ON THE SELECTION OF GUIDED SCRAMBLING SEQUENCES THAT PROVIDE GUARANTEED MAXIMUM RUN-LENGTH CONSTRAINTS

Adriaan J. van Wijngaarden and Kees A. Schouhamer Immink

Abstract – Guided scrambling is a coding technique that generates for each possible source word a fixed set of candidate codewords and subsequently selects the "best" word subject to given channel constraints. The resulting codes are often referred to as "weak" constraint codes because they do not strictly guarantee that the specified constraints will be fulfilled. In this paper, it will be shown that for the important class of maximum run-length constraints a proper selection of the set of guided scrambling sequences can actually guarantee that the code satisfies "strong" constraints. Sequence set selections and code constructions that result in "strong" maximum run-length constraints will be presented.

INTRODUCTION

Guided scrambling generates for each possible source word a unique set of candidate codewords and selects the "best" word subject to certain given channel constraints [1, 2, 4]. This technique has proven to be very effective to generate codewords that satisfy single or multiple channel constraints with high probability. The probability that none of the candidate codewords fulfills the constraints reduces considerably when the set of candidate sequences is increased. The corresponding codes are therefore often referred to as weak constrained codes if they do not strictly guarantee that the specified constraints are not violated. The guided scrambling technique is particularly suited for the construction of weakly constrained high-rate codes, where a relatively small set of candidate sequences

Adriaan J. (de Lind) van Wijngaarden is with Bell Laboratories, Lucent Technologies, MH 2C-360, 600 Mountain Ave, Murray Hill, NJ, U.S.A. E-mail: alw@research.bell-labs.com.

Kees A. Schouhamer Immink is with the Institute for Experimental Mathematics, University of Essen, Ellernstr. 29, 45326 Essen, Germany and with Turing Machines Inc., Willemskade 15 b–d, 3016 DK Rotterdam, The Netherlands. E-mail: immink@exp-math.uni-essen.de.

is very likely to contain a sequence that satisfies the constraints. Typical constraints are maximum run-length constraints and spectral constraints, e.g., to suppress low frequencies. Constrained codes, such as run-length limited codes and DC-free codes, have found widespread application in data storage and communication systems, see [2] and references therein. Runlength-limited sequences, also known as (d, k) sequences, are binary sequences with at least d and at most k 'zeros' between consecutive 'ones'. Of particular interest is the subclass of $(0, k)$ sequences, used extensively in recording and digital communication systems for clock recovery and synchronization. These sequences are generally constructed by concatenating words of a binary block code of length n, each fulfilling $(0, k, k_l, k_r)$ constraints, where k_l and k_r denote the maximum number of leading and trailing zeros ($k_l + k_r \leq k$). These codes are referred to as maximum run-length limited (MRL) codes.

In this article, we are interested in the question how weak the constraints are, and whether it is possible to choose the scrambling sequences in such a way that one can actually impose strong constraints. We concentrate on finding strong maximum run-length constraints.

This paper is organized as follows. We first introduce the notation that is used throughout this article and further specify maximum runlength-limited sequences. Next, we analyze the guided scrambling technique and assess the "strong" constraints that can be imposed on the maximum run-length with a proper choice of the set of scrambling sequences.

DEFINITIONS

To describe the properties of the code construction techniques and for error assessment we need to introduce some notation. Let $\mathcal{A}_2^w$ denote the set of sequences of w symbols from the binary alphabet $\mathcal{A}_2$. A sequence $X \in \mathcal{A}_2^w$ is represented as a string of symbols $x_1 x_2 \ldots x_w$ of length w. The concatenation of two sequences X and Y is denoted by XY, and $P\mathcal{S}^{(w)} = \{PX | X \in \mathcal{S}^{(w)}\}$ denotes a set of sequences of length $p+w$ with prefix $P \in \mathcal{A}_2^p$ and suffix $X \in \mathcal{S}^{(w)}$, where $\mathcal{S}^{(w)} \subseteq \mathcal{A}_2^w$. Similarly, $\mathcal{S}^{(w)}P = \{XP | X \in \mathcal{S}^{(w)}\}$. The null string, denoted by Λ, represents a string of length 0 for which $\Lambda X = X \Lambda = X$. A run of w consecutive symbols a is written as a^w. We use $*$ and $(*)^w$ to denote any element of $\mathcal{A}_2$ and $\mathcal{A}_2^w$, respectively.

Maximum Run-Length Constrained Sequences

Let $\mathcal{G}^{(n)}_{k,k_l,k_r}$ denote the set of binary sequences of length n that fulfill the $(0, k, k_l, k_r)$ constraint. This set can be written as

$$\mathcal{G}^{(n)}_{k,k_l,k_r} = \bigcup_{s=0}^{k_l} 0^s 1 \mathcal{G}^{(n-s-1)}_{k,k,k_r}. \tag{1}$$

The disjoint subsets $\mathcal{G}^{(m)}_{k,k,k_r}$ are determined using recursion relations. For $0 \leq m \leq k_r$, $\mathcal{G}^{(m)}_{k,k,k_r} = \mathcal{A}_2^m$, and for $k_r < m \leq k$, $\mathcal{G}^{(m)}_{k,k,k_r} = \{P\mathcal{G}^{(k_r+1)}_{k,k,k_r} | P \in \mathcal{A}_2^{m-k_r-1}\}$. For $k < m < n$, $\mathcal{G}^{(m)}_{k,k,k_r}$ satisfies the recursion relation

$$\mathcal{G}^{(m)}_{k,k,k_r} = \bigcup_{s=0}^{k} 0^s 1 \mathcal{G}^{(n-s-1)}_{k,k,k_r}. \tag{2}$$

It follows that the cardinality of this set, denoted by $G^{(m)}_{k,k_l,k_r}$, is determined by two equations:

$$G^{(m)}_{k,k,k_r} = \begin{cases} 2^m & \text{if } 0 \leq m \leq k_r \\ 2^{m-k_r-1}(2^{k_r+1} - 1) & \text{if } k_r < m \leq k \\ G^{(m-1)}_{k,k,k_r} + \cdots + G^{(m-k-1)}_{k,k,k_r} & \text{if } m \geq k+1. \end{cases}$$

and

$$G^{(m)}_{k,k_l,k_r} = \sum_{s=0}^{k_l} G^{(m-s-1)}_{k,k,k_r}. \tag{3}$$

The set of $(0, k)$ constrained sequences of length m, where $k_l = k_r = k$, is denoted by $\mathcal{G}^{(m)}_k = \mathcal{G}^{(m)}_{k,k,k}$, and its size by $G^{(m)}_k = G^{(m)}_{k,k,k}$.

GUIDED SCRAMBLING SEQUENCE SETS

Let $\mathcal{Q}^{(n)}_r$ denote the set of r guided scrambling sequences of length n. For a given codeword $X \in \mathcal{A}_2^n$, the set of candidate codewords is given by $\{X + S | S \in \mathcal{Q}^{(n)}_r\}$ of size r.

The initial distribution of the maximum k constraint for the set of source words can be determined exactly. Let $K_q(\mathcal{A}_2^n)$ denote the number of sequences for which the maximum run-length equals q. It follows directly that $K_q(\mathcal{A}_2^n)$ is given by $K_q^{(n)}(\mathcal{A}_2^n) = G_q^{(n)} - G_{q-1}^{(n)}$, for $q \geq 1$, and $K_0^{(n)}(\mathcal{A}_2^n) = 1$. The probability that a random word of length n has run-length q is consequently give by $K_q(\mathcal{A}_2^n)/2^n$. Note that we consider isolated constrained sequences here, where $k_l = k_r = k$. Expressions for $(0, k, k_l, k_r)$ codes can be derived in a similar fashion.

As an example, for sequences of length $n = 16$, the vector $K^{(n)}(\mathcal{A}_2^n) = (K_0^{(n)}, \ldots, K_n^{(n)})$ is given by

$$K^{(16)}(\mathcal{A}_2^{16}) = (1{,}2583{,}16929{,}20135{,}13008{,}6792{,}3277{,}1531{,}704{,}320{,}144{,}64{,}28{,}12{,}5{,}2{,}1).$$

Consider the situation where a set of 2^r scrambling sequences of length n are used. It can be shown that a lower bound on $k_{\max}$ is given by

$$k_{\max} \geq \lfloor \frac{n}{2^r} \rfloor. \tag{4}$$

An upper bound on $k_{\max}$ for an arbitrary set $\mathcal{Q}_r^{(n)}$ is $k - r$. We will present several constructions where a careful choice of the set of scrambling sequences reduces the value of $k_{\max}$ to $n/(r+1)$ or lower.

It can be shown that if $\mathcal{Q}_2^{(n)}$ contains an arbitrary scrambling sequence $A \in \mathcal{A}_2^n$ and its complement, the distribution will be identical to the case where $\mathcal{Q}_2^{(n)} = \{0^n, 1^n\}$. More generally, it can be shown that for any set $\mathcal{Q}_r^{(n)}$, one can add an arbitrary sequence $A \in \mathcal{A}_2^n$ to every sequence in $\mathcal{Q}_r^{(n)}$ to obtain a new set $\{S + A | S \in \mathcal{Q}_r^{(n)}\}$ with an identical maximum run-length distribution.

For example, if $r = 2$ and $\mathcal{Q}_2^{(n)} = \{0^n, 1^n\}$, the distribution is, for $n = 16$, as follows:

$$K^{(16)}(\mathcal{A}_2^{16}, \mathcal{Q}_2^{(16)}) = (2{,}5164{,}30666{,}22246{,}6096{,}1154{,}182{,}24{,}2{,}0{,}0{,}0{,}0{,}0{,}0{,}0{,}0),$$

i.e., $k_{\max} = 8$.

Notice that for the above example the codewords do not contain an indicator to identify which scrambling sequence was used. Without this information, the mapping is not unique. One option is to send this information, r bits, separately, and, in the case when information is affected by noise, to protect this information well to avoid error propagation. Otherwise a single bit error may cause error propagation that affects up to n bits due to the substraction of the wrong scrambling sequence at the receiver.

The sequence index can also be embedded in the codeword in several ways. The first step is to augment the source word $X \in \mathcal{A}_2^m$ by r positions with indices $\mathcal{P}^{(r)} = \{p_1, p_2, \ldots, p_r\}$ to obtain an intermediate word X' of length n. These positions are preferably interspersed with the source word. It is easy to see that these positions remain unaffected by guided scrambling with a set $\mathcal{Q}_r^{(n)}$, where at each position $p_i \in \mathcal{P}^{(r)}$ of every element of $\mathcal{Q}_r^{(n)}$, the value is fixed to a constant value v_i that is known to the decoder. If $v_1 = v_2 = \cdots = v_r = 0$, the positions $\mathcal{P}^{(r)}$ are called *unconstrained positions.* Note that unconstraint positions can also be of use to make a code resilient against errors. Effective error control can be provided using techniques described in [3] and references therein.

CONCLUSIONS

Guided scrambling is an effective technique to generate long and highly efficient codes that satisfy given channel constraints with high probability. We have demonstrated that by carefully selecting the set of guided scrambling sequences one can impose "strong" constraints on the maximum run-length. Although the constraints are generally loose for large values of n, they may in many cases be sufficient to avoid problems with clock recovery.

REFERENCES

[1] I.J. Fair, W.D. Grover, W.A. Krzymien and R.I. MacDonald, "Guided scrambling: a new line coding technique for high bit rate fiber optic transmission systems," *IEEE Trans. Commun.*, vol. 39, no. 2, pp. 289–297, Feb. 1991.

[2] K.A.S. Immink, *Codes for Mass Data Storage Systems,* Shannon Foundation Publishers, Geldrop, The Netherlands, 1999.

[3] A.J. van Wijngaarden and K.A.S. Immink, "Maximum run-length limited codes with error control capabilities," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 4, pp. 602–611, Apr. 2001.

[4] B. Vasic, G. Djordjevic and M. Tosic, "Loose composite constraint codes and their application in DVD," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 4, pp. 765–773, Apr. 2001.

[5] A.J. van Wijngaarden and E. Soljanin, "A combinatorial technique for constructing high-rate MTR-RLL codes", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 4, pp. 582–588, Apr. 2001.