A Novel Distributed RFID Architecture for Secure Data Communications

Kazuya Sakai*, Min-Te Sun†, Wei-Shinn Ku‡, and Ten H. Lai*
*Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio 43210
†Department of Computer Science and Information Engineering, National Central University, Taoyuan 320, Taiwan
‡Department of Computer Science and Software Engineering, Auburn University, Auburn, Alabama 36849–5347

***Abstract*—Privacy protection is the primary concern when RFID applications are deployed in our daily lives. Due to the computational power constraints of passive tags, non-encryption-based singulation protocols have been recently developed, in which wireless jamming is used. However, the existing private tag access protocols without shared secrets rely on unpractical physical layer assumptions and thus they are difficult to deploy. To tackle this issue, we first redesign the architecture of RFID system by dividing an RF reader into two different devices, an RF activator and a trusted shield device (TSD). Then, we propose a novel coding scheme, namely Random Flipping Random Jamming (RFRJ) to protect tags' content. Unlike the past work, the proposed singulation protocol utilizes only the physical layer techniques that are already implemented. Analyses and simulation results validate our distributed architecture with the RFRJ coding scheme which defends tags' privacy against various adversaries including the random guessing attacks, correlation attacks, ghost-and-leech attacks, and eavesdropping.**

***Keywords*-RFID security; privacy; bit encoding**

I. Introduction

Radio Frequency IDentificaiton (RFID) technologies enable a tremendous amount of applications, such as supply chain management [1], electric transportation payment, and warehouse operations [2]. Objects and their owners are automatically identified by an attached RF tag, which causes the privacy threat to individuals and organizations. Thus, privacy protection is the primary concern when RFID applications are deployed to our daily lives. Since passive tags are computationally weak devices, encryption-based secure singulations [3] are not practical. Instead of relying on the traditional cryptographic operations, recent works [4]–[6] employ physical layer techniques, i.e, jamming [7], to protect tags' data. With this approach, tags could be securely identified without pre-exchanged shared keys.

The issue of the existing solutions, the privacy masking [4], RBE [5], and DBE/ODBE [6], is the impractical assumption. In these solutions, all the bits transmitted by a tag are masked (jammed) under the assumption of an additive channel, where the receiver can read a bit only when two bits (the data bit and mask bit) are the same. When the two bits are different, it is assumed that the receiver is unable to recover the corrupted bit. However, this assumption is too strong since a reader should be able to detect signals from two different sources. In reality, a receiver of a data bit will decode it as either 0 or 1 without knowing the bit collision. If there is a bit collision, either the signal strength of data bits from the tag is stronger than that of jamming bits, or vise versa. In other words, depending on the location of the reader, it can only either read all the data bits or all the jamming bits. Also, masking requires the perfect synchronization between data bits and mask bits, which is difficult to achieve in practice.

In addition to this, DBE and ODBE have two drawbacks. One is encoding collision where two different source data bits could be encoded into the same codeword. This causes the singulation process to fail. The other is more important. Tags' data encoded by DBE or ODBE could eventually be cracked, should an adversary repeatedly listens to the backward channel (i.e., signals from a tag to a reader), called the correlation attacks. Moreover, none of the aforementioned solutions protect tags against ghost-and-leech attacks, i.e., impersonation of RF tags, similar to man-in-the-middle attacks.

To tackle these issues, we put forth a new RFID architecture and a novel coding scheme for privacy protection against various adversary models. The contributions of this paper are as follows:

- We redesign the system architecture of the non-encryption-based private tag access where an RF reader is divided into an RF activator and a TSD. The proposed architecture can be built by the current physical layer technologies, and thus our assumptions are much practical than those of the existing solutions.
- The proposed distributed RFID architecture physically defends tags against ghost-and-leech attacks.
- We propose a novel coding scheme, named Random Flipping and Random Jamming (RFRJ), to protect the backward channel from passive adversaries, i.e., the random guessing attacks, correlation attacks, and eavesdropping. In our scheme, a tag/TSD randomly flips/jams a bit in a codeword and keeps the index of the these bits in secret. RFRJ guarantees that the TSD can recover a tag's content with one of the secrets, but an adversary cannot obtain the content of tags.
- Since the backward channel is protected by the RFRJ coding scheme, we can protect the forward channel (i.e., signals from a reader to a tag) by having an RF activator querying based on encoded data (or pseudo

ID) space by RFRJ.
- We conduct theoretical analyses for security of the proposed scheme, and prove that RFRJ provides perfect protection against passive attacks as long as jamming is successful.
- We evaluate our RFRJ coding scheme with the existing solutions by extensive simulations, and illustrate that the new architecture and coding scheme achieve our design goals.

The rest of this paper is organized as follows. Section II provides background knowledge for this research. We design a new RFID architecture in Section III, and propose the RFRJ coding scheme in Section IV. Security analyses are provided in Section V and simulation results are demonstrated in Section VI. In Section VII, we review existing works for RFID security. Section VIII concludes this paper.

II. Preliminary

A. Physical Layer Security

Jamming is widely used for secure communications at the physical layer level, in which jamming signals corrupt receiving signals. Although this indicates that a legitimate receiver cannot decode received signals due to jamming, the full-duplex mode of wireless antennas allows the receiver to simultaneously transmit jamming signals and receive data. This can be done by canceling self-interference, in which transmitting signals interrupt receiving signals. According to [8], the current implementation can cancel self-interference up to 45 dB across 40MHz. Therefore, with jamming techniques, an eavesdropper cannot steal communications unless it is in the very proximity of a jamming source node.

It is known that perfect secrecy is possible without shared secrets by degrading the signal at an eavesdropper relative to that at the legitimate receiver [9]. Thus, jamming is a physical layer security technique suitable to wireless sensor network where encryption-based security system is not practical due to the power constraints of sensor nodes. Dialog code [7] is proposed that provides secure communications without a shared secret for wireless sensor networks. In this scheme, each source bit is encoded to a codeword, and jamming is performed during the transmission of the codeword. To achieve this, two assumptions must be held. One is that a bit level jamming is possible; the other is that an eavesdropper cannot know which bit is jammed. Their implementation with sensor motes shows that both assumptions can be held by simulating a byte as a bit.

Another application of the physical layer security with jamming is the protection of medical devices. In [10], a shield is developed to intermediate all the communications between a medical device of a patient and a reader from a doctor. A shield is capable of full-duplex communications, and protects the channel between a medical device and itself by jamming. On the other hand, the shield and the reader communicate with an encrypted channel. On detecting an unauthorized reader's access, the shield interrupts the communication by jamming all transmitted bits. The authors implemented the shield with a small portable device that looks a necklace, and thus eavesdropping is almost impossible since an adversary must be at a very close position to the shield. By doing this, the proposed architecture does not modify the medical devices in the markets.

B. Bit Level Jamming Models

Let b be a source bit, b_j be a jamming bit, and b' be the outcome of a bit b transmitted under jamming b_j. In [7], jamming channel models are categorized as follows.

- **Probabilistic Flipping Model -** no matter what value b_j has, the source bit b flips with the probability p_j, i.e., $P[b' \neq b] = p_j$.
- **AND Channel Model -** the receiver will decode $b' = 1$ when either b or b_j is 1. Otherwise, $b' = 0$.
- **XOR Channel Model -** the receiver will decode $b' = 1$ when $b \neq b_j$. Otherwise, $b' = 0$. It is known that one-time pad in this model can achieve perfect secrecy if the jamming bits are truly random in [11].
- **General Model -** in this model, $P[b'{=}0|b{=}0, b_j{=}0] + P[b'{=}0|b{=}0, b_j{=}1]{=}1$ and $P[b'{=}0|b{=}1, b_j{=}0] + P[b'{=}0|b=1, b_j=1]=1$. The probability of that $b' = 1$ is similar. This jamming model achieves perfect secrecy, since the probability of that the receiver decodes $b' = 0$ is 0.5 whenever the jamming bits are truly random [7].

C. Distributed RFID Systems

In the traditional RFID system, an RF reader has two components, a transmitter (i.e., query transmission/energizing tags) and a listener (i.e., listening to a tag's reply) as shown in Figure 1 (a), where a diamond represents the transmission function of a reader, a circle represents the listening function of a reader, and a rectangle represents a tag. The communication range of the backward channel is much shorter than that of the forward channel, and thus readers must be deployed based on the short-range backward channel to access all tags in the region as shown in Figure 2 (a). A recent study proposes Distributed RF Sensing model [12] that employs two kinds of devices, single RF transmitter and a number of RF listeners, for each function of a reader as shown in Figure 1 (b). The model contributes to cost reduction of RFID system deployment. For example, in Figure 2, the traditional RFID system requires 9 transmitters and 9 listeners, while the distributed RFID system requires 1 transmitter and 9 listeners.

III. Proposed Architecture

In this section, we propose a new RFID system architecture for a secure singulation as shown in Figure 3.

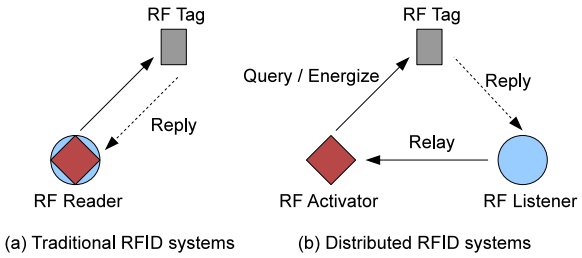

Figure 1. Distributed RFID systems.

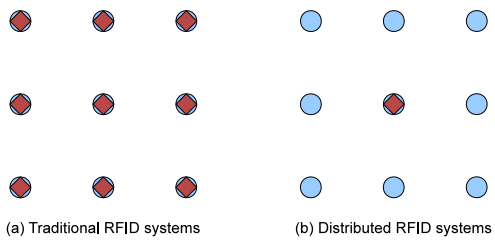

Figure 2. Distributed RFID system deployment.

A. Assumptions

We begin with listing physical layer assumptions as follows.

- A bit level jamming is feasible.
- An eavesdropper does not know if a bit is jammed.
- Probabilistic flipping model is used for a jamming environment.

As we discussed in Section II, the first and third assumptions are already implemented and validated in [7]. On the other hand, there is no implementation of the backward channel protection methods in [4]–[6]. Therefore, our assumptions are much more practical than the past research.

B. New RFID System Architecture

Similar to [12], an RF reader is divided into two components, an *RF activator* and a *Trusted Shield Device (TSD)*. In our new architecture, an RF activator queries a tag with a long-range signal (i.e., the forward channel) and energizes the tag. A TSD receives a tag's reply with a short-range signal (i.e., the backward channel), and it sends the reply to the activator via an encrypted channel which we define as *the relay channel*. In typical RFID applications, a reader forwards tags' data to the back-end server. For simplicity, in this paper we consider the RF activator as the final destination of a tag's data by assuming the activator forwards collected data to the back-end server. A TSD works as an RF listener and it is capable of bit level jamming during reception of a tag's reply. Therefore, our new RFID system architecture consists of three components, an RF activator, a TSD, and RF tags.

In this paper, we introduce a new coding scheme, namely Random Flipping Random Jamming (RFRJ), for the backward channel protection. A tag will send encoded data (i.e., pseudo IDs) to a TSD under the jamming environment. This prevents adversaries from passive attacks, i.e., the random guessing attacks, correlation attacks, and eavesdropping. As we will show later, the RFRJ coding ensures that adversaries cannot decode the original tag's ID from incomplete data due to the jamming while the TSD successfully recovers the data from imperfect information.

A TSD is conceptually similar to the trusted masking device in [5] and a medical device shield implemented in [10], but different in the following functions.

- On overhearing a query from an activator to a tag, a TSD jams a bit in a codeword. As mentioned in the assumption, a bit level jamming is possible.
- If an unauthorized reader tries to access a tag, a TSD jams against all bits of codewords so that the unauthorized reader cannot read the content of the transmitted data. A similar function is implemented in [10], where a shield device jams the whole communication on detecting unauthorized accesses. This can be done by letting an authorized activator communicate with a TSD before a singulation process.

With our new architecture, we can achieve the following design goals:

- The forward channel is protected by having an activator querying tags based on the pseudo ID space encoded by the RFRJ coding scheme.
- The RFRJ coding scheme protects the backward channel against the random guessing attacks, correlation attacks, and eavesdropping, as we will show in Section V.
- Since we assume both an activator and a TSD have computational power, the relay channel can be protected by the traditional cryptographic operations.
- The proposed architecture defends against ghost-and-leech attacks. First, an adversary cannot forward an activator's query to a tag since a TSD blocks all unauthorized accesses. Second, an adversary cannot obtain a tag's reply due to the jamming by TSD. Therefore, an adversary cannot impersonate a tag.
- The physical layer assumptions are much more practical than the existing solutions [4]–[6] as we discussed in Section III-A

IV. Random Flipping Random Jamming Coding

In this section, we present the Random Flipping Random Jamming (RFRJ) coding scheme.

A. Definition

Let r be an RF activator, s be a TSD, and t be an RF tag. An activator which intends to obtain data from a tag sends a query on the forward channel. When the tag replies to the TSD, it encodes every l_b bits in the data into a l_c bits codeword with an encoding function $E(.)$. Note that l_b is not the length of an ID, but the unit to be encoded into a codeword. A coding scheme for private tag access is defined by the parameters, l_b, l_c, and C. Here, C is a set of codewords that could be used for encoding. During

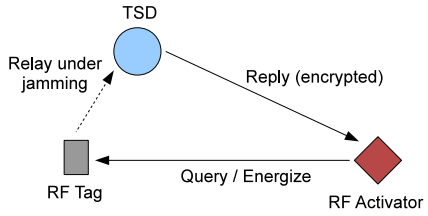

Figure 3. The proposed RFID architecture.

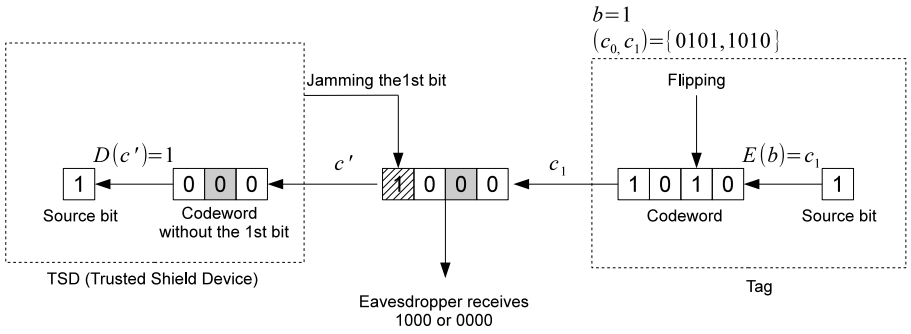

Figure 4. The system model and basic idea.

the transmission of a pseudo ID on the backward channel, the TSD conducts bit level jamming. On receiving the tag's reply, the TSD decodes the received codeword by a decoding function $D(.)$, and forwards the data to the activator via the relay channel.

In general, we call l_b-to-l_c the RFRJ coding scheme. For instance, the coding scheme with $l_b = 1$ and $l_c = 4$ is said to be the 1-to-4 RFRJ coding. The notations utilized in this paper are listed in Table I.

Table I
DEFINITION OF NOTATIONS.

Symbols	Definition
r	The RF Activator r
s	The TSD s
t	The RF tag t
b	The bit b
B	The source bits $\{b_1, b_2, ...\}$.
c	The codeword c
C	A domain of codewords $C = \{c_0, c_1, ...\}$
l_c	The length of a codeword $\|c\|$
l_b	The length of source bits $\|B\|$
I	The index of a bit in a codeword
$E(.)$	The function $E : \{0,1\}^{l_b} \rightarrow \{0,1\}^{l_c}$
$D(.)$	The function $D : \{0,1\}^{l_c} \rightarrow \{0,1\}^{l_b}$
$H(b, b')$	The Hamming distance between b and b'
$H(b, b', i)$	The Hamming distance between b and b' after removing the i-th bit of b and b'
p_j	The probability that a jammed bit is flipped

B. Private Tag Access Protocol

The proposed private tag access protocol works as follows. Suppose an RF activator r plans to read an RF tag t without disclosing the tag's ID to an eavesdropper. For simplicity, we consider the length of the encoding unit l_b to be 1 in this paper. Our idea can be applied to arbitrary values of l_b and l_c, where $l_b < l_c$. On receiving a request, the tag t extends a bit into a l_c-bit codeword, where $l_c \geq 4$ must hold. When the tag transmits data over the backward channel, it randomly selects a bit in a codeword and intentionally flips it. Note that this process is done before the tag sends out the codeword, so the data sent by the tag always contains a one-bit error. On the other hand, the TSD, which is an RF listener with jamming capability, jams a single bit in the codeword. The jamming causes the selected bit to flip. Let p_j $(0 \leq p_j \leq 1)$ be the probability that the bit jammed by the TSD is flipped. We denote I_s and I_t as the indexes of the selected bits by the TSD and the tag, respectively. The TSD randomly selects any bit in the first half of the l_c bits codeword, i.e., $1 \leq I_s \leq \lfloor \frac{1}{2} l_c \rfloor$, while a tag randomly selects a bit in the second half of the codeword, i.e., $\lfloor \frac{1}{2} l_c \rfloor + 1 \leq I_t \leq l_c$. By doing this, we can guarantee that the TSD and the tag do not select the same bit. Thus, the codeword received by the TSD or an eavesdropper contains a two-bit error when jamming flips the I_s-th bit and a one-bit error when jamming fails.

For instance, in Figure 4, a source bit is encoded into a 4-bit codeword. The tag flips the third bit in the codeword, which is colored gray, and the TSD selects the first bit for jamming, which is crossed off.

Assume the original codeword is 1010. Since the tag flips the third bit, it will send 1000 over the backward channel. Meanwhile, the TSD jams the first bit. Hence, the TSD and the eavesdropper will receive $X000$, where X could be decoded to either 0 or 1. The TSD knows I_s, and thus it knows one of the three bits may contain an error after excluding the jammed bit. However, the eavesdropper does not know either which bit the TSD jammed or which bit the tag flipped. For the eavesdropper, two of four bits may contain errors. Thus, the TSD and the eavesdropper have a different amount of information to decode the original codeword. In general, for 1-to-l_c, TSD knows that there is a one-bit error out of $(l_c - 1)$ bits while the eavesdropper knows there is a two-bit error out of l_c bits at best.

Both the TSD and the tag keep the indexes of the bits they jammed/flipped in secret. The TSD has one of the secrets, but the eavesdropper knows neither of them. Therefore, with the coding scheme in which the receiver can decode a source bit when one of $(l_c - 1)$ bits is flipped but not when two of l_c bits are flipped. Our new system architecture and our proposed private access protocol allow for an RF activator to securely collect RF tags' content without shared secrets.

C. A Single Bit RFRJ Coding

We propose the RFRJ coding with the parameter $l_b = 1$ and $l_c = 4$. Note that $l_c = 3$ does not work and $l_c = 4$ is the most efficient in terms of communication cost, which will

be shown later. Let b be a source bit and c be a codeword. The encoding function $E : \{0,1\} \rightarrow \{0,1\}^4$ is defined by $E(b) = c_0$ if $b = 0$ and $E(b) = c_1$ if $b = 1$.

The encoding function $E(.)$ must ensure that the Hamming distance between c_0 and c_1, denoted by $H(c_0, c_1)$, is four. There are 16 such (c_0, c_1) pairs that can be used for private tag access. We call them *valid 4-bit codeword pairs*.

Definition 1 (Valid 4-bit Codeword Pairs) *When $l_c = 4$, a codeword pair (c_0, c_1), corresponding to a source bit pair $(0, 1)$, is said to be valid when the Hamming distance between c_0 and c_1 is four, i.e.,*
$(0000, 1111)$, $(0001, 1110)$, $(0010, 1101)$, $(0100, 1011)$, $(1000, 0111)$, $(0011, 1100)$, $(0110, 1001)$, $(0101, 1010)$, *and* (c_1, c_0).

Let c' be the received codeword in which up to two bits could be flipped. We define the decoding function as $D : \{0,1\}^4 \rightarrow \{0,1\}$. Since a TSD knows the index of the jammed bit, the decoding function ignores the jammed bit. A tag also flips a bit which is unknown to the TSD, and the three bits contain the flipped bit after the TSD removes the jammed bit. Let $H(b, b', i)$ be the Hamming distance between b and b' after removing the i-th bit from b and b'. $D(c')$ outputs 0 when $H(c', c_0, I_s) < H(c', c_1, I_s)$ and 1 when $H(c', c_0, I_s) > H(c', c_1, I_s)$. Note that $H(c', c_0, I_s) = H(c', c_1, I_s)$ never happens.

Next, we prove that the 1-to-4 RFRJ coding successfully achieves our design goal.

Theorem 1 *When the RFRJ coding with a valid codeword pair is used, the receiver can successfully decode the source bit, but the eavesdropper cannot.*

Proof: The TSD knows the value of I_s, so it can exclude the I_s-th bit for the decoding process. Since a tag flips the I_t-th bit where $I_s \neq I_t$, one of the three bits is flipped. Hence, this problem is reduced to whether or not the TSD can recover the original codeword sent by the tag, even if one out of three bits contains an error, while the eavesdropper cannot do it if two out of four bits contain errors.

Let (c_0, c_1) be a codeword pair and c' be the codeword that the TSD and the eavesdropper receive. Since $H(c_0, c_1) = 4$, excluding the I_s-th bit $H(c', c_0, I_s)$ and $H(c', c_1, I_s)$ are three. For instance, after removing the first bit of a codeword pair $(1100, 0011)$, we have $H(100, 011) = 3$. This implies that either c_0 or c_1 must be closer to c' than the other. Thus, the TSD can always decode it.

On the contrary, the eavesdropper does not know both I_s and I_t. All valid codeword pairs have the Hamming distance of four, and the 4-bit codeword received by the eavesdropper may contain a two-bit error. This indicates that $H(c_0, c') = H(c_1, c') = 2$, and the eavesdropper cannot decode it. Therefore, the claim is true. ∎

Example: Consider a bit pair $(0, 1)$ is mapped to one of a valid codeword pair, say $(c_0, c_1) = (0101, 1010)$, as shown in Figure 4. A tag sends a bit 1 which will be encoded to 1010, and it selects the third bit to be flipped, i.e., $I_t = 3$. Along this, the TSD selects the first bit for jamming, i.e., $I_s = 1$. Hence, the TSD will receive $X000$.

Let us mark the jammed bit by X. Since a tag flips a bit in the second half of the codeword, $X000$ contains a one bit error. With the one bit error in the second half of c_0 and c_1, we will have $c_0 = \{X100, X111\}$ and $c_1 = \{X000, X011\}$

Clearly, sets of possible values of c_0 and c_1 are exclusive, and hence $H(X000, c_0, I_s) = H(X000, c_1, I_s)$ never happens. Thus, the TSD can always obtain the original codeword by taking the closer Hamming distance to $X000$. The decoding function takes c_1, and outputs 1.

On the contrary, the eavesdropper can neither derive the original codeword nor the source bit. When two of four bits have errors, i.e., 0000, the eavesdropper cannot distinguish either the second and fourth bits of 0101 or the first and third bits of 1010 are flipped.

D. The 1-to-4 RFRJ Coding

We have illustrated how the RFRJ coding encodes a single source bit to a 4-bit codeword. In general, an RF tag has data with arbitrary length or a constant length ID (e.g., 96-bit defined in EPC Class1 Gen2 [13]). In this section, we elaborate on the complete 1-to-4 RFRJ coding.

In a real RFID applications, a tag is likely to transmit the same data, such as its ID, to a TSD several times. Should an eavesdropper continuously listen, it can recover the content of the tag response by the help of the previous interrogations (the correlation attack [6]). To avoid the attack, we incorporate dependency by using different valid codeword pairs to each source bit.

Let b_k be the k-th source bit that a tag intends to encode. To encode b_k, our coding scheme employs the previous source bits, b_{k-1}, b_{k-2}, b_{k-3}, and b_{k-4}. To be specific, we use the coding table in Table II, where $b_k = 0$ if $k \leq 0$.

For example, the source bits with length four, 1010, will be encoded into four codewords with each having 4 bits, i.e., 1111 1100 1001 1110.

The decoding process is basically the same, but uses different codeword pairs for each source bit. The corresponding codeword for the b_k-th source bit is obtained by Table II. The decoding function $D(.)$ is applied to the received codeword c', computes $H(c', c_0, I_s)$ and $H(c', c_1, I_s)$, and then outputs 0 or 1.

The correctness of RFRJ is given by Lemma 2 and Theorem 3.

Lemma 2 *To successfully decode the k-th source bit, a TSD must successfully decode the $(k-1)$-th source bit.*

Proof: First, note that to decode the k-th source bit, a TSD must know the previous source bits, b_{k-1}, b_{k-2}, b_{k-3}, and

Table II
CODING RULE FOR THE 1-TO-4 RFRJ CODING.

$b_{k-4}b_{k-3}b_{k-2}b_{k-1}$	$b_k = 0$ c	$b_k = 1$ c'
0000	0000	1111
0001	0011	1100
0010	0001	1110
0011	1101	0010
0100	0101	1010
0101	1001	0110
0110	1000	0111
0111	1011	0100
1000	1111	0000
1001	1100	0011
1010	1110	0001
1011	0010	1101
1100	1010	0101
1101	0110	1001
1110	0111	1000
1111	0100	1011

b_{k-4}, which means that the receiver must have successfully decoded the $(k-1)$-th source bit.

The proof is by contradiction. Assume that the TSD does not know b_{k-1} but knows all b_{k-2}, b_{k-3}, and b_{k-4}, then TSD can decode b_k. Let $b_{k-4}b_{k-3}b_{k-2}b_{k-1}$ and $b'_{k-4}b'_{k-3}b'_{k-2}b'_{k-1}$ be two possible previous bit pairs, and the corresponding valid codeword pairs are $c = \{c_0, c_1\}$ and $c' = \{c'_0, c'_1\}$. By the assumption, $b_{k-4} = b'_{k-4}$, $b_{k-3} = b'_{k-3}$, $b_{k-2} = b'_{k-2}$, but $b_{k-1} \neq b'_{k-1}$. To decode b_k without decoding b_{k-1}, the Hamming distance between $H(c_0, c'_0)$, $H(c_0, c'_1)$, $H(c_1, c'_0)$, and $H(c_1, c'_1)$ must be more than two. However, all such codeword pairs have the Hamming distance two as shown in Table II. This indicates the TSD cannot decode when one of two bits is flipped. Therefore, the TSD cannot decode b_k without decoding b_{k-1}, which leads to a contradiction. This concludes the proof.∎

Example: Consider a TSD which successfully decodes $000X$ for $b_{k-4}b_{k-3}b_{k-2}$, but not b_{k-1}, where X could be 0 or 1. The two possible codeword pairs used to encode b_k are $(0000, 1111)$ and $(0011, 1100)$, and their corresponding source bits are 0000 and 0001, respectively. Clearly, $H(0000, 0011)$, $H(0000, 1100)$, $H(1111, 0011)$, and $H(1111, 1100)$ are all two. Thus, the TSD cannot decode the source bit b_k without decoding b_{k-1}.

Theorem 3 *A TSD can successfully decode all source bits encoded by the RFRJ coding scheme.*

Proof: The proof is by induction on k.
Induction base: For the first source bit, the TSD knows the valid codeword pair since the base $b_{k-i} = 0$ $(1 \leq i \leq 4)$ as shown in Table II. From Theorem 1, the TSD successfully decodes the first source bit.

Induction step: Assuming the TSD successfully decodes the k-th source bit, we need to show it can decode the $(k+1)$-th source bit. According to the RFRJ coding scheme, the TSD knows the previous bits for k to $k-4$, when it decodes the k-th source bit. Thus, the TSD knows the valid codeword pair for the $(k+1)$-th source bit from Table II. From Theorem 1, the receiver successfully decodes the $(k+1)$-th source bit. Therefore, the above claim is true. ∎

Theorem 4 *When $l_b = 1$, the RFRJ coding with $l_c = 4$ is the most efficient in term of communication cost.*

Proof: We can prove the above claim by showing that the encoding with $l_c = 3$ does not work. A TSD will receive a 3-bit codeword where one of which is jammed and one of which is flipped. The proof is by contradiction. Assume the RFRJ encoding with $l_c = 3$ is the most efficient in terms of communication cost, then the TSD can decode the original codeword. If the TSD was able to decode the source bit, it would be able to recover the original codeword from the two bits where one of which is flipped after removing the I_s-th bit from consideration. However, the Hamming distance between any pair of two bits is at most two, i.e., $H(00, 11)$, $H(11, 00)$, $H(01, 10)$, or $H(10, 01)$. Thus, when one of two bits is flipped, the TSD cannot recover the original codeword. This is a contradiction. The RFRJ coding with $l_c = 3$ does not work. This completes the proof. ∎

There are 8! coding tables that satisfy the property described in Lemma 2. Therefore, during initialization of an interrogation, an activator can send a query with the coding table number between $[1, 8!]$ to tags to prevent eavesdroppers to utilize the disclosed bits from codewords in the previous interrogations.

V. SECURITY ANALYSIS

In this section, we provide security analysis for the proposed coding scheme. Every source bit is assumed to be 0 or 1 with the same probability 0.5.

A. The 1-to-4 Coding Security

Let X be a random variable that represents the number of flipped bits in a codeword. The I_t-th bit selected by a tag is always flipped with the probability 1 since this is done before the data is transmitted. On the other hand, the I_s-th bit selected by a reader is flipped with the probability p_j, since the jamming does not guarantee that a target bit is flipped. In RFRJ, one or two bits in a codeword could be flipped depending on p_j. The probability that the events $X = 1$ and $X = 2$ occur is obtained by:

$$P[X = 1] = 1 - p_j \quad (1)$$

$$P[X = 2] = p_j \quad (2)$$

Since X is either 1 or 2, $P[X = 1] + P[X = 2] = 1$. In our 1-to-4 RFRJ coding scheme, an eavesdropper cannot decode when two bits are flipped. Thus, the eavesdropper cannot decode the source bit with the probability p_j. This rule is applied to only the first source bit, but not to the k-th bit for $k > 1$ because it is encoded with a dependency.

Let X_k be a random variable that represents the number of flipped bits in the codeword corresponding to the k-th source bit. Again X_k could be 1 or 2. Since a valid codeword pair used for the k-th source bit is defined by the previous source bits, an eavesdropper must decode the $(k-1)$-th source bit to successfully decode the k-th source bit. Thus, the probability that the eavesdropper can decode the k-th source bit is $P[X_k = 1|X_{k-1} = 1]$ with the base $P[X_0 = 1] = 1$. Although the selection of a valid codeword pair is dependent, $X_k = 1, 2$ and $X_{k-1} = 1, 2$ are independent events.

$$\begin{aligned} P[X_k = 1|X_{k-1} = 1] &= P[X = 1] \cdot P[X_{k-1} = 1] \\ &= P[X = 1]^k \\ &= (1 - p_j)^k \end{aligned} \tag{3}$$

Hence, an eavesdropper has a very small chance to successfully decode the k-th source bit when k is large.

B. Random Guessing Attacks

When the eavesdropper cannot decode, they may guess the source bit to be either 0 or 1 with the even probability (i.e., the random guessing attacks). In this subsection, we consider the security of our coding scheme against an eavesdropper with random guessing capability. When a bit flipping by jamming fails, the eavesdropper decodes with the probability 1. Otherwise, it can successfully decode with the probability 0.5 by random guessing. Let b' be the bit decoded by the eavesdropper. Thus, the probability that the eavesdropper successfully decodes the source bit b is given by:

$$P[b = b'] = P[X = 1] + \frac{1}{2}P[X = 2] \tag{4}$$

Let b_k and b'_k be the k-th source bit and a bit decoded by the eavesdropper, respectively. We can obtain the probability that the random guessing succeeds at the k-th source bit as follows.

$$\begin{aligned} P[b_k{=}b'_k] &= P[X_k{=}1|b_{k-1}{=}b'_{k-1}] + \frac{1}{2}P[X_k{=}2|b_{k-1}{=}b'_{k-1}] \\ &= (P[X{=}1] + \frac{1}{2}P[X{=}2]) \cdot P[b_{k-1}{=}b'_{k-1}] \\ &= (P[X{=}1] + \frac{1}{2}P[X{=}2])^k \\ &= (1 - \frac{1}{2}p_j)^k \end{aligned} \tag{5}$$

C. Anonymity Analysis

We will use the entropy based anonymity analysis that has been developed for coding schemes in [6]. Let n_b be the length of source bits (e.g., the data or tag ID length), and n_u be the number of source bits uncompromised by an eavesdropper. Then, the anonymity of the source bit is given by:

$$-\sum \frac{1}{2^{n_u}} log_2(\frac{1}{2^{n_u}}) \cdot \frac{1}{n_b} = \frac{n_u}{n_b} \tag{6}$$

The average anonymity of our 1-to-4 RFRJ coding scheme is computed from the expected number of bits that an eavesdropper will decode. Let Z be the random variable that represents the number of compromised source bits. We will have $n_u = n_b - E[z]$. From Equation 3 and 6, the average anonymity is computed by:

$$\frac{n_b - E[Z]}{n_b} = 1 - \frac{1}{n_b^2}\sum_{k=1}^{n_b} k(1 - p_j)^k \tag{7}$$

D. Analytical Results

According to [6], DBE and ODBE may generate the same pseudo ID form two different source IDs. Although such possibility is very small, pseudo ID collisions cause singulation process to fail and this is not acceptable. Contrarily, our RFRJ coding scheme does not have pseudo ID collision by Lemma 5.

Lemma 5 *When $B \neq B'$ where B and B' are two set of bits, $E(B) \neq E(B')$ always holds.*

Proof: The proof is by contradiction. Assume there exist two set of bits, B and B', such that $E(B) = E(B')$, then there must exist $E(b_k)$ and $E(b'_k)$ where $b_k \neq b'_k$ and $b_{k-i} = b'_{k-i}$ for $1 \leq i \leq 4$. But, according to Table II, this never occurs, since $B \neq B'$, there exists at least one bit pair $b_k \in B$ and $b'_k \in B'$ such that $b_k \neq b'_k$. This is a contradiction. Therefore, the claim must be true. ∎

DBE and ODBE have significantly improved the performance of the privacy masking [4] and RBE [5], especially against correlation attacks. Nevertheless, both DBE and ODBE cannot completely avoid correlation attacks. Hence, eventually the source bits is cracked. According to [6], encoded 96-bit data by ODBE with codeword length 4 and $p_j = 1$ is cracked in 800 interrogation cycles. However, our RFRJ is different. One of the important results in this paper is that the RFRJ coding perfectly protects source bits from passive attacks when $p_j = 1$. This is proved by Theorem 6.

Theorem 6 *When $p_j = 1$, RFRJ achieves the perfect security against passive eavesdropping, random guessing, and correlation attacks.*

Proof: We will prove the above claim by showing that RFRJ coding results in the theoretical upper bound of anonymity and lower bound of random guessing probability.

Eavesdropping - the probability that an eavesdropper can obtain source bits is given Equation 3. When $p_j = 1$, Equation 3 results in $(1 - p_j)^k = 0$. Thus, RFRJ provides the perfect protection against passive eavesdropping.

Random guessing attacks - when all bits in a codeword are disclosed (jamming/masking fails for a codeword), an eavesdropper with the random guessing capability can decode the corresponding source bit with the probability 1. Otherwise, the source bit is successfully guessed with probability 0.5. Hence, the lower bound of the random guessing probability is 0.5^k for k-bit data. The random guessing probability for RFRJ is provided in Equation 5. When $p_j = 1$, we will have $(1 - \frac{1}{2}p_j)^k = 0.5^k$. This validates that RFRJ achieves the lower bound of the random guessing probability.

Correlation attacks - clearly, the upper bound of anonymity is 1. The anonymity of RFRJ for n_b bits source data is obtained by Equation 7. When $p_j = 1$, $P[X = 1] = 0$ and thus $E[Z] = 0$. Hence, the anonymity is 1. This holds for any $n_b \geq 1$, and so encoded data by RFRJ is never cracked as long as $p_j = 1$. Thus, RFRJ avoid the correlation attacks.

Therefore, the claim is true. ∎

VI. Performance Evaluation

In this section, we demonstrate the performance of RFRJ with the existing secure coding schemes for RFID backward channels, including Randomized Bit Encoding (RBE), Dynamic Bit Encoding (DBE), and Optimized Dynamic Bit Encoding (ODBE).

A. Simulation Configurations

We have implemented the $1 - to - 4$ RFRJ coding along with RBE, DBE, and ODBE. For fair comparisons, the codeword length for RBE, DBE, and ODBE is set to be four, which results in the same control overhead as the $1 - to - 4$ RFRJ coding. In this simulation, data exchanged between an RF reader and RF tags are 96-bit tag IDs. Each tag encodes its ID with an encoding scheme and transmits it. 100 RF tags are deployed in the reading range of an RF activator and TSDs. The reader executes a tree-based singulation protocol against encoded IDs. The successful jamming rate p_j varies from 0.1 to 1.0. For correlation attacks, a tag sends its ID under the RFRJ access protocol (or the privacy masking environment for RBE, DBE, and ODBE), and an eavesdropper keeps the scratches of disclosed data from previous interrogations. The number of interrogations for correlation attacks is set to be 1000. For each configuration, 1000 simulations were conducted.

B. Simulation Results

Figure 5 shows the average anonymity of a pseudo ID by different encoding schemes with respect to the successful jamming rate p_j. All encoding schemes except RBE achieve very high anonymity. This implies that RFRJ has a strong protection against eavesdropping. In addition, we would like to emphasize that the physical layer assumptions used in our model are weaker than these in the privacy masking environment.

Figure 6 illustrates the random guessing probability with respect to the successful jamming rate. Although RFRJ has a slightly higher random guessing probability than DBE and ODBE even when p_j is smaller than 0.7, it already provides a very strong protection. To be specific, when $p_j = 0.5$, the random guessing probability of RFRJ is 10^{-28}. It is clear that a random guessing eavesdropper has a very small probability of decoding the source bits.

Figure 7 demonstrates the time required to crack all source bits by the correlation attacks with respect to the successful jamming rate. It is known that data encoded by RBE, DBE, or ODBE is eventually cracked due to their design fault. Contrarily, our RFRJ perfectly protects tags' IDs from the correlation attacks when $p_j = 1$. Note that the figure plots the results for p_j up to 0.95.

Figures 8, 9, and 10 present the average anonymity of a pseudo ID by different encoding schemes with respect to the interrogation cycles for the successful jamming rate, 1.0, 0.9, and 0.8, respectively. For $p_j = 1.0$ (Figure 8), RFRJ always has the maximum anonymity 1.0 because its design completely avoids the correlation attacks. This is one of the significant results of RFRJ. When $p_j = 0.9$, RFRJ achieves a similar anonymity to that of DBE and ODBE, and a much higher anonymity than that of RBE. When $p_j = 0.8$, RFRJ results in a slightly lower anonymity than that of DBE and ODBE. However, the difference is not significant.

VII. Related Work

A. RFID security

In RFID systems, an RF reader must identify individual tags in its proximity by query and response. To effectively read a large number of tags, anti-collision mechanism is critical to the performance of tag singulation protocols. In general, existing tag singulation protocols are classified into two categories, Aloha-based [13] and tree-walking-based [14]. Although these singulaiton protocols successfully identify every tag in a reader's vicinity, both of them does not provide privacy protection for the communications between readers and tags.

While it is desirable that the traditional symmetric and public/private key operations to be used for private tag singulations, such an approach is not practical due to computational power constraint of passive tags. This enforces a number of encryption-based access protocols to use low-cost cryptographic operations [3], such as XOR, concatenations, hash functions, and so on. Although a reader successfully reads a tag without disclosing data to eavesdroppers, encryption-based singulation techniques require large amount of overhead, including key exchanges/distributions [15] and structured key managements [16]. Therefore, in this paper, we focus on private tag authentication without a shared key.

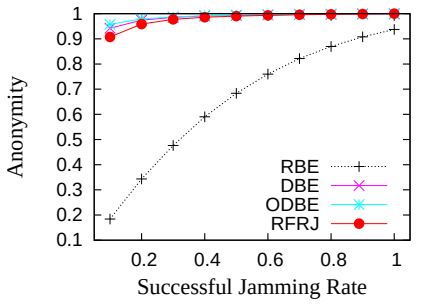

Figure 5. Anonymity.

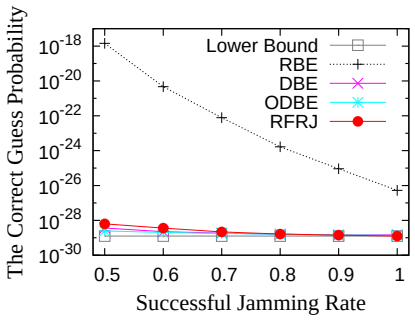

Figure 6. Correct guess probability.

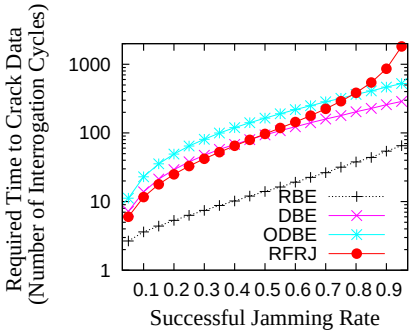

Figure 7. Time to crack tag data.

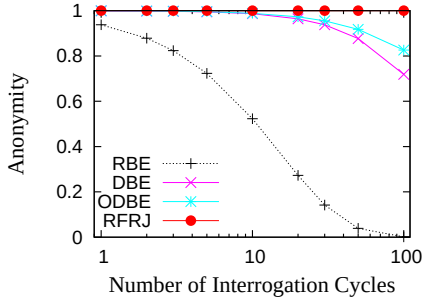

Figure 8. Correlation attacks $p_j = 1.0$.

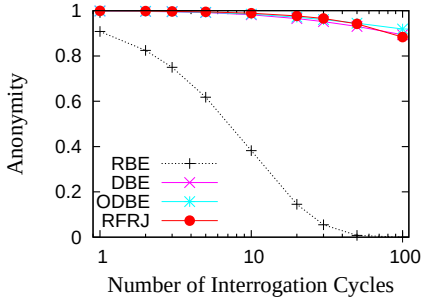

Figure 9. Correlation attacks $p_j = 0.9$.

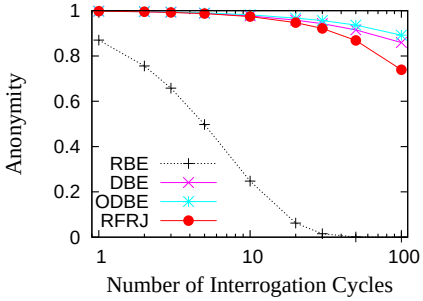

Figure 10. Correlation attacks $p_j = 0.8$.

B. Forward Channel Protection

In tree-walking-based protocols, each node is mapped to a leaf node of a binary tree consisted of the entire ID space, and a reader travels the tree in depth-first or breadth-first order by querying a prefix corresponding to an internal node in the tree. Thus, by eavesdropping the query, an adversary may obtain the tag's ID, and at least tag's ID is partially disclosed. To protect forward channel, the blinded tree-walking [17] and the randomized tree-walking [18] are proposed. In the blinded tree-walking, instead of querying with a prefix that could be the entire ID in worst case, a reader sends a next ID bit to avoid all bits in an ID to be sent. In the randomized tree-walking, each tag maintains two IDs, a read tag ID and a pseudo ID generated by manufacturers or by the tag itself. A reader traverses the tree with prefix of pseudo ID and tags reply with their real ID. These techniques protect the forward channel, but not the backward channel.

C. Backward Channel Protection

The most related studies to this paper is secure tree-walking-based singulations. Since tags can perform only simple functions, the protection of tag's reply is much more difficult than the forward channel protection. To protect the backward channel without shared secrets, the physical layer security techniques are incorporated to the private tag access [4]–[6]. In the privacy masking [4], tag's reply is intentionally corrupted by mask bits, (i.e., jamming under the additive channel). However, should the data sent by a tag and the mask bits are exactly the same, an adversary successfully eavesdrops the tag's content, called *the same bits problem*. Randomized Bit Encoding (RBE) [5] alleviates the same bits problem by encoding by source bit to a codeword with longer length. Nevertheless, RBE is vulnerable to *the correlation attacks*, where an adversary listens tag's reply over several interrogations and recovers the source bits from scratches. To tackle this issue, Dynamic Bit Encoding (DBE) and Optimized DBE (ODBE) [6] utilize the dependency among the source bits during their encoding process, and the information obtain in the previous interrogation is meaningless for the current interrogation. Note that RBE, DBE, and ODBE are used under the privacy masking, and a reader composes a binary tree with pseudo IDs generated by these encoding schemes.

D. Ghost-and-Leech Attacks

Forward/backward channel protection techniques defend tag's ID from passive adversaries, but not active adversaries. Ghost-and-leech attacks [19] is one of active attacks, in which an adversary impersonates a tag by forwarding a reader's query to the tag and the tag's reply to the reader. This attack is similar to the man-in-the-middle attacks in the study of cryptography. In [19], the author proposed Secret Handshake where the user of a tag owner defines a motion signature, e.g., motion of a circle, a triangle, an alpha, etc., and unlocks the tag before a reader accesses it. However, this solution works for only the applications in which a tag is used for owner's identification, such as ID cards, since the motion signature must be defined for individual tags. Hence, this approach cannot be applied to RFID systems where tags are attached to products, e.g., supermarkets, library, supply chains, and more.

VIII. Conclusion

RFID systems serve as an enabling technology for Internet of Things. However, security concerns of existing RFID systems have become a major obstacle for their wide adoption. The RFID protection mechanisms in the literature either work for only a few specific attacks or have unrealistic physical layer assumptions. In this paper, we first propose a novel distributed RFID architecture which divides the RF reader into two parts: an RF activator and a TSD, each tailoring for a specific function of an RF reader. In addition, we propose the RFRJ coding scheme, which when incorporating with the new architecture, works against a wide range of adversaries including the random guessing attacks, correlation attacks, ghost-and-leech attacks, and eavesdropping. The physical layer assumptions of the proposed RFID architecture and the encoding scheme are readily available. In addition, the hardware cost of the new architecture is theoretically cheaper than the existing RFID systems. We believe the proposed architecture will serve as the foundation of the next-generation RFID systems.

References

[1] Y. Li and X. Ding, "Protecting RFID Communications in Supply Chains," in *ASIACCS*, 2007, pp. 234–241.

[2] H. K. H. Chow, K. L. Choy, W. B. Lee, and K. C. Lau, "Design of a RFID Case-based Resource Management System for Warehouse Operations," *Expert Syst. Appl.*, vol. 30, no. 4, pp. 561–576, 2006.

[3] A. Juels, "RFID Security and Privacy: A Research Survey," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, pp. 381–394, 2006.

[4] W. Choi, M. Yoon, and B. hee Roh, "Backward Channel Protection Based on Randomized Tree-Walking Algorithm and Its Analysis for Securing RFID Tag Information and Privacy," *IEICE Transactions*, vol. 91-B, no. 1, pp. 172–182, 2008.

[5] T.-L. Lim, T. Li, and S.-L. Yeo, "Randomized Bit Encoding for Stronger Backward Channel Protection in RFID Systems," in *PerCom*, 2008, pp. 40–49.

[6] K. Sakai, W.-S. Ku, R. Zimmermann, and M.-T. Sun, "Dynamic Bit Encoding for Privacy Protection Against Correlation Attacks in RFID Backward Channel," *IEEE Transactions on Computers*, 2011, in Press.

[7] L. Sang, "Designing Physical Primitives for Secure Communication in Wireless Sensor Networks," *Ph.D. Dissertation, The Ohio State University*, 2010.

[8] M. Jain, J. L. Choi, T. M. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, Real-time, Full Duplex Wireless," in *MOBICOM*, 2011.

[9] A. D. Wyner, "The Wire-Tap Channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[10] S. Gollakota, H. Hassanieh, B. Ransford, D. Katabi, and K. Fu, "They Can Hear Your Heartbeats: Non-Invasive Security for Implantable Medical Devices," in *SIGCOMM*, 2011, pp. 2–13.

[11] H. Delfs and H. Knebl, *Introduction to cryptography: Principles and applications. Springer, 2nd Edition*, 2007.

[12] D. D. Donno, F. Ricciato, L. Catarinucci, A. Coluccia, and L. Tarricone, "Challenge: Towards Distributed RFID Sensing with Software-Defined Radio," in *MOBICOM*, 2010, pp. 97–104.

[13] EPCglobal, "EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860MHz-960MHz version 1.0.9." [Online]. Available: http:www.epcglobalinc.org/standards

[14] J. Myung, W. Lee, J. Srivastava, and T. K. Shih, "Tag-Splitting: Adaptive Collision Arbitration Protocols for RFID Tag Identification," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 6, pp. 763–775, 2007.

[15] A. Juels, R. Pappu, and B. Parno, "Unidirectional Key Distribution Across Time and Space with Applications to RFID Security," in *USENIX Security Symposium*, 2008, pp. 75–90.

[16] M. E. Hoque, F. Rahman, and S. I. Ahamed, "AnonPri: An Efficient Anonymous Private Authentication Protocol," in *PerCom*, 2011, pp. 102–110.

[17] S. A. Weis, S. E. Sarma, R. L. Rivest, and D. W. Engels, "Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems," in *SPC*, 2003, pp. 201–212.

[18] S. A. Weis, *Security and Privacy in Radio-Frequency Identification Devices*. Masters Thesis, MIT, 2005.

[19] A. Czeskis, K. Koscher, J. R. Smith, and T. Kohno, "RFIDs and Secret Handshakes: Defending against Ghost-and-Leech Attacks and Unauthorized Reads with Context-Aware Communications," in *CCS*, 2008, pp. 479–490.