LUBLIN TECHNICAL UNIVERSITY

II INTERNATIONAL SYMPOSIUM

NEET' 2001

NEW ELECTRICAL AND ELECTRONIC TECHNOLOGIES AND THEIR INDUSTRIAL IMPLEMENTATION

Symposium Proceedings

This Symposium was supported by the Committee for Scientific Research (Poland)

Sponsors:

PP-H Elektromontaż – Export S.A.
Oddział Elektromontaż Lublin
20-447 Lublin; ul. Diamentowa 1

Hurtownia Kabli i Przewodów
„BYCHOWO"; Witold Zgubiński
20-479 Lublin, ul. Ciepłownicza 6

Kazimierz Dolny, Poland
February, 14-17, 2001

Information scrambler/descrambler based on combination of data compression and error-correcting codes

Pavel Urbanovich [1), 2)], *Natalia Patsei* [2)]
[1)] *Catholic University of Lublin, Poland,*
[2)] *Belorussian State Technological University, Belarus*

Abstract

Method of scrambler/descrambler construction that provides efficient error free information transferring is presented. The paper introduces scrambling technique based on addition to transforming unites compression coding and error correcting coding. Because of sufficient amount of redundancy in transmitted message pseudorandom sequence generator can be abolished on the receiving side.

Scrambling of digital signals is routinely used in certain communications applications. In others, such as data communications system, scrambling is typically done to assume transition in the received data signal and thereby avoid loss of synchronization in the data recovery process. But some times it is important to ensure that data transmission and recovery is performed without errors. In the real channels the signal by transmission quite often is distorted, and the message data is reproduced with errors. The reason of such errors is the distortions, which are brought in by channel, and interference effecting on a signal.

In accordance with experimental results of error distribution character in telephone data transfer channels in most cases single errors and packet of error length from 2 to 4 bit are appeared [1, 2]. The most efficient way of their elimination is to use sufficient amount of redundancy or Forward Error Correction (FEC). FEC is generally based on the use of available error detection and correction codes. Error-correcting codes in turn increase input message size and as consequence reduce throughput of communication channels.

Here we will consider some aspects of lossless scrambling data transferring without message size increasing, improving both throughput and reliability in the face of noise.

Let define information source S as alphabet $A = \{a_1,...,a_l\}$ together with probabilities, $P(a_1), P(a_2),..., P(a_l)$ satisfying:

$$\sum_{i=1}^{l} P(a_i) = 1, \qquad 0 \le P(a_i) \le 1 \quad \text{for} \quad \forall i .$$

The goal of compression technique is the message transformation in border of one alphabet in such way that message size (the number of alphabet letters) became less and it is possible to reconstruct initial message without any additional information. From the securing point of view the main advantage of compression algorithms is that they modify output message statistic aside of straighten. But the compressed data is highly sensitive the propagation of channel errors. Let choose a simple compression algorithm, for example Huffman code, known as primitive example of autosophy lossless communication. The Huffman code assumes "prior knowledge" of the relative character frequencies stored in a table or library.

During the first phase of information transformation input message data x is splitting into blocks of appropriate size n as follows

$\{x_j\}(j = 1tom)$.

In order to compress information block of length n we need to construct code K with almost constant average compression ratio (it can be achieved for enough large block size)

$R = L(S) / L(K)$.

The length of coded data for each block is variable. Let $|K(a_i)| = d_i$, then average length L of the code words is:

$$L(K) = \sum_{i=1}^{l} d_i * P(a_i),$$

note that $L(K) > H(S)$, where $H(S)$ - the entropy of the source.

Order A so that $P(a_1) \geq P(a_2) \geq ... \geq P(a_l)$. If $l > 2$ then let S' be a source with alphabet $A' = \{a_1,...,a_{l-2},b\}$

and probabilities

$P'(a_i) = P(a_i), i \leq l-2$,

$P'(b) = P(a_{l-1}) + P(a_l)$.

Compute a binary Huffman coding K' for S' and set

$K(a_i) = K'(a_i), \quad i \leq l-2,$

$K(a_{l-1}) = K'(b)0,$

$K(a_l) = K'(b)1.$

Evident that $|K(a_1)| \leq |K(a_2)| \leq ... \leq |K(a_l)|$.

In the capacity of compression technique can be used any other algorithm: dynamic library, dictionary methods, LZ-77, LZW or serial autosophy data trees [3,4].

In common form, if the input message block $x_j = \{x_1,...,x_n\} \in A$ is compressed with binary code K mapping $A \to B$, then the result message block $y = \{y_1,...,y_k\} \in B$ can be expressed as:

$y = K(x),$

with total number of $n = R \cdot k$ bits in code word.

Thus the input block x preliminary compress on 20-100 per cent. So the use of such compression coding makes it possible to use error-correcting codes with different amount of redundancy and correcting capacity in scrambler/descrambler system. Correcting coding in turn increase message size of compressed data without increasing of input message size.

Depending on channel characteristics, type of channel errors and compression ratio a different type of error-correcting codes may be used. As error-correcting code C can be the Hamming code $(2^m - 1, 2^m - 1 - m)$ or one of the important classes of linear codes - cyclic code, where m - any integer. It may be the intermittent code, correcting packets length t; tree-like $(n_0.k_0)$ or convolution (n,k) code; Fajer code $((2^m - 1)(2t-1), (2^m - 1)(2t-1) - m - 2t + 1)$. C code can be constructed basing on the quadratically-residued code (n_1n_2, k_1k_2, d_1d_2) correcting $(2t_1t_2 + t_2 + t_1)$ errors or BCH code [5-7]. It may be parallel concatenation of two codes, known as turbo codes. Turbo codes represent a more recent development in the coding research field [6]. These codes are suited to communications systems where large coding gain is required.

Let define a linear block error-correcting code C of length $n = R \cdot k$ (mentioned that k - length (bits) of compressed data block) on *GF(2)* with minimum distance d_m, described by generator and parity check matrixes $G(k \times n)$ - and $H((n-k) \times n)$ corresponding, that can corrects at most t errors.

During the second phase code adds redundancy (or check) bits r to each message block y in order to complete data block up to n bit. For converting variable length block to fixed length the parameters n and k must be known in prior for both the encoder and decoder and the following equation must be satisfied

$n = k + r$.

This can be achieved by enlargement, supplement, lengthening, shortening or omitting code C or by using a plurality of error-correcting codes. Then the error-correcting coding process can be produce as:

$z = G \cdot y$,

where $z = \{z_1,...,z_n\} \in C$.

Appropriately input message blocks are pre-processed. Received code word z are exclusive-or'ed with the output of pseudorandom sequence generator W:

$f = z \oplus W$, $f = \{f_1,...,f_n\}$,

and whole scrambler process can be represents as:

$f = G \cdot K(x) \oplus W$.

Scrambled data f during transmission due to effect of channel noise reproduced on received side as:

$f' = f \oplus e$.

Receiver side apply only error-correcting decoding without exclusive-or'ed process. Due to large surplus information we can pass pseudo-random sequence W and error vector e influence on coding data. Decoding process can be represented as

$f'H^T = (f \oplus e \oplus W)H^T$,

if design $e' = e \oplus W$ then

$f'H^T = fH^T \oplus e'H^T = s$,

where and s - syndrome; in most cases $s \neq 0$. The errors free decoding process as follow. On the basis of s we can find corresponding leader adjacent class v. This leader is difference between received f' and code word. Thus by subtraction from $f'-v$ it is possible to correct all error combinations for chosen code and restore y block of message, from which in turn through decompression the input block x can be restored.

This method of data transformation can withstand even very noisy and unpredictable lines. From the security point of view the consider method is not strong and cannot provide compromise protection. But introduction some cryptography elements in unite of scrambler/descrambler and veritable block length makes this technique suitable for confidentially cryptography encryption. The other disadvantages of this technique are the significant memory requirement, the delay involved in the procedures of encoding and decoding and additionally complex, causing increased opportunity for failure. In spite of this there are many applications include the transmission of speech, images, or video over cellular networks, weak radio links, or noisy telephone lines where some low transformation rate is preferable to loss of fidelity.

References

[1] Urbanovich P.P., Patsei N.V. The overall diffusion characteristics of cryptography block codes// CIS, Minsk, 1998, V.2, №2, p.139-140.

[2] Урбанович П.П., Пацей Н.В., Спиридонов В.В. Распределение ошибок в телефонных каналах передачи дискретной информации//Известия Белорусской Инженерной Академии, Минскs, 1997, № 1 (3) \1, с.24-26.

[3] A.E. Morh, E.A. Riskin, R. Ladner. Bit allocation for wavelet image compression and uniform bit loss. - In *Proceeding of CISS.*- 1998.

[4] T. Bell, I. Witten, J. Cleary. Text compression.-Prentice Hall. Englewood Cliffs, NJ.-1990.

[5] Ritter T. The efficient generation of cryptographic confusion sequences// Cryptologia, 1991, 15(2), p.81-139.

[6] A. Michelson, A. Levesque. Error-Control Tech. for Digital Comm.- John Wiley & Sons Inc.- 1985, p.45.

[7] R. E. Blahut. Theory and practice of error control codes. - Addison-Wesley pub., Massachusetts.- 1984.

[8] S. Benedetto, G.Montorsi. Design of parallel concatenated convolutional codes. //IEEE Transaction on Communication, May, 1996, vol.44, no.5, p. 591-600.

[9] D. W. Redmill. Image an Video Coding for Noisy Channels, Ph.D Thesis, Cambridge Univ. Eng. Dept.,1994.