editor coordinating the review of this manuscript and approving it for publication was Giorgio Montisci.

fingerprint to classify the target tags from others. When it comes to authenticate a large number of tags, physical fingerprint approaches usually introduce a long delay because they need to deploy a group of tags to a singleton item and scan the same tag several times to obtain reliable low-level feature sequences.

Hash index-based authentication schemes [7] are the most related work to our paper. It is a lightweight authentication scheme that takes advantage of the slot index picked up by the tag during the communication frame to authenticate a tag [11]. The slot index is generated by an irreversible one-way hash function with the input of a tag *EPC* and a random seed r. If each tag has a distinct hash function, the attackers cannot know the hash function and thus fail to make a perfect counterfeit tag with the same slot index sequence produced by the original one. However, the hash-based scheme assumes the authenticator exactly knowing the hash functions embedded in each individual tag [12], [13], which is very challenging and bring much complexity to the RFID system. A central authentication server is needed for managing the embedded hash functions of each tag, and distribute the hash functions to trust user at needed time.

To reduce the system cost of RFID tag authentication system and support fast authentication without prior knowledge, we propose a practical group-based slot constraint (GSC) scheme for lightweight counterfeit detection. A major objective of the proposed method is to authenticate tags without knowing the hash function embedded in each tag. To achieve this goal, we leverage the correlated hash indexes among different tags within the same trusted community to detect counterfeit tags. This idea is inspired by our observation that a group of tags usually coexists in the supply chain from top-tier suppliers to end-tier distributors due to the packaging and delivery policy [6]. Hence, this group of tags can be regarded as a trusted community, and the tags in the same community can be used to authenticate each other. However, the implementation of the trusted community faces the challenge of building a verifiable slot index relationship between tags in the same group. As in classical RFID system, each tag uses completely independent random hash functions, and there is no relationship between their slot indexes.

To fill this gap, this paper proposes leverage shift hashing to build a strong correlation between tags in the same community. The shift-hashing is a hash function that can be used to encode extra metadata with an extra shift component [14], [15]. The tags in the same trust community adopt the same shift hash function $f(\cdot) = h(\cdot) + o_i(\cdot)$ to choose their slot, which includes an anonymous hash part $h(\cdot)$ to generate a shared anchor slot for this trusted community, and a public linear offset part $o_i(\cdot)$ to generate a distinct offset for a specific tag. To generate the shared anchor slot, the tags in the same community i should compute the anchor slot $h(g_i)$ with a shared group id g_i. The shared group ID is common in the supply chain, since when a batch of items belonging to the same category are packaged in a large box for convenient transportation, they usually have a shared prefix ID, which can be chosen as the group ID. Meanwhile, to generate a distinct offset for each individual tag, a straightforward solution is to assign each tag in the group g_i a unique sequence index d_j, which generate a linear offset value with linear offset function $o_i(d_j) = c * d_j$, and the offset between arbitrary tag j and tag k can be determined as an constant offset $off_{jk} = o(d_j) - o(d_k) = c * (d_j - d_k)$. With the shifting hash function, the slot index picked by arbitrary tag j in group i can be represented as $f(id_j) = h(g_i) + o(d_j)$, The major benefits of shifting hash function is that it always maintains constant offset between two tags in the same community. By leveraging this constant slot correlation constraint, we can identify counterfeit tags since they fail to build the pattern of anonymous hash function, and fail to keep the constant slot-correlation with other tags in the same community.

The major contributions of this paper are summarized as follow:

- First, we leverage the shift-hashing functions to build a verifiable index correlation among tags in the same trusted community. This is achieved through the integration of an encoding rule and the shift hashing function. The encoding rule extracts two segments from the original EPC ID. The first segment works as a group ID, which is shared by all tags in the same trusted community, and maps them to a shared anchor slot with the shared anonymous hash function. The second segment work as an index ID, which specify the inner identity in the trusted community, which define the constant offset between any two tags. This makes tags in same trusted community map to correlated slots with constant offset.
- Second, we design a practical group-based slot constraint (GSC) scheme to take advantage of the constant offset between correlated tags to identify counterfeit tags. We investigate how to integrate the shift hashing into the current Aloha-based RFID communication protocol to efficiently detect counterfeit tags in each aloha round.
- Third, we provide a detailed theoretical analysis to optimize the parameter settings of the proposed GSC scheme and show the advantage of the proposed GSC over the state-of-the-art hash-based authentication approach.
- Finally, we conduct extensive simulations to evaluate the performance of the proposed GSC scheme. We vary a number of parameter settings and impact factors to evaluate its performance under a wide range of scenarios. Simulation results demonstrate that the proposed GSC schemes can achieve high accuracy in detecting counterfeit tags.

The remainder of the paper is organized as follows. Section II discusses related work, followed by an introduction to the system model, prior knowledge, and the definition of problems in Section III. Section V presents the detailed design of the proposed method, while Section VI presents the performance analysis of the proposed method. Section VII

evaluate the proposed method under various system settings. Finally, Section IX concludes the paper.

II. RELATED WORK

A. CRYPTOGRAPHY-BASED AUTHENTICATION

Cryptography is commonly used to provide security for RFID systems, and several cryptographic-based RFID authentication protocols have been proposed in recent years to develop RFID authentication protocols with varying levels of security and efficiency. Gildas et al. proposed an RFID authentication protocol that uses symmetric key cryptography and hash functions to provide security against various attacks [16]. Ouaissa et al. reduces the complexity of the authentication process due to the use of Elliptic Curve Cryptography (ECC) [9]. Alzahrani propose an efficient and secure TMIS-based protocol that employs lightweight symmetric key operations [17]. Dinarvand and Barati proposed an RFID authentication protocol that is efficient and secure with lightweight elliptic curve cryptography that has lower computational overhead and higher security level, and that is resistant to various attacks, including replay and man-in-the-middle attacks [10]. Dinarvand and Barati [8] examined the latest RFID authentication protocols based on elliptic curve cryptography in terms of security and performance. Although the cryptography provides high security level and significantly reduce the cost in recent year. They have not been deployed to passive RFID tags due to cost considerations. The mutual authentication process would increase the hardware complexity of the tags, results in significant increases on the manufacturing difficulty and cost.

B. PHYSICAL FINGERPRINT-BASED TAG AUTHENTICATION

Physical-layer tag fingerprint has been investigated for years. A typical solution is to uses the time-related feature, such as back-scatter link frequency (BLF) as the fingerprint [18], which caused due to the clock-drifts of different tag circuit. Meanwhile, many prior work explored how to use the phase shift caused by tag diversity to build a fingerprint. Since the distinction of tag diversity is limited, the current scheme generally needs to adopt multiple tags to build federated groups to identify tags [4]. Another popular direction is to use power-related features as the fingerprint, including the minimal activating power for tag activation at varying distance [19], the power distribution in different frequency bands [20] and the discharging time after losing power from the reader [5] However, a singleton physical fingerprint is usually distinguishable on a small scale. In large-scale RFID systems, there are always tags with fingerprint collision due to the limited resolution of the fingerprint feature. Although some solutions can integrate multiple features to obtain reliable fingerprints [21], [22], they introduce an extra delay to collect and identify fingerprints. In addition, it is very hard to manage such a large volume of fingerprint databases in supply chain systems that need to be distributed and exchanged between suppliers and distributors.

C. HASH-BASED DETERMINISTIC AUTHENTICATION

The hash-based authentication approach can also be further classified into two categories: *deterministic authentication* and *probabilistic authentication*. The deterministic authentication approach [23], [24] is to authenticate each individual counterfeit tag in a per-tag manner, which is reliable but introduces extensive communication costs for ID transmission. These protocols focus on improving the utilization of the frame-slotted aloha through dynamical online frame size optimization. For example, Li et al. applies the hash function to map each tag to a specific slot [25] with a certain hash function. Knowing the hash function, the reader can compute the expected slots picked by all the registered tags and determine whether it is a valid tag. To further improve frame utilization, some recent research tries to leverage multi-hash to resolve hash collisions. Liu et al. applied multi-seed hashing to reconcile collision slots to improve the time frame utilization [26]. Xie et al. applied the perfect hash function [27] to Yu et al. exploited how to apply Gen2 commands to identify missing tags [28] in commercial RFID systems with well-designed selective reading, significantly reducing the polling overhead of the tags on commercial devices. Xie et al. exploited how to leverage redundant multiple tags to improve the reliability of tag identifications [29]. The considerable extra cost of resolving hashing collisions is the major deficiency of deterministic authentication approaches. To authenticate tags in a per-tag manner, the tag needs to be assigned exclusive slots free-of collisions. Besides, some assumption needs to be met, *e.g.*, without the existence of inference tags. Otherwise, the accuracy of deterministic results will be affected and will fail to provide a guarantee of authentication results.

D. HASH-BASED PROBABILISTIC AUTHENTICATION

On the other hand, the probabilistic authentication approach significantly improves time efficiency at the expense of limited reliability loss. Instead of authenticating RFID tags in a per-tag manner, the probabilistic approach evaluates the counterfeit risk and rate of the total tag population, which aims to detect counterfeits attacks when the number of counterfeit tags exceeds some pre-defined threshold. Tan et al. initiated prior research on probabilistic authentication and proposed a hash-based TRP approach to [30] to trigger a warning message when missing tags exceed a user-defined error threshold with high reliability. MSMD scheme introduces sampling techniques [31] to achieve reliable missing tag detection using a small set of sample tags, which significantly improves the time efficiency. The main idea is to leverage the birthday paradox to detect rare events more efficiently. Xie et al. introduces a method to implement personalized sampling rate for commercial tags with C1G2 commands [32]. Moreover, some advanced data structures, e.g., bloom filter [33] and minimal are also introduced to detect counterfeit tags distributed in multiple overlapped regions. However, the probabilistic approach only provides an overview of the system's status.

TABLE 1. Compression with different type of authentication protocols.

Method	Exec time	Reader/Tag modif.	C1G2-compliance	Pros.	Cons.
Cryptography	Medium	Cryptography hash	No	high security	tag complexity
Fingerprint	Slow	None	Yes	c1g2-compliance	long exec. time
Hash	Fast	Light hash	No	time efficient	hash leakage
Hash Offset (Our)	Fast	Light hash	No	time efficient	group knowledge

E. COMPARISON WITH ABOVE METHODS

The differences between the different types of protocols are summarized in Table 1. Cryptography methods have a medium execution time due to the mutual authentication process between the reader and the tags and also provide high security protection. However, they need to implement a computation-intensive enciphering algorithm on tags, which may significantly increase the cost of tags. Physical fingerprint is C1G2-compliant solution, which means it can be easily deployed on current devices. However, the fingerprint is easily to be influenced by the environment, and the fingerprint from different tags could be indistinguishable, which makes it is hard to be applied to verify a large number of tags. Finally, the hash-based solution is a lightweight authentication approach that can be executed very quickly since it integrates the authentication process into the communication process. However, it assumes that each tag holds a private hash function, which should be known to the verifier but keep private to the attacker. However, when the hash function is transmitted to the verifier, the hash function may be leaked, compromising the protection provided by the hash function. This paper proposes to design an efficient hash method to defend against the hash leakage problem. Our solution uses the offset of the slot between the correlated tags to verify their membership. This method is self-explanation, and thus the private hash function does not need to be transmitted, which reduce the attack surface of the system, reduce the network transmission and provide higher security of hash credentials.

III. SYSTEM MODEL AND PROBLEM DEFINITION

A. SYSTEM MODEL

This paper considers common RFID systems consisting of a large number of tags to label items tagged, a single reader to access data stored on the tags, and a back-end server to handle and process data captured by the reader [3]. The reader and server can be seen as a unified central integrator unit, one for communication-intensive data access tasks and one for handling computation-intensive data retrieving tasks. The reader is controlled by the server with the Low-Level Reader Protocol (LLRP), and uploads all the sensing data to the server; this enables us to obtain valuable information from the RFID data stream with very low latency.

B. COUNTERFEIT ATTACK MODEL

We mainly face counterfeit RFID tags in RFID systems, which occurs in a practical scenario where attackers try to steal valuable tagged items for inappropriate benefits. To avoid being detected by the system manager, the attacker tries to hide the theft activity and replace the stolen tags with counterfeit tags with the same ID. If the manager uses the tag identity as the tag's fingerprint, theft tags cannot be detected, which introduces significant economic loss and becomes one of the major risks for RFID systems.

C. COMMUNICATION MODEL

We assume the reader communicates with the tags through a frame-slotted aloha specified by the EPC C1G2 standard. In the frame-slotted Aloha protocol, the entire time frame is divided into f slots. Each tag id_i will use anonymous hash functions $h(\cdot)$ to choose a slot index $h(id_i)$ within a given time frame. To minimize the cost of RFID tags, RFID tags cannot support complex collision avoidance mechanisms. Therefore, tag collision is a major challenge for RFID communication systems. A tag can successfully send its information to the reader only when the tag selects an exclusive slot. When multiple tags send their data at the same time, their signal responses will collide and nothing will be detected by the reader. Although some optimization algorithms have been proposed to improve the utilization of time frames, the optimal upper bound is below $1/exp(1) \approx 0.368$. To break this bottleneck, we need to develop novel techniques to resolve collision slots.

IV. PROBLEM DEFINITION

Let $N = \{x_1, x_2, \cdots, x_n\}$ be the set of candidate tags to be verified by the reader, whose IDs are unknown to the reader. Moreover, tags are divided into k logical groups $\{G_1, G_2, \cdots, G_k\}$ in prior [34], [35] according to their EPC. The tags within each group G_i are assigned to correlated slot indexes with shared hash functions h_i. The attacker in the supply chain may steal m ($m \leq n$) tags $M = \{y_1, y_2, \cdots, y_m\}$ from candidate tags N and replace them with the same number of counterfeit tags $C = \{z_1, z_2, \cdots, z_m\}$. Besides, $z_i \in C$ and $y_i \in M$ have the same EPC ID and cannot be classified with each other by identity in the tag memory. However, since tags can be assigned to random slots with the anonymous hash function $h(\cdot)$, counterfeit tags z_i will always be assigned to a slots different with the candidate tags y_i, namely $h(z_i) \neq h(y_i)$, which provides an interface for the detection of counterfeit tags.

The tags set that exist in the system are updated to $N' = N - C + M$ after suffering counterfeit attacks. The problem with *counterfeit tags detection* is to find counterfeit tags $C = \{z_1, z_2, \cdots, z_m\}$ from the set of attacked candidates N' with misidentified counterfeit tags meet the required accuracy $\epsilon * m$ with $1 - \alpha$ reliability. Meanwhile, the main objective is to

TABLE 2. The notation used in this paper.

symbol	Notation
$N = \{x_1, x_2, \cdots, x_n\}$	the set of integrated tags x_i to be authenticated
$M = \{y_1, y_2, \cdots, y_m\}$	the set of missing tags y_i
$C = \{z_1, z_2, \cdots, z_m\}$	the set of counterfeit tags y_i
$N = \sum_{i=1}^{k} G_i$	the groups of tags G_i
$G_i = \{t_1, t_2, \cdots, t_j\}$	the tag set in G_i
$f_i(id_j, r) = [h_i(g_i, r) + o(id_j)]\%f$	the shift hashing function of group G_i
$h_i(g_i, r)$	the anonymous hash function of group G_i
$o(id_j)$	the open offset function
f	the number of slots in a time frame
w	the length of mask in rough estimation phase
$n = \|N\|$	$\|N\|$ return the # of elements in set N
$\hat{n}$	the estimator of a n
P_{FP}	the false negative error probability of counterfeit detection
$D_i(k_i, k_i)$	the expected $k_i \times k_i$ offset matrix of group G_i
$S_i(k_i, k_i)$	the observed $k_i \times k_i$ offset matrix of group G_i
L_x	the length of time frame in phase x
n_e, n_s, n_c	the number of empty/singleton/collision slots in time frame

minimize the total execution time, including the communication time of the reader T_{com} and the computation time of the server T_{comp} to minimize the impact on the application layer. The main notation used in this paper is summarized in Table 2.

V. GROUP-BASED COUNTERFEIT TAG DETECTION

In this section, we present a more detailed design of the GSC scheme. Specifically, the GSC initializes a single long time frame to achieve counterfeit tag detection by leveraging the slot constraints between tags from the same group. The slot constraints are built with a shared shift hashing function, which maps tags in the same group to correlated slots determined by the group index of the tag. If the slot correlation between two tags violate the expected offset correlation, we can regard at least one of them is counterfeit tag. To identify all counterfeit tags in a group, we can use a straightforward majority vote mechanism to regard subcommunities that achieve a consensus as valid tags. Meanwhile, other outlier tags are regarded as the counterfeit tags.

A. OVERVIEW OF GSC

The entire process of the proposed GSC can be divided into three phases, including *tag estimation*, *tag identification*, and *counterfeit detection*. The tag estimation scheme is to roughly estimate the tag population, which is required to optimize the number of slots in the time frame. The optimized time frame enables us to trade-off between accuracy and time delay of counterfeit detection. Then, with the optimized time frame, the reader initial a time frame for tag identification. Each tag will map to a certain time slot with a hash function, and respond its fingerprint in that slot. By observing the time frame, the reader can gather identification information and transmit the data to the server. Finally, the server analyze the collected time frame to find the trusted sub-communities with consensus. All the tags in the sub-communities can be regarded as valid tags.

B. TAG ESTIMATION PHASE

Since the information on candidate tags N is completely unknown to the reader, the first step is to obtain the estimated cardinality $\hat{n} \approx |N|$ in order to avoid serious collisions and slot waste in the next phase. Our tag estimation consists of two steps. In the first *rough estimation step*, the reader issues a select command with a gradually increasing mask field to adaptively estimate the number of tags in the region. Specifically, in the i-th slot, the length of *mask* is of i bits, and only the tags whose EPC segment matches the *mask* will respond to the reader. Thus, the number of matched tags will gradually reduce with increasing mask length. Once there are no tags that respond to the reader in the w-th slot, the reader will terminate the rough estimation. Since we assume the target EPC segments of tags are randomly generated number, the probability it matches with the w-bit mask can be represented as $1/2^w$ [36]. Thus, a rough estimation of the cardinality of the tag can be represented as:

$$\hat{n}_r = 2^{w-1} \tag{1}$$

In the second *accurate estimation* step, the reader initializes a time frame of $\hat{n}_r$ slots for accurate tag estimation. In that time frame, all tags will be randomly mapped to a slot and will respond as binary bits to the reader. The reader counts the number of empty slots n_e as well as the number of busy slots n_b and uses its difference $n_b - n_e$ to estimate the number of tags. Specifically, the expected number of empty slots and busy slots can be represented as follows [37]:

$$E(n_e) \approx \hat{n}_r(1 - \frac{1}{\hat{n}_r})^n \tag{2}$$

$$E(n_b) \approx \hat{n}_r - \hat{n}_r(1 - \frac{1}{\hat{n}_r})^n \tag{3}$$

Thus, the accurate estimator can be represented as:

$$\hat{n} = \ln\left(\frac{\hat{n}_r + n_e - n_b}{2\hat{n}_r}\right) / \ln\left(1 - \frac{1}{\hat{n}_r}\right) \tag{4}$$

The tag cardinality estimation results provided by Eq. 4 is leveraged by the following phase as the actual tag population $n = |N|$ for parameters optimization.

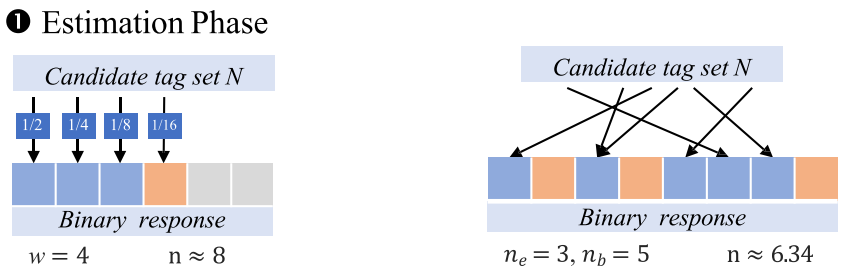

FIGURE 1. The illustrative example of the tag estimation phase in the GSC protocol: with 6 candidate tags $\{t_1, \cdots, t_6\}$ from two groups $\{G_1, G_2\}$. The first rough estimation phases use an efficient estimator implemented with a decreasing exponential sampling rate. The second accurate estimation phase is relayed on the rough estimator to initialize a time frame for linear estimation, each tag randomly maps to a slot with hash function, and the estimator is based on the offset between empty slots and busy slots.

C. TAG IDENTIFICATION PHASE

To identify counterfeit tags using the correlation of slots, the reader first needs to identify their associated groups according to the EPC ID of the tags. To simplify, we assume that all tags in the same group $id =< g_i, j >$ must have two EPC segments, the first segment g_i called group ID, which will allow the reader to categorize tags into k groups after tag identification; the second segment j called group index, which allow the reader to knows its relative order in the group G_i. In the identification phase, the reader will issue a query command to initialize a time frame of f slots for tag identification. The value of f is determined by the accuracy required ϵ, the risk α, and the total population of tags $\hat{n}$.

Specifically, the tags with the same group ID g_i is assumed to use a same anonymous hash function $f_i(id, r) = [h_i(g_i, r) + o(j)]\%f$ to choose a slot to respond to the reader. The hash function consists of two parts: (1) $h_j(\cdot)$ is a private hash function unknown to others, which maps the same group of tags to the same anchor slots $h_i(g_i, r)$ in different rounds, but the anonymous feature of the hash function $h_j(\cdot)$ prevents the attacker from creating a perfect counterfeit tag with the same mapping behavior of the anchor slot. (2) $o(\cdot)$ is a public offset of the tag to the reference slot offset, a straightforward design is uses the prdoduct of index j in the group $G_i = \{t_1, t_2, \cdots t_j\}$ and a const value c as the public offset. Therefore $f_i(id, r) = [h_i(g_i, r) + c * j]\%f$

Although the reader may not know the hash slot $f(t)$ of a specific tag t, given two tags in the same group $t_j, t_k \in G_i$, the reader knows the expected slot difference between them $f(t_j) - f(t_k)$, which should be a constant value, since their anonymous hash functions generate $h(\cdot)$ the same value and the offset function $o_i(\cdot)$ generate a constant offset $c(j-k)$. The reader will check the slots, one by one, to record the identified information, including the ID id, group identity g_i and slot index $f_i(id, r)$ of the identified tags. These identified information will upload to the server for counterfeit tag detection with correlation validation.

D. COUNTERFEIT DETECTION PHASE

Given the identified information, the reader can then detect counterfeit tags group by group. For the tags $\{t_1, \cdots t_j\}$ in the i-th group G_i, the reader will leverage the public offset functions $o(\cdot)$ to compute the offset of each tag $\{o(t_1), \cdots, o(t_j)\}$, and then obtain a $j \times j$ slot difference matrix D_i. Let w and m be the index of two tags in the group $t_w, t_m \in G_i$, the offset matrix value $D_i(w, m) = o(t_w) - o(t_m)$.

Meanwhile, based on the observed information in the identification phase, the reader can obtain another $j \times j$ slot difference matrix S_i, where $S_i(w, m)$ denote the observed slot offset between tags t_w and t_m. For arbitrary two tags t_w and t_m, if $S_i(w, m) \neq D_i(w, m)$, at least one of them should be counterfeit tags. We assume the counterfeits tags is minority, thus the majority of tags are valid ones. Therefore, if a subset of tags in the group G_i meets the slots difference constraints specified in D_i, they can be considered a trust alliance A_i. Other tags outside the alliance, namely $G_i - A_i$, are considered counterfeit. If counterfeit tags and valid tags coexist in the system, the reader will obtain two tags with the same ID but identified in different time slots; the valid tag should have constant constraints with other tags in the same group, while counterfeit tags will map to illegal slots, which can be identified as attackers.

E. ILLUSTRATIVE EXAMPLE

Figures 1, 2, and 3 show an illustrative example of the GSC scheme with two groups of candidate tags $G_1 = \{t_1, t_2, t_3\}$ and $G_2 = \{t_4, t_5, t_6\}$. First, without knowing any information about the candidate tags, the reader first executes a two stage estimation phase to obtain the tag number. The rough estimation algorithm counts the index of the first empty slots in the time frame and computes the estimation results as $\hat{n}_r = 2^{4-1} = 8$. Then, the accurate estimation algorithms initializes a $\hat{n}_r$-slot time frame for accurate tag estimation and counts the number of number of busy slots n_b and empty slots n_e, respectively. By submitting the observed $n_e - n_b = -2$ into Eq. 4, we can obtain the accurate tag number as $\hat{n} = 6.3$.

After knowing the tag population, the reader will initialize a time frame for tag identification, which should be long enough to avoid tag collisions. The reader will record the identified information in the $id : slot$ format, thus obtaining the set $I = \{t_1 : 1, \cdots, t_6 : 8\}$. The identified information I will categorized into two groups I_{G_1} and I_{G_2} based on the group IDs embedded in t_i. Each group of tags will be checked in sequence to find counterfeit tags. For instance, for the tags

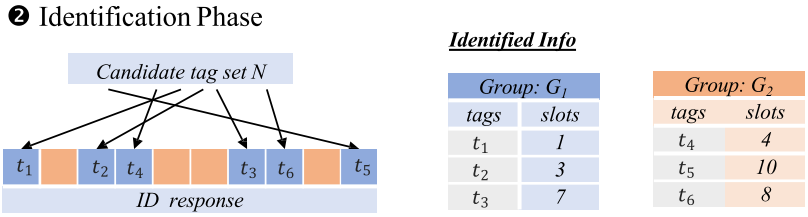

FIGURE 2. The illustrative example of each phase of the GSC protocol: with 6 candidate tags $\{t_1, \cdots, t_6\}$ from two groups $\{G_1, G_2\}$, where $G_1 = \{t_1, t_2, t_3\}$ and $G_2 = \{t_4, t_5, t_6\}$. The tags in the same group are assigned to a slot with a shifting hash function $f_i(id_j, r)$, consisting of an anonymous part that maps the tags $h_i(g_i, r)$ in the same group to the same reference slot index, and a offset part to map the different tags in g_i to the fixed slot offset. Through this way, we build the slot correlation between tags.

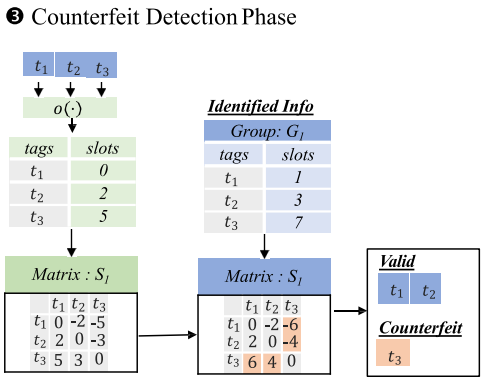

FIGURE 3. The illustrative example of each phase in the GSC protocol: with 6 candidate tags $\{t_1, \cdots, t_6\}$ from two groups $\{G_1, G_2\}$. The server will compute the expected offset between tags and obtain a expected offset matrix D_i. Meanwhile, the server will also derive the observed offset matrix S_i based on the identification phase collected from the data. Through the offset valuation between two matrixes, we can identify outlier counterfeit tags without knowing the explicit hash function.

in G_1, we will compute the slots offset $O_1 = \{t_1 : 0, t_2 : 2, t_3 : 5\}$ and obtain the expected slot difference matrix $D_{3\times3}$. Meanwile, the reader can also construct an observed slot difference matrix $S_{3\times3}$ based on I_{G_1}. By comparing $S_{3\times3}$ and $D_{3\times3}$, we can find that t_1 and t_2 meet the slot offset constraints while t_3 has a different offset with $\{t_1, t_2\}$, and thus it is regarded as a counterfeit tag.

VI. PERFORMANCE ANALYSIS

A. COMMUNICATION TIME OF ESTIMATION PHASE

In this section, we try to derive the theoretical execution time of GSC to understand its performance. First, let L_{est} be the number of slots in the estimation phase. According to Section V-B, it consists mainly of two parts: L_{RE} and L_{AE}, which denote the number of slots in the rough and accurate estimation stage, respectively. In the rough estimation stage, the sampling probability is reduced at the factor of $1/2$, the probability that the i-th slot is an empty slot can be represented as

$$(1 - 1/2^i)^n \approx e^{-n/2^i}$$

Let n denote the number of tags and ϵ denote the probability of estimation error, the number of required slots should meet the following equation constraints [38]:

$$L_{RE} \leq (\ln n + \ln \ln \epsilon)/\ln 2 \tag{5}$$

Meanwhile, by subsisting Eq. 5 into $L_{AE} = 2^{L_{RE}-1}$, we can obtain:

$$L_{AE} \leq -n/2 \ln \epsilon \tag{6}$$

Since in the tag estimation phase, the tag only needs to send binary response, the total exectuion time of estimation phase can be represented as:

$$T_{est} \approx (-n/2 \ln \epsilon + (\ln n + \ln \ln \epsilon)/\ln 2) * t_B \tag{7}$$

where t_B denotes the slot length for the binary response.

B. COMMUNICATION TIME OF IDENTIFICATION PHASE

During the identification phase, the reader needs to initialize a time frame of f slots for tag identification. Specifically, there are three types of slots in the time frame: collision slots containing more than two tags, singleton slots containing exactly one tag and empty slot containing no tags. Let n_e, n_s, and n_c denote the number of empty singleton slots and collision slots, respectively. We have [39]:

$$\begin{aligned} n_e &\approx f(1 - 1/f)^n \\ n_s &\approx n(1 - 1/f)^n \\ n_c &\approx f - (f + n)(1 - 1/f)^n \end{aligned}$$

To avoid tag collisions, the number of singleton slots n_s should meet the following equation $n_s \geq n(1 - \epsilon)$. Since the exact value of population of tag n is unknown to the reader and only the estimated tag population $\hat{n}$ is known, we substitute $\hat{n}$ into the above equation, and obtain the setting of f:

$$f \geq \frac{1}{1 - (1 - \epsilon)^{1/\hat{n}}} \tag{8}$$

Let t_e, t_s, t_c denote the length of empty, singleton and collision slots, the total execution time of identification phase can be represented as:

$$T_{idf} \approx n(1 - \epsilon) * t_s + n * \epsilon * t_c + (L_{idf} - n) * t_e \tag{9}$$

According to the C1G2 standard, the lengths of each type of slots are shown in Table 1. The link frequency of the reader is set to BLF=256 Khz and the adopted encoding method is set to Miller-4, and the transmission rate is 64 KB/s. Thus, we have $t_e : t_c : t_s \approx 1 : 4 : 30$. By subsisting this relationship into Eq. 9, we have:

$$T_{idf} \approx (L_{idf} + 29n) * t_e \tag{10}$$

TABLE 3. The length of each type of slot in the common frame slotted alpha scheme, where t_e denotes the length of the empty slot, t_s denotes the length of a singleton slot, t_c denotes the length of the collision slot and t_B denotes the length of binary slots.

slots	Waitting time	Response	Command
t_e	$T_1 + T_3$	NA	QueryRep
(100 us)	(40 us)	0	(4 bit/ 60 us)
t_s	$2T_1 + 2T_2$	RN16 +EPC+CRC	QueryRep+ACK
(3060 us)	(160 us)	(160 bit/ 2500 us)	(26 bit/ 400 us)
t_c	$T_1 + T_2$	RN16	QueryRep
(390 us)	(80 us)	(16 bit/ 250 u)s	(4 bit/ 60 us)
t_B	$T_1 + T_2$	Binary	QueryRep
(160 us)	(80 us)	(1 bit/ 16 u)s	(4 bit/ 60 us)

C. ERROR CASES OF COUNTERFEIT DETECTION

The proposed counterfeit detection protocol would face both false negative error and false positive error. The false negative means that some counterfeit tags would be miss identified by GSC. It happens when the counterfeit tags is assigned a slot meet the following two conditions: First, if the counterfeit tag maps to the collision slots, its identity cannot be identified by the reader, since the reader does not know to which group it belongs to, whose probability can be represented as

$$P_{collision} = 1 - (1 + n/f)(1 - 1/f)^n \tag{11}$$

Second, if the counterfeit tag maps to a singleton slot that happens to be same as the associate candidate tag, such a probability can be represented as

$$P_{singleton} = n/f^2(1 - 1/f)^n \tag{12}$$

Thus, the total rate of mis-identified counterfeit tags can be approximated as:

$$P_{FN} = 1 - (1 + n/f - n/f^2)(1 - 1/f)^n \tag{13}$$

Meanwhile, false positive error also may happens when the number of counterfeit tags exceed the number of valid tags in a group. In such case, the alliance build by counterfeit tags would mislead the majority vote algorithm, and falsely be identified as valid tags, while other valid tags are misidentified as counterfeit tags. For a given proportion of counterfeit tags in a group p, the false positive error occurs when the number of counterfeit tags exceeds the number of valid tags. Since there are k tags in a group, the number of counterfeit tags should be $k * p$ and the number of valid tags should be $k * (1 - p)$. The RFID tag identification rate can be represented as $q = (1 - 1/f)^n$, which is determined by the time frame f and the total population of tags n. Let i and j denote the number of counterfeit and valid tags. Thus, the counterfeit tags are falsely identified as valid tag when:

$$P_{FP} = \sum_{i=1}^{k*p} q^i (1-q)^{k*p-i} * \left[\sum_{j=0}^{j<i} q^j (1-q)^{k*(1-p)-j} \right] \tag{14}$$

At least, the proportion of counterfeit tags p in any group should no more than half of the group population. To control the risk of counterfeit tags detection, we need to make sure the proportion of counterfeit tags p is small enough, to ensure $P_{FP} < \alpha$, where α is reliability requirements set by the user. In additional, we can also increase the length of time frame f to obtain better reliability. This is because the tag identification rate q increase with the increases of f, resulting in smaller P_{FP}.

D. ALGORITHM COMPUTATION COMPLEXITY ANALYSIS

The major computation cost of the proposed methods consists of two parts: (1) tag estimation phase (2) counterfeit detection phase. In the rough tag estimation phase, the algorithm first needs to iterate the status of the slot to detect the first empty slot, which takes around $O(logn)$ judgment operations. Then, in the accurate tag estimation phase, the reader needs to further loop the time frame to count the number of collision, singleton, and empty slots to feed into the estimation algorithm, which introduces $O(n)$ addition operations to count the type of slots.

Meanwhile, the counterfeit detection phase account for the majority of computation cost, in each round the reader need to record the slot index of identified tags, compute the offset between their slot indexes and compare with the expected slot offset. During the counterfeit detection process, the reader needs to take O(n) copy operations to record the index of the identified slot. To obtain the expected slot offset, the reader takes O(n) hash operations to compute the slot index of each tag. Let m denote the number of tags in each group; thus the reader takes O($m * n$) subtraction operations to obtain the offset in the slot, as well as O($m * n$) judgment operations to identify counterfeit tags. Therefore the total complexity would grow with the increase of number of tags in a group.

E. EXECUTION TIME OF COUNTERFEIT DETECTION PHASE

Since the estimation and counterfeit detection phases are a low-rate online communication process between the reader and the tag, while the counterfeit detection phase is a purely high-speed computation process executed on the server, The first two phases account for most of the execution time of the GSC protocol, which usually takes tens of thousands of times in large-scale RFID systems. Meanwhile, the counterfeit detection task on the server can be performed in milliseconds on the personal PC. Therefore, the main cost of counterfeit detection lies in the communication process.

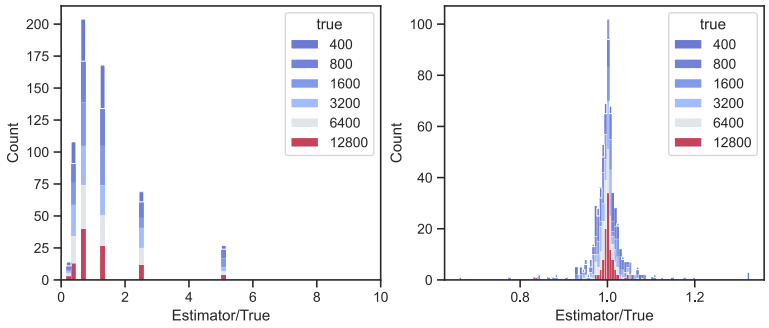

FIGURE 4. The density distribution of estimator/true value with varying distribution.

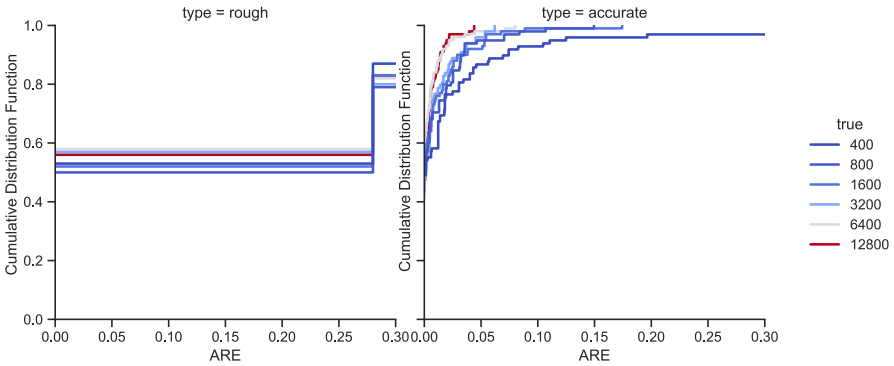

FIGURE 5. The cumulative distribution of ARE with varying tag population. The left figure is the CDF of rough estimation phase and the right figure is the CDF of accurate estimation phase.

VII. SIMULATION RESULTS

A. SIMULATION SETTINGS

We evaluated the time efficiency and precision of the GSC scheme through extensive simulations with various parameters and settings. Since GSC is the only protocol that can achieve batch counterfeiting tag identification without knowing hash functions, we focused mainly on evaluating the performance of GSC in *execution time* and *accuracy* For simplicity, we assume that each group has the same number of k tags. For reliability, each result is reported by averaging 100 individual simulations with different random seed.

B. ESTIMATION ACCURACY

In Figure 4, we vary the number of tags from 400 to 12800 to evaluate the ratio of the estimation value to the true value. The results is better when the ratio is close to 1, and a ratio value larger than 1 means overestimation, while a small ratio value smaller than 1 means under estimation. We evaluate the estimation performance of both the rough estimation phase and the accurate estimation phase. Under each tag population setting, we run the protocol 100 times to obtain the ratio value and count the histogram of the estimation value. We can find that the estimate value of the accurate estimation phase follows a normal distribution and is very close to 1. Most of the algorithm results within a small ratio range ranging from 0.95 to 1.05. Meanwhile, rough estimation can produce overestimation, namely $ratio > 6$, due to the random nature of the rough estimator. When tag population is small, the rough estimator have a higher chance for overestimation. This simulation illustrates why we need a two-phase estimation and shows that the estimation results are accurate enough for the following optimization.

Figure 5, shows the relative error of the protocol with varying population of tags, the relative error ARE is calculated as the relative difference in the estimation value, which could be calculated as:

$$\text{ARE} = \frac{est - true}{true} \tag{15}$$

We can find that the rough estimation phase shows a sparse ARE accumulation pattern, and about half of algorithm execution return accurate estimation results with small error, while introduce large error (ARE > 0.3) in 20% of the execution round. Meanwhile, the accurate estimation phase shown in the figure on the right has significantly higher accuracy than the rough estimation phase. Specifically, 90% of the estimation rounds return a value of $ARE < 0.05$. In addition, we can find that the accuracy improves with increasing tag population. For example, if the tag number is 6400, the ARE is around 0.02 within 90% of execution, Meanwhile, the ARE is increased to around 0.08 if tag number is 400. This is because when estimating a large number of tags, the reader initializes a long time frame and enjoys a smaller variance.

C. TIME-EFFICIENCY PERFORMANCE

In Fig. 6, we set the threshold error $\epsilon = 0.05$ and vary the number of tags to evaluate the execution time of the GSC protocol. Each point represents the execution time of a single round, and the area with a deeper color represents the region of the primary execution time distribution. We can see that the execution time is proportionate to the increased number

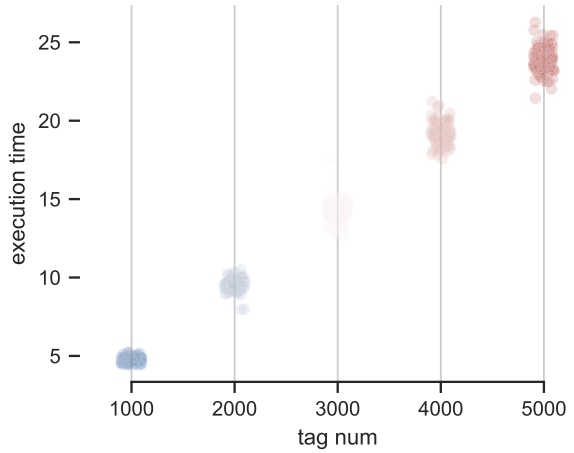

FIGURE 6. Execution time with varying tag number. Execution time increases linearly with increasing number of tags population.

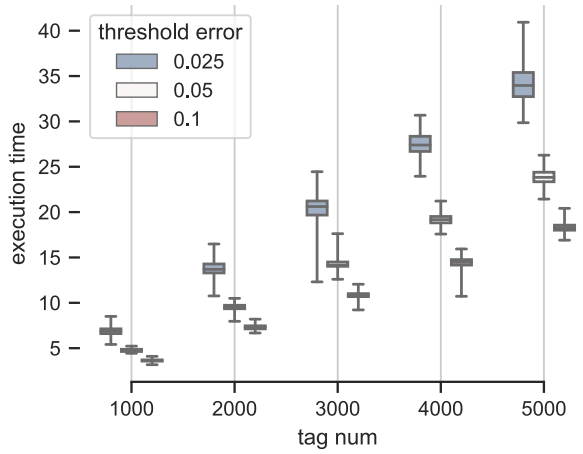

FIGURE 7. Execution time with a varying threshold error value ϵ.

of tags, this is because GSC protocols have to identify all the tags, and thus the length of time frame is determined by the number of tags. However, GSC still can be regarded as a time-efficiency solution, since it can detect counterfeit tags in a batch manner, which takes only 25 seconds to verify up to 5000 tags. While the state-of-the-art physical layer energy feature scheme [5] need to verify tag in a per tag manner by powering the tags and waiting the time it transforms to power off. It takes about 4 seconds/per tags to verify a tag, which takes too much execution time in large-scale RFID system.

In Fig. 7, we vary the threshold setting ϵ from 0.025 to 1.0. We observe that as the decrease of ϵ, the execution time of GSC protocol significantly increases. For example, when the number of tags is equal to 5000, the execution of GSC ($\epsilon = 0.025$) takes 50% and 70% more execution time, compared to GSC ($\epsilon = 0.05$) and GSC ($\epsilon = 0.1$), receptively. The underlying reason is that the length of the time frame in the identification stage significantly increases as the decreases of ϵ. The total number of slots of GSC ($\epsilon = 0.025$) is about $2\times$ of GSC ($\epsilon = 0.05$) and $4\times$ of GSC ($\epsilon = 0.1$).

D. ACCURACY PERFORMANCE

Fig. 8 shows the corresponding number of counterfeits missed with respect to the rate of counterfeits and the

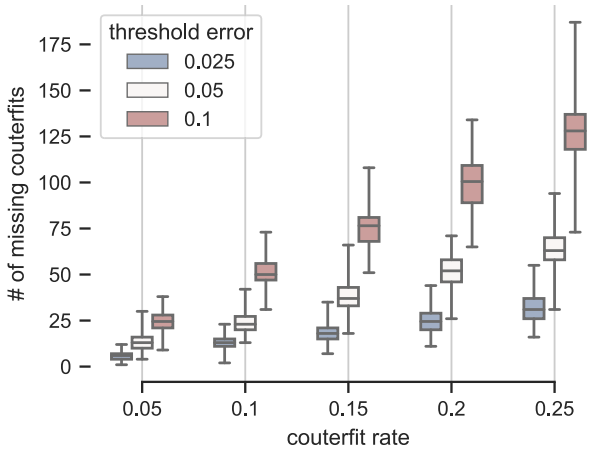

FIGURE 8. The number of missing counterfits with varying threshold error ϵ.

threshold error ϵ. We set the number of candidate tags to be 5000 and vary the rate of counterfeit tags from 0.05 to 0.25. We can observe that the missed counterfeit tags increases with increasing counterfeit rate. This is obviously because with more counterfeit tags, the reader is more likely to misidentify them due to the unavoidable tag collisions and coincidental time slot selection. Besides, we can also find that the number of missed counterfeit tags significantly increase as the increases of ϵ. The underlying reason is that the GSC with small ϵ has a shorter time frame L_{idf}, which increases the probability of tag collisions, as well as the chance of coincidental slot selection. For example, when the counterfeit rate is 25%, the number of unidentified counterfeit tags of GSC ($\epsilon = 0.025$) is approximately 2.4% of the total counterfeit population, which is only 1/4 of GSC($\epsilon = 0.05$) and 1/4 of GSC($\epsilon = 0.1$).

VIII. DISCUSSION ON PRACTICAL ISSUES

A. ANONYMOUS HASH FUNCTIONS

The proposed method assumes that the tag uses anonymous hash functions to choose its slot, which is only known by the manufacturer, and no authentication or attacker should know about this hash function. Such assumption is aligned with the common commercial-off-the-shelf system (COTS), which uses a random number generator to produce unpredictable slot index. The verifiable credential is protected by the anonymous hash function, if such hash function is leaked, the proposed method fails to work since the attackers may produces a perfect counterfeit tag with the same slot index pattern. In additional, the proposed method is not compatible with some recent design with high slot utilization, which assume there is a uniform public hash function shared by all the tags and reader. Under such an assumption, the reader can predict the slot picked by each tag and issue some commands to resolve or skip collision slots for higher efficiency. If the tag want to supports both high slot utilization and anonymous hash authentication, a possible solution is to hold two types of hash functions private $h_1(\cdot)$ and public $h_2(\cdot)$, where $h_1(\cdot)$ is used to generate verifiable slot index when verifying tags

and $h_2(\cdot)$ is used to generate predictable slot index when collecting tags.

B. MULTI-READER SCENARIO

In a multiple-reader scenario, we can use an aggregation strategy or a segmentation strategy to coordinate the responses of multiple readers. With the aggregation strategy, all readers use the same configuration to initialize a same length of time frame. In this setting, no matter where the tag is located, it always responds to the reader on the same slot index. Therefore, the tag ID and the associated slot index from multiple readers could be aggregated and analyzed in the back-end, and the reader can use the counterfeit detection algorithm directly on the aggregated results to verify their slot offset. When the group size of trust community is large, the segmentation strategy can be applied to verify the tags. in this setting, each reader uses an individual configuration to initialize time frame to fit for the tag population within the coverage area of the reader. Therefore, the frame length under the segmentation strategy could be much shorter than that under the aggregation strategy. However, since each reader uses different communication settings, the slot index collected by different readers is incomparable. To address this problem, the trust community will be split by their locations, each sub-group of tags within each reader is considered independently. The segmentation strategy usually has better time efficiency because it can adapt to the population of tags with optimized frame setting. However, dividing tags into subgroups will decrease the accuracy of the authentication. If several counterfeit tags are located in the same reader, it may conduct a byzantine attack to cheat the authentication system. Therefore, the segmentation strategy is a time-efficient solution that is only applicable when the trust community is very large, while the aggregation strategy is a safe choice without accuracy delay.

C. C1G2-COMPATIBILITY

The proposed method is designed following the specification of the C1G2 standard. For example, the message exchange format within each communication round is compatible with the C1G2 protocol. However, it cannot be directly implemented in commercial RFID systems for the following reasons: First, on the tag side, the proposed method requires RFID tags to select their slot with a shift hashing function, while current C1G2 tags use a random number generator to choose their slot. Second, on the reader side, the proposed method needs to record the index of the identified ID, which can be supported by the reader theoretically but there is no API for that information at present. Therefore, to implement this method on the RFID system, we need to carry out two modifications: First of all, the tag should support hash functions, which may add extra hardware complexity and cost compared to the C1G2 random number generator; Meanwhile, the reader should provide more low-level information, which may add software complexity.

IX. CONCLUSION

In this paper, we propose a Group-based Slot Constraint (GSC) scheme for lightweight counterfeit detection in large-scale RFID systems. GSC uses a hybrid hash function with both private and public parts to map tags to a set of random and correlated slots with a fixed value constraint, which can be leveraged to identify counterfeit tags in a batch manner. Theoretical analysis and simulations demonstrate that GSC offers desirable accuracy and outperforms existing physical functional solutions in terms of time efficiency. The major contributions of GSC lies on offer an effective batch authentication solution without exactly knowing the hash functions required by the tags. These features simplify the authentication process, and enable effective authenticating group tags without a complicated certificate database. The major limitations include that it only works in systems with naturally groups tags and fails to work if the counterfeit tags become the majority of the group. In future work, we will investigate how to overcome the above limitations and try to further extend the applications hybrid hash functions to various fields of applications to speeds up the tag identification and event detection in RFID systems.

REFERENCES

[1] L. Xie, H. Han, Q. Li, J. Wu, and S. Lu, "Efficiently collecting histograms over RFID tags," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2014, pp. 145–153.
[2] 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 Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3145–3155, Dec. 2014.
[3] L. Xie, Y. Yin, A. V. Vasilakos, and S. Lu, "Managing RFID data: Challenges, opportunities and solutions," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1294–1311, 3rd Quart., 2014.
[4] L. Yang, P. Peng, F. Dang, C. Wang, X.-Y. Li, and Y. Liu, "Anti-counterfeiting via federated RFID tags' fingerprints and geometric relationships," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1966–1974.
[5] X. Chen, J. Liu, X. Wang, H. Liu, D. Jiang, and L. Chen, "Eingerprint: Robust energy-related fingerprinting for passive RFID tags," in *Proc. 17th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2020, pp. 1101–1113.
[6] X. Xie, X. Liu, S. Guo, H. Qi, and K. Li, "A lightweight integrity authentication approach for RFID-enabled supply chains," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2021, pp. 1–10.
[7] Q. Lin, L. Yang, and Y. Guo, "Proactive batch authentication: Fishing counterfeit RFID tags in muddy waters," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 568–579, Feb. 2019.
[8] N. Dinarvand and H. Barati, "A survey and comparing RFID authentication protocols based on elliptic curve cryptography," *Majlesi J. Telecommun. Devices*, vol. 5, no. 1, pp. 1–5, 2016.
[9] M. Ouaissa, M. Ouaissa, A. Rhattoy, and M.-I. University, "An efficient and secure authentication and key agreement protocol of LTE mobile network for an IoT system," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 4, pp. 212–222, Aug. 2019.
[10] N. Dinarvand and H. Barati, "An efficient and secure RFID authentication protocol using elliptic curve cryptography," *Wireless Netw.*, vol. 25, no. 1, pp. 415–428, Jan. 2019.
[11] X. L. Liu, X. Xie, and X. B. Zhao, "Fast identification of blocked RFID tags," *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 2041–2054, Jan. 2018.
[12] X. Chen, K. Yang, X. Liu, Y. Xu, J. Luo, and S. Zhang, "Efficient and accurate identification of missing tags for large-scale dynamic RFID systems," *J. Syst. Archit.*, vol. 124, Mar. 2022, Art. no. 102394.
[13] Y. Wang, J. Liu, X. Wang, X. Chen, Y. Yan, and L. Chen, "Time-efficient missing tag identification in an open RFID system," *ACM Trans. Sensor Netw.*, vol. 16, no. 3, pp. 1–27, Aug. 2020.

[14] T. Yang, A. X. Liu, M. Shahzad, D. Yang, Q. Fu, G. Xie, and X. Li, "A shifting framework for set queries," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 3116–3131, Oct. 2017.
[15] X. Xie, X. Liu, H. Qi, S. Guo, and K. Li, "A tag-correlation-based approach to fast identification of group tags," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3096–3109, Sep. 2022.
[16] G. Avoine, M. A. Bingöl, X. Carpent, and S. B. O. Yalcin, "Privacy-friendly authentication in RFID systems: On sublinear protocols based on symmetric-key cryptography," *IEEE Trans. Mobile Comput.*, vol. 12, no. 10, pp. 2037–2049, Oct. 2013.
[17] B. A. Alzahrani, "Secure and efficient cloud-based IoT authenticated key agreement scheme for e-Health wireless sensor networks," *Arabian J. Sci. Eng.*, vol. 46, no. 4, pp. 3017–3032, Apr. 2021.
[18] D. Zanetti, B. Danev, and S. Apkun, "Physical-layer identification of UHF RFID tags," in *Proc. 16th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2010, pp. 353–364.
[19] S. C. G. Periaswamy, D. R. Thompson, and J. Di, "Fingerprinting RFID tags," *IEEE Trans. Depend. Sec. Comput.*, vol. 8, no. 6, pp. 938–943, Nov./Dec. 2011.
[20] 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 Trans. Netw.*, vol. 24, no. 2, pp. 846–858, Apr. 2016.
[21] Q. Pan, Z. An, X. Yang, X. Zhao, and L. Yang, "RF-DNA: Large-scale physical-layer identifications of RFIDs via dual natural attributes," in *Proc. 28th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 419–431.
[22] K. Zhang, J. Zhang, X. Xie, X. Tong, X. Liu, and K. Li, "Frequency- and orientation-related phase fingerprints for RFID tag authentication," in *Proc. 19th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Sep. 2022, pp. 307–315.
[23] S.-R. Lee, S.-D. Joo, and C.-W. Lee, "An enhanced dynamic framed slotted Aloha algorithm for RFID tag identification," in *Proc. 2nd Annu. Int. Conf. Mobile Ubiquitous Syst., Netw. Services*, 2005, pp. 166–172.
[24] M. A. Bonuccelli, F. Lonetti, and F. Martelli, "Tree slotted aloha: A new protocol for tag identification in RFID networks," in *Proc. Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, 2006, pp. 603–608.
[25] T. Li, S. Chen, and Y. Ling, "Efficient protocols for identifying the missing tags in a large RFID system," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1974–1987, Dec. 2013.
[26] X. Liu, K. Li, G. Min, Y. Shen, A. X. Liu, and W. Qu, "A multiple hashing approach to complete identification of missing RFID tags," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 1046–1057, Mar. 2014.
[27] X. Xie, X. Liu, K. Li, B. Xiao, and H. Qi, "Minimal perfect hashing-based information collection protocol for RFID systems," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2792–2805, Oct. 2017.
[28] J. Yu, W. Gong, J. Liu, L. Chen, K. Wang, and R. Zhang, "Missing tag identification in COTS RFID systems: Bridging the gap between theory and practice," *IEEE Trans. Mobile Comput.*, vol. 19, no. 1, pp. 130–141, Jan. 2020.
[29] X. Xie, X. Liu, H. Qi, and K. Li, "Fast identification of multi-tagged objects for large-scale RFID systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 992–995, Aug. 2019.
[30] C. C. Tan, B. Sheng, and Q. Li, "How to monitor for missing RFID tags," in *Proc. 28th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2008, pp. 295–302.
[31] 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 Trans. Netw.*, vol. 22, no. 4, pp. 1079–1091, Aug. 2014.
[32] X. Xie, X. Liu, X. Zhao, W. Xue, B. Xiao, H. Qi, K. Li, and J. Wu, "Implementation of differential tag sampling for COTS RFID systems," *IEEE Trans. Mobile Comput.*, vol. 19, no. 8, pp. 1848–1861, May 2019.
[33] J. Yu, L. Chen, R. Zhang, and K. Wang, "On missing tag detection in multiple-group multiple-region RFID systems," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1371–1381, May 2016.
[34] X. Liu, K. Li, J. Wu, A. X. Liu, X. Xie, C. Zhu, and W. Xue, "Top-k queries for multi-category RFID systems," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.
[35] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "A group-based binary splitting algorithm for UHF RFID anti-collision systems," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 998–1012, Feb. 2020.
[36] M. Shahzad and A. X. Liu, "Fast and accurate estimation of RFID tags," *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 241–254, Feb. 2015.
[37] M. M. Hasan, S. Wei, and R. Vaidyanathan, "Estimation of RFID tag population size by Gaussian estimator," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
[38] C. Qian, H. Ngan, Y. Liu, and L. M. Ni, "Cardinality estimation for large-scale RFID systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1441–1454, Sep. 2011.
[39] M. Kodialam and T. Nandagopal, "Fast and reliable estimation schemes in RFID systems," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2006, pp. 322–333.

BING ZHANG received the B.Sc. degree in computer science from the College of Computer Science and Technology, Tianjin University, China, in 2005, the M.Sc. degree in computer systems engineering from the Technical University of Denmark, Denmark, in 2008, and the Ph.D. degree in computer science from the College of Computer Science and Technology, Tianjin University, in 2016. She is currently a Lecturer with the Tianjin University of Commerce. Her current research interests include artificial intelligence, biomedical computing, medical image processing, and optimization methods.

XU HU received the B.Sc. degree in computer science from the College of Computer Science and Technology, Tianjin University, China, in 2005, and the M.Sc. degree in economics and management from the School of Economics and Management, Tsinghua University, China, in 2020. He is currently pursuing the Ph.D. degree in computer science with the College of Computer Science and Technology, Tsinghua University. His current research interests include smart grids, the Internet of Things, cloud computing, and green energy.

SHAN GAO received the Ph.D. degree in computer science from the College of Computer Science and Technology, Tianjin University, China, in 2012. She is currently a Lecturer with the Tianjin University of Commerce. Her research interests include decision cognition, decision reasoning, and knowledge model construction.