Blind estimation of block interleaver length and encoder parameters
Roland Gautier, Gilles Burel, Mélanie Marazin, Crépin Nsiala-Nzáza

To cite this version:
BLIND ESTIMATION OF BLOCK INTERLEAVER LENGTH 
AND ENCODER PARAMETERS

ROLAND GAUTIER¹
GILLES BUREL¹
MELANIE MARAZIN¹
CREPIN NSIALA NZEZA²

Abstract: Interleavers are key devices in most digital transmission systems. An approach for blind estimation of the interleaver length, as well as the encoder rate and constraint length, is proposed in this paper. The approach is based on linear algebra. We show that the rank of a matrix built from the intercepted interleaved stream falls when the number of rows of this matrix is a multiple of the interleaver length. Furthermore, the values of the ranks allow to estimate the encoder rate and constraint length. Typical applications are transmission surveillance and self-recovering receivers.

Keywords: cognitive-radio applications, convolutional encoder, interleaver, blind estimation, non-cooperative communications, interception.

1. Introduction

As shown on figure 1, in many digital transmission systems, the binary data stream is first encoded for sparse error protection (using, for instance a block code or convolutional code). Then the binary stream is interleaved by reordering the bits of the stream in a certain way. The same permutation is usually made on blocks of length \( t_c \) bits, where \( t_c \) is called the length of the interleaver. The interleaved binary data is finally fed to a digital transmitter which performs the carrier modulation, filtering and amplification [17].

¹ Université Européenne de Bretagne; UBO/Lab-STICC, UMR CNRS 3192, 6 avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex 3, France, e-mails: roland.gautier@univ-brest.fr, gilles.burel@univ-brest.fr

² Université Lille Nord de France, INRETS, LEOST, 59650, Villeneuve d’Ascq, France, e-mails: crepin.nsiala@inrets.fr
On the receiver side, the signal is demodulated to recover the interleaved binary data. Then the data is de-interleaved and decoded. In cooperative applications, the classical receiver can perform data de-interleaving because it knows the permutation to apply on the received data. On the contrary, in non-cooperative applications, the interleaver is unknown and it is necessary to recover its parameters. Typical applications are multistandard adaptive receivers and spectrum surveillance [19] but the literature about self-recovering receivers is not very rich. In [8], we have proposed a self-recovering receiver for encoded and scrambled binary data streams. In this paper, we propose a method which is able to recover the length of the interleaver and also the parameters of the encoder used by the transmitter. These parameters are very important to characterize the interleaver and also the encoder.

Interception of communications is attempted for a variety of reasons including reconnaissance, surveillance, and other intelligence gathering activities, as well as position fixing, identification, and communications jamming [7]. For example, an aircraft might attempt to intercept the communications between a submarine or ship and a satellite, or a satellite might attempt to intercept ground-to-ground communications. Typically, the interceptor has no knowledge of modulation, encoder and interleaver characteristics [6].

In the context of interception, techniques that automatically identify the modulation type of a received signal have been proposed [12] [16]. The complex problem of detecting the presence of direct sequence spread-spectrum signals hidden in noise was also considered [13] [15] [18], either using techniques based on cyclostationarity or on statistical fluctuations of correlation estimators [5]. In [4] we have proposed an approach which is able to estimate the spreading sequence used by the transmitter. An algorithm for blind spreading sequence discovery for DS-CDMA signal interception has also been proposed in [9]. A cyclostationarity-based method for estimation of the spreading sequence was described in [19]. Once spreading sequences and modulation are estimated, the interceptor is able to despread, synchronize, and demodulate the signal. Software Defined Radio architectures for electronic signal interception, identification and jamming are now considered [20].

To go further, there is a need for techniques to determine which encoder and which interleaver were used by the transmitter, in order to decode and to de-
interleave the binary data. For the moment, there is still little published work in this area. In [2] an efficient algorithm for recovery of unknown constraint length and generator polynomials for linear convolutional encoders was proposed, and extended to noisy data in [3]. The algorithm is powerful, but its complexity increases exponentially with the encoder constraint length \((K)\), because it requires construction and manipulations of a transition table with \(2^{2K-1}\) rows and \(2^{2K}\) columns. The algorithm could probably be extended to global estimation of the encoder-interleaver block. However, in that case, its complexity would increase exponentially with the length of the interleaver \(l_i\) (instead of \(K\)). Even with very short interleavers, such a complexity is orders of magnitude above the capabilities of today available hardware. This is the reason why, in this paper, we propose a different approach, which is based on linear algebra instead of manipulation of transition tables.

The paper is organized as follows. In Section 2, the proposed method is described. Then, in Section 3, experimental results using standard encoders and interleavers are provided to illustrate the approach. Finally, a conclusion is drawn in Section 4.

In the sequel, for illustration purpose, we provide examples related to HIPERLAN Type 2 and CDMA2000 standards. However, the method is not limited to these particular protocols.

2. Proposed method

2.1. General idea

The basic idea of the method is first of all to use some classical linear algebra properties to determine the length of the interleaver and then to take profit of redundancy of the encoded data stream to determine the encoder rate. For instance, consider an \(N_e\)-dimensional vector containing \(N_e\) successive encoded and interleaved bits. The principle of our approach is to reshape columnwise the interleaved data stream vector to obtain matrices \(H_i\) for different number \(i\) of rows and to compute for each matrix its rank in Galois Field GF(2). For each matrix \(H_m\) with \(m=q\times l_e\) multiple of \(l_e\) (the length of the interleaver) and under some hypotheses detailed in Subsection 2.2, the rank is \(\min(m\times r+K-1,m)\) (with \(r\) the encoder length and \(K\) its constraint length, see 2.2), while it is \(m\) for the other values not multiple of \(l_e\). The first fall of rank occurs for \(m=l_e\) (if \(m>m\times r+K-1\)) which gives the length of the interleaver. The rate and the constraint length of the encoder is obtained by linear regression from different values of the rank of \(H_m\) for \(m=q\times l_e\), where \(l_e\) is the length of the interleaver just estimated.
2.2. Mathematical model and hypotheses

Let us note $N_e$ the number of observed binary symbols, $r = \frac{k}{n}$ the encoder rate with $k$ the length of a binary message word, $n$ the length of the code words and $K$ the encoder constraint length if we consider a convolutional encoder (for a classical block encoder $K = 1$). Consider the binary vectors below:

- $h$: the vector containing $N_e$ samples of the interleaved data stream;
- $g$: the corresponding outputs of the encoder (dimension $N_e$);
- $g_0$: the corresponding unknown information data (dimension $N_e \times r + K - 1$);

From the principle of linear encoders and interleavers, it can be shown that there exist matrices $G$ and $E$ such that $g = Gg_0$ and $h = Eg$. According to figure 2, which summarizes the mathematical model, we have:

$$h = EGg_0$$  \hspace{1cm} (1)

Figure 2. Mathematical model

If we know the encoder, we can compute the matrix $G$ as described in the Appendix where a practical example taken from the CDMA2000 Standard is used to explain the construction of $G$. In our approach the encoder is unknown so that parameters $n$, $k$ or $K$ are also unknown. Nevertheless, the two important hypotheses are:

- These encoders add redundancy such that $n > k$;
- The encoder is generally well chosen which implies that matrix $G$ is of full rank in GF(2).

The matrix $E$ is only a permutation matrix. In this study, we only consider block interleavers in which the interleaver does not operate globally on all the encoded data stream bits, but only block by block. In each block ($l_e$ bits), the permutation made on the data bits is the same and the size $l_e$ of each block is named the length of the interleaver and has to be recovered in our case. We can represent the interleaver as a square matrix $E$ which is constructed as a square block diagonal matrix where each block is a permutation matrix $E_0$ ($dim(E_0) = l_e$) as it is shown below:
Blind Estimation of Block Interleaver Length and Encoder Parameters

Interleavers are used to increase burst error correction capability of a simple code which is generally designed for sparse errors correction. If we consider a linear code $C(n,k)$ in GF(2) and an integer $n_e$, we can show that the set of $n_e$ interleaved code words is a code of length $n \times n_e$, with dimension $k \times n_e$ in GF(2), which is able to correct burst errors of length $t \times n_e$ if the code $C$ corrects burst errors of length $t$. According to this property, the length $l_e$ of the interleaver is a multiple of $k$, because it operates on a block of code words and not on a block of individual bits to respect this property.

In our non cooperative context, only vector $h$ given by equation 1 is observed. Vectors $g_0$, $g$, matrix $G$ and submatrix $E_0$ (also matrix $E$ which is based on submatrix $E_0$) are unknown. To estimate the length of the interleaver, we have to determine the size of $E_0$.

2.3. Mathematical description of the method

For pedagogical purpose, suppose that we decide to reshape columnwise the interleaved data stream in such manner to obtain a matrix $H_{l_e}$ of $l_e$ rows and so $N_e/l_e$ columns. By construction each column vector of this matrix is always the product of a certain matrix $F$ of size $(l_e, l_e \times r + K - 1)$ and a vector consisting of a block of original data stream bits of size $l_e \times r + K - 1$. The rank in GF(2) of the matrix $H_{l_e}$ is therefore equal to the minimum between $l_e \times r + K - 1$ and $l_e$ if $H_{l_e}$ is not degenerate (which is the case when $G$ is full rank). More generally, we can say that, if we reshape columnwise the interleaved data stream in such manner to obtain a matrix with a number of rows multiple of the size of the interleaver $q \times l_e$, the rank in GF(2) of this matrix $H_{q \times l_e}$ will be $\min(q \times l_e \times r + K - 1, q \times l_e)$, because all the columns of this matrix are the product of the different vectors of the original data stream by a same matrix $F$ of size $(q \times l_e, q \times l_e \times r + K - 1)$.

In the other cases, with a number of rows $i \neq q \times l_e$, no matrix $F$ exists such that each column vector of matrix $H_i$ is always the product of $F$ and a vector consisting of a block of original data stream bits. So, in these cases, the rank of $H_i$ is equal to $i$ (the number of rows of $H_i$).

To determine the length of an interleaver, we only have to construct the
matrices $H_j$ by reshaping columnwise the interleaved data stream bits for all $j$ (number of rows) in the interval $[1, nb_{\text{max}}]$, with $nb_{\text{max}}$ sufficient and to compute for each matrix $H_j$ its rank in GF(2).

If the encoder rate $r$ and the constraint length verify $K < l_r \times (1 - r) + 1$, the length of the interleaver is obtained at the first fall of rank. If it is not the case, the length of the interleaver is equal to the difference between the respective number of rows corresponding to two consecutive falls of rank of $H_j$. Finally, we can also determine the encoder rate and its constraint length by linear regression from values of the rank of $H_m$ for $m = q \times l_r$, with $q \in [1, \frac{n_r}{K}]$, because $\text{rank}(H_m)$ is given by the equation below:

$$\text{rank}(H_m) = m \times r + K - 1$$

if $K < m \times (1 - r) + 1$.

3. Simulation results

First of all, we illustrate the method by showing the content of various matrices for a given encoder and interleaver (Subsection 3.1). Thereafter, in Subsections 3.2 and 3.3 we give some simulation results in the context of HIPERLAN/2 and CDMA2000 standards. The choice of these standards is for illustration purpose only: it is clear that blind estimation is not required in these cases because the interleaver characteristics used by these standards are in the public domain. Obviously, in the interception context, the transmitter is not likely to use a public domain interleaver. However, our blind approach would still work, because it is not based on particular characteristics.

3.1. Illustration of the method using matrices graphical representations

Let us consider a rate 1/2 convolutional encoder with constraint length $K = 6$ and the two generating polynomials below ([14] p.907):

$$\text{polyG} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 53 \\ 75 \end{pmatrix}_{\text{octal}}$$

This yields to matrix $G_{\text{sub}}$ shown on figure 3 (black=1, white=0).

**Figure 3. Graphical representation of $G_{\text{sub}}$**
For illustration purpose, we consider an particular block interleaver named usually “rows/columns interleaver”. This name comes from the construction of this interleaver where the inputs data are written row by row in an array of \( N1 \) rows and \( N2 \) columns, and read column by column to form the interleaved data stream. This interleaver has a length equal to \( N1 \times N2 \). Note that four different interleavers can be obtained depending of the beginning array corner of reading. Others conventions are used in the literature to construct such block interleavers (see [14] p. 374-388, p. 398-407 or [10] p. 35-63 for example).

We have chosen to start reading the data in the interleaver array at the top left corner and to take \( l_e = 12 \) for the interleaver length (\( N1 = 4 \) and \( N2 = 3 \)). To see the effect of the interleaver, we can construct for example the matrix \( G \) shown in figure 4 to encode \( l_e \times r \times 3 = 18 \) bits (see the Appendix for more details about the construction of a matrix \( G \)).

![Figure 4. Graphical representation of \( G \) to encode 18 bits](image)

The permutation matrix \( E_0 \) of size \( 12 \times 12 \) representing one block of the interleaver is given graphically on figure 5 and the global interleaver matrix \( E \) on figure 6.

![Figure 5. Graphical representation of \( E_0 \)](image)
We can see on figure 7 the effect of the interleaver on the encoder matrix $G$, where the product $EG$ is plotted (compare to $G$ on figure 4).

We will try to estimate the interleaver length from $N_e = 200 \times 12 = 2400$ interleaved bits. First of all, we construct successively the matrices $H_i \; \forall i \in [1;nb_{\text{max}}]$ with $nb_{\text{max}} = \sqrt{N_e} \in \mathbb{N}$, and to compute their ranks in GF(2) which is represented on figure 8. We can see that the first fall of rank is obtained for
\[ i = 12 \] which is exactly the length of the interleaver. Graphically we can also see on figure 8 that the line joining all the points for \( i = q \times l \), has a slope equal to \( \frac{1}{2} \) and an intersection with the vertical axis (for \( i = 0 \)) equal to 5, which give us from equation 3 the encoder rate \( r = \frac{1}{2} \), and constraint length \( K = 6 \) (because \( K - 1 = 5 \)). These results are exactly the parameters of the encoder used.

**Figure 8.** Representation of \( \text{rank}\{H_i\} \) versus \( i \)

### 3.2. Example with HIPERLAN Type 2 Standard

HIPERLAN Type 2 standard uses a rate \( \frac{1}{2} \) and constraint length \( K = 7 \) convolutional encoder with the two generating polynomials below:

\[
polyG = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 133 \\ 171 \end{pmatrix}_{\text{octal}}
\]

All encoded data bits are interleaved by a block interleaver with a block size corresponding to the number of bits in a single OFDM symbol which is given in [11] (p. 11). In this example we have taken \( l_e = 48 \), but the method works with the others values of \( l_e \) defined in this standard. The interleaver is defined by a two step permutation which is described in [11] (p. 16). The matrix \( E_0 \) for the block interleaver with \( l_e = 48 \) for the HIPERLAN Type 2 standard is represented graphically on figure 9.
Figure 9. Graphical representation of $E_0$ (Hiperlan2 example)

We try to estimate the interleaver length from $N_c = 400 \times 48 = 19200$ interleaved bits. The rank in GF(2) of the matrices $H_i \quad \forall i \in [1; nb_{\text{max}}]$ with $nb_{\text{max}} = 138$ is represented on figure 10. We can see that the first fall of rank is obtained for $i = 48$ which is the length of the interleaver. From the line joining all points for $i = q \times l_c$ (with $l_c = 48$ just found), we can give the encoder rate $r = \frac{1}{2}$ and its constraint length $K = 7$, which are exactly the parameters of the encoder used.

Figure 10. Representation of $\text{rank} \left| H_i \right|$ versus $i$ (Hiperlan2 example)
3.3. Example with CDMA2000 Standard

For example, the Enhanced Access Channel for spreading rate 1 (for the Data) in the CDMA2000 standard uses a rate $\frac{1}{4}$ and constraint length $K=9$ convolutional encoder with generating polynomials below (see [1] p. 2-95):

\[
polyG = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 765 \\
1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 671 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 513 \\
1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 473 \\
\end{pmatrix}
= \begin{pmatrix}
765 \\
671 \\
513 \\
473 \\
\end{pmatrix}_{\text{octal}}
\]  

In this standard, all the encoded data bits are interleaved by block with different block size specified in [1] (p. 2-107). We have taken $l_e = 768$ and the matrix $E_0$ for the block interleaver with $l_e = 768$ for the Enhanced Access Channel in the CDMA2000 standard is represented graphically on figure 11. This matrix can be computed from the rule described in [1] (p. 2-106 - 2-107).

Figure 11. Graphical representation of $E_0$ (CDMA2000 example)

We try to estimate the interleaver length from $N_e = 3100 \times 768 = 2380800$ interleaved bits. The rank in GF(2) of the matrices $H_i \forall i \in [1; nb_{max}]$ with
$nb_{\text{max}} = 1543$ is represented on figure 12. The method provides a perfect estimation of the interleaver length, encoder rate and its constraint length which are $l_e = 768$, $r = \frac{1}{4}$ and $K = 9$.

![Figure 12. Representation of $\text{rank}(H_i)$ versus $i$ (CDMA2000 example)](image)

4. Conclusion

In this paper, we have proposed an approach for blind estimation of the interleaver length and the encoder rate and constraint length. The intercepted data stream is written columnwise in a matrix and we have shown that the rank of this matrix falls when its number of rows is a multiple of the interleaver length. Furthermore, from the plot of the rank with respect to the number of rows of this matrix, we have shown that the encoder rate and constraint length can be determined.

This approach provides essential data in the context of spectrum surveillance applications, as well as for the purpose of building a self-recovering receiver. Since it is based on linear algebra, it is easy to program using a matrix-oriented language, such as Matlab. Further work will include evaluation of the robustness of the approach when the intercepted interleaved stream contains errors.
Appendix: Computation of matrix $G$ using the model of the linear encoder

The objective of this appendix is to help understanding what is the matrix $G$ used in our mathematical model, and what is its content. Please note that computation of matrix $G$ by this mean is not used in our blind method, because the encoder is unknown.

As an illustration, we will consider a convolutional encoder, but the method can be easily extended to any linear code, such as block codes (Hamming, BCH, etc.). A convolutional encoder is defined by its generating polynomials. For instance, the CDMA2000 Reverse Fundamental Channel (for Radio configuration 1) uses a rate $1/3$ convolutional code, with constraint length $K = 9$ and the generating polynomials coefficients below ([1] p. 2-95 - 2-96):

$$
\begin{align*}
poly_1 &= (101101111) \\
poly_2 &= (110110011) \\
poly_3 &= (111001001)
\end{align*}
$$

The octal form is often used for concision:

$$
polyG = \begin{pmatrix}
557 \\
663 \\
711
\end{pmatrix}_{\text{octal}}
$$

The coded stream $z$ is then ([1] fig. 2.1.3.1.4.1.2-1 p. 2-96):

$$
\begin{align*}
z_{3n} &= y_n + y_{n-1} + y_{n-2} + y_{n-3} + y_{n-5} + y_{n-6} + y_{n-8} \\
z_{3n+1} &= y_n + y_{n-1} + y_{n-4} + y_{n-5} + y_{n-7} + y_{n-8} \\
z_{3n+2} &= y_n + y_{n-3} + y_{n-6} + y_{n-7} + y_{n-8}
\end{align*}
$$

where $y$ stands for the uncoded data stream.

Let us note $g$ an $N_\epsilon$-dimensional vector containing $N_\epsilon$ successive samples of the coded stream and $g_0$ a vector containing the information sequence which generated $g$ (dimension $N_\epsilon \times r + K - 1$). We can write:

$$
g = Gg_0
$$

where
Note that $G$ contains the submatrices $G_{sub}$ below (dimension $n \times K$):

$$G_{sub} = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}$$

Each row of $G_{sub}$ contains the coefficients of a generating polynomial in reverse order. In our application, matrices $G$ and vectors $g$ and $g_0$ are unknown. But, only the linear algebra properties of such computed matrix are important in this case, without the exact knowledge of the generating polynomials.

**References**


correlation estimators”, *IEEE Int. Symp. on Intelligent Signal Processing and Communication Systems (ISPACS’2000)*, November 5-8, 2000, Honolulu, Hawaii, USA


[11] ETSI TS 101 475 (v1.3.1), *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer*, december 2001


