Transmitter authentication using hierarchical modulation in dynamic spectrum sharing

Vireshwar Kumara,*, Jung-Min (Jerry) Parka, Kaigui Bianb

a Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA
b School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

ABSTRACT

One of the critical challenges in dynamic spectrum sharing (DSS) is identifying non-conforming transmitters that violate spectrum access rules prescribed by a spectrum regulatory authority. One approach for facilitating identification of the transmitters in DSS is to require every transmitter to embed an uniquely-identifiable authentication signal in its waveform at the PHY-layer. In most of the existing PHY-layer authentication schemes, the authentication signal is added to the message signal as noise, which leads to a tradeoff between the message signal’s signal-to-noise ratio (SNR) and the authentication signal’s SNR under the assumption of constant average transmitted power. This implies that one cannot improve the former without sacrificing the latter, and vice versa. In this paper, we propose a novel PHY-layer authentication scheme called Hierarchical Modulation with Modified Duobinary Signaling for Authentication (HMM-DSA), which relaxes the constraint on the aforementioned tradeoff. HMM-DSA utilizes a modified duobinary filter to introduce some controlled amount of inter-symbol interference into the message signal, and embeds the authentication signal in the form of filter coefficients. Our results show that the proposed scheme, HMM-DSA, improves the error performance of the message signal as compared to the prior art.

1. Introduction

It is widely believed that dynamic spectrum sharing (DSS) is one approach for significantly increasing spectrum utilization efficiency. In a DSS environment, secondary users (SUs) opportunistically access fallow radio spectrum that is not utilized by primary (a.k.a. incumbent) users (PUs). SUs are required to follow a set of spectrum access rules or regulations prescribed by a spectrum regulatory authority (e.g., the Federal Communications Commission (FCC) in the U.S.A.) to protect the PUs from interference, and to minimize inter-SU interference. Therefore, to ensure the viability of the spectrum sharing model, effective and low-cost spectrum (rule) enforcement measures must be adopted (Park et al., 2014; Dutta and Chiang, 2016). Spectrum rule enforcement is an especially critical issue when federal government (including military) systems share spectrum with non-government systems, such as the case in the 3.5 GHz band in which commercial small-cell networks are expected to coexist with incumbent military radar systems (Altamimi et al., 2013; FCC, 2015).

In spectrum enforcement and security, one of the critical challenges is identifying, and if possible authenticating, non-conforming (“rogue”) or malfunctioning SU transmitters that have violated spectrum access rules. To authenticate transmitters, cryptographic mechanisms at the higher layers have been used. However, the ability to authenticate and/or uniquely identify SU transmitters at the PHY-layer is especially useful in heterogeneous coexistence environments, where incompatible systems (i.e., systems with different protocol stacks) cannot decode each others’ higher-layer signaling—e.g., IEEE 802.22 and 802.11af systems coexisting in TV white space (Feng et al., 2013). In a PHY-layer authentication scheme for spectrum enforcement, all SUs are mandated to employ a mechanism for embedding an authentication signal (which contains the identity of the transmitter, and possibly a certificate of compliance) into the message signal (which contains the data that the transmitter wants to send). Tamper resistance mechanisms are employed to prevent the circumvention of the authentication mechanism by hacking (Smith et al., 2012; Xiao et al., 2009).

In this paper, we define two types of intended receivers—unaware and aware receivers (Yu et al., 2008a). An unaware receiver is able to correctly demodulate and decode the message signal, but cannot authenticate the received signals, either because it has no knowledge of the authentication scheme or does not know the key required for
authenticating the transmitter. Also, a receiver that does not intend to authenticate the received signals is classified as an unaware receiver. On the other hand, a receiver that needs to recover the message signal as well as the authentication signal (embedded into the message signal) in order to identify the transmitter and authenticate its signals is called an aware receiver.

A conventional PHY-layer authentication scheme should embed the authentication signal into the message signal such that it enables the aware receiver to extract the message and the authentication signals from its received signal, while at the same time, enables the unaware receiver to recover the message signal from its received signal without requiring the unaware receiver to change its demodulation or decoding procedure. One approach to achieve this is to add the authentication signal to the message signal as noise (Yu et al., 2008a). To limit the detrimental effects of the authentication signal on the message signal, the principle of hierarchical modulation (Yu et al., 2008b, Tan et al., 2011) is often applied—i.e., the authentication signal (low priority signal) is carried on the low-power, high-resolution constellation while the message signal (high priority signal) is embodied by the high-power, low-resolution constellation.

In such an approach, both the aware receiver and the unaware receiver decode the message signal in the presence of the authentication signal, 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, 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 embedded signal at the receiver (Jiang et al., 2012). This means that there is a fundamental tradeoff in the existing schemes between the SNRs (and the error performances) of the message signal and the authentication signal.

In this paper, we propose a novel PHY-layer authentication scheme, called Hierarchical Modulation with Modified Duobinary Signaling for Authentication (HMM-DSA), that can be used by the aware receivers to identify rogue SU transmitters without significantly affecting the error performance of the message signal at the aware and the unaware receivers. The proposed scheme is based on duobinary signaling, a waveform shaping technique that has been traditionally used to increase bandwidth efficiency (Pasupathy, 1977), and hierarchical modulation, a technique to enable multi-resolution signaling (Ramchandran et al., 1993). In HMM-DSA, a hierarchically modulated duobinary signal is generated by inducing controlled inter-symbol interference (ISI) into the message signal. The controlled ISI is introduced by utilizing a modified duobinary filter whose coefficients are generated by using the authentication signal. In this way, HMM-DSA embeds the authentication signal into the message signal as well as relaxes the constraint on the aforementioned tradeoff that plagues the existing schemes.

The main contributions of this paper are summarized below.

- We propose the PHY-layer authentication scheme, HMM-DSA, in which the intended receiver can be either an aware receiver (which extract both the message and authentication signals) or an unaware receiver (which only extracts the message signal).
- We show that our approach enables significant improvement in the error performance of the message signal at the aware receiver when compared to that at the unaware receiver. We also show that HMM-DSA outperforms the prior art in terms of the detection performance of the message signal at the aware receiver.
- We have implemented HMM-DSA on Universal Software Radio Peripheral (USRP) radio boards, and provided testbed experiment results that corroborate our simulation results.

The rest of the paper is organized as follows. We provide the related work in Section 2. We describe the problem statement for PHY-layer authentication in Section 3, and discuss HMM-DSA in Section 4. We analyze the error performance of HMM-DSA in Section 5, and compare HMM-DSA with the prior art in Section 6. We discuss the experimental validation of HMM-DSA in Section 7, and conclude the paper by highlighting the main contributions in Section 8.

2. Related work

Based on the definitions of the aware and unaware receivers, the PHY-layer authentication schemes in the existing literature can be broadly divided into two categories. The schemes in the first category do not enable the intended receivers to function as the unaware receivers (Goergen et al., 2010; Yang et al., 2012; Miller and Trappe, 2011; Kumar et al., 2016; Jin et al., 2015). This means that in these schemes, all the intended receivers need to know the employed authentication mechanisms to demodulate and decode the message signals. In other words, these schemes require every intended receiver to be an aware receiver.

The schemes in the second category enable the intended receivers to be unaware receivers (Yu et al., 2008a, 2008b; Tan et al., 2011; Jin et al., 2015; Kumar et al., 2014). This means that in these schemes, the intended receivers are able to decode and demodulate the message signals without the knowledge of the employed PHY-layer authentication mechanisms. In Yu et al. (2008a), the authentication signal is added to the message signal as noise. In Yu et al. (2008b), Tan et al. (2011), Jin et al. (2015), the technique of hierarchical modulation is employed, and the authentication signal is carried on the high-resolution constellation while the message signal is embodied by low-resolution constellation where the average power of the embedded signal remains the same as the original message signal (with unmodified constellation). In these schemes, this embedding procedure leads to a fundamental tradeoff between the SNRs of the message signal and the authentication signal. The scheme proposed in Kumar et al. (2014) avoids the aforementioned tradeoff, but has very low authentication rate (i.e., the rate at which the authentication bits are embedded into the message bits).

3. Problem description

3.1. Model

In this paper, we assume the following authentication scenario. Alice, Bob, and Charlie share the same wireless medium. Alice (a.k.a “transmitter”) intends to transmit messages to Bob (a.k.a. “aware receiver”) and Charlie (a.k.a. “unaware receiver”) via the wireless medium as per the rules established for DSS. Alice and Bob have agreed on an authentication scheme that allows Bob to verify the messages he receives from Alice. Charlie does not know the authentication scheme, and cannot authenticate Alice’s messages at the PHY-layer, but can demodulate and decode the message signal.

3.2. Challenges

In the above model, the operations performed by Alice can be decomposed into two parts—generation of the authentication signal, and embedding of the authentication signal into the message signal. Similarly, the operations performed by Bob can be decomposed into two parts—extraction of the authentication signals from the received signal, and verification of the authentication signal. Hence, there are two distinct technical problems in devising a PHY-layer authentication scheme: (1) generating the authentication signal that later needs to be verified by an aware receiver; and (2) embedding the authentication signal into the message signal that later needs to be extracted by an aware receiver. To solve the first problem successfully, various threats need to be considered and mitigated (Tan et al., 2011; Goergen et al., 2010; Yang et al., 2012; Kumar et al., 2016, 2014). In this paper, we do not consider the first problem, and only focus on the second problem.
Our aim is to devise a scheme that enables Alice to embed an authentication signal into the message signal at the PHY-layer while causing no or minimal degradation in Bob’s ability to demodulate and decode the message signal. The primary technical challenges in devising such a scheme are as follows.

1. Alice should be able to embed the authentication signal into the message signal without affecting the message throughput.
2. Bob should be able to extract the message and authentication signals from the received signal.
3. Charlie should be able to extract the message signal without knowing the authentication signal’s embedding mechanism.
4. The embedding of the authentication signal into the message signal should not significantly impact the detection performance of the message signal at Bob and Charlie.

4. Hierarchical Modulation with Modified Duobinary Signaling for Authentication (HMM-DSA)

In this section, we provide a detailed description of Hierarchical Modulation with Modified Duobinary Signaling for Authentication (HMM-DSA). In HMM-DSA, we generate hierarchically modulated signals by introducing controlled ISI into the message signal using modified duobinary filtering. Further, we utilize the coefficients of the filter to embed the authentication signal.

In the following discussions, we use MS and AS to denote the message signal and the authentication signal generated by Alice in the baseband, respectively. We use $\tilde{M}$ and $\tilde{A}$ to denote the message signal and the authentication signal estimated by Bob, respectively. Finally, let $\tilde{M}$ denote the message symbol estimated by Charlie.

The message signal to be transmitted by Alice, MS, is assumed to be a sequence of bits which are statistically independent and identically distributed. The message bit sequence is represented by $\{d\}$. Using modulation, e.g., quadrature phase shift keying (QPSK), the message bit sequence, $\{d\}$, is mapped to a message symbol sequence, $\{x\}$. The authentication signal to be transmitted by Alice, AS, is considered as a sequence of bits which are statistically independent and identically distributed. It is represented by $\{a\}$. Using non-return to zero (NRZ) encoding, the authentication bit sequence, $\{a\}$, is mapped to an authentication symbol sequence, $\{u\}$. Hence, the authentication bits, 1 and 0, are mapped to the authentication symbols, +1 and −1, respectively.

Further, we assume that the number of symbols in the authentication symbol sequence, $\{u\}$, is represented by $K$, and the number of symbols in the message symbol sequence, $\{x\}$, is represented by $KN$. The message symbol sequence of length $K\cdot N$ is divided into $K$ blocks, each of length $N$ symbols. In HMM-DSA, an authentication symbol is embedded into each block of the message symbols, and hence $N$ is the number of message symbols transmitted for each authentication symbol. Further, the $n^{th}$ message symbol in the $k^{th}$ block is represented by $x_{k,n}$, where $k = 1, 2, \ldots, K$ and $n = 1, 2, \ldots, N$, and the authentication symbol corresponding to the $k^{th}$ block is represented by $u_k$.

4.1. Transmitter (Alice)

4.1.1. Embedding of the authentication signal into the message signal

In the $k^{th}$ block, for each message symbol, $x_{k,n}$, a duobinary sample, $z_{k,n}$, is generated using the modified duobinary filter shown in Fig. 1. The sample, $z_{k,n}$, is represented as

$$z_{k,n} = x_{k,n} + u_k \cdot \delta \cdot x_{k,n-1}.$$  \hspace{1cm} (1)

where $0 < \delta < 1$. Here, $\delta$ represents the weight of the delayed message signal, $x_{k,n-1}$. Hence, the ISI introduced to each duobinary sample, $z_{k,n}$, corresponding to the message symbol, $x_{k,n}$, comes only from the preceding message symbol, $x_{k,n-1}$. The amount of ISI is controlled by $\delta$, and leads to a hierarchically modulated duobinary sample. Moreover, in the $k^{th}$ block, the authentication symbol, $u_k$, determines whether the ISI is added to $x_{k,n}$ or subtracted from $x_{k,n}$, which is the core idea behind HMM-DSA. This implies that the authentication symbol, $u_k$, is embedded into the $k^{th}$ block of the message symbols in the form of the filter coefficient of the modified duobinary filter.

For the $k^{th}$ block, we observe that the duobinary sample, $z_{k,1}$, is given by $z_{k,1} = x_{k,1} + u_k \cdot \delta \cdot x_{k,0}$. Hence, we require an extra symbol, $x_{k,0}$, to initialize the duobinary filtering of the message symbols. $x_{k,0}$ is called the initialization symbol. We assume that the same initialization symbol is utilized for all $k = 1, 2, \ldots, K$.

4.1.2. RF front-end processing

We assume that Alice utilizes orthogonal frequency-division multiplexing (OFDM) for transmitting the generated duobinary samples. Hence, having generated the duobinary sample sequence, $\{z\}$, Alice generates OFDM symbols by taking the inverse fast fourier transform (IFFT) of the duobinary samples. Finally, the signal is up-converted and transmitted. Note that HMM-DSA does not rely on nor is it constrained by any particular RF front-end processing method (such as the single antenna based OFDM system assumed in this paper). For instance, multiple-input and multiple-output (MIMO) and cyclic prefix based OFDM can be utilized in HMM-DSA for transmitting the duobinary samples, and the RF front-end processing at Bob and Charlie can be modified correspondingly.

4.1.3. Illustration

Table 1 illustrates the above embedding process through an example with $\delta = 0.3$ and $N = 3$. For this illustration, we assume that the message symbol is the bit sequence represented by $\{01011\}$, and the authentication symbol to be embedded into the message symbol is the bit sequence represented by $\{01\}$. We assume that Alice utilizes binary phase shift keying (BPSK) modulation to map the message bits to the message symbols, i.e., $x_{k,n} = \pm 1$. Also, the authentication bits $a_0 = 0$ and $a_1 = 1$ are mapped to $u_1 = -1$ and $u_2 = +1$, respectively.

To perform HMM-DSA, the message symbol sequence is divided into two blocks of message symbols, and one authentication symbol is embedded into each of the two blocks. Here, we assume that the initialization symbol, $x_{k,0} = -1$, for $k=1,2$. The duobinary samples are computed for the authentication symbol $a_1 = -1$ for the first block, and $a_1 = +1$ for the second block. Hence, after duobinary filtering, we obtain a four-level hierarchically modulated duobinary sample—i.e., $z_{k,n}$ has one of the four possible values: $+1 + \delta = +1.3$, $+1 - \delta = +0.7$, $-1 - \delta = -1.3$, $-1 + \delta = -0.7$.

<table>
<thead>
<tr>
<th>$d$</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>−1</td>
<td>−1</td>
<td>+1</td>
<td>+1</td>
<td>−1</td>
<td>−1</td>
</tr>
<tr>
<td>$u$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>−0.7</td>
<td>+1.3</td>
<td>+0.7</td>
<td>−1.3</td>
<td>+0.7</td>
<td>+1.3</td>
</tr>
</tbody>
</table>

Table 1

An example illustrating duobinary filtering in HMM-DSA with $\delta = 0.3$ and $N = 3$. 
Note that the sequences of the message bits in both the blocks are the same, but the corresponding duobinary sample sequences in both the blocks are different. In this way, we have embedded $a_i = 0$ and $a_1 = 1$ into the first block of the message bits and the second block of the message bits, respectively. We note that the four-level output of $z_{k,n}$ is used to express one of the two binary values of the message symbols, $z_{k,n} = \pm 1$, and hence there is an inherent redundancy in this process which is utilized to embed the authentication symbols.

4.2. Unaware receiver (Charlie)

4.2.1. RF front-end processing

After down-converting and sampling the received signal, Charlie generates the estimated duobinary samples by taking the fast fourier transform (FFT) of OFDM symbols. Hence, each of the estimated duobinary samples is represented by $r_{k,n} = z_{k,n} + w_{k,n}$ for $k = 1, 2, \ldots, K$, and $n = 1, 2, \ldots, N$. Here, the additive noise $w_{k,n}$ is assumed to be independent of $z_{k,n}$, and is a circularly-symmetric complex Gaussian random variable with mean equal to zero and variance equal to $\sigma_w^2$.

4.2.2. Detection of the message signal

Since Charlie is an unaware receiver, he is only interested in recovering the message signal without verifying the authentication signal. Hence, Charlie utilizes the conventional symbol-by-symbol detection (SSD) method to estimate the message signal, $\hat{M}$. This means that Charlie directly performs demodulation of each sample, $r_{k,n}$, to obtain the estimated message bit sequence represented by $\{d\}$.

4.2.3. Illustration

Table 2 provides an example, illustrating the results of utilizing SSD, and performing BPSK demodulation of the duobinary samples generated in Table 1 in the absence of any noise.

4.3. Aware receiver (Bob)

4.3.1. RF front-end processing

After down-converting and sampling the received signal, Bob generates the estimated duobinary samples by taking the FFT of the OFDM symbols. Hence, Bob divides the estimated duobinary samples into blocks, where each of the estimated duobinary samples is represented by $z_{k,n} = \hat{x}_{k,n} + w_{k,n}$ for $k = 1, 2, \ldots, K$, and $n = 1, 2, \ldots, N$. Here, the additive noise $w_{k,n}$ is assumed to be independent of $\hat{x}_{k,n}$, and is a circularly-symmetric complex Gaussian random variable with mean equal to zero and variance equal to $\sigma_n^2$.

4.3.2. Extraction of the message and authentication signals

In the $k$th block, having estimated the received duobinary sample sequence as $\{\hat{x}\}$, the SSD method can be utilized to estimate the received message bit sequence, $\{d\}$. However, Bob has knowledge of the authentication signal embedding process. Hence, Bob can improve the signal detection performance over SSD using the following detection procedure.

We note that the duobinary sample sequence is generated from the message symbol sequence and has memory of length $1$—i.e., the current duobinary sample is related to the current message symbol as well as the previous message symbol. Hence, Bob utilizes the maximum likelihood sequence detection (MLSD) for each block of the estimated duobinary samples (Forney, 1972). As a result, the length of the trellis in MLSD is equal to the length of each block of the estimated duobinary samples, i.e., $N$. The MLSD determines the sequence of the message symbols that generates the sequence of the possible duobinary samples, represented by $\{\hat{x}\}$, which is the closest to the sequence of the estimated duobinary samples, $\{\hat{z}\}$, in terms of Euclidean distance over the whole trellis. In HMM-DSA, the previous message symbol is weighted by $\delta$, and either added to the current message symbol if the authentication symbol is $+1$, or subtracted from the current message symbol if the authentication signal is $-1$. Hence, Bob follows the following steps to extract the message signal, $\hat{M}_b$, and the authentication signal, $\bar{A}_S$, from the estimated duobinary sample sequence, $\hat{z}$.

1. For the $k$th block, Bob generates two trellis structures for the two possible values of the authentication symbols, i.e., $+1$ and $-1$.
2. He separately computes the Euclidean distance between the sequence of the possible duobinary samples, $\{\hat{z}\}$, and the estimated duobinary samples, $\{\hat{z}\}$, over the whole trellis of each of the structures.
3. The trellis structure with the minimum Euclidean distance is selected, and the corresponding estimate of the authentication symbol, $\hat{a}_k$, is determined. The estimated authentication bit, $\hat{a}_k$, is obtained by the NRZ decoding of $\hat{a}_k$.
4. The estimated sequence of the message symbols, $\hat{l}$, is obtained using the selected trellis structure. The estimated message bit sequence, $\{\hat{l}\}$, is generated from $\{\hat{x}\}$ by demodulation.
5. Finally, by concatenating the estimated message bits, and the estimated authentication bits of all the $K$ blocks, Bob obtains the estimated message signal, $\hat{M}_b$, and the estimated authentication signal, $\bar{A}_S$, respectively.

4.3.3. Illustration

Fig. 3 shows the trellis structures used for the MLSD for estimating the message and authentication symbols embedded in Table 1. Recall that in the illustration, the message symbols are generated using BPSK modulation, and have two possible values, $-1$ and $+1$. Hence, the two trellis structures (Figs. 3a and b) corresponding to two possible values of the authentication symbol ($-1$ and $+1$) are generated by considering all possible transitions from each of the two possible message symbols. For example, in Figs. 3b, an arrow from symbol $+1$ with the label $+1/1+1+\delta$ represents a transition to the next message symbol indicated by the left number, $+1$. The right number, $+1 + \delta$, denotes the resultant duobinary sample. Also, recall that the initialization symbol is $-1$.

Hence, from Fig. 3, we can readily observe that the duobinary sample of $-1 + \delta$ or $-1 - \delta$ can be detected for the first message symbol with value $-1$ only if the authentication symbol is $-1$ or $+1$, respectively. But the authentication symbol can be decided to be $-1$ if the first duobinary sample is $+1 + \delta$ along with the first message symbol with value $+1$. Similarly, if the first duobinary sample is $+1 - \delta$ along with the first message symbol with value $+1$, the authentication symbol has to be $+1$.

5. Analysis of HMM-DSA

5.1. Average energy

Assume that the message symbols are modulated using $M$th order quadrature amplitude modulation ($M$-QAM), and the average energy of
the transmitted signal is $E_{av}$. Hence, for the $M$-QAM modulated signal transmitted without any ISI, the average energy per message symbol is $E_{av}$. Also, since each message symbol carries $\log_2 M$ bits, the average energy per message bit is given by $E_b = \frac{1}{\log_2 M} E_{av}$.

For HMM-DSA, assume that the average transmitted energy is kept unchanged from the conventional $M$-QAM. Hence, the average energy per duobinary sample is $E_{av}$. When the message symbols are filtered using the modified duobinary filter, given by Eq. (1), the average energy per duobinary sample becomes $(1 + \delta^2)$ multiplied by the average energy per message symbol. Hence, the average energy per message symbol in HMM-DSA is given by $E_{av}(1 + \delta^2)$. The rest of the average energy $E_{av}\delta^2/(1 + \delta^2)$ is captured by the ISI which is utilized to embed the authentication symbols. Also, since each duobinary sample corresponds to $\log_2 M$ message bits, the average energy per message bit is $E_{av}\delta^2/(1 + \delta^2) = E_b(1 + \delta^2)$. Further, since each duobinary sample carries $\log_2 M$ bits of information corresponding to the ISI, the average bit energy corresponding to the ISI is $E_{av}\delta^2/(1 + \delta^2) = E_b\delta^2/(1 + \delta^2)$. Note that the average energy per bit is a factor that needs to be considered when we analyze the error performance of the message and authentications signals.

5.2. Error performance

We can follow the above discussions and description of generalized QAM in Vitthaladevuni and Alouini (2001) to obtain the bit error rate (BER) of the message and authentication signals. Here, we provide the expressions for the BER in a particular scenario where the message signal is modulated using QPSK.

In HMM-DSA, the controlled ISI added to the QPSK modulated message symbols results in the hierarchically modulated duobinary samples which can be represented using the constellation with 16 possible symbols as shown in Fig. 4a. In the figure, we observe that the message and authentication signals are embodied in two different constellations, i.e., the message signal is carried in the low-resolution constellation, and the authentication signal is carried in the high-resolution constellation. The effect of this multi-resolution modulation can be observed when we compare the BER of the message signal with that of the authentication signal with different sets of parameters.

5.2.1. BER of $\hat{M}_S$

Recall that SSD is utilized for estimating the message signal at the unaware receiver, $\hat{M}_S$. Hence, the BER of $\hat{M}_S$ is given by

$$P_{\hat{M}_S} = \frac{1}{4} \text{erfc} \left( \frac{E_b}{\sqrt{N_0}} \frac{1 - \delta}{\sqrt{1 + \delta^2}} \right) + \frac{1}{4} \text{erfc} \left( \frac{E_b}{\sqrt{N_0}} \frac{1 + \delta}{\sqrt{1 + \delta^2}} \right)$$

(2)

where erfc, $E_b$ and $N_0$ represent the complementary error function, the average bit energy, and the noise power spectral density, respectively. Also, recall that $\delta$ represents the controlled ISI embedded into the message symbols.

5.2.2. BER of $\hat{M}_S$

The message signal at the aware receiver, $\hat{M}_S$, is detected by using MLSD instead of SSD. It is prohibitively complex to derive an exact expression for the BER of $\hat{M}_S$. However, its upper bound can be readily derived as Forney (1972)

$$P_{\hat{M}_S,\text{upper}} = \text{erfc} \left( \frac{E_b}{\sqrt{N_0}} \right)$$

(3)

The BER of $\hat{M}_S$ can also be lower bounded by the BER of the QPSK modulated message signal without any ISI which is given by

$$P_{\hat{M}_S,\text{lower}} = P_{\text{QPSK}} = \frac{1}{2} \text{erfc} \left( \frac{E_b}{\sqrt{N_0}} \right)$$

(4)

5.2.3. BER of $\hat{A}_S$

By using the results from Vitthaladevuni and Alouini (2001), we can derive the lower bound of the BER of the authentication signal at the aware receiver, $\hat{A}_S$, as

$$P_{\hat{A}_S,\text{lower}} = \frac{1}{2} \text{erfc} \left( \sqrt{2N_0} \frac{E_b}{\sqrt{N_0}} \frac{\delta}{\sqrt{1 + \delta^2}} \right)$$

(5)
Recall that \( N \) represents the length of the trellis utilized in MLSD at the aware receiver. Notation \( N \) also represents the number of message symbols transmitted for each authentication symbol. Hence, the above equation represents the BER of the authentication signal when the same authentication symbol is embedded into \( N \) message symbols using hierarchical modulation, and is estimated using soft-decision decoding.

5.3. Effect of \( E_b/N_0 \)

Fig. 5 shows the BER vs. \( E_b/N_0 \) curves in HMM-DSA with \( \delta = 0.3 \) and \( N=3 \). We utilize the BER of the QPSK modulated message signal without ISI as the benchmark which is labeled as “QPSK” in Fig. 5. The curve for the BER of “QPSK” is expressed by \( P_{\text{QPSK}} \) in Eq. (4). In the figure, we observe the BER of \( M_S \), expressed by \( P_{M_S} \) in Eq. (2), is significantly higher than that of QPSK. However, compared to the BER of \( M_S \), the BER of \( M_S \) is closer to that of QPSK. This clearly demonstrates that the error performance of message signal at the aware receiver, \( M_b \), is improved by using MLSD instead of SSD.

Further, in Fig. 5, we can readily see that the BER of \( M_S \), although higher than that of \( M_S \), is noticeably lower than that of \( A_S \) when we consider \( E_b/N_0 \) in the range from 0 dB to 8 dB. This means that when we have the ISI, \( \delta = 0.3 \), and the length of each block, \( N=3 \), the shift in the constellation symbols in Fig. 4(a) from their conventional positions is not significant enough to cause a significant drop in the error performance of \( M_S \). However, this relatively small shift makes decoding of the authentication signal difficult.

5.4. Effect of \( N \)

Fig. 6 shows the BER vs. \( N \) curves in HMM-DSA with \( \delta = 0.3 \) and \( E_b/N_0 = 6 \text{ dB} \). In the figure, we observe that changing \( N \) does not affect the error performance of \( M_S \). This is due to the fact that \( M_S \) is decoded using SSD. However, by increasing \( N \), we observe that the BER of \( A_S \) decreases. Recall that each of the authentication symbols is estimated using \( N \) duobinary samples. Hence, increasing \( N \) leads to more number of duobinary samples utilized for the estimation of \( A_S \), which results in a lower BER of \( A_S \). Further, in Fig. 6, we observe that larger \( N \) leads to a lower BER of \( M_S \). This can be attributed to the fact that as \( N \) increases, the trellis length for MLSD increases, resulting in better detection of \( M_S \).

Note that as one authentication symbol is inserted in each block of \( N \) QPSK modulated message symbols, the authentication rate (the rate at which the authentication bits are embedded into the message bits) is given by \( \frac{1}{N} \log_2 \text{M} \). In general, the authentication rate in HMM-DSA is given by \( \frac{1}{N} \log_2 \text{M} \) where \( M \) represents the order of the modulation scheme, e.g., \( M=4 \) for QPSK. Hence, increasing \( N \) leads to a lower authentication rate. Here, we discuss two special cases—(1) \( N=1 \), and (2) \( N \gg 1 \).}

5.4.1. Case, \( N=1 \)

As shown in Fig. 6, when \( N=1 \), the BER of \( M_S \) is close to the BER of \( M_S \). This means that HMM-DSA with \( N=1 \) provides no significant advantage to the aware receiver over the unaware receiver in terms of decoding the received message signal. Here, although the authentication rate is high (=1/2), the BER of the authentication signal, \( A_S \), is also significantly high (>0.1).

5.4.2. Case, \( N \gg 1 \)

As shown in Fig. 6, if HMM-DSA with sufficiently large \( N \) (i.e., \( N \gg 1 \)) is used, the BER of \( M_S \) gets closer to the value 2.4 \times 10^{-3} \text{ which is the BER of the QPSK modulated signal without any ISI at } E_b/N_0 = 6 \text{ dB}. This means that although the minimum Euclidean distance between the constellation symbols in HMM-DSA is smaller than that in QPSK modulation, the BER of \( M_S \) in HMM-DSA can be made asymptotically equal to that of QPSK by increasing the length of each block, i.e., \( N \). In other words, after adding the controlled ISI, \( M_S \) in HMM-DSA can be detected using MLSD with nearly the same error performance as the message signal without any ISI. This is one of the most important features of HMM-DSA. Further, when \( N \gg 1 \), although the authentication rate is very low, the BER of the authentication signal approaches to 0. Here, it is important to note that a low authentication rate (due to setting \( N \) to an arbitrarily large value) is acceptable in transmitter authentication (which is considered in this paper), but may not be acceptable for message authentication.

5.5. Effect of \( \delta \)

Fig. 7 shows the BER vs. \( \delta \) curves in HMM-DSA with \( N=3 \) and \( E_b/N_0 = 6 \text{ dB} \). In the figure, we observe that while the BER of \( M_S \),

\[ E_b/N_0 = 6 \text{ dB} \] In the figure, we observe that changing \( N \) does not affect the error performance of \( M_S \). This is due to the fact that \( M_S \) is decoded using SSD. However, by increasing \( N \), we observe that the BER of \( A_S \) decreases. Recall that each of the authentication symbols is estimated using \( N \) duobinary samples. Hence, increasing \( N \) leads to more number of duobinary samples utilized for the estimation of \( A_S \), which results in a lower BER of \( A_S \). Further, in Fig. 6, we observe that larger \( N \) leads to a lower BER of \( M_S \). This can be attributed to the fact that as \( N \) increases, the trellis length for MLSD increases, resulting in better detection of \( M_S \).

Note that as one authentication symbol is inserted in each block of \( N \) QPSK modulated message symbols, the authentication rate (the rate at which the authentication bits are embedded into the message bits) is given by \( \frac{1}{N} \log_2 \text{M} \). In general, the authentication rate in HMM-DSA is given by \( \frac{1}{N} \log_2 \text{M} \) where \( M \) represents the order of the modulation scheme, e.g., \( M=4 \) for QPSK. Hence, increasing \( N \) leads to a lower authentication rate. Here, we discuss two special cases—(1) \( N=1 \), and (2) \( N \gg 1 \).
increases by increasing the ISI—i.e., δ, the BER of $\overline{AS}$ decreases. It is also evident from the Eqs. (2) and (5) that the BER of $M_{S}$ increases, and the BER of $\overline{AS}$ decreases when we increase the ISI by increasing δ. This means that as the presence of the authentication signal becomes more dominant (by increasing δ) in HMM-DSA, the error performance of $M_{S}$ degrades. This phenomenon can be attributed to the fact that the message signal’s detection at the unaware receiver in HMM-DSA is constrained by the tradeoff between the message signal’s SNR and the authentication signal’s SNR. However, in the approaches utilized in the prior art for PHY-layer authentication, this tradeoff is unavoidable for both, the unaware as well as the aware receivers. On the other hand, in HMM-DSA, the aware receiver can overcome the loss in the error performance of the message signal by utilizing MLSD. Hence, we note that the increase in ISI also increases the BER of $M_{S}$, but the increase in the BER of $\overline{AS}$ is less than that of $M_{S}$. In fact, the BER of $M_{S}$ can be further reduced by using larger value of N—i.e., larger block length, as shown in Fig. 6. This is an advantageous feature of HMM-DSA compared to the prior art.

5.6. Selection of values for N and δ

We can select values for N and δ to meet the performance requirements of a given deployment scenario. For instance, in the scenario where the BER of $M_{S}$ in HMM-DSA needs to be close to the conventional message signal without any ISI, we consider the following cases.

1. If we need to achieve a particular authentication rate, we proceed by first determining a corresponding value for N as discussed in Section 5.4. Further, we select the value for δ based on the trade-off between the BERs of $M_{S}$ and $\overline{AS}$ as shown in Fig. 7.

2. If we need to achieve a particular BER of the message signal at the unaware receiver, $M_{S}$, we proceed by first determining a value for δ followed by selecting a value for N. Here, we determine the corresponding value for δ by observing the BER curve of $M_{S}$ in Fig. 7. Then, we select the value of N by considering the trade-off between the authentication rate and the BER of $\overline{AS}$ as discussed in Section 5.4.

6. Comparison with the prior art

In this section, we compare HMM-DSA against a benchmark that is representative of the prior art: Authentication Tagging using Modulation (ATM) (Tan et al., 2011). ATM utilizes the phase based hierarchical modulation to embed the authentication signal. In ATM, an authentication bit of 1 is embedded by shifting the phase of a message constellation symbol towards the Q-axis by $\theta$. An authentication bit of 0 is embedded by shifting the phase towards the I-axis by $\theta$. Further, successive decoding is utilized in ATM, i.e., the decoding of the message signal is followed by the decoding of the authentication signal.

In the following discussion, we compare HMM-DSA and ATM when the message signal is modulated using QPSK. Fig. 4b illustrates 8 possible constellation symbols when ATM is utilized with QPSK modulated message signal. For comparison with HMM-DSA, in ATM, the same authentication bit is repeatedly embedded into N message symbols, where N is the same number as the block length in HMM-DSA. This means that the number of message symbols transmitted for each authentication bit are the same for ATM and HMM-DSA.

6.1. Resource overhead

Embedding the authentication signal into the message signal requires applying changes to the message signal itself, and thus may incur some PHY-layer resource overhead. Examples of this overhead include increase in average transmission power, increase in bandwidth, decrease in message throughput, and increase in complexity of the transmitter and/or receiver.

The overall average transmission power and bandwidth are kept unchanged from the conventional message signaling. By design, ATM and HMM-DSA have the same message throughput as the conventional message signaling. ATM and HMM-DSA also have the same authentication rate.

In terms of transmitter’s and aware receiver’s computational complexity, ATM is advantageous compared to HMM-DSA. To implement ATM, Alice and Bob only need to modify how the message and the authentication symbols are mapped to the constellation symbols. However, the implementation of HMM-DSA is more complex—Alice needs to add controlled ISI to the message symbols using the modified duobinary filter, and Bob requires the use of MLSD to extract the message and the authentication signals.

Specifically, at Bob, the computational complexity of decoding each message signal is $O(M^{2})$ due to MLSD employed by HMM-DSA, where M represents the order of the modulation scheme of the message signal. In ATM, the corresponding computational complexity is $O(M)$. Note that in spite of its high computational complexity, MLSD is a standard technique for decoding signals in modern communication systems due to its error performance advantage. Also, note that given a particular modulation scheme, the computational complexity of demodulating and decoding one message symbol does not change by a change in N, the length of the trellis in MLSD.

6.2. Error performance

ATM as well as HMM-DSA intentionally corrupts the message signal to insert the authentication signal. Hence, the error performance of $M_{S}$ degrades as the authentication signal becomes more prominent. Here, we compare the BER expression of $M_{S}$ in HMM-DSA given by Eq. (2) with the BER expression of $M_{S}$ in ATM given in Tan et al. (2011). In order to limit the degradation in the error performance of $M_{S}$ to the same extent, i.e., to achieve the same BER of $M_{S}$ in ATM and HMM-DSA, we obtain the relationship between $\theta$ (phase shift in ATM) and δ (ISI in HMM-DSA), to be $\delta = \tan \theta$.

Fig. 8 shows the BER vs. $E_{b}/N_{0}$ curves for the message and authentication signals in HMM-DSA with $\delta = 0.3$, and those in ATM with $\theta = \arctan \delta$. We use $N=3$ for both, HMM-DSA and ATM. In the figure, we observe that the BER of $M_{S}$ is higher in ATM as compared to HMM-DSA. Note that since we use $\theta = \arctan \delta$ for ATM, the BER curve labeled “$M_{S}$ in ATM” in Fig. 8, would also correspond to the BER curves of $M_{S}$ in ATM as well as $\overline{AS}$ in HMM-DSA. Hence, we deduce that in ATM, the technique of SSD is employed at both the aware and the unaware receivers, and there is no way by which the aware receiver can outperform the unaware receiver in terms of the error performance of the message signal. However, in HMM-DSA, the aware receiver can improve the error performance of the message signal by utilizing
MLSD. Note that this performance can be further improved by decreasing the authentication rate, i.e., by increasing $N$. Hence, from Fig. 8 and the above discussion, we infer that HMM-DSA has a significant error performance advantage over ATM in terms of $\tilde{M}_\delta$.

However, in Fig. 8, we observe that the BER of the authentication signal, $\tilde{M}_\delta$, is lower in ATM as compared to HMM-DSA. In fact, the curve labeled “$\tilde{M}$ in ATM” is given by the Eq. (5) which is the lower bound for the BER of $\tilde{M}_\delta$ in HMM-DSA. We note that each transmitted sample is independent in ATM. This means that the message and authentication symbols are independently estimated for each received sample in ATM. However, in HMM-DSA, due to modified duobinary filtering, each received sample is correlated to the previous received sample. This correlation decreases the BER of the message signal as MLSD can be utilized to estimate the sequence of the message symbols. On the other hand, this correlation increases the BER of the authentication signal as the estimation of the authentication symbol is affected by errors in the estimation of not only the current message symbol, but also the previous message symbol (see Eq. (1)).

Table 3 provides an illustrative example of the comparison between ATM and HMM-DSA with $N=3$, $E_b/N_0 = 6$ dB, $\delta = 0.3$, and $\theta = \arctan \delta$.

<table>
<thead>
<tr>
<th></th>
<th>BER of $\tilde{M}_\delta$</th>
<th>BER of $\tilde{M}_\delta$</th>
<th>BER of $\tilde{M}_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM Tan et al. (2011)</td>
<td>0.0148</td>
<td>0.0148</td>
<td>0.0249</td>
</tr>
<tr>
<td>HMM-DSA</td>
<td>0.0148</td>
<td>0.0055</td>
<td>0.0382</td>
</tr>
</tbody>
</table>

From the above discussion, we observe that, on one hand, HMM-DSA performs better than ATM in terms of the error performance of the message signal at the aware receiver, $\tilde{M}_\delta$; on the other hand, HMM-DSA performs worse than ATM in terms of the error performance of the authentication signal, $\tilde{M}_\delta$.

To remove the uncertainty in the comparison of HMM-DSA and ATM, we compare them using the curve proposed in Tan et al. (2011), which we call as the receiver error characteristics (REC) curve. The REC curve for an aware receiver is generated by using the BER of the message signal on the Y-axis, and the BER of the authentication signal on the X-axis. In Fig. 9, we plot REC curves of HMM-DSA and ATM. The curve for HMM-DSA is obtained by varying $\delta$, and the curve for ATM is obtained by varying $\theta$. We utilize $N=3$ and $E_b/N_0 = 6$ dB for both, HMM-DSA and ATM. According to the REC curves shown in the figure, we can observe that HMM-DSA clearly outperforms ATM.

For a PHY-layer authentication scheme to be viable, the aware receiver must be able to decode both the message and the authentication signals with sufficiently good error performance. Considering this requirement, and the REC curves of HMM-DSA and ATM shown in Fig. 9, we conclude that HMM-DSA enjoys a significant advantage over ATM in terms of the error performance of the aware receiver.

6.3. Comparison between HMM-DSA and the scheme proposed in Yu et al. (2008b)

It is important to note that HMM-DSA is different from the 4/16-QAM based scheme proposed in Yu et al. (2008b) which we refer to as Amplitude based Hierarchical Modulation for Authentication (AHMA). Although both, HMM-DSA and AHMA, employ hierarchical modulation resulting in the same constellation symbols as shown in Fig. 4a, each of them uses a different approach for generating the embedded signal. HMM-DSA uses the modified duobinary filtering to generate the embedded signal, whereas AHMA embeds the authentication signal into the message signal as noise. From the expressions of BER for HMM-DSA, AHMA in Yu et al. (2008b) and ATM in Tan et al. (2011), it can be inferred that when $\theta = \arctan \delta$, HMM-DSA, AHMA and ATM have the same error performance for the message signal at the unaware receiver, $\tilde{M}_\delta$. Here, $\delta$ in HMM-DSA and AHMA, and $\theta$ in ATM are the deviations of the constellation symbols of the message signal from the optimal positions to embed the authentication signal. Further, when $\theta = \arctan \delta$, AHMA and ATM have the same error performance for the message and authentication signals at the aware receiver. Hence, the above discussions on the comparison of ATM and HMM-DSA also apply to the comparison of AHMA and HMM-DSA.

7. Experimental validation

To evaluate the performance characteristics of HMM-DSA in a testbed environment, we implemented HMM-DSA using three USRP radios: (1) Alice (transmitter), (2) Bob (aware receiver), and (3) Charlie (unaware receiver). We used National Instruments’ LabVIEW as the system-design platform to configure the USRPs.

7.1. Design

The bits in the message signal are generated using a long message text, and transmitted without any error correction coding. The authentication signal is also generated using an authentication text without any error correction coding. Alice utilizes QPSK as the modulation scheme for the message signal. The authentication signal is embedded into the message signal using HMM-DSA with a block length of $N=3$, and ISI of $\delta = 0.3$. Alice utilizes cyclic prefix based OFDM for transmitting the embedded signal over 1 MHz bandwidth. Some of the sub-carriers in OFDM are used for transmitting pilot symbols. The pilot symbols are used by Bob and Charlie to estimate the channel. The conventional processes like performing inverse fast fourier transform (IFFT), and adding cyclic prefix are performed to generate the OFDM symbols. Further, OFDM frames are generated by adding the preamble symbols. The preamble symbols are appended to facilitate Bob and Charlie with the time and frequency synchronization. Finally, the embedded signal is transmitted over-the-air at the center frequency of 915 MHz.

We start the USRP radios of Bob and Charlie before starting Alice’s transmission. Hence, Bob and Charlie receive all the transmitted samples. After achieving time and frequency synchronization using the preamble and pilot symbols, Bob and Charlie demodulate and decode the received signal. Bob extracts the message and authentication signals, and Charlie extracts only the message signal. At Bob, the received message and authentication bits are compared with the transmitted message and authentication bits to calculate their BERs, respectively. Similarly, at Charlie, the received message bits are compared with the transmitted message bits to calculate the BER.
curves for the message and authentication scheme referred to as HMM-DSA relaxes this constraint, and improves the error performance of the message signal at the aware receiver as compared to that at the unaware receiver. However, this advantage over the prior art is achieved at the cost of higher computational complexity of the aware receiver.

8. Conclusion

In this paper, we proposed a novel PHY-layer transmitter authentication scheme referred to as Hierarchical Modulation with Modified Duobinary Signaling for Authentication (HMM-DSA). One of the biggest drawbacks of most existing schemes is that the error performance of the message signals at the aware and unaware are the same. HMM-DSA relaxes this constraint, and improves the error performance of the message signal at the aware receiver as compared to that at the unaware receiver. However, this advantage over the prior art is achieved at the cost of higher computational complexity of the aware receiver.

References


FCC, 2015. Shared commercial operations in the 3550–3650 MHz band, Federal Register, vol. 80, no. 120, June.


