Anti-counterfeiting via Federated RFID Tags' Fingerprints and Geometric Relationships

Lei Yang*, Pai Peng*, Fan Dang*, Cheng Wang‡, Xiang-Yang Li*†, Yunhao Liu*
* School of Software and TNLIST, Tsinghua University, China
† Department of Computer Science, Illinois Institute of Technology, USA
‡ Department of Computer Science, Tongji University, China
E-mail:{young,pai,fan}@tagsys.org, chengwang@tongji.edu.cn, xiangyang.li@gmail.com, yunhao@greenorbs.com

***Abstract*—RFID has been widely adopted as an effective method for anti-counterfeiting. Legacy systems based on security protocol are either too heavy to be affordable by passive tags or suffering from various protocol-layer attacks, *e.g.* reverse engineering, cloning, side-channel. In this work, we present a novel anti-counterfeiting system, *TagPrint*, using COTS RFID tags and readers. Achieving a *low-cost* and *offline* genuineness validation utilizing *passive* tags has been a daunting task. Our system achieves these three goals by leveraging a few of federated tags' fingerprints and geometric relationships. In TagPrint, we exploit a new kind of fingerprint, called *phase fingerprint*, extracted from the phase value of the backscattered signal, provided by the COTS RFID readers. To further solve the separation challenge, we devise a geometric solution to validate the genuineness. We have implemented a prototype of TagPrint using COTS RFID devices. The system has been tested extensively over 6,000 tags. The results show that our new fingerprint exhibits a good fitness of uniform distribution and the system achieves a surprising Equal Error Rate of 0.1% for anti-counterfeiting.**

***Keywords*—RFID, Anti-counterfeiting, Phase fingerprint, TagPrint**

I. Introduction

Counterfeiting has been viewed as one of the most severe economic crimes in modern society. It leads to serious damages to companies, governments, and individuals with a unique set of problems. World Customs Organization reports that counterfeiting has been detected in around 140 countries in 2008 [1]. A frequently cited estimation from the World Health Organization [1] claims that 10% of the global medicine supply is counterfeit, while rising to 30% in the developing countries. The fight against counterfeiting becomes a global challenge. A large number of technologies have been proposed, ranging from very simple to highly sophisticated ones [1], such as watermark, security ink, and laser code, barcode, *etc.* Unfortunately, these tranditional anti-counterfeiting methods are easy to be cracked in practice [1]. Worse, majority of them depend on human eyes to distinguish, lacking a high-efficient and automatic validation method.

To improve the efficiency, RFID technology has been introduced for anti-counterfeiting in recent years. The common way of RFID enabled anti-counterfeiting is to attach an RFID tag on an item. A serial number is stored in both the tag's memory and the database of manufacturer [2], [3]. When validating, the consumer employs an RFID reader to obtain the serial number from the tag and then sends the number to the manufacturer. If the number exists in the database, the tag (and the related item) is considered to be genuine. Otherwise, it is fake. This approach is based on the assumption that the tag manufacturing is kept under strict control, and tampering with tag memory is not allowable. Nevertheless, this assumption no longer holds true nowadays due to the development of semiconductor technology and rapidly falling tag price. It is not difficult for counterfeiters to acquire the genuine serial numbers and write them into forged tags. To enhance the security, a variety of protocols are designed for tag authentication, such as [4]–[7]. For example, encrypted message are transmitted between reader and tag for anti-eavesdrop. However, these methods have been rarely employed in practice for three reasons. First, their computing overhead are too heavy to be affordable by the passive tags with low cost and limited power budget. Second, they are challenged by various protocol-layer attacks, such as reverse engineering, side-channel, replay attack, cloning, and so on. Especially there is no way of distinguishing the original and the cloned in protocol layer. Third, online validation at server-side is required for determining whether the serial number exists in database. The manufacture has to invest more and more in their authentication servers to cope with the increasing validation requests.

Motivated by the above limitations, the third way is explored in this paper: *fingerprint based anti-counterfeiting*. The tag' fingerprints are the unique spectral- or time-domain features obtained by analyzing its communication at physical layer, resulting from tag's physical characteristics, such as coil size, antenna length, impedance, *etc.* The key appeal of applying fingerprint for anti-counterfeiting is twofold. First, the tags' fingerprints are unique and unforgeable, thereby they can provide high security guarantees against various protocol-layer attacks. Second, no upgrades of hardware or firmware on existing systems are required such that millions of deployed RFID tags are generally being compatible. Although many tag fingerprints have been proposed in [8]–[15], their limited fingerprint ranges and dedicated extraction device required restrict their universal adoption for anti-counterfeiting in practice. For example, [10] uses a purpose-built oscilloscope with a extremely high sampling rate to extract SN16. In this paper, we observe that the reader's transceiver and tag's reflection characteristic will all introduce some additional phase shifts besides the RF phase rotation over distance between reader and tag [16]. We exploit this additional phase shifts as a new kind of fingerprint, called *phase fingerprint*, for identifying a pair of reader and tag. As a step towards understanding phase fingerprint, we conduct a large-scale experiment involving

978-1-4799-8381-0/15/$31.00 ©2015 IEEE

6,000 tags, which demonstrates that phase fingerprint has a good fitness of uniform distribution across different tags.

Transforming the phase fingerprint into a practical anti-counterfeiting system, however, requires addressing multiple challenges. First, it is hard to extract the fingerprint from measured phase due to the dependency of distance, especially when facing thousands of tags. Second, phase fingerprint suffers a joint influence from reader and tag, making usage of different readers for fingerprint acquisition and validation stage impossible. We call this issue *separation challenge*. Third, the uniqueness of phase fingerprint is subject to the limited phase resolution. To address these issues, we design an anti-counterfeting system, called *TagPrint*, including three distributed components deployed in tag provider, product manufacturer and consumer respectively. First, we devise a fast, reliable and automatic approach for tag provider to acquire the tags' phase fingerprints in a batch mode. Second, the product manufacturer attaches m tags ($m \geq 4$) as a federated meta tag on each product to fingerprint the product genuineness. Third, the consumer, as a purchaser of product, leverages a geometric approach to validate the genuineness.

To summarize, we made the following contributions:

• We exploit a new kind of fingerprint for a pair of reader and tag from their backscatter signals. To best of our knowledge, we are the first to propose the phase fingerprint. In addition, a fast and reliable approach is devised to automatically acquire these fingerprints.

• A large-scale experiment involving 6,000 tags is performed to demonstrate the stability and randomness of phase fingerprint. The results show that our new fingerprint exhibits a good fitness of uniform distribution over tags, outperforming the accuracy of tag classification over ∂_{TIE} [10], $\bar{P}_B$ [10], GenPrint [13] and MinPower [11] by $39.46\%\times$, $130\%\times$, $25.69\%\times$ and $7.5\%\times$.

• We jointly utilize federated tags' fingerprints and geometric relationships for the genuineness validation. Our approach is a totally offline solution without any communication between consumer and product manufacturer.

• We design and implement the phase fingerprint based anti-counterfeiting system, purely based on COTS RFID devices, which makes the fast adoption and deployment possible. We systematically evaluate the system with extensive experiments and it achieves an Equal Error Rate of 0.1%.

The rest of the paper is organized as follows. The main design of TagPrint is overviewed in Section II. We exploit the new fingerprint extracted from the backscatter signals in Section III. The approaches to fingerprinting and validating genuineness are presented in Section IV and Section V respectively. The implementation and evaluation are given in Section VI. The related work is reviewed in Section VII. Lastly, we conclude our work in Section VIII.

II. Overview

This section presents the threat model and system architecture.

A. Threat Model

Ultra-low-cost UHF tags (5-10 cents each) have become the preferred choice of many industries for anti-counterfeiting. Following the common practices, we focus on UHF tags in this work. There are four kinds of entities: tag provider, product manufacturer, consumer and counterfeiter in the system. The *tag provider* manufactures the RFID tags, like Alien [17] or ImpinJ Corp [18]. We do not make any assumptions on the production and purchase of tags, which are totally uncontrollable, such that both the product manufacturer and counterfeiter are able to purchase any number of tags from tag provider. The *product manufacturer* utilizes the technique of RFID to protect their products from being counterfeited. The *consumer*, as a purchaser of product, desires to know whether the product is genuine. The *counterfeiter* is to pursue huge profits through forging products and making them pass the genuineness validation.

Counterfeiter's capability. We hardly put limitations on the counterfeiter's capabilities. The counterfeiter can: (*i*) eavesdrop any wireless communications between the reader and tags, and read or write any tags' memory; (*ii*) copy a tag's memory to another one (clone tag), which behaves as the same as the copied one in terms of communications and computations; (*iii*) find a tag with the fingerprint as same as the genuine one's from numerous candidates at a price; (*iv*) not recycle the tags from products and re-attach them on the forged product. To avoid such behaviors, the product manufacturer usually adopts the self-destructive design so that the tags attached to the product are destroyed after the product's cover is opened [2].

Counterfeiting purpose. We make a reasonable and practical assumption on the purpose of counterfeiting as follows. *The goal of the counterfeiting is to pursue huge profits. There is no motivation for counterfeiter if the counterfeiting is unprofitable.* This suggests that our system should use the minimum cost to achieve the relative security as long as making the counterfeiting unprofitable.

Consumer capability. The consumer who attempts to validate genuineness of product, can guarantee his reader not to be hacked. The reader can correctly execute our validation codes. Since this paper concentrates on anti-counterfeiting, the consumer's privacy protection is beyond our discussion here. The consumer can simply destroy the tags when the product is purchased.

B. System Architecture

TagPrint is a distributed system as shown in Fig. 1, involving the following three main components.

• *Fingerprinting tag*: Tag provider extracts the phase fingerprint from the backscatter signals of tag through a fast, reliable and automatic approach. Both the tags and their phase fingerprints are offered to downward manufacturers.

• *Fingerprinting genuineness*: When manufacturing the product, the manufacturer attaches a group of m federated tags ($m \geq 4$) on each product. These tags' locations are randomized but fulfill a geometric constraint. For offline validation, these tags' fingerprints and geometric relationships are encrypted and stored in tags' memories.

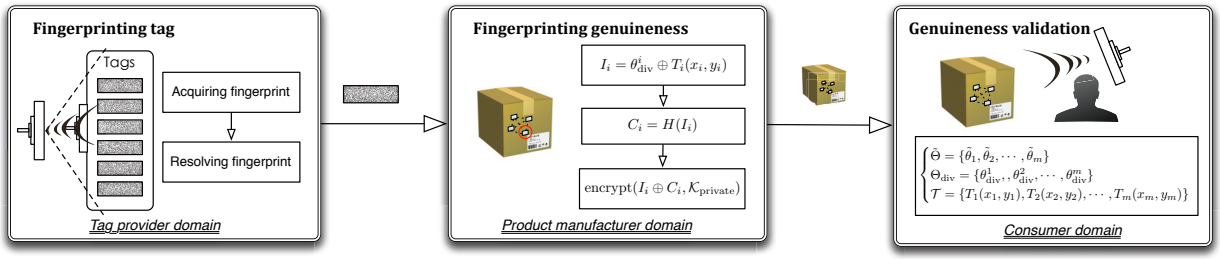


Fig. 1: The system architecture. Three components distributed in the system of tag provider, product manufacturer and consumer.

• *Validating Genuineness*: The consumer employs any COTS reader to obtain the phase fingerprints and geometric relationships from the federated tags' memories, as well as measure the phase values from backscattered signals. Eventually, the product's genuineness is validated through determining whether the inferred tags geometric relationships agree with the stored ones.

Below we describe these components in detail.

III. Fingerprinting Tag

In this section, we present the phase fingerprint and the acquisition approach, then conduct a large-scale experiment to study its statistical features.

A. Defining Phase Fingerprint

The RF phase is a common parameter supported by COTS readers. Suppose d is the distance between the reader antenna and the tag, the signal traverses a total distance of $2d$ back and forth in backscatter communication [16]. The total phase rotation output by the reader can be expressed as

$$\begin{cases} \theta = \left(\frac{2\pi}{\lambda} \times 2d + \theta_{\text{ant}} + \theta_{\text{tag}}\right) \bmod 2\pi \\ \theta_{\text{div}} = \theta_{\text{ant}} + \theta_{\text{tag}} \end{cases} \tag{1}$$

where λ is the wavelength. In addition to the RF phase rotation over distance, the reader's transceiver and the tag's reflection characteristic will all introduce some additional phase rotation θ_{ant} and θ_{tag} respectively, as shown in Fig. **??**. We exploit this additional phase shifts as a new kind of fingerprint, called phase fingerprint, for a pair of reader and tag.

Definition 1 (Phase fingerprint): The phase shift, introduced by the hardware characteristics from a pair of reader and tag, is defined as the phase fingerprint, denoted as θ_{div}.

B. Acquiring Phase Fingerprint

An RF phase θ is measured each time when a tag is successfully interrogated. We can not directly extract θ_{div} from θ because of the existence of d. The naive method is to measure d for each time. Although this method is simple and mathematically sound, one practical problem arises in working system, when facing thousands of tags, that it will consume enormous time and human power to do this. Even worse, θ is a Gaussian random variable instead of an accurate value [19]. Thus, we need to design a *fast*, *reliable* and *automatic* approach to finish this task.

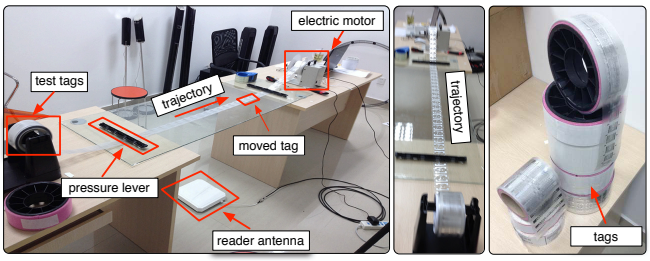


Fig. 2: Illustration of acquisition of phase fingerprint.

Fig. 2 displays our hardware setup. We build a "conveyor" style arrangement where one reel of test tags are conveyed along a pre-defined trajectory with a uniform linear motion. The tags' movements are powered by an electric motor. Two pressure levers are used to ensure the tags over the glass taking stable movements. A reader is deployed under the glass[1]. When passing through the glass area, each test tag will be interrogated for several times.

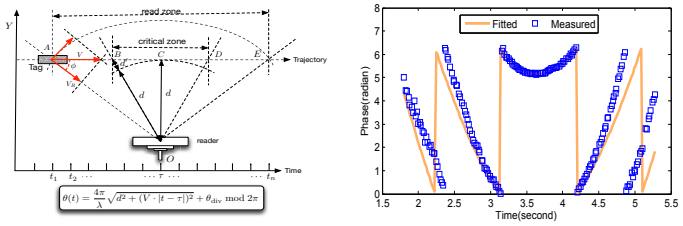
We model the tag's movement as shown in Fig. 3(a). For clarity, the read zone is assumed within the fan AOE. All test tags are conveyed from position A to E along a pre-defined linear trajectory at a constant speed of V. The A and E are the positions where the tag is firstly and lastly interrogated. The reader is deployed at position O with a vertical distance of d to the tag's trajectory. Let τ be the time that the tag moves at position C where $OC \perp AE$. Then the theoretical phase measured at arbitrate any time t can be given by:

$$\theta(t) = \frac{4\pi}{\lambda}\sqrt{d^2 + (V \cdot |t - \tau|)^2} + \theta_{\text{div}} \bmod 2\pi \tag{2}$$

where λ, v, and d are known parameters.

Suppose we obtain n phase measurements from a particular tag, denoted as $\{\tilde{\theta}(t_1), \tilde{\theta}(t_2), \cdots, \tilde{\theta}(t_n)\}$ where $\tilde{\theta}(t_i)$ is the i^{th} phase measurement at time t_i. It is not difficult to resolve the θ_{div} given a measured phase $\tilde{\theta}(t)$ and time t. However, the challenges come from the thermal noise and doppler effect, resulting in inaccurate phase measurement in practice. In addition, the electromagnetic wave is hard to control and EPC C1Gen2 protocol adopts a random anti-collision protocol, so the time base τ is also unknown besides θ_{div} in Eqn. 2. To address these challenges, TagPrint adopts the *nonlinear least squares* to estimate θ_{div}. The objective function is formalized

[1]The glass is widely considered as the perfect material that does not affect the signal propagation in UHF band.



(a) Mathematical model (b) Experimental result

Fig. 3: Resolving phase diversity. (a) Mathematical model on the detected phase when the test tag moves along the predefined trajectory. (b) The fitted $\theta_{\text{div}} = 3.98$ radians and $\tau = 3.67$ seconds.

as follows:

$$\min \sum_{i=1}^{n} |\theta(t_i) - \tilde{\theta}(t_i)|^2 \tag{3}$$

Subject to:

$$\begin{cases} \{\tilde{\theta}(t_1), \tilde{\theta}(t_2), \cdots, \tilde{\theta}(t_n)\} \\ t_1 \leq \tau \leq t_n \\ 0 \leq \theta_{\text{div}} \leq 2\pi \end{cases}$$

The Gauss-Newton method, which is based on a linear approximation of the objective function in the neighborhood of parameter vector, is employed here. We start with an initial approximation of the parameter vector and iteratively update this parameter vector until it converges to a minimum of an objective function.

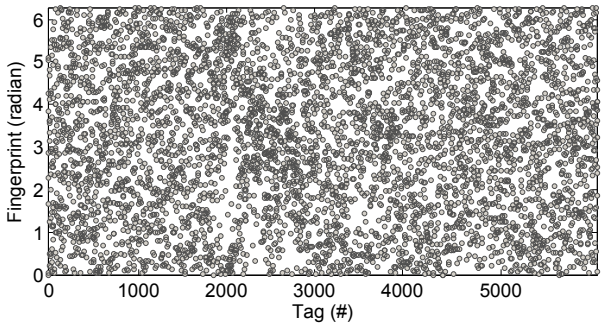
Fig. 3(b) shows an example to estimate the θ_{div} and τ. We can observe that the critical zone (the period that is closest to τ), is well matched, but the measured phase beyond this zone is usually greater than the theoretical value because of the doppler effect. The further position has much more noticeable deviation, because the radical velocity from tag to reader becomes larger when tag leaves reader.

C. Phase Fingerprint Characterization

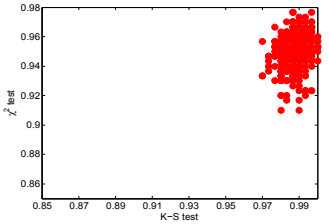
Our initial experiments are conducted over total 6,000 tags using the conveyor. The detailed hardware setup is described in §VI. We attempt to study the basic statistic characteristics of phase fingerprint.

Hypothesis 1: The phase fingerprint follows uniform distribution over tags.

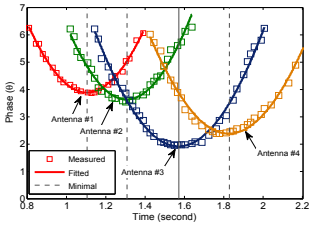
We acquire phase fingerprints from 6,000 COTS tags using a same reader. The results are illustrated in Fig. 4(a). None apparent patterns are observed by visual inspection. We use Kolmogorove-Smirnov Test (K-S test) and χ^2 Test of 0.95 significance level to further test their randomness. K-S test and χ^2 test are widely adopted tool to test the goodness-of-fit of a specific distribution. In each test, we shuffle the 6,000 phase fingerprints into 100 equal sized groups. For each group, K-S and χ^2 test are independently applied to test the goodness-of-fit of the uniform model. 95% groups should pass the test if the phase fingerprint follows the uniform model when the significance level is set to 0.95. The test repeats 300 times and each time outputs a pass percentage. Fig. 4(b) shows the test results of goodness-of-fit on uniform distribution. The x-value



(a) Resolved phase fingerprints over $6,000$ COTS tags



(b) Randomness test



(c) Reader impacts

Fig. 4: Statistics exhibition on phase fingerprint.

and y-value of the points are the pass percentages performed by K-S and χ^2 test respectively. We expect that if these points are clustered at the right corner, the fitness of the uniform model is good. As a result, we find all the K-S tests demonstrate that the phase fingerprint has a perfect fitness. The χ^2 test is usually more rigorous than K-S test. In our results, all pass rates of χ^2 test are above 0.9, and the mean pass rate is 0.95. These two randomness tests verify that phase fingerprint has a good fitness of uniform distribution over tags.

Hypothesis 2: Different readers take different impacts on the phase fingerprint.

Re-checking the Eqn. 1, we notice that when $t = \tau$, *i.e.* the tag arrives at position C, the θ obtains the minimum in the critical zone, *i.e.* $\theta_{\text{min}} = \frac{4\pi}{\lambda} d + \theta_{\text{div}} \bmod 2\pi$. Since d is a constant, then the minimum value indirectly reflects the phase fingerprint. To investigate the reader's impact on the phase fingerprint, we employ different readers to acquire the fingerprints. We change the reader's transceiver characteristics through using different antennas. We think the reader is different when it connects to different antennas. Therefore, we deploy four different antennas in a row under the glass to collect the phase from a same tag, keeping the same vertical distance to trajectory. The measured and fitted results are shown in Fig. 4(c). The visual inspection tells that four antennas take different impacts on the phase fingerprint, where the θ_{min}s are 3.7, 3.6, 1.8 and 2.4 radians. Thus, the phase fingerprint is not a 'good' metric to identify a tag in terms of the dependence of reader.

IV. Fingerprinting Genuineness

In spite of the good fitness of uniform distribution, it is challenging to employ phase fingerprint for anti-counterfeiti ng for three reasons. First, the phase resolution of COTS reader limits the fingerprint range. For example, an ImpinJ Reader [16] uses 12 bits to encode phase value leading to

$2^{12} = 4,096$ unique phase fingerprints at most. There is a 0.024% of probability to find two tags with same fingerprints. Second, phase fingerprint θ_{div} is a combination of θ_{ant} and θ_{tag}, which cannot be separated in theory[2]. It requires to use the same reader in the stages of fingerprint acquisition and validation, which is imfeasible in practice. Third, being different from fingerprint acquisition, it is neither convenient nor user-friendly to ask the consumer to build a 'conveyor' device for resolving θ_{div}.

To deal with the issues above, the product manufacturer attaches m tags ($m \geq 4$) as a federated group on a each product. These tags' positions are randomized but fulfill the following geometric constraint.

Definition 2 (Geometric constraint): The Euclidean distance between any two tags is less than $\lambda/2$.

A tag coordinate system is built to describe these tags' locations. These tags' relative positions among each other in the coordinate system is called as *geometric relationships*, which are assumed to be known after attached. We denote the i^{th} tag's coordinate as $T_i(x_i, y_i)$.

The tags' phase fingerprints are transferred from the tag provider to product manufacturer after purchased. We do not need to know how the provider acquires these fingerprints, but only require their acquisition to fulfill the following constraint.

Definition 3 (Acquisition constraint): The m tags' fingerprints relevant to a same product must be measured using a same antenna such that their fingerprints contain the same influence of reader.

In fact, we can leverage the 'separation challenge' to protect the tag's fingerprint from being cracked, because the counterfeiter has no way to find a same tag with desired fingerprint, without using the private reader of tag provider.

Definition 4 (Product fingerprint): The product fingerprint is composed of m tuples, $\{< \theta^1_{\text{div}}, T_1 >, \cdots, < \theta^m_{\text{div}}, T_m >\}$, each of which contains a tag's fingerprint and coordinate.

In summary, we define the sequence of pairs of tag fingerprint and coordinate as the product fingerprint, formally defined in Defn. 4. To support the offline validation, we adopt the technique of asymmetric encryption. Suppose the product manufacturer has a public/private key pair, it uses the private key $\mathcal{K}_{\text{private}}$ to encrypt the i^{th} tag's fingerprint and coordinate as follows:

$$\text{cipher}_i = \text{encrypt}(\theta^i_{\text{div}} \oplus T_i(x_i, y_i) \oplus C_i, \mathcal{K}_{\text{private}}) \quad (4)$$

where $\oplus$ is the concatenate operator, and $C_i = H(\theta^i_{\text{div}} \oplus T_i(x_i, y_i))$ is the checksum to ensure the integrity of cipher text. The encrypted information is stored in the relevant tag's memory. Although the cryptographic method is employed here, we must claim that no crypto operations are performed by the tag. It is only used for the purpose that these information retrieved from the tags, indeed comes from the product manufacturer. Note that the encrypted information is completely public. Everyone including the counterfeiter is able to decrypt them if they know the product manufacturer's public key. We will discuss the security analysis in next section.

[2]No matter how to adjust parameters in Eqn. 1, these two variables always take the same changes.

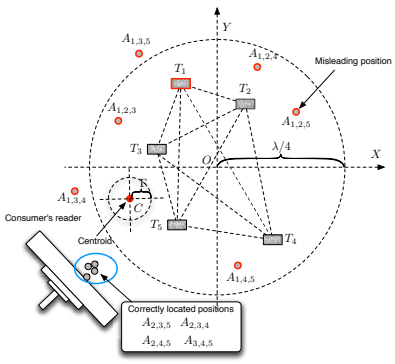


Fig. 5: An example of genuineness validation. Five tags are attached on the products where T_1 is a forged tag.

V. Genuineness Validation

This section presents a geometric approach to validate the genuineness, and give analysis on its security.

A. Validating Product Genuineness

Targeting to offline genuineness validation, the consumer employs his/her reader to execute the following steps:

Step 1: The consumer places the reader at a random location, which is kept unchanged during data acquisition.

Step 2: TagPrint obtains two encrypted values, θ^i_{div} and $T_i(x_i, y_i)$ from i^{th} tag's memory and records the backscatter signal's phase $\tilde{\theta}_i$ at the same time. If it fails to decrypt the two values using product manufacturer's public key $\mathcal{K}_{\text{public}}$, the product is declared to be fake. Otherwise go to *Step 3*.

Step 3: TagPrint enumerates all possible combination of 3 tags from the m tags. For each combination $\{T_i, T_j, T_k\}$, TagPrint utilizes hyperbola based method (explain later) to locate the reader's position with the inputs of $\{\tilde{\theta}_i, \tilde{\theta}_j, \tilde{\theta}_k\}$, $\{\theta^i_{\text{div}}, \theta^j_{\text{div}}, \theta^k_{\text{div}}\}$ and $\{T_i(x_i, y_i), T_j(x_j, y_j), T_k(x_k, y_k)\}$. Finally, there are total $\binom{m}{3}$ locations calculated.

Step 4: In theory, these $\binom{m}{3}$ reader locations would coincide with each other at the ground truth. With regards to noise, we calculate the centroid of these resolved locations and define an empirical threshold Γ. Let $A_{i,j,k}$ denote the resolved reader location by tag T_i, T_j and T_k, and C be the centroid. If $|A_{i,j,k}C| < \Gamma$ for all $A_{i,j,k}$, the product is declared to be genuine. Otherwise, it is fake.

For clear presentation, we assume reader and tags locate in a same plane, and describe the system in two dimensions, but it is easy to be extended to the three dimensions. Fig. 5 illustrates an example of genuineness validation. There are 5 tags attached on the product and T_1 is a forged tag (defined in Defn. 5). The $A_{i,j,k}$ is the position in tag coordinate system located by tag T_i, T_j and T_k. Any localization involving the T_1 does not correctly eliminate the influence of θ^1_{div}, leading to the incorrect locations. Therefore, only four positions located by $\{T_2, T_3, T_4, T_5\}$ are correct and clustered around the ground truth. The calculated centroid is at position C. As a result, all distances $|A_{i,j,k}C|$ from the resolved reader position to the

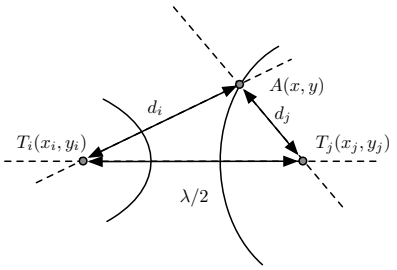


Fig. 6: A hyperbola with two different branches.

centroid is larger than the threshold of Γ, so the product is validated as being fake.

Definition 5 (Forged tag): The tag is called *forged tag* when its actual phase fingerprint or coordinate does not agree with the stored values in its memory.

B. Hyperbola based Localization

Suppose there are two tags, $T_i(x_i, y_i)$ and $T_j(x_j, y_j)$, and a reader at position $A(x, y)$. The distances from two tags to the reader are denoted as $|T_iA| = d_i$ and $|T_jA| = d_j$. The basic idea of hyperbola based position is to build a hyperbola utilizing the distance difference, *i.e.* $\Delta d_{i,j} = d_i - d_j$ to confine the reader location. Namely:

$$|\sqrt{(x-x_i)^2+(y-y_i)^2} - \sqrt{(x-x_j)^2+(y-y_j)^2}| = \Delta d_{i,j}$$

The reader position is confined on the curve as illustrated in Fig. 6. Finally, the reader location can be discovered when three hyperbolas are constructed using three tags's geometric relationships. The question is how to infer the $\Delta d_{i,j}$ using the three measured phase, phase fingerprint and tags' coordinates? Due to the space limitation, we only sketch the result.

Let $\Delta\tilde{\theta}_{i,j} = \tilde{\theta}_i - \tilde{\theta}_j$ and $\Delta\theta^{i,j}_{\text{div}} = \theta^i_{\text{div}} - \theta^j_{\text{div}}$. We have, base on Eqn. 1, the following equation:

$$\Delta\tilde{\theta}_{i,j} = \tilde{\theta}_i - \tilde{\theta}_j = (\frac{4\pi}{\lambda}\Delta d_{i,j} + \Delta\theta^{i,j}_{\text{div}}) + 2\pi K$$

where K is an integer ensuring the $\Delta\tilde{\theta}_{i,j}$ within $(0, 2\pi]$. After simple derivation, we have

$$\Delta d_{i,j} = \frac{\lambda}{4\pi}(\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}) - \frac{\lambda}{2}K \tag{5}$$

Both $\Delta\tilde{\theta}_{i,j}$ and $\Delta\theta^{i,j}_{\text{div}}$ are known parameters: $\Delta\tilde{\theta}_{i,j}$ can be calculated using the measured phases and $\Delta\theta^{i,j}_{\text{div}}$ can be obtained from two tags' memories. It worth noting that $\Delta\theta^{i,j}_{\text{div}} = \theta^i_{\text{div}} - \theta^j_{\text{div}} = \theta^i_{\text{tag}} - \theta^j_{\text{tag}}$. Because of the acquisition constraints (Defn. 3), the influence of reader is eliminated by the difference of phase fingerprint. On the other hand, thanks to the geometric constraints (Defn. 2), the value of K is confined into a small candidate $\{-1, 0, 1\}$. Specifically, considering the sign of $\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}$, there exist four cases .

(1) When $0 \le \Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}$, the hyperbola is depicted as follows.

$$\Delta d_{i,j} = \begin{cases} \frac{\lambda}{4\pi}(\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}) & \Delta d \ge 0 \\ \frac{\lambda}{4\pi}(\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}) - \frac{\lambda}{2} & \Delta d < 0 \end{cases} \tag{6}$$

(2) When $\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}} < 0$, we have the similar result as follows:

$$\Delta d_{i,j} = \begin{cases} \frac{\lambda}{4\pi}(\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}) + \frac{\lambda}{2} & \Delta d \ge 0 \\ \frac{\lambda}{4\pi}(\Delta\tilde{\theta}_{i,j} - \Delta\theta^{i,j}_{\text{div}}) & \Delta d < 0 \end{cases} \tag{7}$$

In a particular instance, both $\Delta\tilde{\theta}_{i,j}$ and $\Delta\theta^{i,j}_{\text{div}}$ are deterministic so only one case should happen. Actually, the piecewise equation describes a hyperbola which has two different branches, as illustrated in Fig. 6. The most prominent characteristic of hyperbola based localization is that any tiny measurement on Δd will lead to a significant localization error. As the Defn. 5 says, the forged tags cannot provide the correct fingerprints and coordinates that agree with the actually measured. The mismatch between $\Delta\tilde{\theta}_{i,j}$ and $\Delta\theta^{i,j}_{\text{div}}$ makes the localization results 'run out' leading to unclustered $\binom{m}{3}$ positions.

C. Security Analysis

Considering the genuineness validation from another perspective, the process likes performing a hash operation as follows:

$$h(\{\theta^i_{\text{tag}}, \theta^j_{\text{tag}}, \theta^k_{\text{tag}}\}, \{\tilde{\theta}_i, \tilde{\theta}_j, \tilde{\theta}_k\}, A(x,y)) = \{T_i, T_j, T_k\} \tag{8}$$

where $A(x, y)$ can be considered as a random seed and $\{\theta^i_{\text{tag}}$, θ^j_{tag},$\theta^k_{\text{tag}}\}$ is a secret key that cannot be comprised due to the 'separation challenge'. Even if the counterfeiter knows the output, he/she cannot reversely infer the secret key because there exist a $\bmod$ operation in Eqn. 1. The genuineness is validated through comparing the resolved tags' locations with the stored values in their memories. Next, we discuss three major attacks potentially threat genuineness validation.

Impersonation by cloning attack: The counterfeiter cannot tamper fingerprints in tags' memories because he/she does not know the product manufacturer's private key. His/her only method is to find out m tags, whose differences of fingerprint (*i.e.*, $\Delta\theta^{i,j}_{\text{div}}$) for all pairs are the same as the tags' from a genuine product, for cloning. Being different from the forged tag, the cloned tag has the same fingerprints and memory values as the original. Let N denote the cardinality of fingerprint space. Thanks to the 'separation challenge', the counterfeiter has to collect enough candidate tags such that their fingerprints can cover the entire fingerprint space. The question is how many tags should the counterfeiters to purchase? We can reduce this question to the *Coupon Collectors Problem*. Consequently, the counterfeiter must purchase about $N(\ln N)$ tags as candidates at least. Second, the counterfeiter must perform $\binom{N(\ln N)}{m}$ trails to choose m tags from the candidates to find out the correct combination. In practice, $N = 2^{12}$ and $m \ge 4$, so the counterfeiter must purchase $34,070$ tags (≈ 3407 dollar) and conduct 3.8243×10^{20} trails at least to find the correct tags. Both the cloning cost and huge computations make it hard and unprofitable, if not possible, to clone a genuine product!

Impersonation by forging attack: Can the counterfeiter clone small parts of tags and allow others to be forged? As Defn. 5, the fingerprints of forged tags do not agree with that in their memories. This will result in the mismatch between $\Delta\theta_{i,j}$ and $\Delta\theta^{i,j}_{\text{div}}$ in Eqn. 6. Thus, the influences caused by θ_{div} are not well-eliminated, leading to the incorrect distance

TABLE I: The tag models

#	Model	Company	Chip	Antenna size(mm)
1	AZ-9610	Alien	H3	44.45×10.325
2	AZ-9620	Alien	H3	27×9.7
3	AZ-9629	Alien	H3	22.5×22.5
4	AZ-9630	Alien	H3	73×12.7
5	AZ-9634	Alien	H3	44×46

TABLE II: Classification success rate

Fingerprint	Size	CSR(%) (Min;Max)	N_{tag}
∂_{TIE}	2^6	71.4 (69.7;73.0)	50
$\bar{P}_B$	2^5	43.2 (38.6;47.7)	50
Spectral	–	97.57(–;–)	50
GenPrint	–	76.94 (71.4;79.42)	150
MinPower	–	92.55 (90.7;94.4)	100
TagPrint	2^{12}	99.58(91.81;100)	50
TagPrint	2^{12}	96.71(90.63;100)	150
TagPrint	2^{12}	80.39(74.00;100)	1,000
TagPrint	2^{12}	55.31(13.00;100)	6,000

difference Δd. Every small error in Δd will lead to different branches, incurring huge location error as mentioned before.

Impersonation by replay attack: The counterfeiter can record signals from a genuine tag, and let the forged tag later retransmit an identical signal to the reader. The reader cannot distinguish the retransmitted signals from the genuine ones, if the counterfeiter can successfully make them identical. To our best knowledge, none exiting fingerprint based work can effectively defend against such an attack except ours. Besides the information retrieved from tags, our validation also utilizes the reader position, which is unknown to tags. The replay attack will fail because the reader position is randomly chosen. The consumer can select two different positions to check whether the output positions are identical. If yes, there is a replay attack.

VI. Implementation and Evaluation

In this section, we firstly present the system implementation and then give an extensively evaluation.

A. Implementation

Reader: We adopt an ImpinJ [18] Speedway modeled R420 reader without any hardware or firmware modification. The reader supports four directional antennas at most, being compatible with EPC Gen2 standard. The whole RFID system operates in the $920 \sim 926$ MHz band with frequency hopping. The size of antenna is $225mm \times 225mm \times 40mm$. **Tag**: There are 12 total different models of EPC Gen2 tags employed, which come from two of the biggest tag providers, Alien Corp [17] and ImpinJ Corp [18]. The rightmost picture illustrated in Fig. 2 shows parts of tags which are grouped in rollers. The tags' detailed information[3] is listed in Table I. Each model contains 50 tags and total $6,000$ tags are tested. **Software**: We adopt LLRP protocol to communicate with the reader. The ImpinJ reader extends this protocol for supporting the phase report. We adjust the configuration of reader to immediately report reading whenever tag is detected. The software is implemented using Java.

B. Methodology

We evaluate the TagPrint's performance in terms of phase fingerprint based tag recognition and genuineness validation.

[3]Only 5 kinds of tag models listed in this table due to page limit.

• *Tag recognition*: We adopt the *Classification Success Rate (CSR)* to evaluate the classification accuracy of tag's phase fingerprint. The CSR is defined as the percentage of correctly assigned testing fingerprints to their respective class. Each individual tag is considered as one class. The k-Nearest Neighbor is employed here for our classifier.

• *Genuineness validation*: When validating, we employ the difference of phase fingerprints to deal with the 'separation challenge'. To demonstrate the feasibility of this method, we collect phase fingerprints of 2 tags using 6 different antennas. Second, we attach $4 \sim 7$ tags on a product and employ the TagPrint to validate its genuineness. The results are expressed using Equal Error Rate (EER) as our metric for genuineness validation.

C. Accuracy of Tag Recognition

Table II shows the final classification accuracies compared with other 5 kinds of fingerprints proposed in recent work, where the size is the cardinality of fingerprint space and the N_{tag} is the number of testing tags. We directly display the classification accuracy claimed in [9]–[11], [13] as the benchmar, because we are limited by the lack of corresponding hardwares.

∂_{TIE} and $\bar{P}_B$: These two fingerprints are proposed in [10]. The authors design their purpose-built oscilloscope whose sampling rate is as high as $100MS/s \sim 1GS/s$. ∂_{TIE} is the Time Interval Error (TIE), a feature in time domain, which describes how the edge of tag's clock varies from the theoretical clock. $\bar{P}_B$ is the average power of an acquired RN16 preamble. In their evaluation, total 50 tags are tested. Consequently, they obtain mean accuracy of 71.4% and 43.2% using ∂_{TIE} and $\bar{P}_B$. They also find that their method has a better performance when classifying tags with different models, and suggest to jointly use these two fingerprints to improve the accuracy. However, their fingerprint space is limited within 2^6.

Spectral: Two types of features are extracted from spectral domain, modulation-shape and spectral PCA, in [9]. They test the fingering on a set of 50 RFID smart cards which work at HF band. Their method can obtain an extremely high CSR, 97.57%. Unfortunately, their method depends on dedicated monitor devices and works for HF band only.

GenPrint: Two features are developed in [13], the covariance based pulse feature (Cov) and the power spectrum density based signal feature (PSD), which are measured using USRP. GenPrint has a mean accuracy of 76.94%. The number of tested tag increases to 150.

MinPower: [11] employ a dedicated device called *Voyantic Tagformance Lite System* to acquire the minimum power required for a tag to respond at multiple frequencies as the tag's fingerprint. The MinPower can obtain a mean accuracy of 92.55% over 100 tags.

TagPrint: For accuracy estimation, we use the data set shown in Table I. A 5-fold cross validation is used to calculate the classification success rate. For each tag, we collect 100 acquisitions of its phase fingerprint using the conveyor-style approach. The resolved set is split into 5 independent folds. One fold (20 acquisitions) is used for training and the

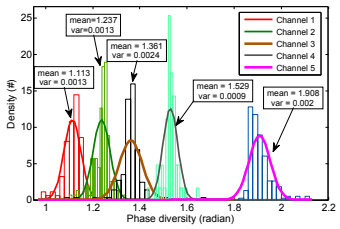


Fig. 7: Impact of frequency

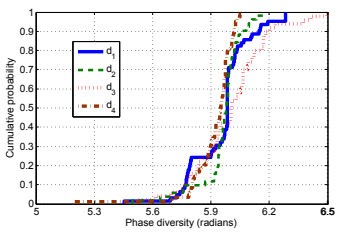


Fig. 8: Impact of distance

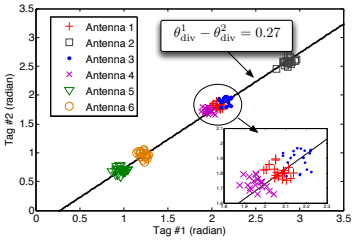


Fig. 9: Impact of antenna

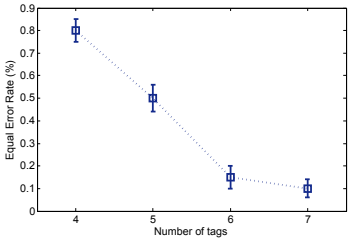


Fig. 10: Validation

remaining four folds (80 acquisitions) are used to form the testing tag fingerprints. The average CSR is recorded for each tag and the average CSR over all tags is reported. As a result, utilizing phase fingerprint for classification can achieve a very good CSR when the number of testing tag is less than $1,000$. Specifically, in our case, when $N < 50$, the mean CSR is 99.58%. Even the N increases to 150, it still achieves a mean CSR of 96.71%, which is higher than previous work. However, when $N > 5,000$, the rate falls to around 55%. This is because the the cardinality of fingerprint space equals $2^{12} = 4,096$, resulting that the fingerprints collide among $1,904$ tags at least.

Summary: None of large-scale fingerprint acquisitions and classifications are conducted before. Our experiment demonstrates that small numbers of samples may output an inaccurate statistics. The phase fingerprint exhibits a good statistic feature and capability of accurate classification on the minor scale. It improves the accuracy of tag classification over the fingerprints of ∂_{TIE}, $\bar{P}_B$, GenPrint and MinPower by $39.46\%\times$, $130\%\times$, $25.69\%\times$ and $7.5\%\times$.

D. Stability of Tag Recognition

One of the most important feature of fingerprint is the stability. This requires us to perform a benchmark for estimating the stability of phase fingerprint with respect to different configurations, including frequency, distance, orientation and transmission power.

Impact of frequency: A typical UHF reader has 16 channels working at $920 \sim 926$ MHz ISM band. To show how the frequency affects the fingerprint, we resolve the phase fingerprint through the conveyor over 5 frequency channels using a same reader. Five set of the resolved phase fingerprints are plot in Fig. 7. We can see that (1) the resolved phase fingerprint varies over the frequency, so the frequency is a leading factor affecting the fingerprint. This can be understood that the different activation frequency incurs different impedance matchings on the tag's antenna, outputting different phase rotations. This result suggests us to label the tag's fingerprint with channel number. When validating, the same channel would be employed. (2) Although the phase fingerprint obeys a typical Gaussian distribution, the variance is so small (< 0.002 radians) that it can be considered stable enough in practice.

Impact of distance: We vary the vertical distance d between the reader and tag's trajectory to observe the changes on resolved phase fingerprint. In the experiment, four distances are tested, $d_1 = 0.759m$, $d_2 = 0.727m$, $d_3 = 0.695m$ and $d_4 = 0.631m$. The results are plotted in Fig. 8 in which the mean phase fingerprints are 5.9570, 5.9591,5.9998 and 5.922 radians corresponding to the four distances. As expected, none apparent differences are found among these results. All variances are under an acceptable level. This demonstrates that the distance is not a factor affecting the fingerprints. We also study other two parameters, orientation and transmission power. The similar results are obtained as that of distance. We omit their result figures due to space limitation.

Summary: The parameters including distance, orientation and transmission power take bare impacts on tag's phase fingerprint, which exhibits good stability in practice. However, the frequency has an obvious influence on the phase fingerprint. Fortunately, the communication frequency is completely determined by the reader's configuration. We can label the fingerprint with channel number to inform the validation client to choose the correct frequency.

E. Genuineness Validation

Impact of reader diversity: We firstly conduct a serial of experiments to demonstrate whether the reader's diversity θ_{ant} can be well-eliminated through the difference of fingerprints. In the experiment, we employ 6 different antennas to acquire 2 tags' fingerprints. For each pair of antenna and tag, 20 fingerprints are acquired. The results are plotted in Fig. 9. In the figure, the x-axis and y-axis are the resolved fingerprints θ^1_{div} and θ^2_{div} for Tag#1 and Tag#2. Each point $(\theta^1_{\text{div}}, \theta^2_{\text{div}})$ corresponds to a same antenna. From the figure, we can observe that (1) for each antenna, the results are highly clustered, which validates the stability of phase fingerprint again. (2) The results of Antenna#1, Antenna#3 and Antenna#4 are very close because these three antennas comes from a same manufacturer. It is probable that the fingerprint is not random among antennas, because the antenna is only a part of reader's transceiver. (3) The mean value of 6 clusters locates on a straight line, *i.e.*, $\theta^1_{\text{div}} - \theta^2_{\text{div}} = 0.27$. This phenomenon can be explained as follows. Because $\theta^1_{\text{div}} = \theta_{\text{ant}} + \theta^1_{\text{tag}}$ and $\theta^2_{\text{div}} = \theta_{\text{ant}} + \theta^2_{\text{tag}}$, we can get that $\theta^1_{\text{div}} - \theta^2_{\text{div}} = \theta^1_{\text{tag}} - \theta^2_{\text{tag}} = 0.27$ by subtracting previous equations. The 0.27 is the difference of two tags' diversities. This demonstrates that the difference is able to eliminate the reader's diversity, so there is no issues to employ one reader for fingerprint acquisition but use another reader for validation.

Validation accuracy: Secondly, we evaluate the accuracy of genuineness validation. We estimate the EER as follows. We separate 20 validating products into two categories: genuine and faker. The genuine category includes genuine products whose federated tags are genuine. The faker category contains fake products on which one of federated tags are forged. For a given threshold, the *False Reject Rate (FRR)* is the percentage of false rejected tests, while the *False Accept Rate (FAR)* is

the percentage of false accepts in the faker category. The EER is the error rate where both FAR and FRR are equal. The value of the threshold at the EER is our threshold Γ for an accept/reject decision. The results of analysis for $m = 4 \sim 7$ tags attached on each item are shown in Fig. 10. The EER dramatically decreases with higher tag number reaching an EER $= 0.1\%$ approximately. This means that TagPrint can verify the genuineness of product with an accuracy of 99.9% (genuine accepts), while allowing 0.1% of false rejects when $m > 6$. It seems our solution needs $4 \sim 6$ more tags than the existing fingerprints based. However, the benefits are obvious: (1) no dedicate equipment required, which saves thousands of money on the acquisition device. (2) total offline validation. Product manufacture does not invest on building and maintaining validation servers; (3) higher security guarantee than other methods; (4) applying in COTS tags, without need of developing new kinds of chips. Lastly, we must claim that our solution is not replacing existing all counterfeiting methods, like barcode and laser, but to complement those. In fact, we will not increase cost that much when the product itself is not very cheap while RFID is cheap. For example, the tags cost is $1/1000$ of the price of a bottle of wine.

VII. Related Work

RFID security. Considerable conventional cryptographic based models have been proposed to prevent unauthorized reading of tag [2], [5], [6]. The security of these protocols is based on application and communication layers of the RFID tag. These protocols are too heavy to be affordable by the passive tags. Even worse, they face challenges from reverse-engineering, side-channel and relay attacks. After creation of a cloned tag with the same data, there is no mechanism to differentiate the original and the cloned. TagPrint fingerprints a passive RFID tag based on the physical characteristics that are difficult to clone, which does not depend on the resources of the RFID tag.

Fingerprinting. RFID fingerprinting refers to utilizing the physical layer information to identify digital device, which has received considerable attention in recent years [8]–[15]. [9] employed spectral and time-domain features, like the clock skew, as fingerprint and yield low error rates. [10] studied the physical layer identification of UHF tags with a population of 70 UHF RFID tags collected at varying tag-reader distances. [11] proposed to create an electronic fingerprint of a tag based upon the physical attributes of the tag. richter2008fingerprinting reported the possibility of detecting the country that issued a given passport by looking at the bytes that an e-passport sends. [15] they proposed to use the initialization state of SRAMs in RFID tags to create a physical fingerprint.

Anti-counterfeiting. Several work studied on anti-counterfeiting using physical uncloneable function (PUF) technique [20]. The PUH is physical structures but hard to predict. Such kind of tags are expensive and need to embed special chips in the tag. It also cannot be identified by the COTS readers. [2] proposed a method to authenticate a batch of RFID tags.

VIII. Conclusion

We present TagPrint for anti-counterfeiting using COTS RFID readers and tags. A key innovation is to exploit a novel hardware fingerprint for RFID tags acquired from the phase values of the backscatter signals. We wish our studies were able to promote RFID applying in anti-counterfeiting.

Acknowledgement

The research of Fan Dang is supported in part by NSF China Project No. 61472217. The research of Yunhao Liu is supported in part by the NSF China Distinguished Young Scholars Program 61125202. The research of Li is partially supported by NSF NSF ECCS-1247944, NSF CMMI 1436786, NSFC under Grant No. 61170216, No. 61228202.

References

[1] WHO, "Anti-counterfeit technologies for the protection of medicines."

[2] L. Yang, J. Han, Y. Qi, and Y. Liu, "Identification-free batch authentication for rfid tags," in *Proc. of IEEE ICNP*, 2010.

[3] Y. Zheng and M. Li, "Fast tag searching protocol for large-scale rfid systems," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 3, pp. 924–934, 2013.

[4] S. A. Weis, S. E. Sarma, R. L. Rivest, and D. W. Engels, "Security and privacy aspects of low-cost radio frequency identification systems," 2004.

[5] L. Lu, J. Han, L. Hu, Y. Liu, and L. M. Ni, "Dynamic key-updating: privacy-preserving authentication for rfid systems," in *Proc. of IEEE PerCom*, 2007.

[6] T. Dimitriou, "A secure and efficient rfid protocol that could make big brother (partially) obsolete," in *Proc. of IEEE PerCom*, 2006.

[7] T. Li, W. Luo, Z. Mo, and S. Chen, "Privacy-preserving rfid authentication based on cryptographical encoding," in *Proc. of IEEE INFOCOM*, 2012.

[8] H. Richter, W. Mostowski, and E. Poll, "Fingerprinting passports," in *Spring Conference on Security*, vol. 1, no. 1, 2008.

[9] B. Danev, T. S. Heydt-Benjamin, and S. Capkun, "Physical-layer identification of rfid devices." in *Proc. of Usenix Security Symposium*, 2009.

[10] D. Zanetti, B. Danev *et al.*, "Physical-layer identification of uhf rfid tags," in *Proc. of ACM MobiCom*, 2010.

[11] S. C. G. Periaswamy, D. R. Thompson, and J. Di, "Fingerprinting rfid tags," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 6, pp. 938–943, 2011.

[12] D. Zanetti, P. Sachs, and S. Capkun, "On the practicality of uhf rfid fingerprinting: how real is the rfid tracking problem?" in *Proc. of Springer Privacy Enhancing Technologies*, 2011, pp. 97–116.

[13] D. Ma, C. Qian, W. Li, J. Han, and J. Zhao, "Geneprint: Generic and accurate physical-layer identification for uhf rfid tags," in *Proc. of IEEE ICNP*, 2014.

[14] H. P. Romero, K. A. Remley, D. F. Williams, and C.-M. Wang, "Electromagnetic measurements for counterfeit detection of radio frequency identification cards," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 5, pp. 1383–1387, 2009.

[15] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up sram state as an identifying fingerprint and source of true random numbers," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1198–1210, 2009.

[16] "Speedway revolution reader application note: Low level user data support," in *Speedway Revolution Reader Application Note*, 2010.

[17] "Alien corp." http://www.alientechnology.com/tags.

[18] "Impinj," http://www.impinj.com/.

[19] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile tag to high precision using cots devices," in *Proc. of ACM MobiCom*, 2014.

[20] P. Tuyls and L. Batina, "Rfid-tags for anti-counterfeiting," in *Topics in Cryptology–CT-RSA 2006*. Springer, 2006, pp. 115–131.