WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. **00**: 1–15 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/wcm.0000

On Cracking Direct-Sequence Spread-Spectrum Systems †

Youngho Jo and Dapeng Wu*
Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, U.S.A.

Summary

Secure transmission of information over hostile wireless environments is desired by both military and civilian parties. Direct-sequence spread spectrum (DS-SS) is such a covert technique resistant to interference, interception and multipath fading. Identifying spread-spectrum signals or cracking DS-SS systems by an unintended receiver (or eavesdropper) without a priori knowledge is a challenging problem. To address this problem, we first search for the start position of data symbols in the spread signal (for symbol synchronization); our method is based on maximizing the spectral norm of a sample covariance matrix, which achieves smaller estimation error than the existing method of maximizing the Frobenius norm. After synchronization, we remove a spread sequence by a cross-correlation based method, and identify the spread sequence by a matched filter. The proposed identification method is less expensive and more accurate than the existing methods. We also propose a zigzag searching method to identify a generator polynomial that reduces memory requirement and is capable of correcting polarity errors existing in the previous methods. In addition, we analyze the bit error performance of our proposed method. The simulation results agree well with our analytical results, indicating the accuracy of our analysis in additive white Gaussian noise (AWGN) channel. By simulation, we also demonstrate the performance improvement of our proposed schemes over the existing methods. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: Direct-Sequence Spread-Spectrum; PN sequence; code generator polynomial; linear feedback shift register; cracking; probability of error

1. Introduction

In this paper, we consider the problem of eavesdropping on the adversary's communication, which uses Direct-Sequence Spread-Spectrum (DS-SS). The DS-SS is a covert communication technique; the information symbols are modulated by a pseudorandom noise (PN) sequence prior to transmission. This results in a wideband signal, which is resistant to interference, jamming, interception and multipath fading [1, 2, 3].

To eavesdrop on the adversary's communication, one needs to (a) identify the start position of data symbols in the intercepted spread signal for the purpose of symbol synchronization, (b) estimate data symbols, (c) estimate the PN sequence, and (d) estimate the code generator polynomial of the PN sequence.

To identify the start position of data symbols, we present a method based on the spectral norm which achieves smaller estimation error in Section 4. After the symbol synchronization, we remove a PN sequence from the intercepted signals by a correlation method to estimate data symbols without a priori knowledge about that PN sequence in Section 5. Identification of a PN sequence is processed by a matched filter between the intercepted signal and the estimated data symbols in Section 6.

One of the harder problems in eavesdropping

*Correspondence to: Prof. Dapeng Wu, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, U.S.A. wu@ece.ufl.edu, http://www.wu.ece.ufl.edu.

Prepared using wcmauth.cls *[Version: 2007/01/09 v1.00]*

on DS-SS signals is the polarity ambiguity of the estimated spread sequence and data symbols: Erroneous reversal of polarity of each chip in the estimated PN sequence compared to the true PN sequence is a major source of the performance degradation of an eavesdropper. Therefore, we need to estimate a code generator polynomial to mitigate this polarity problem. We propose a searching method to identify the code generator polynomial in order to correct polarity errors as well as to reduce memory requirement of an eavesdropper in Section 7. Saving hundreds or thousands of sequence bits in the memory of an eavesdropper is very expensive.

The probability of error performance of an eavesdropper is a function of signal-to-noise ratio (SNR), the number of data symbols, and the length of the spread sequence of the intercepted signal. Therefore, we need to study the analytical probability of error performance of the eavesdropper with respect to these parameters. By doing so, we can efficiently predict the performance of the eavesdropper.

First, we use a Gaussian approximation method in Ref. [4] in order to find a marginal probability density function of the symbol estimator in Section 5 and the sequence estimator in Section 6, respectively. Second, we find the probability of error of the symbol detector without the proposed code generator estimator as a sum of products of error functions and frequencies of the number of errors in the estimated spread sequence. Finally, we compare the probability of error with and without the proposed generator polynomial estimator in Section 7.

The contributions of this paper are: (i) a generator polynomial estimator which can identify a code generator polynomial and can correct polarity errors in the estimated PN sequence and estimated data symbols, (ii) a theoretical verification of the probability of error of a code generator estimator with respect to signal-to-noise ratio (SNR), the number of data symbols, and the length of the spread sequence of the intercepted signal, and (iii) the accuracy of performance prediction of an eavesdropper.

The remainder of this paper is organized as follows: Related works are discussed in Section 2. Section 3 describes the signal model. Section 4 introduces our method of identifying the start position of a data symbol in the spread signal. Section 5 presents how to remove data symbols from the intercepted signal. Then, estimation of a spread sequence is presented in Section 6. Section 7 discusses how to identify a PN code generator polynomial and how to correct polarity errors. Section 8 presents simulation results to show the effectiveness and to validate the analytical probability of error of our approaches. Section 9 concludes this paper.

2. Related Works

Wireless communications are very common both for military and commercial parties. The ability to use communication while mobile has great benefits for both parties. However, wireless communication has many security issues, since communication takes place over a wireless channel while the users are usually mobile. Such a wireless channel suffers from a number of vulnerabilities: (i) The channel is vulnerable to eavesdropping. (ii) The data can be altered. (iii) The absence of wired link makes it much easier to cheat on identities. (iv) The channel can be overused. (v) Finally, the channel can be jammed, notably in order to perpetrate a denial-of-service (DoS) attack [5, 6].

To eavesdrop on the adversary's communication which uses DS-SS, the estimation of the spread sequence from the intercepted signal is a key to crack on these DS-SS systems and is a challenging problem. The literature on this subject is not rich. We briefly discuss some related works which have studied this problem.

First, an eavesdropper needs to detect any transmission of DS-SS signals in order to crack a secure DS-SS signal. A method based on the fluctuation of an autocorrelation estimator, instead of on the autocorrelation itself, was proposed in [7]. The fluctuation of the autocorrelation estimator was used to estimate an accurate spread code period. Since the intercepted signal may experience delay, the interceptor must find the start position of a data symbol in the spread signal. To identify the start position of a data symbol in the spread signal, Ref. [8] proposed a correlation-based method. A method of maximizing the Frobenius norm of a covariance of the intercepted signal was proposed in [9]. However, the Frobenius norm may result in the increase of estimation error as the period of the PN sequence increases; hence, their method does not work well for the PN sequence of a long period. To address this limitation, a method based on the spectral norm which achieves smaller estimation error than the Frobenius norm based method is proposed in Section 4.

Second, to identify the PN sequence, several methods were proposed in the literature [8, 10, 11, 12]. In Ref. [10], a method based on a multichannel identification technique was proposed to recover the convolution between the PN sequence and the

DOI: 10.1002/wcm

channel response for blind channel estimation; the limitation of this method is high computational complexity. In Ref. [12], a method based on principal component analysis (PCA) was used to estimate the PN sequence from eigenvectors corresponding to the first and the second largest eigenvalues of the sample covariance matrix; however, the computational complexity required by PCA is high. Ref. [8] suggested the use of chip-by-chip detection to estimate the PN sequence and to use a parallel processing to combat the polarity ambiguity in successive demodulation and decoding. However, their parallel processing approach has a limitation: It increases memory requirement and does not mitigate the polarity error. Ref. [11] proposed a multiple subsection cross-correlation averaging method to estimate the PN sequence; however, the method used only half of the captured symbols. It is known that the more data symbols used, the more accurate the estimation is. In Section 6, we propose a cross-correlation based method that uses all captured symbols and achieves higher estimation accuracy.

Third, to correct the polarity ambiguity in the estimated spread sequence and data symbols, an eavesdropper needs to estimate a code generator polynomial. The estimators used in Refs. [11, 12] did not consider the problem of polarity errors in the estimated PN sequence, i.e., erroneous reversal of polarity of each chip in the estimated PN sequence (compared to the true PN sequence). Therefore, the probability of correct estimation of the PN sequence, using their estimators, may be less than 50%. This leads to significant performance degradation in terms of bit error rate (BER) or symbol error rate (SER). We solve this problem by identifying the PN code generator polynomial in Section 7. Not only is it important to estimate the PN sequence, but we also need to identify the PN code generator. Identifying the PN code generator polynomial improves the accuracy of estimating the PN sequence and data symbols by a factor of two, over the methods proposed in Refs. [11, 12].

In [8], the probability of error of their sequence estimator was analyzed for each chip of a spread sequence. They found a marginal probability density function of that sequence estimator by a numerical integration. However, they did not consider the polarity ambiguity in their correct estimation probability analysis. Therefore, their analysis had a limitation. In Section 7, we consider the polarity ambiguity in the analysis of the probability of error of a sequence estimator and a symbol detector. We also provide a complete expression for the probability of error of a symbol detector in Section 7.

3. Signal Model

A baseband representation of a DS-SS signal is given by [10, 12]:

$$y(t) = \sum_{l=-\infty}^{+\infty} a_l h(t - kT_s) + n(t) \qquad (1)$$

$$h(t) = \sum_{k=0}^{P-1} c_k p(t - kT_c) \qquad (2)$$

where T_s is the symbol duration, and a_l is a QPSK or BPSK modulated symbol transmitted at time kT_s[†]. We assume the symbols a_l are centered and uncorrelated. Let $n(t)$ denote the noise at the output of the received filter and the noise is additive white Gaussian noise (AWGN) and uncorrelated with the information signal a_l. The effect of the transmitter filter, the reception filter, the channel response and the pseudo-random sequence c_k is represented by $h(t)$. Let $p(t)$ denote the convolution of all filters of the transmission chain. T_c is the chip duration and $\{c_k\}_{k=0,\cdots,P-1}$ is the pseudo-random sequence of length P where $P = T_s/T_c$. In this paper, we assume the symbol duration T_s can be estimated by the method in Ref. [7] for simplicity. Note that we consider the AWGN channel only in this paper. Our study can be extended to multipath environments if we use a blind channel estimation method proposed in Ref. [10]. However, in the current work, we limit ourselves to the AWGN channel for simplicity. Table I lists the notations used in this paper.

4. Symbol Synchronization

The captured signal $y(t)$ in (1) is sampled and divided into non-overlapping windows with the eavesdropper's sampling duration T_{ev}. We assume the sampling duration of an eavesdropper is the chip duration for simplicity, however this is not a requirement of our method. Therefore, $P \cdot (L+1)$ samples are available by sampling $(L+1) \cdot T_s$ long signal with the sample duration T_c. Rewriting $P \cdot$

[†] For reasons of simplicity and clarity of presentation, we only focus on the QPSK/BPSK modulation. If we adopt a blind modulation detection method, our work can be applied to higher order modulation like 64-QAM with little modification.

Table I. List of Notations.

$\Pr(\cdot)$	probability of $(\cdot)$	P_b	probability of error
$(\cdot)^H$	conjugate transpose of $(\cdot)$	$(\cdot)^T$	transpose of $(\cdot)$
$\mathrm{Re}(\cdot)$	real part of $(\cdot)$	$\mathrm{Im}(\cdot)$	imaginary part of $(\cdot)$
GF	Galois field or finite field	F	Field
$\mathbf{E}(\cdot)$	expectation of matrix $(\cdot)$	$\mathrm{sgn}(x)$	sign of x
$\mathbf{a}$	a vector (or set) of sequence (symbol)	$\mathbf{a}^*$	$-\,\mathrm{sgn}(\mathbf{a})$
$\lfloor\cdot\rfloor$	the nearest integer less than equal $(\cdot)$	$\hat{a}$	estimation of a
$<\mathbf{x},\mathbf{y}>$	inner product of $\mathbf{x}$ and $\mathbf{y}$	$\mathrm{erfc}(\cdot)$	complementary error function of $(\cdot)$
$\|\cdot\|_2$	spectral norm of $(\cdot)$	$\|\cdot\|_F$	Frobenius norm of $(\cdot)$
$\mathcal{N}$	normal distribution	$\mathcal{B}$	binomial distribution
λ	eigenvalue	$a_{i,j}$	the $(i,j)^{th}$ entry of a matrix $\mathbf{A}$
L	the number of synchronized data symbols	P	the length of the spread sequence

$(L+1)$ samples as a matrix $\mathbf{y}^k$ with dimension $P \times (L+1)$, we have:

$$\mathbf{y}^k = \begin{bmatrix} \mathbf{y}_{-1}^k & \cdots & \mathbf{y}_{l-1}^k & \mathbf{y}_l^k & \cdots & \mathbf{y}_{L-1}^k \end{bmatrix} \quad (3)$$

where the superscript k represents the kT_c time-delayed desynchronized signal of (1) for $k = 0, \cdots, P-1$. Let $\mathbf{y}_l^k$ denote a column of the desynchronized $\mathbf{y}^k$. We may write $\mathbf{y}_l^k$ as follows:

$$\mathbf{y}_l^k = \begin{bmatrix} y_{l,k} & \cdots & y_{l,P-1} & y_{l+1,0} & \cdots & y_{l+1,k-1} \end{bmatrix}^T = \begin{bmatrix} a_l h_{l,k} + n_k \\ \vdots \\ a_l h_{l,P-1} + n_{P-1} \\ a_{l+1} h_{l+1,0} + n_0 \\ \vdots \\ a_{l+1} h_{l+1,k-1} + n_{k-1} \end{bmatrix} \quad (4)$$

where $[\cdot]^T$ denotes the transpose, $y_{l,k}$ is the k^{th} entry of a column y_l^k and $h_{l,k}$ is the spreading sequence of $y_{l,k}$. Now, we can modify (4) as follows:

$$\begin{aligned} \mathbf{y}_l^k &= \begin{bmatrix} h_{l,k} & 0 \\ \vdots & \vdots \\ h_{l,P-1} & 0 \\ 0 & h_{l+1,0} \\ \vdots & \vdots \\ 0 & h_{l+1,k-1} \end{bmatrix} \begin{bmatrix} a_l \\ a_{l+1} \end{bmatrix} + \begin{bmatrix} n_k \\ \vdots \\ n_{P-1} \\ n_0 \\ \vdots \\ n_{k-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{h}_l^e & \mathbf{h}_{l+1}^b \end{bmatrix} \mathbf{a}_l^k + \mathbf{n}^k \\ &= \mathbf{h}_l^k \, \mathbf{a}_l^k + \mathbf{n}^k \end{aligned} \quad (5)$$

where $\mathbf{h}_l^e$ denotes a vector containing the end of the spreading waveform for a duration of $T_s - kT_c$ followed by zeroes for a duration kT_c; $\mathbf{h}_{l+1}^b$ is a vector containing zeroes for a duration $T_s - kT_c$ followed by the beginning of the spreading waveform for a duration kT_c; $\mathbf{a}_l^k$ denotes a vector containing two desynchronized symbols a_l and a_{l+1}; $\mathbf{n}^k$ stands for the noise. Therefore, it is necessary to make a column $\mathbf{y}_l^k$ have only one data symbol a_l. That is:

$$\begin{aligned} \mathbf{y}_l^0 &= \begin{bmatrix} y_{l,0} \cdots y_{l,k-1} \; y_{l,k} \cdots y_{l,P-1} \end{bmatrix}^T \\ &= \mathbf{h}_l^0 a_l + \mathbf{n}^0 \end{aligned} \quad (6)$$

and (3) becomes:

$$\mathbf{y}^0 = \begin{bmatrix} \mathbf{y}_0^0 \cdots \mathbf{y}_{l-1}^0 \; \mathbf{y}_l^0 \cdots \mathbf{y}_{L-1}^0 \end{bmatrix} \quad (7)$$

Note that samples which belong to a_{-1} and a_L in (3) are truncated in the synchronized intercepted signal (7). Let $\mathbf{R}$ denote the covariance matrix of (5).

$$\begin{aligned} \mathbf{R} &= \mathbf{E}[\mathbf{y}_l^k {\mathbf{y}_l^k}^H] \\ &= \mathbf{h}_l^k \mathbf{E}\left[\mathbf{a}_l^k {\mathbf{a}_l^k}^H\right] {\mathbf{h}_l^k}^H + \sigma_n^2 \mathbf{I}_P \end{aligned} \quad (8)$$

where $[\cdot]^H$ is the conjugate transpose, $\mathbf{I}_P$ represents a $P \times P$ identity matrix, $\mathbf{E}[\cdot]$ denotes expectation, and σ_n^2 is the noise variance.

To place the starting spread sequence $h_{l,0}$ in the proper position in (4), we search for a maximum of the spectral norm of the sample covariance matrix of (8). The spectral norm of a matrix is the square root of the largest eigenvalue of $\mathbf{R}$ in Ref. [13]. Let $\|\mathbf{y}\|_2$ denote the spectral norm of the square covariance matrix.

$$\|\mathbf{y}\|_2 = \sqrt{\lambda_{\max}(\mathbf{R})} \quad (9)$$

where $\lambda_{\max}(\mathbf{R})$ stands for the largest eigenvalue of the covariance matrix. Then, the spectral norm of (7) is:

$$\|\mathbf{y}_l^0\|_2 = \|\mathbf{h}_l^0\|^2 E\left[|a_l|^2\right] + \sigma_n^2 \tag{10}$$

However, the spectral norm of (5) is:

$$\|\mathbf{y}_l^k\|_2 = \begin{cases} \|\mathbf{h}_l^e\|^2 E\left[|a_l|^2\right] + \sigma_n^2 & \text{if } kT_c \le \frac{T_s}{2} \\ \|\mathbf{h}_{l+1}^b\|^2 E\left[|a_{l+1}|^2\right] + \sigma_n^2 & \text{if } kT_c > \frac{T_s}{2} \end{cases} \tag{11}$$

if the singular values are expressed in decreasing order. Since $\|\mathbf{h}_l^0\|^2 \ge \|\mathbf{h}_l^e\|^2$ or $\|\mathbf{h}_l^0\|^2 \ge \|\mathbf{h}_{l+1}^b\|^2$, we can determine the synchronized version of (3) by maximizing the spectral norm in (9) with respect to $k = 0, \cdots, P-1$ as follows:

$$\begin{aligned} \hat{\mathbf{y}}^0 &= \underset{k\in[0,P-1]}{\operatorname{argmax}} \|\mathbf{y}^k\|_2 \\ &= \underset{k\in[0,P-1]}{\operatorname{argmax}} \sqrt{\lambda_{\max}(\mathbf{R})} \\ &= \underset{k\in[0,P-1]}{\operatorname{argmax}} \sqrt{\lambda_{\max}\left(\mathbf{E}\left[\mathbf{y}^k\mathbf{y}^{k^H}\right]\right)} \end{aligned} \tag{12}$$

In Ref. [9], the Frobenius norm was used to search for the start position of a data symbol. Note that the square of the Frobenius norm $\|\mathbf{y}\|_F^2$ is the sum of squares of the singular values of $\mathbf{y}$. There are errors in the eigenvalue decomposition of the sample covariance $\hat{\mathbf{R}}$ due to the noise according to matrix perturbation theory [13]. The expected value of the perturbation error of the Frobenius norm is $P^2 \cdot \sigma_n^2$, while that of the spectral norm is σ_n^2 [13]. The Frobenius norm has a tendency to increase the mean square error (MSE) as the spread sequence length increases. Thus, their method does not perform well for long length sequences. To mitigate this limitation, we use the spectral norm in (12).

Fig. 1 shows the theoretical and simulated squared spectral norm $\|\mathbf{y}^k\|_2$ in (12). For the calculation, 10,000 trials are carried out and averaged together. In the simulation, we use QPSK. The PN sequence is an m-sequence [1, 14, 15] with the length $P = 31$ and with a generator polynomial $f(x) = 1 + x^2 + x^5$. The SNR is -5dB. When $k = 0$, the spectral norm has a peak. Note that the more samples, the more accurate estimation of $\hat{\mathbf{y}}^0$ in (12) can be achieved.

We also compare the MSEs in the estimation of the time-delay kT_c, $\mathbf{E}[(\hat{k} - k)^2]$, between the spectral norm and the Frobenius norm. The same simulation parameters are used in Fig. 1, except that SNR is varied from -20dB to 5dB. Fig. 2(a) shows that the spectral norm has smaller MSEs than the Frobenius norm, when the sequence length $P = 31$ is fixed and the number of symbols L is 128, 256 and 512. We can synchronize the captured signal with fewer symbols by the spectral norm. Fig. 2(b) shows the case with the fixed number of symbols $L = 128$ and with the varied length of spread sequence P =15, 31, and 63. As the length P increases, the MSE is increased by a factor of P^2; that is, the MSE, normalized by the squares of the sequence length P^2, is almost the same.

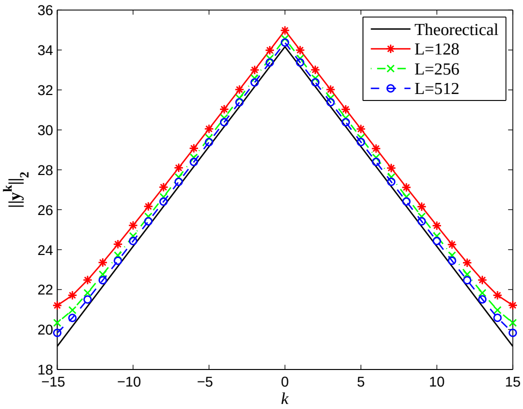
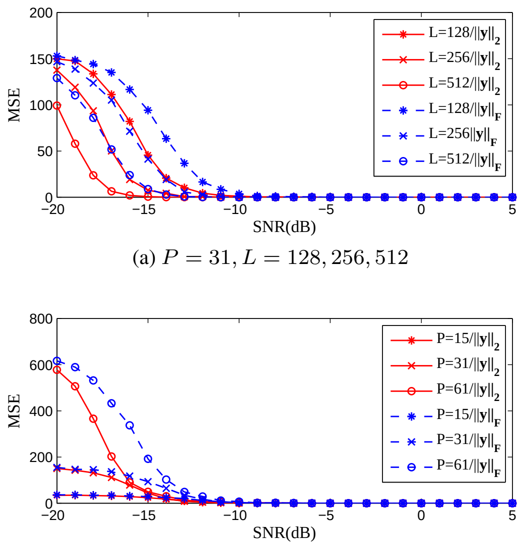


Fig. 1. Theoretical and simulated spectral norm, $\|\mathbf{y}^k\|_2$, in (12) with $P = 31$ and SNR=-5dB



(a) $P = 31, L = 128, 256, 512$

(b) $L = 128, P = 15, 31, 63$

Fig. 2. Comparison of MSE, $\mathbf{E}[(\hat{k} - k)^2]$, by the spectral norm vs. Frobenius norm

DOI: 10.1002/wcm

5. Symbol Estimation

After symbol synchronization, we need to remove the spread sequence in (6) to estimate the information symbol a_l from the synchronized signal $\mathbf{y}_l^0$ in (7). With the property of strong self-correlation and weak cross-correlation of spread spectrum, we use a method based on a cross-correlation between a test column, say $\mathbf{y}_t^0$, and a column of a data symbol a_l, say $\mathbf{y}_l^0$, of the synchronized signal in (7). Then, we have:

$$C_{\mathbf{y}_t^0\mathbf{y}_l^0}(\tau)=\mathbf{y}_t^0\mathbf{y}_l^{0H}(\tau) \tag{13}$$

If the spread sequence is an m-sequence [1, 14, 15], $C_{\mathbf{y}_t^0\mathbf{y}_l^0}(\tau=0) \geq C_{\mathbf{y}_t^0\mathbf{y}_l^0}(\tau \neq 0)$. Then,

$$C_{\mathbf{y}_t^0\mathbf{y}_l^0}(0)=\sum_{k=0}^{P-1} y_{t,k}y_{l,k}^* \tag{14}$$

Now we can estimate the symbol a_l from (14) as follows:

$$\begin{aligned}\hat{a}_l = \text{sgn}\left[\text{Re}\left(C_{\mathbf{y}_t^0\mathbf{y}_l^0}(0)\right)\right] \\ + j \cdot \text{sgn}\left[\text{Im}\left(C_{\mathbf{y}_t^0\mathbf{y}_l^0}(0)\right)\right]\end{aligned} \tag{15}$$

where $\text{Re}(\cdot)$ takes the real part and $\text{Im}(\cdot)$ takes the imaginary part of a complex. $\text{sgn}(x)$ is the sign function with value 1, if $x > 0$, and -1 otherwise. Note that the estimated symbol $\hat{a}_l$ in (15) is estimated up to an unknown multiplicative factor. Therefore, the sign of the symbol in (15) can be reversed by this multiplicative factor. This problem was not considered in Refs. [11, 12]. We will solve this problem by estimating a code generator polynomial in Section 7.

In order to analyze the performance of the symbol estimator in (14), we use a Gaussian approximation of the sum of products of two random variables in order to find a marginal probability density function of the symbol estimator in (14). Let $q_{l,k} = y_{t,k}y_{l,k}^*$, $a_t = a_0$, and $h_{0,k} = h_{l,k} = c_k$ in (14) for simplicity.

$$\begin{aligned} q_{l,k} &= (a_0c_k + n_{0,k})\,(a_lc_k + n_{l,k})^* \\ &= c_k\left(a_0 + \frac{n_{0,k}}{c_k}\right)c_k^*\left(a_l^* + \frac{n_{l,k}^*}{c_k^*}\right) \\ &= |c_k|^2\left(a_0 + \frac{n_{0,k}}{c_k}\right)\left(a_l^* + \frac{n_{l,k}^*}{c_k^*}\right)\end{aligned} \tag{16}$$

Assume $a_0 = 1$ and $a_l = 1$ for generality. Then,

$$q_{l,k} = \epsilon\,(1 + \sigma\bar{n}_{0,k})\left(1 + \sigma\bar{n}_{l,k}^*\right) \tag{17}$$

where $\epsilon = |c_k|^2$, $\sigma^2 = \sigma_n^2/\epsilon$, and $\bar{n}_{l,k}$ is a Gaussian random process with zero mean and a unit variance, i.e., $\bar{n}_{l,k} \sim \mathcal{N}(0,1)$. From the above definitions, we have the SNR $\rho = \epsilon/\sigma_n^2 = 1/\sigma^2$.

Let $\alpha^+ = 1 + \sigma\bar{n}_{0,k}$, $\beta^+ = 1 + \sigma\bar{n}_{l,k}$, and $\gamma^+ = \alpha^+\beta^+$. Since $\bar{n}_{l,k} \sim \mathcal{N}(0,1)$, $\alpha^+ \sim \mathcal{N}(1,\sigma^2)$ and $\beta^+ \sim \mathcal{N}(1,\sigma^2)$. We want to find the marginal probability density $f(\gamma^+)$ which is a product of two normal distributions. To find the marginal density $f(\gamma^+)$, we need to integrate the product of a conditional distribution $f(\gamma^+|\beta^+)$ and a marginal distribution $f(\beta^+)$ with respect to β^+.

$$\begin{aligned} f(\gamma^+) &= \int_{-\infty}^{\infty} f(\gamma^+|\beta^+)f(\beta^+)d\beta^+ \\ &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\frac{1}{|\beta^+|}\exp\left[-\frac{1}{2\sigma^2}\left(\frac{\gamma^+ - \beta^+}{\beta^+}\right)^2\right]d\beta^+ \\ &+ \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\frac{1}{|\beta^+|}\exp\left[-\frac{1}{2\sigma^2}(\beta^+ - 1)^2\right]d\beta^+ \end{aligned} \tag{18}$$

The integration in (18) can be obtained using a numerical integration, a Monte Carlo, or a Gaussian approximation with a given ρ [4]. Among these three methods, we use an approximation of products of two normal distributions. Let μ_{γ^+} denote the mean and $\sigma_{\gamma^+}^2$ denote the variance of γ^+. Since α^+ and β^+ are independent of each other, the mean and variance of γ^+ are:

$$\begin{aligned}\mu_{\gamma^+} &= \mathbf{E}[\alpha^+\beta^+] \\ &= \mathbf{E}[\alpha^+]\mathbf{E}[\beta^+] = \mu_{\alpha^+}\mu_{\beta^+} = 1\end{aligned} \tag{19}$$

$$\begin{aligned}\sigma_{\gamma^+}^2 =& \mathbf{E}[(\alpha^+\beta^+)^2] - (\mathbf{E}[\alpha^+\beta^+])^2 \\ =& \mathbf{E}[(\alpha^+)^2]\mathbf{E}[(\beta^+)^2] - \mu_{\alpha^+}^2\mu_{\beta^+}^2 \\ =& \mu_{\alpha^+}^2\sigma_{\beta^+}^2 + \mu_{\beta^+}^2\sigma_{\alpha^+}^2 + \sigma_{\alpha^+}^2\sigma_{\beta^+}^2 \\ =& 2\sigma^2 + \sigma^4\end{aligned} \tag{20}$$

The mean and variance of $\epsilon\gamma^+$ are $\epsilon\mu_{\gamma^+}$ and $\epsilon^2\sigma_{\gamma^+}^2$. Since data symbols $\{q_{l,k}\}$ are independent of each other and have the same distribution, the distribution of $\hat{a}_l$ will be approximately normally distributed according to the central limit theorem [16]. Therefore, the distribution of $\hat{a}_l$ is a normal distribution $\mathcal{N}(\epsilon P\mu_{\gamma^+}, \epsilon^2 P\sigma_{\gamma^+}^2)$ for $P \gg 1$.

DOI: 10.1002/wcm

Under the condition that $a_0 = 1$, the probability of the error estimation of $\hat{a}_l$ is

$$\begin{aligned} & p(\hat{a}_l < 0|a_l = +1) \\ & \quad = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\epsilon^2 P\sigma_{\gamma+}^2}} \exp\left[-\frac{(x-\epsilon P\mu_{\gamma+})^2}{2\epsilon^2 P\sigma_{\gamma+}^2}\right] dx \\ & \quad = \frac{1}{2}\,\mathrm{erfc}\left(\mu_{\gamma+}\sqrt{\frac{P}{2\sigma_{\gamma+}^2}}\right) \end{aligned} \tag{21}$$

where

$$\mathrm{erfc}\,(x) = \frac{2}{\sqrt{\pi}}\int_{x}^{\infty}\exp\,(-t^2)dt \tag{22}$$

For $a_l = -1$, with similar notations $\mu_{\gamma-}$ and $\sigma_{\gamma-}^2$ and using the same procedure, we have:

$$\begin{aligned} & p(\hat{a}_l > 0|a_l = -1) \\ & \quad = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi\epsilon^2 P\sigma_{\gamma-}^2}} \exp\left[-\frac{(x-\epsilon P\mu_{\gamma-})^2}{2\epsilon^2 P\sigma_{\gamma-}^2}\right] dx \\ & \quad = 1 - \frac{1}{2}\,\mathrm{erfc}\left(\mu_{\gamma-}\sqrt{\frac{P}{2\sigma_{\gamma-}^2}}\right) \\ & \quad = \frac{1}{2}\,\mathrm{erfc}\left(\mu_{\gamma+}\sqrt{\frac{P}{2\sigma_{\gamma+}^2}}\right) \end{aligned} \tag{23}$$

since $\mu_{\gamma-} = -\mu_{\gamma+}$ and $\sigma_{\gamma-}^2 = \sigma_{\gamma+}^2$. The bit-error rate of estimation of the symbol a_l is:

$$\begin{aligned} P_b^a &= p(\hat{a}_l > 0|a_l = -1)p(a_l = -1) \\ &+ p(\hat{a}_l < 0|a_l = +1)p(a_l = +1) \\ &= \frac{1}{2}\,\mathrm{erfc}\left(\mu_{\gamma+}\sqrt{\frac{P}{2\sigma_{\gamma+}^2}}\right) \end{aligned} \tag{24}$$

6. Spread Sequence Estimation

To recover the PN sequence in (6), we use a matched filter operation between the synchronized intercepted signal $\mathbf{y}^0$ in (7) and the estimated data symbols $\hat{\mathbf{a}}$ in (15). That is:

$$\hat{\mathbf{c}} = \mathrm{sgn}\left(<\mathbf{y}^0, \hat{\mathbf{a}}>\right) \tag{25}$$

where $\langle\cdot,\cdot\rangle$ denotes inner product, $\hat{\mathbf{a}} = [\hat{a}_0, \cdots, \hat{a}_{L-1}]^T$ is a vector of the estimated symbol, and $\hat{\mathbf{c}} = [\hat{c}_0, \cdots, \hat{c}_{P-1}]^T$ stands for a vector of the estimated spread sequence. The sign of the sequence in (25) can also be reversed by a multiplicative factor in (15).

To analyze the performance of the sequence estimator in (25), we use the same Gaussian approximation method in Section 5. The sequence estimator (25) can be rewritten as follows:

$$\begin{aligned} \hat{c}_k &= \sum_{l=0}^{L-1} y_{l,k}\cdot\hat{a}_l \\ &= \sum_{l=0}^{L-1} (a_l h_{l,k} + n_{l,k})\cdot\hat{a}_l \end{aligned} \tag{26}$$

Let $\omega_{l,k} = y_{l,k}\bar{a}_l$, and $h_{l,k} = c_k$ for simplicity. Assume $c_k = 1$ for generality. Then we have:

$$\begin{aligned} \omega_{l,k} &= (a_l c_k + n_{l,k})\,\hat{a}_l \\ &= a_l\left(c_k + \frac{n_{l,k}}{a_l}\right)\hat{a}_l \\ &= \sqrt{\epsilon}\,(1+\sigma\bar{n}_{l,k})\,\hat{a}_l \end{aligned} \tag{27}$$

where $\sigma^2 = \sigma_n^2/\epsilon$, $\epsilon = |a_l|^2$. Let $u^+ = 1 + \sigma\bar{n}_{l,k}$, $\gamma^+ = \bar{a}_l$, and $w^+ = u^+\gamma^+$. Since $\bar{n}_{l,k} \sim \mathcal{N}(0,1)$, $u^+ \sim \mathcal{N}(1,\sigma^2)$. We need to find a marginal probability density $f(w^+)$ which is a product of two normal distributions as (18):

$$f(w^+) = \int_{-\infty}^{\infty} f(w^+|\gamma^+)f(\gamma^+)d\gamma^+ \tag{28}$$

Let μ_{w+} denote the mean and σ_{w+}^2 denote the variance of w^+. Since u^+ and γ^+ are independent, the mean and variance of w^+ are:

$$\begin{aligned} \mu_{w+} &= \mathbf{E}[u^+\gamma^+] \\ &= \mathbf{E}[u^+]\mathbf{E}[\gamma^+] = \mu_{u+}\mu_{\gamma+} = 1 \end{aligned} \tag{29}$$

$$\begin{aligned} \sigma_{w+}^2 =& \mathbf{E}[(u^+\gamma^+)^2] - (\mathbf{E}[u^+\gamma^+])^2 \\ =& \mathbf{E}[(u^+)^2]\mathbf{E}[(\gamma^+)^2] - \mu_{u+}^2\mu_{\gamma+}^2 \\ =& \mu_{u+}^2\sigma_{\gamma+}^2 + \mu_{\gamma+}^2\sigma_{u+}^2 + \sigma_{u+}^2\sigma_{\gamma+}^2 \\ =& 3\sigma^2 + 3\sigma^4 + \sigma^6 \end{aligned} \tag{30}$$

Therefore, the mean and variance of $\sqrt{\epsilon}w^+$ are $\sqrt{\epsilon}\mu_{w+}$ and $\epsilon\sigma_{w+}^2$. Since the sequence $\{\omega_{l,k}\}$ is independent of each other and has the same distribution, the distribution of $\hat{c}_k$ will be approximately normally distributed according to the central limit theorem [16]. Therefore, the distribution of $\hat{c}_k$ is a normal distribution $\mathcal{N}(\sqrt{\epsilon}L\mu_{w+}, \epsilon L\sigma_{w+}^2)$ for $L \gg 1$.

DOI: 10.1002/wcm

Under the condition that $c_k = 1$, the probability of the error estimation of $\hat{c}_k$ is:

$$p(\hat{c}_k < 0 | c_k = +1)$$
$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\epsilon L\sigma_{w^+}^2}} \exp\left[-\frac{(x-\sqrt{\epsilon}L\mu_{w^+})^2}{2\epsilon L\sigma_{w^+}^2}\right] dx$$
$$= \frac{1}{2}\,\mathrm{erfc}\left(\mu_{w^+}\sqrt{\frac{L}{2\sigma_{w^+}^2}}\right) \tag{31}$$

For $c_k = -1$, with similar notations μ_{w^-} and $\sigma_{w^-}^2$ and using the same procedure, we have:

$$p(\hat{c}_k > 0 | c_k = -1)$$
$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\epsilon L\sigma_{w^-}^2}} \exp\left[-\frac{(x-\sqrt{\epsilon}L\mu_{w^-})^2}{2\epsilon L\sigma_{w^-}^2}\right] dx$$
$$= \frac{1}{2}\,\mathrm{erfc}\left(\mu_{w^+}\sqrt{\frac{L}{2\sigma_{w^+}^2}}\right) \tag{32}$$

since $\mu_{w^-} = -\mu_{w^+}$ and $\sigma_{w^-}^2 = \sigma_{w^+}^2$.

The probability of error P_b^c in estimation of the sequence c_k is:

$$\begin{aligned} P_b^c &= p(\hat{c}_k \neq c_k) \\ &= p(\hat{c}_k > 0 | c_k = -1)p(c_k = -1) \\ &+ p(\hat{c}_k < 0 | c_k = +1)p(c_k = +1) \\ &= \frac{1}{2}\,\mathrm{erfc}\left(\mu_{w^+}\sqrt{\frac{L}{2\sigma_{w^+}^2}}\right) \end{aligned} \tag{33}$$

Note that the probability of error in (24) and (33) does not account for the polarity error. We will address this problem in the succeeding section.

7. Identification of Generator Polynomial

It is expensive to save hundreds or thousands of sequence bits, say $\hat{\mathbf{c}} = [\hat{c}_0, \cdots, \hat{c}_{P-1}]^T$, in the memory of the eavesdropper. This motivates us to estimate a generator polynomial of the estimated spread sequence from (25).

Shift registers are the practical and efficient implementation for the spread sequence. We considered a linear feedback shift register (LFSR) as the implementation technique for PN sequence. The correct selection of the n-tuple tap-weights (or n feedback stages) will result in a maximal sequence of the length $N = 2^n - 1$ [1, 14].

Let $F = GF(q)$ where q is a prime or a power of a prime where GF[‡] denotes a finite field and q is called the order of the field F [14]. If the feedback function $f(x_0, \cdots, x_{n-1})$ is a linear function; that is, if it can be expressed as

$$\begin{aligned} &f(x_0, \cdots, x_{n-1}) = \\ &\qquad w_0x_0 + \cdots + w_{n-1}x_{n-1}, \quad w_i \in F \end{aligned} \tag{34}$$

where w_i denotes a tap weight of the LFSR for $i = 0, \cdots, n-1$ over F. Then, sequences have the linear recursion relation [14]:

$$c_{k+n} = \sum_{i=0}^{n-1} w_i c_{k+i}, \quad k = 0, 1, 2, \ldots \tag{35}$$

If we have $2n$ successive sequence bits, we can estimate the generator polynomial of sequence $\mathbf{c} = (c_0, \cdots, c_{P-1})$ over $F = GF(q)$. We may rewrite the recursion relation (35) into the following matrix representation [14]:

$$\begin{bmatrix} c_n \\ c_{n+1} \\ \vdots \\ c_{2n-1} \end{bmatrix} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_1 & c_2 & \cdots & c_n \\ \vdots & \ddots & \ddots & \vdots \\ c_{n-1} & c_n & \cdots & c_{2n-2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{n-1} \end{bmatrix} \tag{36}$$

We can solve the recursion relation in (36) over $GF(q)$ to obtain tap weights $\mathbf{w} = (w_0, \cdots, w_{n-1})$. The next successive sequence bit can be generated and tested with the estimated tap weights $\hat{\mathbf{w}}$ and n successive sequences using transform matrix M of LFSR as follows [14]:

$$M = \begin{bmatrix} 0 & 0 & \cdots & 0 & \hat{w}_0 \\ 1 & 0 & \cdots & 0 & \hat{w}_1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \hat{w}_{n-1} \end{bmatrix} \tag{37}$$

and

$$\begin{aligned} &(\hat{c}_{k+1}, \hat{c}_{k+2}, \cdots, \hat{c}_{k+n}) \\ &\quad = (\hat{c}_0, \hat{c}_1, \cdots, \hat{c}_{n-1})M^{k+1} \end{aligned} \tag{38}$$

Note that $det(M) = (-1)^n \hat{w}_0$ and thus M is invertible if and only if $\hat{w}_0 \neq 0$.

The method we propose is called a "*zigzag estimator*" which searches for a generator polynomial primarily based on (36) and (38) from the estimated

[‡] Finite fields are called Galois fields. See [14].

DOI: 10.1002/wcm

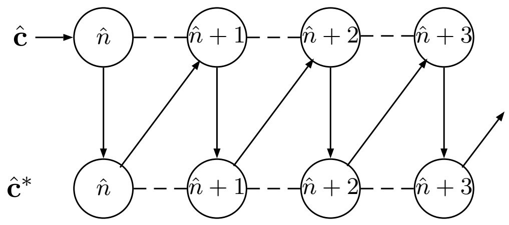


Fig. 3. A graphical illustration of the zigzag estimator where $\hat{n} = \lfloor \log_2(P+1) \rfloor$ and $\hat{\mathbf{c}}^* = -\operatorname{sgn}(\hat{\mathbf{c}})$

sequence $\hat{\mathbf{c}}$ in (25), and also corrects the polarity error in the estimated sequence $\hat{\mathbf{c}}$ and data symbol $\hat{\mathbf{a}}$. A graphical representation of the zigzag estimator is given in Fig. 3. A fundamental idea in correcting signs of the estimated sequence is that the zigzag estimator returns the non-sign reversed sequence if the zigzag estimator can find a code generator polynomial from the estimated sequence $\hat{\mathbf{c}}$. Let us introduce some mathematical notations. Let $\mathbf{c} = (c_0, c_1, \cdots)$ be the set of sequence or symbols. Let $\hat{\mathbf{c}}$ denote the estimated version of $\mathbf{c}$ and $\mathbf{c}^*$ be the sign-flipped version of $\mathbf{c}$, i.e., $\mathbf{c}^* = -\operatorname{sgn}(\mathbf{c})$. Let $\lfloor n \rfloor$ denote the nearest integer less than or equal to n. Following is an algorithm for the zigzag estimator. Note that we do not claim the algorithm herein is optimal or sub-optimal.

(Step 1) Estimate $\hat{\mathbf{c}}$ by (25). Store $\mathbf{s} \leftarrow \hat{\mathbf{c}}$ ($\mathbf{s}$ temporal memory of $\hat{\mathbf{c}}$). Initialize TestFlag $\leftarrow$ 0, FlipCount $\leftarrow$ 0, and $\hat{n} \leftarrow \lfloor \log_2(P+1) \rfloor$

(Step 2) Estimate $\hat{f}_{\hat{n}}(x)$ by (36).

(Step 3) Generate $2\hat{n}$ successive $\hat{\alpha}_i$ by (38). Increase FlipCount← FlipCount+1.

(Step 4) Test $\hat{c}_i = \hat{\alpha}_{i,i=2\hat{n},\cdots,4\hat{n}-1}$.

- (4a) If $\hat{c}_i = \hat{\alpha}_{i,i=2\hat{n},\cdots,4\hat{n}-1}$, set TestFlag $\leftarrow$ 1 and go to (Step 6).
- (4b) If $\hat{c}_i \neq \hat{\alpha}_{i,i=2\hat{n},\cdots,4\hat{n}-1}$, set TestFlag $\leftarrow$ 0 and go to (Step 5).

(Step 5) If TestFlag=0, flip the sequence $\hat{\mathbf{c}} \leftarrow \hat{\mathbf{c}}^*$

- (5a) If FlipCount=1, increase FlipCount← FlipCount+1. Go to (Step 2).
- (5b) If FlipCount=2, increase $\hat{n} \leftarrow \hat{n}+1$ and reset FlipCount $\leftarrow$ 0. Go to (Step 2).

(Step 6) Check polarity errors. If $\mathbf{s} \neq \hat{\mathbf{c}}$, $\hat{\mathbf{c}} \leftarrow \hat{\mathbf{c}}^*$. Store $\hat{\mathbf{c}}_{zigzag} \leftarrow \hat{\mathbf{c}}$.

Note that the method proposed in this section can also be applied to Gold codes. Some pairs of *m*-sequences with the same degree can be used to generate Gold Codes by linearly combining two *m*-sequences with different offsets in Galois field. If the estimated generator polynomial can be decomposed into two preferred pairs of m-sequence, we can decompose the estimated generator polynomial into two m-sequences. For example, a Gold code generator $f(x) = 1 + x + x^2 + x^3 + x^4 + x^5 + x^6$ can be factored into $f_1(x) = 1 + x + x^3$ and $f_2(x) = 1 + x^2 + x^3$[§].

Finally, we use a matched filter operation between the intercepted signal $\mathbf{y}^s$ and the sign corrected estimated sequence $\hat{\mathbf{c}}_{zigzag}$ in (Step 6). That is:

$$\hat{\mathbf{a}}_{zigzag} = \operatorname{sgn}\left(< \mathbf{y}^0, \hat{\mathbf{c}}_{zigzag} >\right) \tag{39}$$

However,

$$\hat{\mathbf{a}}_{non-zigzag} = \operatorname{sgn}\left(< \mathbf{y}^0, \hat{\mathbf{c}} >\right) \tag{40}$$

Now we are ready to find the probability of error of the zigzag estimator in (39). First, we will find the probability of error of the symbol detector without the zigzag estimator of (40). After that, the probability of the symbol detector by the zigzag estimator (39) will be analyzed. A symbol detector of a cooperative receiver (*Rx*) can be written as follows:

$$\begin{aligned} \hat{\mathbf{a}} &= \operatorname{sgn}\left(\langle \mathbf{y}, \mathbf{c} \rangle\right) = \operatorname{sgn}\left(\langle \mathbf{c}\mathbf{a} + \sigma\bar{\mathbf{n}}, \mathbf{c} \rangle\right) \\ &= \operatorname{sgn}\left(\mathbf{a}\langle \mathbf{c}, \mathbf{c} \rangle + \sigma \langle \bar{\mathbf{n}}, \mathbf{c} \rangle\right) \end{aligned} \tag{41}$$

where $\bar{\mathbf{n}} \sim \mathcal{N}(0,1)$. Then, signal-to-noise ratio ρ_{Rx} of the cooperative receiver is:

$$\rho_{Rx} = \frac{\langle \mathbf{c}, \mathbf{c} \rangle^2}{\sigma^2 \|\mathbf{c}\|^2} = \frac{P}{\sigma^2} \tag{42}$$

Therefore, the probability of error of the cooperative receiver with the known spread sequence $\mathbf{c}$ is [17]:

$$P^a_{b,Rx} = \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{P}{2\sigma^2}}\right) \tag{43}$$

However, signal-to-noise ratio ρ_{Ev} of the eavesdropper (*Ev*) is:

$$\rho_{Ev} = \frac{\langle \mathbf{c}, \hat{\mathbf{c}} \rangle^2}{\sigma^2 \|\hat{\mathbf{c}}\|^2} = \frac{\langle \mathbf{c}, \hat{\mathbf{c}} \rangle^2}{\sigma^2 P} \tag{44}$$

[§] See 4.4 Decomposition of LFSR sequences in [14].

DOI: 10.1002/wcm

Table II. Relationship among $\Pr(K=k)$ in (45), $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$, and $\frac{1}{2}\operatorname{erfc}\left(\sqrt{\rho_{Ev}}\right)$ in (47) where k denotes the number of errors in the estimation of the spread sequence of the non-zigzag method.

k	0	1	$\cdots$	P
$\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$	P	$P-2$	$\cdots$	$-P$
$\frac{1}{2}\operatorname{erfc}\left(\sqrt{\rho_{Ev}}\right)$	$\frac{1}{2}\operatorname{erfc}\left(P\sqrt{\frac{1}{2\sigma^2 P}}\right)$	$\frac{1}{2}\operatorname{erfc}\left((P-2)\sqrt{\frac{1}{2\sigma^2 P}}\right)$	$\cdots$	$\frac{1}{2}\operatorname{erfc}\left(-P\sqrt{\frac{1}{2\sigma^2 P}}\right)$
$\Pr(K=k)$	$f(0;P,P_b^c)$	$f(1;P,P_b^c)$	$\cdots$	$f(P;P,P_b^c)$

In Section 6, we find the probability of error P_b^c in the estimation of each chip in (33). The number of incorrect estimations of each chip c_k in the spread sequence $\mathbf{c}$ is an independent yes/no experiment with a fail probability P_b^c. Let K denote the number of errors in the estimation of the spread sequence $\mathbf{c}$ of the length P. Then, we can write $K \sim \mathcal{B}(P, P_b^c)$. The probability of getting exactly k errors in the estimation of the spread sequence $\mathbf{c}$ with the length of P is given by:

$$\begin{aligned}\Pr(K=k) &= f(k;P,P_b^c)\\ &= \binom{P}{k}\left(P_b^c\right)^k\left(1-P_b^c\right)^{P-k}\end{aligned} \tag{45}$$

for $k = 0, 1, 2, \cdots, P$ with the binomial coefficient

$$\binom{P}{k} = \frac{P!}{(P-k)!k!} \tag{46}$$

If there is no error in the estimation of the spread sequence $\mathbf{c}$, i.e., $\mathbf{c} = \hat{\mathbf{c}}$ or $k = 0$, $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P$ and $\rho_{Ev}(K=0) = 1/\sigma^2$. If there is only one error in the estimation of the spread sequence $\mathbf{c}$, i.e., $k = 1$, $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P-2$ and $\rho_{Ev}(K=1) = (P-2)^2/(\sigma^2 P)$. Table II shows the relationship among $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$, ρ_{Ev} and $\Pr(K=k)$ in (47) and (45). Therefore, the probability of error in the estimation of symbol without the zigzag estimator can be written as follows:

$$P^a_{b,non-zigzag} = \sum_{k=0}^{P} \frac{1}{2}\operatorname{erfc}\left((P-2k)\sqrt{\frac{1}{2\sigma^2 P}}\right)\Pr(K=k) \tag{47}$$

Since $f(k;P,P_b^c) = f(P-k;P,1-P_b^c)$, the probability of k errors in the estimation of the spread sequence is the same as that of $P-k$ correct estimations of the spread sequence. Let Q denote the number of correct estimations of the spread sequence and $q = P-k$. If the length of the spread sequence P is odd,

$$\begin{aligned}P^{a,odd}_{b,non-zigzag} &= \\ &\sum_{k=0}^{\lfloor P/2 \rfloor} \frac{1}{2}\operatorname{erfc}\left((P-2k)\sqrt{\frac{1}{2\sigma^2 P}}\right) f(k;P,P_b^c)\\ &+ \sum_{q=0}^{\lfloor P/2 \rfloor} \frac{1}{2}\operatorname{erfc}\left((2q-P)\sqrt{\frac{1}{2\sigma^2 P}}\right) f(q;P,1-P_b^c)\end{aligned} \tag{48}$$

If P is even,

$$\begin{aligned}P^{a,even}_{b,non-zigzag} &= \\ &\sum_{k=0}^{P/2-1} \frac{1}{2}\operatorname{erfc}\left((P-2k)\sqrt{\frac{1}{2\sigma^2 P}}\right) f(k;P,P_b^c)\\ &+ \binom{P}{P/2}\left(P_b^c\right)^{P/2}\left(1-P_b^c\right)^{P/2}\\ &+ \sum_{q=0}^{P/2-1} \frac{1}{2}\operatorname{erfc}\left((2q-P)\sqrt{\frac{1}{2\sigma^2 P}}\right) f(q;P,1-P_b^c)\end{aligned} \tag{49}$$

The meanings of (48) and (49) are (i) the probability distribution of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$ is symmetric, (ii) the probability distribution of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle > 0$ follows $f(k;P,P_b^c)$, and (iii) the probability distribution of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle < 0$ follows $f(q;P,1-P_b^c)$.

The zigzag estimator in (39) can identify the PN code generator polynomial to correct the polarity error in the estimation of the spread sequence if $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = -P$ or $k = P$ or $q = 0$. Then, the probability of error in the estimation of symbols with the zigzag estimator $P^a_{b,zigzag}$ is:

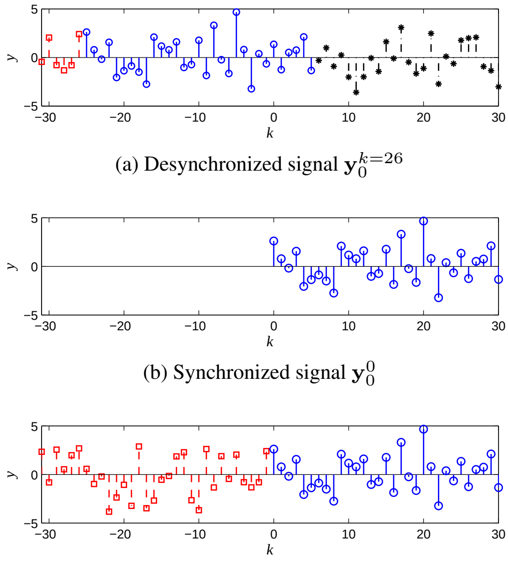
$$\begin{aligned}P^a_{b,zigzag} &= \\ &\sum_{k=0}^{P-1} \frac{1}{2}\operatorname{erfc}\left((P-2i)\sqrt{\frac{1}{2\sigma^2 P}}\right)\Pr(K=k)\\ &+ \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{P}{2\sigma^2}}\right)\Pr(K=P)\end{aligned} \tag{50}$$

DOI: 10.1002/wcm

The probability of error of (50) by the zigzag estimator is enhanced by a factor of two, compared to (47) without the zigzag estimator, if the zigzag estimator can estimate a code generator polynomial. We will validate the probability of error of the symbol detector in (39) and (40) in the following Section 8.

8. Simulation and Validation

In this section, we present a complete simulation example to illustrate our approaches. We consider information symbols modulated by an m-sequence with the generator polynomial $f(x) = 1 + x^2 + x^5$ of length $N = P = 31$. Signal constellation is BPSK and the number of data symbols is $L = 128$. The received signal is corrupted by AWGN noise with SNR=-10dB. We assume the sampling rate of an eavesdropper is the chip rate T_c, however this is not required by our method.



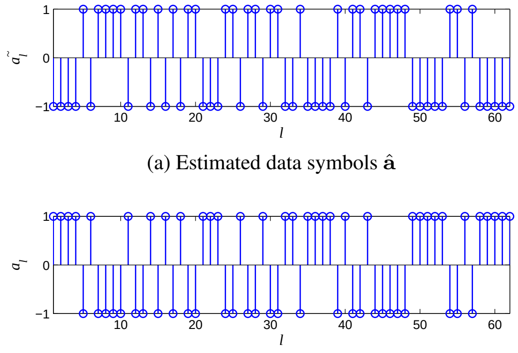
(a) Desynchronized signal $\mathbf{y}_0^{k=26}$

(b) Synchronized signal $\mathbf{y}_0^0$

(c) True signal $\mathbf{y}$

Fig. 4. Symbol synchronization by the spectral norm where the dashed □ denotes a_{-1}, the solid ○ denotes a_0, the dashed-dotted $*$ denotes a_1, respectively

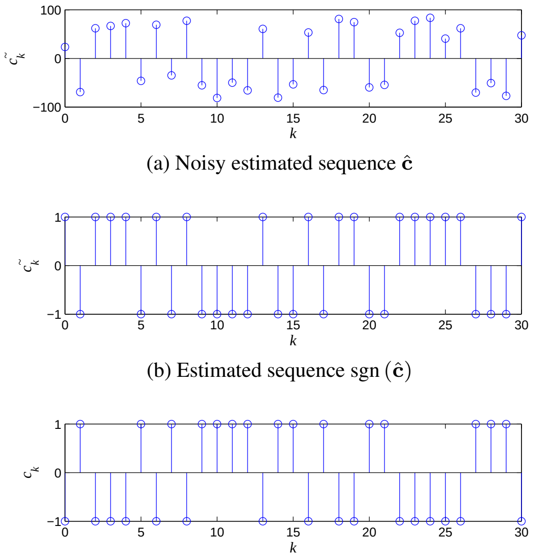
First, we need to determine a synchronized version of the intercepted signal by (12). Fig. 4(a) shows a desynchronized signal $\mathbf{y}^{k=26}$ delayed by $26T_c$. Therefore, a sample window $\mathbf{y}_{-1}^{k=26}$ contains the end of a symbol a_{-1} for a duration of $5T_c$ followed by the beginning of next symbol signal a_0 for a duration $26T_c$. A synchronized signal $\mathbf{y}_0^{k=0}$ by (12) is shown in Fig. 4(b). Note that the desynchronized samples which belong to a_{-1} are truncated. Fig. 4(c) shows



(a) Estimated data symbols $\hat{\mathbf{a}}$

(b) True data symbols $\mathbf{a}$

Fig. 5. Data symbols estimation by (15)

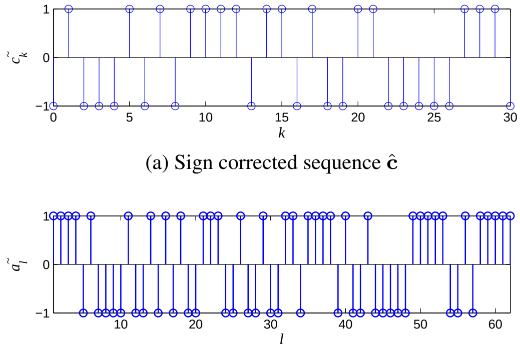


(a) Noisy estimated sequence $\hat{\mathbf{c}}$

(b) Estimated sequence sgn $(\hat{\mathbf{c}})$

(c) True sequence $\mathbf{c}$

Fig. 6. Spread sequence estimation by (25)



(a) Sign corrected sequence $\hat{\mathbf{c}}$

(b) Sign corrected data symbol $\hat{\mathbf{a}}$

Fig. 7. Sign correction by the zigzag estimator

DOI: 10.1002/wcm

two synchronized sample windows for the purpose of comparison.

Second, estimations of data symbols and spread sequences are followed by the symbol synchronization. Fig. 5 shows the first 62 estimated data symbols $\hat{\mathbf{a}}$ by (15). Fig. 6 shows the noisy estimated sequence $\hat{\mathbf{c}}$ by (25). Note that the estimated data $\hat{\mathbf{a}}$ in Fig. 5(a) and the estimated sequence in Fig. 6(b) are sign reversed versions of the true symbol $\mathbf{a}$ and the true sequence $\mathbf{c}$, respectively. Therefore, $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = -P$. We can correct polarity errors in the estimated data symbol $\hat{\mathbf{a}}$ and the spread sequence $\hat{\mathbf{c}}$ by the proposed zigzag estimator.

Third, we can correct polarity errors in the estimated data symbol $\hat{\mathbf{a}}$ and the spread sequences $\hat{\mathbf{c}}$ by the zigzag estimator. The zigzag estimator in Section 7 searches and tests a generator polynomial from the estimated sequence $\hat{\mathbf{c}}$ by the recursion relation in (36) and the transform matrix in (37). Fig. 7 shows the sign corrected sequence $\hat{\mathbf{c}}_{zigzag}$ and data symbol $\hat{\mathbf{a}}_{zigzag}$ by the proposed zigzag estimator.

We also conduct a simulation to evaluate the performance of the proposed zigzag estimator. The generator polynomial used in this simulation is an m-sequence with $f(x) = 1 + x + x^{11} + x^{12} + x^{14}$ and the sequence is truncated, $P < N = 2^n - 1$ for comparison. We randomly seed initial conditions and randomly generate data symbols corrupted by AWGN noise. The signal constellation is BPSK. 10,000 simulation trials are carried out and averaged.

First, we compare the histogram of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$ between with the zigzag estimator and without the zigzag estimator to verify the relation between (47) and (50). In this simulation, we use the number of data symbol $L = 128$, the length of the spread sequence $P = 64$, and SNR=-5dB. Fig. 8 shows the comparison of histograms $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$ between with the zigzag estimator and without the zigzag estimator. The horizontal axis is the value $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$ and the vertical axis represents the normalized occurrence frequency of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$. The occurrence frequency of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P$ is 0.9991 and 0.4939 for with the zigzag estimator and without the zigzag estimator, respectively. However, the occurrence frequency of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = -P$ with the zigzag estimator and without the zigzag estimator is 0.0000 and 0.5052. The zigzag estimator in Section 7 can identify the PN code generator polynomial to correct the polarity error in the estimation of the spread sequence when $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = -P$. Therefore, the occurrence frequency $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P$ with the zigzag estimator is the sum of that of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P$ and $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = -P$ without the zigzag estimator. This validates the relation between (47) and (50).

Second, we compare $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ between with the zigzag estimator and without the zigzag estimator in (47) and (50), respectively. Fig. 9 shows the $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ with different combinations of the number of data symbols L and the length of the spread sequence P. Note that the $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ corresponds to $\Pr(K = 0)$ without the zigzag estimator and $\Pr(K = 0) + \Pr(K = P)$ with the zigzag estimator. The $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ increases as the number of the intercepted symbols L increased and also increases as the length of the spread sequence P increased. Therefore, the $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ obtained by the zigzag estimator is almost two times greater than that without our proposed zigzag estimator.

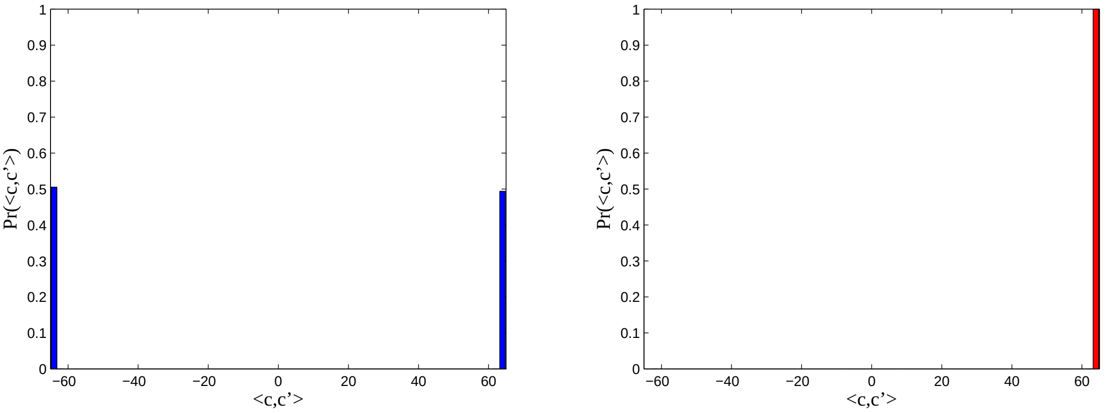
Third, we compare the simulated probability of error of the symbol detector in (39) and the analytical probability of error in (50). Fig. 10 shows the $P^a_{b,zigzag}$ with different combinations of the number of data symbols L and the length of the spread sequence P. Note that $P_{b,Rx}/P = 64$ denotes the probability of error of a cooperative receiver in (43) with $P = 64$ for comparison. The $P^a_{b,zigzag}$ is enhanced as the number of samples L increases and as the length of sequence P increases. Moreover, the analytical performance $P^a_{b,zigzag}$ of the symbol detector in (50) is almost the same as that of the simulated probability of error in (40). Since the simulated $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P) \simeq 1$ for $P = 64$, $L = 128$, SNR=-5dB in Fig. 9 over 10,000 trials, $P^a_{b,zigzag} \simeq P_{b,Rx}$. Therefore, the analytical probability of error in (50) is a good approximation of the performance of an eavesdropper. This analytical performance can provide an efficient prediction of the performance of our proposed zigzag estimator.

Finally, we conduct a simulation with the n-tuple code generator polynomial to validate the performance of the proposed zigzag method with the long length sequence. The length of the spread sequence is $P = 2^n - 1$. We randomly seed initial conditions and randomly generate data symbols corrupted by AWGN noise with SNR = -10dB. The signal constellation is BPSK and 10,000 simulation trials are carried out and averaged. Fig. 11 shows the probability of the correct estimation of the spread sequence $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ with $n = 6, \cdots, 13$ with the number of data symbols $L = 256$. The $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ increases as the length of the spread sequence $P = 2^n - 1$ increased.

9. Conclusion

In this paper, we consider the problem of eavesdropping on the adversary's communication, which uses Direct-Sequence Spread-Spectrum (DS-SS). To

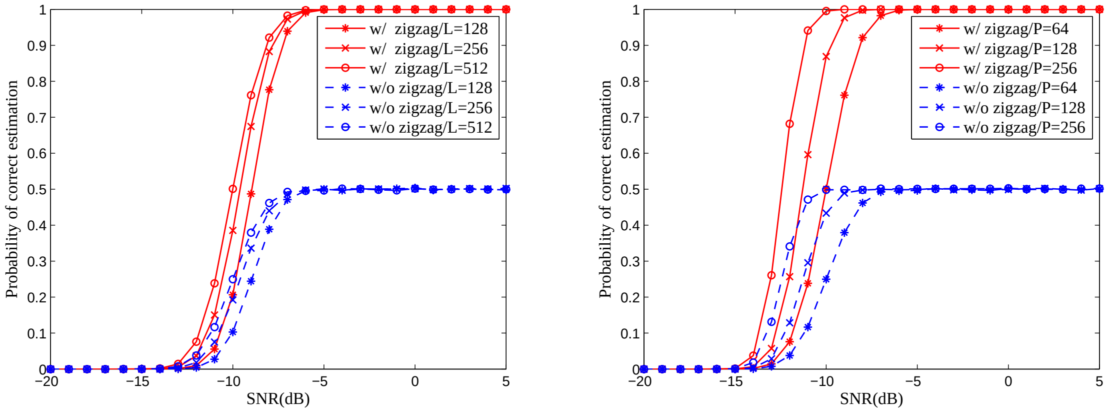
DOI: 10.1002/wcm



(a) w/o the zigzag estimator

(b) w/ the zigzag estimator

Fig. 8. Histogram of $\langle \mathbf{c}, \hat{\mathbf{c}} \rangle$ with $P = 64$, $L = 128$, and SNR=-5dB



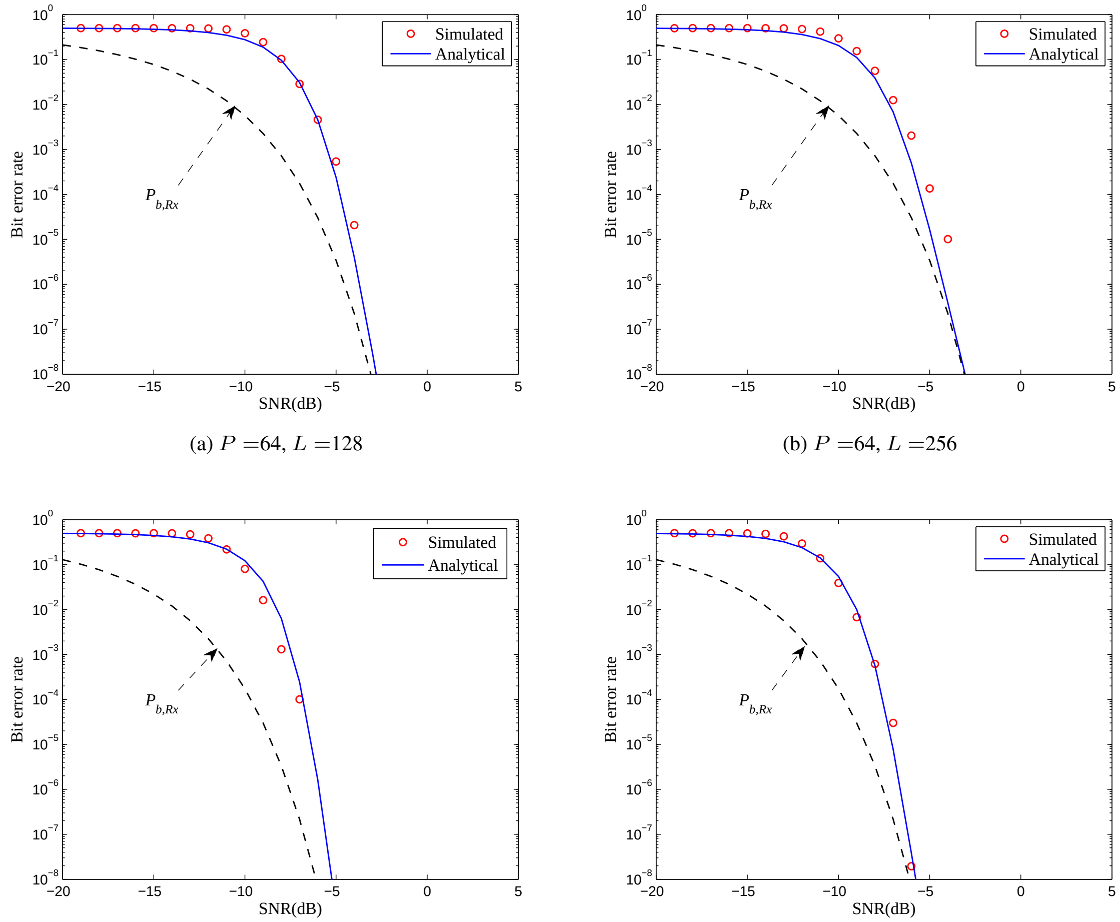
(a) P =64, L =128, 256, 512

(b) L =512, P =64, 128, 256

Fig. 9. Comparison of the probability of correct estimation of the spread sequence $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$

intercept the adversary's communication, one needs to (a) identify the start position of a data symbol in the spread signal for symbol synchronization purpose, (b) remove the PN sequence, (c) estimate the PN sequence, and (d) estimate the generator polynomial. In this paper, we propose effective methods to address these four problems. To identify the start position of a data symbol, we developed a method that uses the spectral norm of the sample covariance matrix. After symbol synchronization, a method based on the cross-correlation was used to estimate data symbols up to an unknown multiplicative factor. These estimated symbols were used by a matched filtering operation for identifying the PN sequence from the intercepted signal. In addition to obtaining the PN sequence and the data symbols, we also proposed a zigzag estimator to identify the PN code generator polynomial and proposed a method to identify the polarity in the received signal. Our method improves the probability of correct estimation by a factor of two, compared to the previous method. We also analyze the probability of error of the zigzag estimator in terms of SNR, the number of intercepted data symbols, and the length of the spread sequence. Our validation by simulation and theoretical analysis show the effectiveness of our proposed method. Our proposed method can be used by an interceptor to eavesdrop on an adversary's communication. Other applications of our method

DOI: 10.1002/wcm



(a) P =64, L =128

(b) P =64, L =256

(c) P =128, L =256

(d) P =128, L =512

Fig. 10. Comparison of the simulated and analytical probability of bit error $P^a_{b,zigzag}$ with the zigzag estimator in (50) where $P_{b,Rx}$ is the probability of error of an cooperative receiver (Rx)

include altering an adversary's information, isolating an adversary's communication link, and jamming by a denial-of-service (DoS) attack by our smart eavesdropper. Furthermore, our method can be used for the synchronization by estimating the received spread code by an intended receiver.

References

1. Simon MK, Omura JK, Scholtz RA, Levitt BK. *Spread spectrum communications handbook (revised ed.)*. McGraw-Hill, Inc.: NY, USA, 1994.
2. Scholtz RA. The Origins of Spread-Spectrum Communications May 1982; **COM-30, No.5**:822–854.
3. Peterson RL, Borth DE, Ziemer RE. *An Introduction to Spread-Spectrum Communications*. Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1995.
4. Ware R, Lad F. Approximating the Distribution for Sums of Products of Normal Variables. *Research Report UCDMS 2003/15*, University of Canterbury, Christchurch, New Zealand 2003.
5. Buttyan L, Hubaux JP. *Security and Cooperation in Wireless Networks*. Cambridge University Press: Cambridge.
6. Poisel AR. *Modern Communications Jamming Principles and Techniques*. Artech House Publishers, November 30, 2003.
7. Burel G. Detection of Spread Spectrum Transmissions using fluctuations of correlation estimators. *IEEE Int. Symp. on Intelligent Signal Processing and Communication Systems (1SPACS'2000)* 2000; .
8. Zhan Y, Cao Z, Lu J. Spread-spectrum sequence estimation for DSSS signal in non-cooperative communication systems. *IEE Proceedings Communications* Aug 2005; **152**(4):476–480, doi:10.1049/ip-com:20045197.
9. Bouder C, Azou S, Burel G. A robust synchronization procedure for blind estimation of the symbol period and the timing offset in spread spectrum transmissions. *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications* 2002; **1**:238–241, doi:10.1109/ISSSTA.2002.1049322.
10. Tsatsanis MK, Giannakis GB. Blind estimation of direct

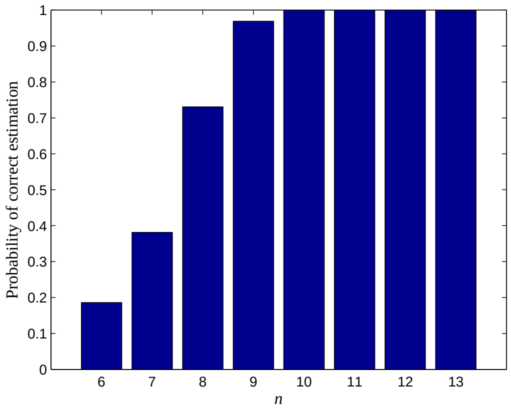


Fig. 11. Probability of correct estimation of the spread sequence $\Pr(\langle \mathbf{c}, \hat{\mathbf{c}} \rangle = P)$ by the zigzag estimation with n-tuple generator polynomial, $P = 2^n - 1$, SNR=-10dB and L=256

sequence spread spectrum signals in multipath. *IEEE Transactions on Signal Processing* 1997; **45**:1241–1252.

11. Jiang L, Ji H, Li L. A Blind Estimation Algorithm for PN Sequence in DS-SS Signals. *8th International Conference on Signal Processing* 16-20 2006; **3**:–, doi:10.1109/ICOSP.2006. 345866.
12. Burel G, Bouder C. Blind estimation of the pseudo-random sequence of a direct sequence spread spectrum signal. *21st Century Military Communications Conference Proceedings* 2000; **2**:967–970, doi:10.1109/MILCOM.2000.904074.
13. Stewart GW. Perturbation Theory for the Singular Value Decomposition. *Technical Report CS-TR-2539* 1990.
14. Golomb SW, Gong G. *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge University Press: NY, USA, 2004.
15. Sarwarte DV, Pursley MB. Cross-Correlation Properties of Pseudorandom and Related Sequences May 1980; **68, No.5**:593–619.
16. Peebles PZJ. *Probability, random variables and random signal principles*. McGraw Hill, New York, 2001.
17. Verdu S. *Multiuser Detection*. Cambridge University Press, 1998.

DOI: 10.1002/wcm