PHY-Layer Authentication by Introducing Controlled Inter Symbol Interference

Vireshwar Kumar*, Jung-Min "Jerry" Park*, T. Charles Clancy*, Kaigui Bian†
*Bradley Department of Electrical and Computer Engineering, Virginia Tech
†School of Electronics Engineering and Computer Science, Peking University

***Abstract*—Spectrum security and enforcement is one of the major challenges that need to be addressed before spectrum-agile and opportunistic spectrum access technologies can be deployed. Rogue transmitters are a major threat to opportunistic spectrum access. One approach for deterring rogue transmissions is to enable receivers to authenticate or uniquely identify secondary transmitters. Although cryptographic mechanisms at the higher layers have been widely used to authenticate transmitters, the ability to authenticate transmitters at the physical (PHY) layer has a number of key advantages over higher-layer approaches. In existing schemes, the authentication signal is added to the message signal in such a way that the authentication signal appears as noise to the message signal and vice versa. Hence, existing schemes are constrained by a fundamental tradeoff between the message signal's signal-to-noise ratio (SNR) and the authentication signal's SNR. In this paper, we propose a novel PHY-layer authentication scheme called *Precoded Duobinary Signaling for Authentication* (P-DSA). P-DSA introduces some controlled amount of inter-symbol interference (ISI) into the data stream. The addition of the controlled ISI introduces redundancy in the message signal which can be utilized to embed the authentication signal. In this way, P-DSA relaxes the constraint on the aforementioned tradeoff. Our results show that P-DSA achieves superior detection performance compared to the prior art without sacrificing message throughput or increasing power.**

I. Introduction

It is widely believed that a transition from the legacy "command-and-control" spectrum regulatory model (where spectrum is parceled and allocated to specific stakeholders and applications) to a more flexible model of *shared dynamic spectrum access* is necessary to achieve more efficient spectrum usage. In this model, primary (a.k.a incumbent) users and secondary users operate in the same bands. One of the critical challenges that needs to be addressed to realize the spectrum sharing model is the development of technologies for *spectrum security and enforcement* [1]. Security is an especially critical consideration considering the recent calls in the U.S.A. for sharing of federal government spectrum, including military spectrum, with non-government systems. One of the specific challenges in addressing this problem is the ability to uniquely identify rogue transmitters with high confidence. Here, rogue transmitters denote transmitters that violate prescribed spectrum access rules.

While cryptographic mechanisms at the higher layers have been widely used to authenticate transmitters, the ability to authenticate and/or uniquely identify transmitters at the PHY-layer has a number of key advantages over higher-layer approaches. A PHY-layer scheme enables a receiver to quickly distinguish between legitimate and rogue transmitters without having to complete higher-layer processing, which is unnecessary and wasteful since we do not need to authenticate the data contained in the messages, but instead authenticate the transmitter or its waveform. PHY-layer authentication is especially useful in heterogeneous coexistence environments, where incompatible systems (i.e., systems with different protocol stacks) may not be able to decode each others' higher-layer signaling—e.g., IEEE 802.22 and 802.11af systems coexisting in TV white space.

In most of the existing schemes [2]–[4], the authentication signal is added to the message signal in such a way that the authentication signal appears as noise to the message signal and vice versa—we refer to this approach as the "*blind signal superposition*" method [5]. In such an approach, the authentication signal is fully present when the message signal is decoded, thus resulting in decreased signal-to-noise ratio (SNR) for the message signal, assuming average transmission power has not been increased to embed the authentication signal. Hence, there is a fundamental tradeoff between the message signal's SNR and the authentication signal's SNR—it is impossible to improve the former without worsening the latter and vice versa. This means that the degradation in the message signal's SNR is significant when the authentication signal's SNR is increased to a level sufficient for authenticating the received signal at the receiver [6].

To overcome this trade-off, we propose a novel scheme that employs pulse shaping to authenticate signals. Our approach exploits the inherent redundancy in pulse shaping to embed the authentication signal into the message signal. Specifically, our approach uses the redundancy in *Precoded Duobinary Signaling* (P-DS). P-DS is a waveform shaping technique that has been traditionally used to increase bandwidth efficiency [7], [8]. In P-DS, a controlled amount of inter symbol interference (ISI) is introduced in the transmitted pulses, and the detection procedure at the receiver cancels out the ISI. The proposed approach, called *Precoded Duobinary Signaling for Authentication* (P-DSA), uses the controlled ISI to embed the authentication bits. The main contributions of this paper are summarized below.

- We propose a novel PHY-layer authentication scheme, and show that it achieves performance advantage, compared to prior art, without sacrificing message throughput or increasing power.
- We discuss a set of fundamental criteria to evaluate the proposed scheme. These criteria are applicable to the vast majority of the existing PHY-layer authentication schemes, and can be used to compare them both qualitatively and quantitatively.

978-1-4799-0895-0/13/$31.00 ©2013 IEEE

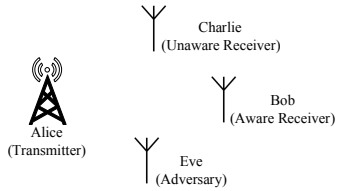


Fig. 1: Assumed authentication scenario.

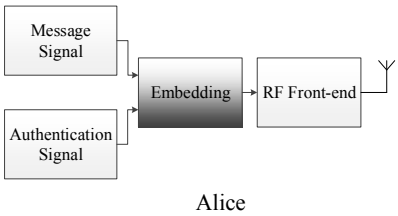


Fig. 2: Transmitter model.

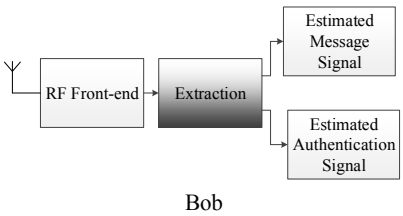


Fig. 3: Receiver model.

The rest of the paper is organized as follows. Section II describes the problem. We describe our scheme in Sections III. We establish the performance criteria of a PHY-layer authentication scheme and evaluate our scheme through comparison with the prior art in Section IV. Section V provides the related work and Section VI concludes the paper.

II. Problem Statement

A. Overview of the Problem

In the spectrum sharing paradigm, a heterogeneous mix of cognitive radio devices/networks, with secondary access priority, opportunistically access the same band while avoiding interference to the primary users and minimizing interference to each other. In this scenario, malicious secondary users that violate spectrum access rules pose a serious threat. Malicious users that effectively hijack spectrum resources or disturb peaceful coexistence need to be identified and thwarted. The first step in thwarting rogue transmitters is enabling regulators to uniquely identify or authenticate them. This can be achieved by requiring all secondary user radios to incorporate a mechanism for authenticating their waveforms and employ tamper resistance mechanisms to prevent the circumvention of the authentication mechanism by hacking. In this approach, PHY-layer authentication is ideal because it enables a receiver to quickly distinguish between legitimate and rogue transmitters without having to complete higher-layer processing, which is unnecessary and wasteful. Note that the objective of PHY-layer authentication is to uniquely identify the transmitter that has transmitted a given waveform by authenticating the waveform itself, which is different from authenticating the message carried by the waveform. The latter is handled at the application layer.

B. Network Model

We assume a scenario model illustrated in Figure 1 [2]. In this model, Alice, Bob, Charlie, and Eve are four secondary user systems which share the same wireless medium. Alice is a transmitter and intends to transmit messages to Bob and Charlie via the wireless medium. Suppose Alice and Bob have agreed on a keyed authentication scheme (implemented at the PHY layer) that allows Bob (a.k.a. "aware receiver") to authenticate the waveforms he receives from Alice. To enable authentication, Alice embeds an authentication signal into the message signal. In this model, Bob represents a regulator that needs to ensure compliance with spectrum rules. Bob also represents a regular receiver that intends to authenticate Alice's message signal. Charlie (a.k.a. "unaware receiver") does not know the authentication scheme and cannot authenticate Alice's waveforms at the PHY-layer, but should be able to demodulate and decode the message signal that can be authenticated at upper layers. Note that it is important for a PHY-layer authentication scheme to enable Charlie to recover the message signal even though he has no knowledge of the authentication scheme. Eve, the adversary, has knowledge of the authentication scheme but does not know the key, and hence cannot forge Alice's authentication signal.

In this paper, we use MS_a and AS_a to denote the message signal and the authentication signal generated by Alice in the baseband, respectively. We use MS_b and AS_b to denote the message signal and authentication signal estimated by Bob, respectively. We use MS_c to denote the message signal estimated by Charlie. Further, we use the term "*embedded signal*" to denote a message signal that has been embedded with an authentication signal. Also, we use bit error rate (BER) and signal-to-noise ratio (SNR) as the detection performance metric.

C. PHY-Layer Authentication

The operation of PHY-layer authentication can be decomposed into two processes [4]: (1) generation of the authentication information and (2) transmission and reception of the embedded signal.

1) Authentication Information Generation: Suppose that the time duration of Alice's transmission is divided into time windows of length, T, where the ith time window is represented by $[t_{i-1}, t_i]$. In order to enable the authentication of each time window, Alice generates the following one-way hash chain using a cryptographic hashing algorithm (e.g., SHA-3):

$$h_n \xrightarrow[HASH]{} h_{n-1} \xrightarrow[HASH]{} \cdots\cdots h_1 \xrightarrow[HASH]{} h_0, \qquad (1)$$

where each h_i is represented by K_h bits. Each h_i, for $i = 1, \cdots n$, is considered valid only in i^{th} time window. Moreover, each h_i of K_h bits is channel-coded into K_c bits with an error-detecting code (e.g., CRC). For each time-window, K_s synchronization bits and K_g guard bits are appended to the K_c bits to create the authentication signal, AS_a, of $K = K_c + K_s + K_g$ bits. If the rate at which authentication bits are transmitted by Alice is R_a bits/s, the length of a time window can be calculated as $T = t_i - t_{i-1} = K/R_a$. The end value of the hash chain, h_0, should be published widely so that all the aware receivers have knowledge of it. For example, it can be published using a database maintained by a regulatory entity.

After receiving the bits in the ith time-window, Bob extracts h_i from the estimated authentication signal, AS_b, by removing the synchronization and guard bits, and then carrying out channel decoding for error detection and correction. Bob verifies the authenticity of h_i by computing the hash of h_i,

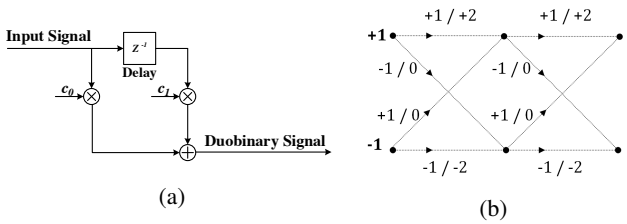


Fig. 4: (a) Duobinary filter, and (b) Trellis used by MLSD in P-DS.

and then comparing the result with h_{i-1}, which was obtained in the $(i-1)^{th}$ time window. If the two values match, AS_b is considered to be valid; otherwise, AS_b is invalid.

2) Transmission and Reception: Alice embeds AS_a into MS_a to generate the transmitted signal (see Figure 2) in such a manner that Bob can extract MS_b as well as AS_b from the received signal (see Figure 3). The main objective of this paper is to put forth the authentication scheme that can achieve *embedding* at the transmitter and *extraction* at the receiver. The details of the embedding and extraction procedures are elaborated in Section III.

III. PHY-Layer Authentication by Adding ISI: P-DSA

A. Background: Precoded Duobinary Signaling (P-DS)

The core idea of P-DS is to introduce a controlled amount of ISI in the transmitted pulses and change the detection procedure at the receiver to cancel out the ISI, and thereby achieve the ideal symbol-rate packing of 2 symbols/s/Hz without using infinitely sharp filters [9]. Let $\{d_n\},\ n = 1, 2 \cdots, N$, be a sequence of bits representing the binary message signal that needs to be transmitted, where N represents the size of a block of the message signal. Using the non-return-to-zero (NRZ) encoding, a bipolar sequence, $\{w_n\}$, is generated from $\{d_n\}$. Further, the duobinary signal, y_n, is generated by adding the weighted and delayed pulse of w_n to itself. It is achieved by using the digital filter (see Figure 4a) with coefficients $c_l,\ l = 0, 1$. Hence, the duobinary signal is represented by

$$y_n = c_0 \cdot\ w_n + c_1 \cdot\ w_{n-1}. \tag{2}$$

If we set $c_0 = 1$ and $c_1 = 1$, we obtain

$$y_n = w_n + w_{n-1}. \tag{3}$$

This equation signifies that the duobinary filter adds to a given discrete bipolar symbol the value of the immediately previous symbol. If $w_n = \pm 1$, this filter results in a three-level output—i.e., y_n has one of three possible values: $+2$, 0 or -2. This three-level output is used to express one of the two binary values of y_n, and hence there is an inherent redundancy in this encoding that we exploit to embed the authentication signal.

However, the encoded signal level can be 0 for two cases—when $w_{n-1} = +1$ is followed by $w_n = -1$ and when $w_{n-1} = -1$ is followed by $w_n = +1$. Therefore, if the receiver decodes w_{n-1} incorrectly, it affects the decoding of y_n and consequently, the detection of w_n is also likely to be in error. This error propagation can be avoided by *precoding* the message sequence at the transmitter. Therefore, before applying the duobinary filter, the message sequence is precoded to produce a new sequence called the precoded sequence. Therefore, we call this signaling as *Precoded Duobinary Signaling* (P-DS).

The precoded sequence, $\{p_n\}$, for the binary message sequence, $\{d_n\}$, is generated using the relation $p_n = d_n \oplus p_{n-1}$, where $\oplus$ represents modulo-2 addition. Further, the bipolar sequence, $\{w_n\}$, is generated from the precoded sequence, $\{p_n\}$, using NRZ encoding. $\{y_n\}$ is generated using (3) and transmitted after RF processing. We note that the first precoded bit is generated as $p_1 = d_1 \oplus p_0$. Also, we observe that the encoded signal, y_1, corresponding to w_1 is given by $y_1 = w_1 + w_0$, where w_0 and w_1 are the bipolar signal levels for p_0 and p_1, respectively. Hence, we require an extra bipolar signal, w_0 and a corresponding bit, p_0, to start the encoding of the message signal, $\{d_n\},\ n = 1, 2 \cdots, N$. Bit p_0 is called an *initialization bit* which is usually given the value of 0. Correspondingly, state w_0 is called an *initialization state* which is usually given the value of -1.

At the receiver, the received signal is estimated as the sequence, $\{\hat{y}_n\}$ in the baseband. Two decoding methods can be utilized—symbol-by-symbol detection (SSD) and maximum likelihood sequence detection (MLSD). Using SSD method, the estimated message, $\{\hat{d}_n\}$, is obtained using the following decoding decision rule.

$$\hat{d}_n = \begin{cases} 0, & \text{if } \hat{y}_n = +2 \text{ or } -2; \\ 1, & \text{if } \hat{y}_n = 0. \end{cases} \tag{4}$$

The BER of the message signal decoded using SSD [9] is given by

$$P_{SS} = \frac{3}{4} \cdot erfc\left(\frac{\pi}{4}\sqrt{\frac{E_b}{N_0}}\right), \tag{5}$$

where $erfc$, E_b and N_0 represent the complementary error function, the average bit energy, and noise power spectral density, respectively.

However, the three-level duobinary signaling incurs an increase in the number of constellation points in Euclidean space compared to binary signaling, which implies that duobinary signaling's performance against noise is inferior to that of binary signaling. However, the P-DS encoded sequence is generated from a bipolar sequence and has memory of length 1—i.e., the current state is related only to the previous state. Hence, we can use the MLSD (based on Viterbi trellis decoding) with two states (i.e., $+1$ and -1) to obtain an estimate of the transmitted bipolar sequence, $\{\hat{w}_n\}$. Figure 4b shows the trellis used by the MLSD, and it is generated by considering all possible transitions from each of the states. For example, an arrow from state $+1$ with the label $+1/+2$ represents a transition to the next state indicated by the left number, $+1$. The right number, $+2$, denotes the resultant signal level.

The received bipolar sequence, $\{\hat{w}_n\}$, is estimated from $\{\hat{y}_n\}$ using MLSD. Further, the estimated precoded sequence, $\{\hat{p}_n\}$, is generated from $\{\hat{w}_n\}$ using NRZ decoding. Finally, to obtain the estimated message sequence, $\{\hat{d}_n\}$, the decoding of the estimated precoded sequence is carried out as $\hat{d}_n = \hat{p}_n \oplus \hat{p}_{n-1}$, where $\oplus$ represents modulo-2 addition. The BER of the message signal using MLSD [10] is upper bounded by

TABLE I: An example illustrating P-DS encoding.

d_n		0	1	0	1	1	0
p_n	0	0	1	1	0	1	1
w_n	-1	-1	$+1$	$+1$	-1	$+1$	$+1$
y_n		-2	0	$+2$	0	0	$+2$

TABLE II: An example illustrating SSD in P-DS.

$\hat{y}_n$	-2	0	$+2$	0	0	$+2$
$\hat{d}_n$	0	1	0	1	1	0

TABLE III: An example illustrating P-DSA encoding (the underlined bits are the authentication bits to be embedded).

d_n		0	1	0		0	1	0
p_n	$\underline{0}$	0	1	1	$\underline{1}$	1	0	0
w_n	-1	-1	$+1$	$+1$	$+1$	$+1$	-1	-1
y_n		-2	0	$+2$		$+2$	0	-2

(a) (b)

Fig. 5: (a) MLSD for P-DS, and (b) Modified MLSD for P-DSA (The bold lines represent the possible paths emanating from the initialization state).

$$P_{ML} = erfc\left(\sqrt{\frac{E_b}{N_0}}\right). \tag{6}$$

Table I provides an example illustrating the results of performing P-DS coding of the message, $\{010110\}$ with the initialization bit, $p_0 = 0$. Table II illustrate the SSD of the message sequence encoded in Table I.

Next, we describe our scheme called *P-DS for Authentication* (P-DSA). In P-DS, a known initialization bit is needed to start the encoding of the message signal. However, we note that this initialization bit can be varied while encoding, with minimal effect on the performance of the message signal's decoding procedure. The core idea of P-DSA is to generate the embedded signal for each block of the message signal in such a way that the initialization bit is varied based on the authentication signal, i.e., P-DSA uses this initialization bit as an authentication bit.

B. Embedding of AS_a into MS_a

We assume that MS_a contains blocks of binary sequences of length N represented by $\{d_n\},\ n = 1, 2 \cdots, N$ and AS_a is a binary sequence of length K generated using the scheme described in II-C1 and represented by $\{a_k\},\ k = 1, 2 \cdots, K$. The encoding procedure of P-DSA is the same as the one for P-DS except that the precoding of each message sequence block is initiated using an authentication bit to be embedded. For each block of MS_a, $\{d_n\}$, we generate the precoded sequence $\{p_n\}$. Next, we generate the bipolar sequence $\{w_n\}$ from $\{p_n\}$ using NRZ encoding. Finally, the encoded sequence $\{y_n\}$ is generated from $\{w_n\}$ using (3).

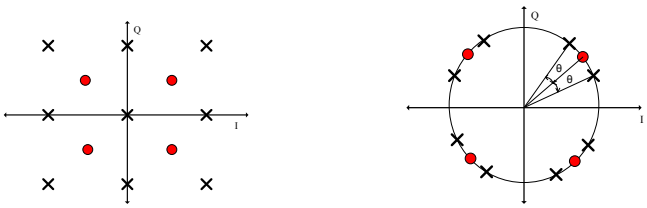
As noted earlier, an initialization bit, p_0, is required to initiate the precoding of $\{d_n\}$ in each block. In P-DS, it is achieved by choosing a standard value for p_0. For different blocks, the same p_0 and hence the same w_0 is repeatedly used to initiate the encoding. The core idea of P-DSA is to replace the bit, p_0, in each block of MS_a with a bit from AS_a, a_k. Hence, the bipolar signal, w_0, for the kth block is generated from the authentication bit, a_k, using NRZ encoding. In the kth block of the embedded signal, the first precoded bit, p_1, is generated by using the first message bit, d_1 and an authentication bit, a_k. As a result, for $a_k = 0$, the resultant precoded bit, p_1, is 0 and 1 for $d_1 = 0$ and $d_1 = 1$, respectively. Similarly, for $a_k = 1$, the resultant precoded bit, p_1, is 1 and 0 for $d_1 = 0$ and $d_1 = 1$, respectively. Table III illustrates an example of P-DSA encoding.

C. Extraction of MS_b and AS_b

In P-DSA, we generate the embedded signal by changing the encoding procedure and change decoding procedure accordingly to extract the message and the authentication signals. Note that P-DSA modifies neither the symbol mapping nor the correlation among the symbols being transmitted. Hence, the SSD is not affected by this change in encoding, while MLSD needs only a slight modification as described below.

In P-DS, at the transmitter, the precoding of each block of N bits of the message signal is started with the predecided initialization bit. At the receiver, MLSD starts with the initialization state which is generated from the same initialization bit. The MLSD decides on the sequence of states that is closest to the received signal in terms of Euclidean distance over the whole trellis. The complexity of this problem is significantly reduced by using the Viterbi algorithm which makes a decision on the possible paths reaching each possible state independent of other states [11]. In this case, the MLSD starts with the two paths from the initialization state to the *possible first states* corresponding to the first received symbol as shown in Figure 5a. Recall that trellis decoding makes a decision on the path reaching a particular state only if there are two or more paths reaching it. Hence, in P-DS, no decision is needed to select the path on each of the possible first states from the initialization state.

In P-DSA, the initialization bit is an authentication bit, and hence it also has to be estimated by the MLSD in order to decode the sequence. Hence, we need to account for the paths emanating from both the possible initialization states as shown in Figure 5b. Out of the two possible paths reaching each of the possible first states, we find the one that pertains to the closest first received symbol. In effect, the receiver performs SSD to determine the first symbol—i.e., it selects the closest signal level among $+2$, 0 and -2, and uses this knowledge to estimate the path from the initialization state to the state corresponding to the first symbol. Note that the signal level of $+2$ (-2) can be detected for the first zero-valued message bit if the authentication bit's state is $+1$ (-1). With the first received signal level as 0, if the first message bit's state is $+1$, the authentication bit's state has to be -1 and vice versa.



(a) QPSK with P-DSA. (b) QPSK with ATM.

Fig. 6: Constellation (red circles represent the message signal and black crosses represent the embedded signal).

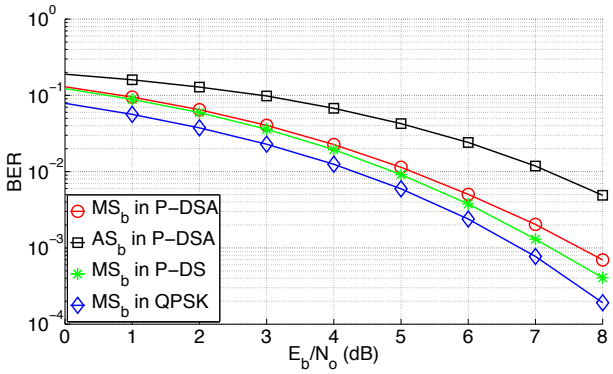


Fig. 7: BER performance of P-DSA.

IV. Performance Evaluation of P-DSA

We define the fundamental criteria that characterize the performance of PHY-layer authentication schemes and analyze the performance of P-DSA using those criteria. We compare P-DSA against a benchmark that is representative of the prior art: *Authentication Tagging using Modulation* (ATM) [4].

We generate the authentication signal using the method proposed in Section II-C1, and apply P-DSA and ATM, respectively, on a quadrature phase-shift keying (QPSK) modulated message signal to obtain the embedded signal. In P-DSA, the controlled ISI added to the QPSK signal results in a constellation with 9 possible symbol positions as shown in Figure 6a. On the other hand, ATM utilizes the phase based hierarchical modulation to embed the authentication signal which leads to a constellation of 8 possible symbol positions as shown in Figure 6b. In ATM, an authentication bit of 1 is embedded by shifting the phase of a QPSK message constellation symbol towards the Q-axis (representing quadrature-phase) by θ. An authentication bit of 0 is embedded by shifting the phase towards the I-axis (representing in-phase) by θ.

A. Resource Overhead

Embedding the authentication signal in the message signal requires applying changes to the message signal itself, and thus incurs some PHY-layer overhead. For instance, the mechanism proposed in [12] results in drop in the message throughput. Other examples of overhead include increase in average transmission power, increase in bandwidth, and increase in complexity of the transmitter and/or receiver.

By design, P-DSA as well as ATM does not change the message throughput. Also, the overall average transmission power is unchanged from standard QPSK. In terms of the transmitter's and the aware receiver's complexity, ATM is advantageous compared to P-DSA. To implement ATM, the transmitter (Alice) and the receiver (Bob) only need to modify how the embedded signal is mapped to the constellation symbols. However, implementation of P-DSA is more complex—Alice needs to add controlled ISI, and Bob requires a MLSD to extract the message and the authentication signals.

In terms of spectrum efficiency, however, P-DSA has an advantage over ATM. The P-DSA scheme can achieve the ideal Nyquist rate of $2W$ symbols/s (where W is the signal bandwidth) while utilizing a physically realizable filter. In contrast, the spectrum efficiency of ATM depends on the roll-off factor, α (typically $0.2 - 0.3$), of the raised cosine filter that is used for waveform shaping. As a result, P-DSA utilizes only $1/(1+\alpha)$ of the bandwidth required by ATM, and hence it is more spectrally efficient.

B. Message Signal's Error Performance

This criterion refers to the achievable error performance (in terms of BER) when decoding the received message signal, MS_b. Figure 7 shows BER vs. E_b/N_0 curves for MS_b in P-DSA, where one authentication bit is embedded into each block of messages bits of length, $N = 16$. For comparison, we also show error performance of MS_b when P-DS is applied to QPSK with $N = 16$ and no authentication signal is embedded. We observe that the performance of P-DS is very close to that of QPSK which is used as the benchmark. This signifies that despite the addition of the ISI in the P-DS waveform, its message signal can be detected with nearly the same error performance as that of QPSK (which does not add ISI) if MLSD, with sufficiently long block length, N, is used at the receiver.

The figure shows that the error performance of MS_b in P-DSA is inferior to that of P-DS. There are two reasons for this degradation. Firstly, in P-DS, the receiver has perfect knowledge of the initialization bit's state; whereas in P-DSA, the initialization bit of each block are the authentication bits, and hence they need to be estimated. Secondly, in P-DSA, Bob employs SSD for detecting the state of the authentication bit and the first message bit of each block, but employs MLSD for rest of the message bits. Hence, the overall detection performance of MS_b in P-DSA is inferior to that of P-DS, which uses MLSD for *all* the bits in a block. As a result, the BER of the message signal in P-DSA can be upper bounded by

$$P_{MS_b}^{P-DSA} = \frac{1}{N} \cdot P_{SS} + \left(1 - \frac{1}{N}\right) \cdot P_{ML}, \tag{7}$$

where P_{SS} and P_{ML} are obtained using (5) and (6), respectively.

Figure 8 shows the error performance of MS_b in P-DSA with $N = 16$, ATM with phase shift of $\theta = \pi/12$ rad and ATM with phase shift of $\theta = \pi/6$ rad. In ATM, the message signal's constellation points are intentionally positioned in non-optimal positions so that the authentication signal's constellation can be superimposed on top of the message signal's constellation. Hence, as the presence of the authentication signal becomes more dominant (by increasing θ) in ATM, the BER performance of the message signal detection degrades as shown in Figure 8. Our scheme, P-DSA, is not constrained by such a

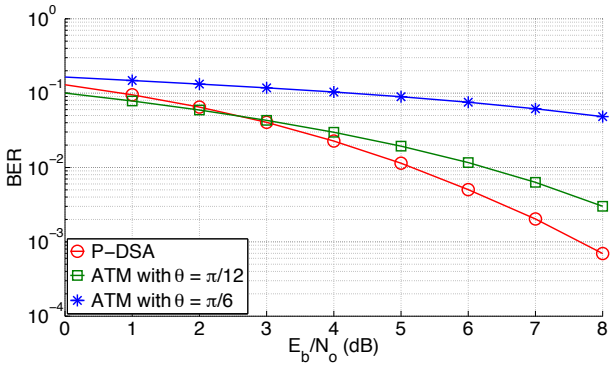


Fig. 8: BER performance of the message signal.

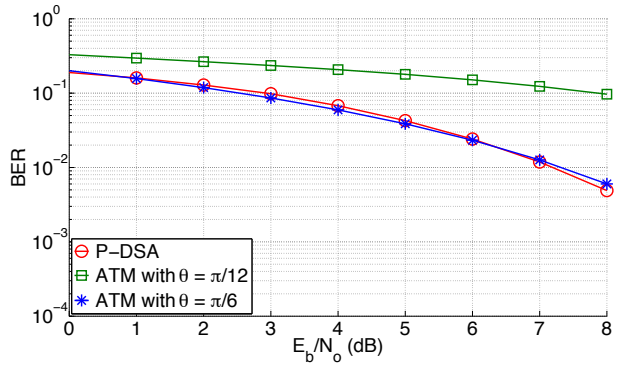


Fig. 9: BER performance of the authentication signal.

tradeoff, and this attribute provides an important advantage in terms of error performance.

C. Authentication Signal's Error Performance

This criterion refers to the achievable error performance when decoding the authentication signal, AS_b. In P-DSA, the state of the authentication signal is determined by each block's first received signal level which, in turn, is estimated through comparison to the three signal levels: $+2$, 0, and -2. In essence, decoding the authentication signal depends on the performance of SSD, and does not benefit from MLSD as shown in Figure 7. Hence, the BER of the authentication signal is given by P_{SS} which is calculated using (5).

Figure 9 shows the error performance of AS_b in P-DSA with $N = 16$, ATM with phase shift of $\theta = \pi/12$ rad and ATM with phase shift of $\theta = \pi/6$ rad. In ATM, the message and authentication signals are embodied in two different constellations (i.e., message signal is carried in the low-resolution constellation and authentication signal is carried in the high-resolution constellation). The effect of this multi-resolution modulation can be observed when we compare ATM's curves in Figure 8 and 9. Comparing the curve of ATM with phase shift of $\theta = \pi/12$ rad in Figure 8 with that of Figure 9, we see that the BER performance of message signal is noticeably better than that of authentication signal. Moreover, we also observe that the exact opposite is true for ATM with $\theta = \pi/6$ rad. When the phase shift is $\theta = \pi/12$ rad, the shift in the constellation points (from their conventional QPSK positions) is not significant enough to cause a significant drop in BER of message signal detection. However, this relatively small shift in phase makes decoding of the authentication signal difficult, because it is carried in a high-resolution constellation. When $\theta = \pi/6$ rad, the situation is reversed.

From Figure 9, we observe that with $\theta = \pi/6$ rad, ATM has comparable BER performance compared to P-DSA for the detection of the authentication signal. However, Figure 8 shows that P-DSA has a significant advantage in terms of BER performance of message signal detection. On the other hand, when ATM with $\theta = \pi/12$ is used, the BER performance of message signal detection is improved (compared to ATM with $\theta = \pi/6$). However, changing from $\theta = \pi/6$ to $\theta = \pi/12$ causes a significant increase in BER for the detection of the authentication signal as shown in Figure 9.

For a PHY-layer authentication scheme to be viable, Bob must be able to decode both the message and the authentication signals with sufficiently good BER. Considering this requirement, we can conclude that P-DSA enjoys a significant advantage over ATM. From Figures 8 and 9, and the above discussions, we can make the following conclusions.

- ATM makes a tradeoff between the message signal's SNR and the authentication signal's SNR under the assumption of constant average power. This implies that one cannot improve the former without scarifying the latter, and vice versa. This attribute is a fundamental drawback of blind signal superposition.
- P-DSA does not make the aforementioned tradeoff, and instead embeds the authentication signal by exploiting the inherent redundancy in the waveform shaping process. The resulting nine-level signal does increase the number of constellation points (thereby decreasing the minimum Euclidean distance between constellation points), but nevertheless manages to outperform ATM in terms of BER performance.

D. Authentication Rate

In PHY-layer authentication, the authentication signal is embedded by altering the message signal in a certain manner so that the receiver can detect the alteration and use it to extract the authentication information. The rate at which the alteration can be made is called the authentication rate.

In P-DSA, one bit of AS_a is transmitted in each block (of length N bits) of MS_a, which leads to an authentication rate of $1/N$. Although the authentication rate in P-DSA can be varied by changing N, decreasing N leads to a lower trellis length for MLSD. This leads to lower error performance for MS_b, which is inferred using equation (7). However, changing N does not affect the error performance of AS_b in P-DSA as the detection of AS_b depends only on the detection of the first received signal level in each block. On the other hand, ATM achieves an authentication rate of $1/2$—one authentication bit can be inserted for every two message bits or one QPSK modulated symbol.

E. Security

There are three facets of security that need to be considered: *integrity*, *impersonation* and *replay*. To ensure integrity, the authentication scheme should not allow Eve to modify

TABLE IV: An example illustrating HM-DSA encoding (the underlined bits are the authentication bits to be embedded).

d_n	$\underline{1}$	0	1	1	$\underline{0}$	0	1	1
w_n	$+1$	-1	$+1$	$+1$	-1	-1	$+1$	$+1$
y_n		-0.7	$+0.7$	$+1.3$		-1.3	$+0.7$	$+1.3$

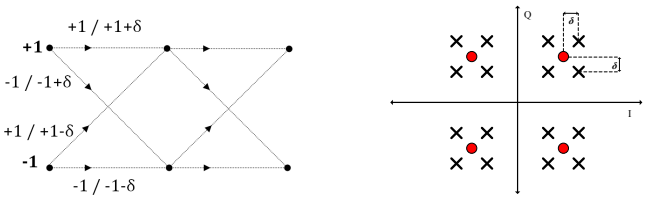


Fig. 10: (a) Trellis used by modified MLSD in HM-DSA, and (b) Constellation of QPSK with HM-DSA.

Alice's messages without being detected by Bob. To thwart impersonation, the scheme should not allow Eve to create valid proofs of authenticity for her messages. To thwart replay attacks, Alice's authentication signal should incorporate countermeasures that deter the interception and replay of her message signal. We do not consider conventional jamming attacks. It is important to realize that security not only depends on the strength of the cryptographic primitives used to create the authentication signal, but also on how the authentication signal is embedded in the message signal as elaborated below.

1) Integrity: To break the authentication scheme, Eve may attempt to either remove or corrupt the authentication signal. In P-DSA as well as ATM, the authentication signal is embedded into the message signal in such a way that the selective removal of the authentication signal is not possible. However, Eve can successfully corrupt the authentication signal by transmitting interference so that Bob cannot verify the authenticity of the message transmitted by Alice. We will refer to such an attack as an *obstruction of authentication* (OOA) jamming attack. Note that the OOA jamming is different from a conventional (or indiscriminate) jamming attack. The objective of conventional jamming is to prevent a targeted receiver from correctly decoding the transmitted message by generating interference of sufficient power. In contrast, the objective of OOA jamming is to generate just enough interference to prevent Bob from verifying the authenticity of the message, yet still enable him to correctly decode the message itself. In certain scenarios, this may encourage Bob to treat the received message as a legitimate message without actually authenticating it. Hence, OOA has obvious security implications.

OOA jamming can be quite effective against hierarchical modulation schemes, including ATM, since in these schemes, the message signal is embodied by a high-power constellation while the authentication signal is carried on a low-power constellation. Eve can emit just enough interference to exploit the power difference between the two constellations, and thus prevent decoding of the authentication signal but enable decoding of the message signal. OOA jamming is difficult to detect because it can readily be mistaken for noise or non-malicious interference.

On the other hand, P-DSA is robust to OOA jamming. To obstruct Bob from decoding the authentication signal, Eve would need to generate interference that is sufficiently powerful to also make decoding of the message signal impossible.

2) Impersonation: In a successful impersonation attack, Eve is able to create proofs of authenticity for her messages that are convincing enough to trick Bob into thinking that those messages have been created by Alice. Therefore, the authentication signal has to be designed to make such exploits infeasible. In the proposed design, Alice embeds the hashes from the generated chain in reverse order. Hence, having received h_i, Eve can only generate h_{i-1}, but not h_{i+1} (we assume that hash function has the required security properties to resist cryptanalytic attacks). Therefore, only Alice can transmit h_{i+1} in the $(i+1)^{th}$ time window, and impersonation attacks are thwarted.

3) Replay: To launch replay attacks, Eve needs to obtain h_i which is valid only in i^{th} time window. However, since h_i is transmitted only once in one time-window, there is no way Eve can replay the received signal.

Note that there are important differences between our scheme and those proposed in [4], [12] in the way that the authentication information is generated. Unlike the scheme of [4], our scheme thwarts replay attacks within a time window. However, the minimum delay after which the received signal can be authenticated at Bob is T in our scheme, which is larger than that in the scheme discussed in [4]. In [12], the transmitter obtains the hash h_n (that starts the hash chain) from the regulator and publishes h_0 (the tail of the hash chain). However, in our scheme, Alice does not share h_n with Bob and only publishes h_0 for all the receivers. Hence, Bob authenticates Alice's messages without any prior coordination between the two.

F. Transparency

This criterion dictates that a PHY-layer authentication scheme should embed the authentication signal into the message signal such that it enables the aware receiver (Bob) to extract the authentication signal, while at the same time, enables the unaware receiver (Charlie) to recover the message signal *without* requiring the unaware receiver to change its demodulation or decoding procedure. In P-DSA, to avoid error propagation, we use precoding at the transmitter and remove the precoding to estimate the message signal at the receiver. Also, the embedded signal transmitted by Alice contains zero-valued signal levels as shown in Figure 6a. Therefore, Charlie must have the knowledge of P-DSA for extracting the message signal. In contrast, in ATM, Charlie does not need to change the demodulation/decoding procedure to recover the message signal—i.e., he simply treats the embedded signal as a regular QPSK modulated signal and the embedded authentication signal as noise. Therefore, ATM has the advantage over P-DSA in terms of transparency. However, in ATM, Bob with knowledge of the embedding scheme, does no better than Charlie in terms of error performance. To address the problem of transparency, we propose a variant of P-DSA called *Hierarchically Modulated Duobinary Signaling for Authentication* (HM-DSA) which has the transparency property.

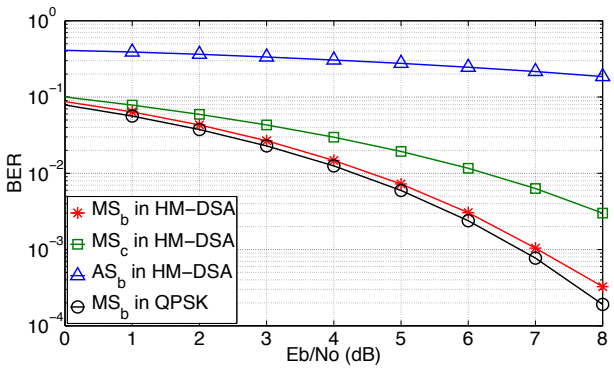


Fig. 11: BER performance of HM-DSA.

HM-DSA: The fundamental encoding by Alice and decoding by Bob in HM-DSA remains the same as those in P-DSA except that we do not use the precoding at the transmitter, and correspondingly modify the decoding at the receiver. Using the NRZ encoding, Alice generates a bipolar sequence, $\{w_n\}$, from the message sequence, $\{d_n\}$. Using $c_0 = 1$ and $c_1 = \delta$, where $0 < \delta < 1$ in the duobinary filter represented by equation (2), we obtain a duobinary signal given by

$$y_n = w_n + \delta \cdot w_{n-1}. \tag{8}$$

The parameter, δ, determines the amount of ISI encountered by the current pulse from the previous pulse. For $w_n = \pm 1$, we obtain a four-level output—i.e., y_n has one of the four possible values: $+1+\delta$, $+1-\delta$, $-1+\delta$ or $-1-\delta$. Note that the four-level encoded signal, y_n, is used to express one of the two binary values, and hence there is an inherent redundancy in this encoding that we can exploit to embed the authentication signal. Also, from (8), we observe that the encoded signal, y_1, corresponding to w_1 is given by $y_1 = w_1 + \delta \cdot w_0$. Here, w_1 is the bipolar signal level for the first message bit, d_1. The other bipolar signal, w_0 is generated from an authentication bit, a_k, to start the encoding of the message signal, $\{d_n\}$, $n = 1, 2 \cdots, N$. Table IV illustrates an example of the HM-DSA embedding process with $\delta = 0.3$.

Finally, the encoded sequence, $\{y_n\}$, is processed through the RF front-end and transmitted over the air. Once Charlie receives the signal as $\{\hat{y}_n\}$, he performs SSD. The estimated message, $\{\hat{d}_n\}$, is obtained by the following decoding decision rule:

$$\hat{d}_n = \begin{cases} 0, & \text{if } \hat{y}_n = -1+\delta \text{ or } -1-\delta; \\ 1, & \text{if } \hat{y}_n = +1+\delta \text{ or } +1-\delta. \end{cases} \tag{9}$$

Bob, with knowledge of the embedding scheme, estimates the first state of the message signal and the authentication bit's state by executing a minimum distance check (SSD) on the first received signal. Further, Bob uses the MLSD (see Figure 10a) with two states (i.e., $+1$ and -1) to obtain an estimate of the received bipolar sequence, $\{\hat{w}_n\}$.

When we apply HM-DSA to a message signal modulated using QPSK, HM-DSA leads to an amplitude based hierarchical modulation popularly known as 4/16 quadrature amplitude modulation (QAM) as shown in Figure 10b. Hence, we can obtain the BER performance of AS_b and MS_c by

$$\begin{aligned} P_{AS_b}^{HM-DSA} &= \frac{1}{4} \cdot erfc\left(\sqrt{\frac{E_b}{N_0} \cdot \left(\frac{1-\delta}{1+\delta^2}\right)}\right) \\ &+ \frac{1}{4} \cdot erfc\left(\sqrt{\frac{E_b}{N_0} \cdot \left(\frac{1+\delta}{1+\delta^2}\right)}\right). \end{aligned} \tag{10}$$

$$\begin{aligned} P_{MS_c}^{HM-DSA} &= \frac{1}{2} \cdot erfc\left(\sqrt{\frac{E_b}{N_0} \cdot \left(\frac{\delta}{1+\delta^2}\right)}\right) \\ &- \frac{1}{4} \cdot erfc\left(\sqrt{\frac{E_b}{N_0} \cdot \left(\frac{2+\delta}{1+\delta^2}\right)}\right) \\ &+ \frac{1}{4} \cdot erfc\left(\sqrt{\frac{E_b}{N_0} \cdot \left(\frac{2-\delta}{1+\delta^2}\right)}\right). \end{aligned} \tag{11}$$

The BER performance of MS_b can be upper bounded by

$$P_{MS_b}^{HM-DSA} = \frac{1}{N} \cdot P_{MS_c}^{HM-DSA} + \left(1 - \frac{1}{N}\right) \cdot P_{ML}. \tag{12}$$

Comparison of ATM and HM-DSA: As stated previously, both of these schemes have the transparency property. Both these schemes have the same average power, bandwidth and message throughput. Additionally, both these scheme are vulnerable to OOA jamming attack. However, implementation of HM-DSA is more complex as Bob requires MLSD to extract the message and the authentication signals. Using the expressions of BER for HM-DSA and ATM in [4] to perform a fair BER comparison between ATM and HM-DSA, we set $\delta = \tan\theta$ so that the two schemes' error performance for the authentication signal is the same. Note that when $\delta = \tan\theta$, the two schemes' message signal error performance achievable by Charlie is also the same. The error performance of the two schemes is identical because they are constrained by the same tradeoff mentioned previously for SSD. However, it is important to note that the two schemes differ in terms of the message signal's error performance achievable by Bob. In HM-DSA, Bob can significantly improve the message signal's error performance by employing MLSD, which is possible because of the way HM-DSA adds signal redundancy through its encoding process. This is not possible with ATM.

Figure 11 shows the error performance of HM-DSA with $\delta = \tan\pi/12$ and $N = 16$. In the figure, a regular QPSK-modulated signal's error performance curve (first curve from the bottom) is used as a benchmark. We can see that message signal's error performance achievable by Bob is close to that of the QPSK signal. If we compare this with ATM's performance with $\theta = \pi/12$ (i.e., message signal's error performance achievable by Bob)—which is identical to the curve labeled "MS_c in HM-DSA" in Figure 11—the advantage of HM-DSA is obvious.

It is important to note that HM-DSA is different than the 4/16-QAM based scheme proposed in [3] which we refer to as *Amplitude based Hierarchical Modulation for Authentication* (AHMA). From the expressions of BER for HM-DSA, AHMA in [3] and ATM in [4], it can be inferred that when $\delta = \tan\theta$, HM-DSA, AHMA and ATM have the same error performance for authentication and message signals using SSD. Here, δ in HM-DSA and AHMA, and θ in ATM are the deviations of the

constellation symbols of the message signal from the optimal positions to embed the authentication signal. Hence, the above discussions on comparison of ATM and HM-DSA also apply to the comparison of AHMA and HM-DSA.

V. Related Work

In essence, PHY-layer authentication is closely related to or is equivalent to radio frequency (RF) fingerprinting [13], [14], electromagnetic signature identification [15]–[17], PHY-layer watermarking [5], [18]–[20], and transmitter identification [21], [22]. These schemes can be broadly divided into the three categories.

Schemes in the first category utilize the idiosyncrasies of the communication system, such as RF signal characteristics or intrinsic channel characteristics [23], as unique signatures to authenticate/identify transmitters. Although this approach has been demonstrated to work in controlled lab environments, its sensitivity to environmental factors—such as temperature changes, channel conditions and interference—limits its efficacy in real-world scenarios.

Schemes in the second category embed an artificial authentication signal in the message signal and extract it at the receiver [2]–[4]. In this approach, the authentication signal is embedded in the message signal in such a way that the authentication signal acts as noise to the message signal and vice versa. The schemes of this category can be collectively referred to as *blind signal superposition*. As mentioned previously, this method is constrained by the unavoidable tradeoff between the message signal's SNR and the authentication signal's SNR.

The third category includes techniques that avoid the drawbacks of blind signal superposition [5], [12], [24]. In [5], the message signal at the transmitter is processed with a synthesized channel-like filter that is generated using the authentication signal. However, since this approach requires estimation of the channel response at the receiver, it may not be a viable approach when the coherence time is short. In [12], the authentication signal is embedded into the transmitted OFDM signal by repeating some message symbols over the sub-carriers to generate a cyclo-stationary signature. However, this scheme achieves authentication at the cost of loss in the data throughput. The PHY-layer authentication scheme in [24] embeds the authentication signal as a frequency shift in the pilots of the message signal. Although the shifts in the pilot signals do not affect the channel estimates at an aware receiver significantly, they affect the channel estimate at an unaware receiver who does not know the authentication embedding scheme.

VI. Conclusion

We proposed a novel PHY-layer authentication scheme referred to as *Precoded Duobinary Signaling for Authentication* (P-DSA). P-DSA is fundamentally different from the prior art, and it relaxes the tradeoff that constrains the blind signal superposition schemes. Although P-DSA increases the number of points in the signal constellation (compared to conventional binary signaling), our simulation results show that it achieves improved error performance over the prior art without sacrificing message throughput or increasing power. P-DSA inherits such desirable attributes at the cost of increased transmitter/receiver complexity.

References

[1] J. Bernhard, J. Reed, J.-M. Park, and A. Clegg, "Final report of the NSF workshop on enhancing access to the radio spectrum (EARS)," http://www.nsf.gov/mps/ast/nsf_ears_workshop_2010_final_report.pdf, 2010, accessed: March 1, 2013.

[2] P. Yu, J. Baras, and B. Sadler, "Physical-layer authentication," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 38–51, Mar. 2008.

[3] ——, "Multicarrier authentication at the physical layer," in *Int. Symp. on World of Wireless, Mobile and Multimedia Networks*, June 2008, pp. 1–6.

[4] X. Tan, K. Borle, W. Du, and B. Chen, "Cryptographic link signatures for spectrum usage authentication in cognitive radio," in *Proc. ACM WiSec*, June 2011, pp. 79–90.

[5] N. Goergen, T. Clancy, and T. Newman, "Physical layer authentication watermarks through synthetic channel emulation," in *IEEE Symp. New Frontiers Dynamic Spectrum*, Apr. 2010, pp. 1–7.

[6] T. Jiang, H. Zeng, Q. Yan, W. Lou, and Y. Hou, "On the limitation of embedding cryptographic signature for primary transmitter authentication," *IEEE Wireless Commun. Letters*, vol. 1, no. 4, pp. 324–327, Aug. 2012.

[7] S. Pasupathy, "Correlative coding: A bandwidth-efficient signaling scheme," *IEEE Commun. Soc. Mag.*, vol. 15, no. 4, pp. 4–11, July 1977.

[8] V. Vadde and S. Gray, "Partial response signaling for enhanced spectral efficiency and RF performance in OFDM systems," in *IEEE Global Telecommun. Conf.*, vol. 5, 2001, pp. 3120–3124.

[9] J. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.

[10] J. Forney, G., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Tran. Info. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[11] J. Forney, G.D., "The viterbi algorithm," *IEEE Proc.*, vol. 61, no. 3, pp. 268–278, March 1973.

[12] L. Yang, Z. Zhang, B. Y. Zhao, C. Kruegel, and H. Zheng, "Enforcing dynamic spectrum access with spectrum permits," in *Proc. ACM MobiHoc*, 2012, pp. 195–204.

[13] J. Hall, M. Barbeau, and E. Kranakis, "Detecting rogue devices in Bluetooth networks using radio frequency fingerprinting," in *Commun. Comput. Netw.*, Oct. 2006.

[14] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Canadian J. Electr. Comput. Eng.*, vol. 32, no. 1, pp. 27–33, 2007.

[15] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. ACM MobiCom*, 2008, pp. 116–127.

[16] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *Int. Conf. Inform. Process. Sensor Netw.*, Apr. 2009, pp. 25–36.

[17] K. Remley et al, "Electromagnetic signatures of WLAN cards and network security," in *Proc. IEEE Int. Symp. Signal Process. Inform. Technol.*, Dec. 2005, pp. 484–488.

[18] I. Cox, M. Miller, and A. McKellips, "Watermarking as communication with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, July 1999.

[19] C. Fei, D. Kundur, and R. Kwong, "Analysis and design of secure watermark-based authentication systems," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 1, pp. 43–55, Mar. 2006.

[20] J. Kleider, S. Gifford, S. Chuprun, and B. Fette, "Radio frequency watermarking for OFDM wireless networks," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process. Proc.*, vol. 5, May 2004, pp. 397–400.

[21] X. Wang, Y. Wu, and B. Caron, "Transmitter identification using embedded pseudo random sequences," *IEEE Trans. Broadcast.*, vol. 50, pp. 244–252, Sep. 2004.

[22] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2571–2579, July 2008.

[23] Y. Liu, P. Ning, and H. Dai, "Authenticating primary users' signals in cognitive radio networks via integrated cryptographic and wireless link signatures," in *IEEE Symp. Security and Privacy*, May 2010, pp. 286–301.

[24] R. Miller and W. Trappe, "Short paper: ACE: authenticating the channel estimation process in wireless communication systems," in *Proc. ACM WiSec*, 2011, pp. 91–96.