Advanced Science and Technology Letters
Vol.95 (CIA 2015), pp.84-93
http://dx.doi.org/10.14257/astl.2015.95.16

A Method for the Construction of Minimum-Scrambling Source Coding in Line Coder

Wan-Pyo Hong[1],

[1] Department of IT, Hansei University, Kyunggido, 453-742, Korea
{wphong@hansei.ac.kr}

Abstract. This paper develops the minimum scrambling source coding method on the basis of 8-bit symbol. The symbol is formed with upper (column) 4-bit sub-symbol and lower (row) 4-bit sub-symbol. This paper shows the combination method of upper sub-symbol and lower sub-symbol to make unscrambling symbols. This paper also presents the way how to make a minimum scrambling character source coding with the symbols. After the unscrambling and scrambling symbol is confirmed by the developed method, each character is coded by a symbol or symbols on the basis of probability of the character use frequency. A character that has high use frequency is given an unscrambling symbol(s) but a character that has low use frequency is coded a scrambling symbol(s). When the number of characters is equal to or less than the number of the unscrambling symbols, the scrambling in the line coder is not generated. But if the number of character is exceeded the number of the unscrambling symbols, the characters are coded based on a probability of the character use frequency. The developed method applies all kinds of source coding system that affects data transmission signal.

Keywords: AMI, Line Coding, Scrambling, Source coding, Symbol

1 Introduction

A characters entered in information devices is coded in basis of a symbol of 7-bit, combination of upper sub-symbol (column) 3-bit × lower sub-symbol (row) 4-bit or 8-bit, combination of upper sub-symbol 4-bit × lower sub-symbol 4-bit[1]. Such a coding is the character source coding. Text files are stored in binary format or through a binary code on information device such as computers. So source coding limits on data storage. The source code is transmitted into the transmission line from information device. Therefore, the source code should be transformed to the line code through the line coding or channel coding. The methods of line coding are RZ, NRZ, Manchester and AMI and so on [2]-[3]. The characteristic of source code effects on selection of the line coding method in the line coder. The AMI coding method has been used into the line coding system for the long distance transmission signal. The AMI line coding system requires a scrambling function to avoid the continual bit 0s in the line code caused by the continual bit 0s in the source code [4]. That is, the scrambling function directly affects the quality and efficiency in the data transmission.

ISSN: 2287-1233 ASTL

Therefore, the symbol of source codes related to the scrambling is very important factor in viewpoint of the signal transmission.

This paper shows the combination method of upper sub-symbol 4-bit and lower sub-symbol 4-bit to make an unscrambling source code. This paper also presents the way how to make a minimum scrambling source coding with the combined codes. After confirms the unscrambling and scrambling codes, each character is given a code on the basis of the probability of the character use frequency. The scrambling and unscrambling code is given to a character of the low and high use frequency respectively. When the number of characters is equal to or less than the number of the unscrambling codes, the scrambling in the line coder is not generated. But if the number of character is over the number of the unscrambling codes, the characters are coded based on a probability of the character using frequency. In this case, the scrambling generation frequency in the line coder is reduced. The developed method in this paper applies to HDB-3 [5]-[6] scrambler that scrambles the sequential 4-bit "0" to a certain bits pattern in the line coder.

On the other, the variable length codes have been used to improve the efficiency of data transmission by means of reducing the data bit rate. The typical variable length codes are Morse code and Huffman code [7]. The concept of the variable length coding is that the small bit symbol to the high use message, the many bit symbol to the low use message [8]. Even though the existing character source code has a fixed or uniformed symbol such as on the basis of 7-bit or 8 bits length, the minimum scrambling source coding developed in this paper is realized with the concept of the variable length symbols. That is, much consecutive bit "0" s code is assigned to a low probability character in the use frequency and small consecutive bit "0" s code is assigned to a high probability character in the use frequency. If this method is applied to the character source coding, the scrambling frequencies in the line coder caused by the source code are ultimately minimized or completely deleted. On the other hand, the existing character source coding is performed by means of a software technical in the presentation layer of OSI standard. Otherwise, the line coding is processed by means of a hardware technical in the physical layer. If the developed method for the minimum scrambling source coding in this paper is applied to the character source coding, the hardware-based scrambling function for the source code in the physical layer is replaced to the software-based source coding technical in the presentation layer. Therefore, the minimum scrambling source coding method also improves the flexibility and operability of the transmission system and increases the efficiency of the data transmission.

2 Minimum Scrambling Source Coding Method

According to the ECMA-35 standard, the 8-bit symbols have total 256 codes by 16 columns and 16 rows. It is divided into two areas; one for the control function code, the other for the letters and symbols code. The column 00 and 01, 08 and 09 areas are used for the main and sub control function codes respectively. The remaining areas are assigned for the graphic character codes. The 7-bit code system is simply

transformed to the 8-bit codes system by adding the highest order bit b8 “0” with together b7 to b1 [9]. ASC11 7-bit code [10] is converted to 8-bit code like this.
The Unicode system is an international standard for the character source codes. The Unicode system is on the basis of the unit combination system, upper sub-symbol 4-bit and lower sub-symbol 4-bit. And then each code consists of 16-bit, lower sub-symbol 4-bit and upper sub-symbol 12-bit [11]. On the other hand, now the most use UTF code system in Website is UTF-8 code system [12]. The Unicode code is transformed to UTF code system when it is transmitted into a transmission line The UTF-8 code is 24-bit code made by adding 8 bits to the Unicode system [13]. Therefore, Unicode system gives definitely an influence to the transmission signal. Like the ASCII character code system, Unicode system also does not have any special rule for code assignment in code area.

2.1 Combination rule between Upper and Lower sub-symbol

The combination method between the upper and lower sub-symbol for the construction of minimum-scrambling source codes is developed as follows. A code means a symbol combined by two sub-symbols, for example X2+X1(8-bit) or X4+X3+X2+X1(16-bit). Here, X is sub-symbol. First, confirm all codes in a coding area which have consecutive 4-bit “0”. To do that, this paper develops the rule to identify an unscrambling code and scrambling code in the coding area. Table 1 & 3 show it. Table 1 shows the rule to make a code by combing the upper and lower sub-symbol on the level of the lower sub-symbol b4~b1 [14]. Table 3 shows the rule to make a symbol with sub-symbols on the level of upper sub-symbol b4~b1. The rule is applied to the HDB-3 scrambling method. Second, draw the minimum-scrambling source coding map as shown in Table 2 & Table 4.

Lower sub-symbol to Upper sub-symbol combination. Table 1 shows rule to make an 8-bit symbol for minimum scrambling source coding. In the Table, the combination limitation in upper sub-symbol means that upper sub-symbol 4-bit, b8~b5 must not be combined to lower sub-symbol b4~b1 because of occurrence of consecutive 4-bit “0”s. For example, Let a lower sub-symbol be 0hex (0000bit), the lower sub-symbol cannot be combined to 0~Fhex in upper sub- symbol because all codes have consecutive bit “0” exceeding 4-bit “0”s due to the low sub-symbol (0000) binary. If a lower sub- symbol is 1hex (0001)binary, it cannot be combined to 0, 2, 4, 6, 8, A, C, Ehex in upper sub-symbols. But it can be permitted to combine to 1, 3, 5, 7, and 9, B, D, and Fhex. In Table 2, mark “X” is expressed the scrambling symbol in a symbol or a code. In Table 2 and Table 4, mark “X” is expressed the scrambling symbol in a symbol or a code. In case of a symbol or code based on 8-bit, there is one double scrambling symbol in sub-symbol 0 plus upper sub-symbol 0. Mark “Z” symbol is double scrambling symbol.

Table 1. Rule to make a symbol for Minimum Scrambling Source Coding. Lower sub-symbol b4~b1 base

Lower Sub-symbol (Hex)	Upper Sub-symbol Combination limit	Combination permit
0	All	-
1	0, 2, 4, 6, 8, A, C, E	1, 3 ,5 , 7, 9 , B, D, F
2	0, 4, 8, C	1, 2, 3, 5,6, 7, 9, A, B, D, F
3	0, 4, 8, C	1, 2, 3, 5,6, 7, 9, A, B, D, F
4	0, 8	1, 2, 3, 4,5, 6, 7, 9, A, B, C, D, E, F
5	0, 8	1, 2, 3, 4,5, 6, 7, 9, A, B, C, D, E, F
6	0, 8	1, 2, 3, 4,5, 6, 7, 9, A, B, C, D, E, F
7	0, 8	1, 2, 3, 4,5, 6, 7, 9, A, B, C, D, E, F
8	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
9	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
A	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
B	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
C	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
D	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
E	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F
F	0	1, 2, 3, 4,5, 6, 7, 8, 9, A, B, C, D, E, F

Table 2. Minimum-Scrambling Source Coding Map from Table 1

LOWER SUB SYMBOL	UPPER SUB-SYMBOL(HEX) 0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	Z	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1	X		X		X		X		X		X		X		X	
2	X	Y			X						X				X	
3	X				X						X				X	
4	X	Y	Y	Y							X					
5	X										X					
6	X	Y									X					
7	X										X					
8	X	Y	Y	Y	Y	Y	Y	Y								
9	X															
A	X	Y														
B	X															
C	X	Y	Y	Y												
D	X															
E	X	Y														
F	X															

Upper sub-symbol to Lower sub-symbol combination. Table 3 shows rule to make an 8-bit symbol for minimum scrambling source coding based on upper sub-symbol to lower sub-symbol. In the Table 3, the combination limitation in upper sub-symbol means that lower sub-symbol 4-bit, b8~b5 must not be combined to upper sub-symbol b4~b1 because of occurrence of consecutive 4-bit "0"s. For example, let an upper sub-symbol be 0hex (0000) binary. In this case, all lower sub-symbol 0-Fhex cannot be combined with upper sub- symbol 0hex because all code have consecutive bit "0" due to upper sub-symbol (0000) binary. When upper sub-symbol is 1hex (0001) binary, it cannot be combined with 0hex in lower sub-symbols. But it can be permitted to combine with all lower sub-symbols except 0hex. The upper sub-symbol limits to combine with O, 1, 2, 3, 4, 5, 6, 7hex low sub-symbol in case of upper sub-symbol 8hex. In Table 4, mark "X" is expressed the scrambling symbol in a symbol or a code. In case of a symbol or code based on 8-bit in Table 3, there is one double scrambling in sub-symbol 0 plus upper sub-symbol 0 same as in Table 1. Mark "Z" symbol is double scrambling symbol.

Table 3. Rule to make a symbol for Minimum Scrambling Source Coding. Upper Sub-Symbols b8~b5 base

Upper sub-symbols (hex)	Lower Sub-Symbol (hex)	
	Combination limitation	Combination permission
0	All	-
1	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
2	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
3	O	1,2,3,4,5,6,7,8,9,A,B,C,D,,E,F
4	O,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
5	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
6	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
7	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
8	O,1,2,34,5,6,7	8,9,A,B,C,D,E,F
9	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
A	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
B	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
C	O,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
D	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
E	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
F	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

Table 4 Minimum Scrambling Source Coding Map from Table 3

Upper Sub	Lower Sub-Symbol (Hex)															
Symbol	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	Z	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

1	X		Y		Y		Y		Y		Y		Y		Y	
2	X	X			Y				Y				Y			
3	X				Y				Y				Y			
4	X	X	X	X					Y							
5	X								Y							
6	X	X							Y							
7	X	X							Y							
8	X		X	X	X	X	X	X								
9	X															
A	X	X														
B	X															
C	X	X	X													
D	X															
E	X	X														
F	X															

2.2 Combination of multi sub-symbol

Most of the characters codes are made of by means of combination of sub-symbols such as additional sub-symbols are added to a basic 8-bit symbol in Table 1 and Table 3. For instance, a symbol X_1+ X_2 in Table 1 and Table 3 is added to additional sub-symbols X_3+X_4, ···, X_n. The additional sub-symbols or following sub-symbols are combined to a preceding symbol according to the combination rule to make a basic symbol in Table 1 and Table 3.

2.3 Combination between Preceding code and Following code

When a code is made of by combination of sub-symbols, combination criteria is the leftmost bit in lower sub-symbol in Table 1 the rightmost bit in upper sub-symbol in Table 3. That is, when the leftmost bit in lower sub-symbol in Table 1 or the rightmost bit in upper sub-symbol in Table 3 in following code is connected to the rightmost bit in upper sub-symbol in preceding code, scrambling is generated because of occurrence of over consecutive 4-bit 0. For example, in Table 1 and Table 2, Let assume there are two codes and are connected when are sequentially entered into line coder to be transformed to transmission signal. The preceding and following code is X_2+X_1= $(1+1)_{hex}$ = $(00010001)_{binary}$ and X_2+X_1= $(2+2)_{hex}$ = $(00100010)_{binary}$ respectively. Here upon, low subscript 1 and 2 is expressed lower and upper sub-symbol respectively. As shown here, each of the two codes does not be scrambled. But when the two codes are combined such as $(2+2+1+1)_{hex}$ = $(0010001000010001)_{binary}$, scrambling is generated. Therefore, combination between preceding code and following code should be considered to make unscrambling code. Mark “Y” in Table 2 and Table 4 is scrambling symbol which is occurred when preceding symbol and following symbol are connected.

2.4 Minimum-Scrambling Source Coding Method

To make a minimum-scrambling source coding, firstly each character is arranged in coding area in accordance with the use frequency and characteristic of the character. For example, characters that have high probability in use frequency and communication protocols are arranged in unscrambling area. Otherwise, characters that have low probability in use frequency and are used for device-own control are arranged in a scrambling occurrence area.

Let suppose as follows. *C (i)* is total number of symbols or codes in coding Table. Here, $i = 2n$, n is the number of bit a symbol. For example, if the bit number of a symbol is 7-bit and 8-bit, the number of symbols is 128ea and 256ea respectively. *Ss (lu)* or *Ss (ul)* is scrambling symbol, *l*=lower sub-symbol, *u*=upper sub-symbol. Therefore, an 8-bit symbol generates maximum two times scrambling. For instance, a symbol which consists of lower sub-symbol 0_{hex} plus upper sub-symbol $_{0hex}$ occurs two scrambling. *Ss (r)* and *Ss (c)* is the number of scrambling symbols in a column or row respectively. Here, *r* and *c* is shown the figure *0~F* in column or row.

( ) ( ) Σ ( ) ( ) (1)

So, there are maximum 16 scrambling in a row or column. Let Ss_T be total number of scrambling symbols in the coding Table.

Σ ( ) ( ) (2)

Let *P (i)* be the probability of use frequency of a character.

Σ ( ) (3)

Let S_{SUT} be total scrambling-unrelated symbols. Then, total number of symbols or codes *C (i)* is,

( ) (4)

Then, average amount of character use frequency is Cf.

( ) ( ) (5)

Table 2 is a minimum source coding map based on Table 1. The coding map has four kinds of symbols, X, Y and Z. Each of these symbols has scrambling weight value. But the scrambling weight values are different in a symbol and a combination symbols. The scrambling weight value of X, Y and Z is 2, 1 and 3 in a symbol respectively. Let assume that scrambling frequency rate in a symbol Sfr be. Then (1) is developed as follows,

( ) ( ) Σ( ) ( ) ( ) (6)

Let C_N be total source coding characters. If $S_{SUT} \geq C_N$, scrambling is not occurred. But If $S_{SUT} \leq C_N$, scrambling is generated. In the latter case, following requirement must be satisfied to minimize the average amount of scrambling frequency rate Sfr_{av}.

( ) ( ) ( ) ( ) (7)

( ) ( ) ( ) ( ) (8)

3 Minimum-Scrambling Source Coding Analysis

Table 2 is a minimum source coding map based on Table 1. The coding map has four kinds of symbols, X, Y and Z. Each of these symbols has scrambling weight value. But the scrambling weight values are different in a symbol and a combination symbols. The scrambling weight value of X, Y and Z is 2, 1 and 3 in a symbol respectively. From (6), the scrambling frequency rate in a column or row in Table 2 is calculated as follows,

( ) ( ) * ( ) ( ) ( )
( ) * ( ) ( ) ( ) ( ) ( ) ( )
( ) ( )

( ) ( ) * ( ) ( ) ( ) ( )
( ) * ( ) ( ) ( ) ( )
.
.
.
( ) ( )

(9)

From (9), the low sub-symbol $_{0hex}$ in row is the worst one to combine to upper sub-symbols and next 1, 2, 3 and so on sequentially. In case of the low sub-symbol $_{0hex}$ in row, scrambling symbols are 16ea and scrambling occurrence is 17times because but scrambling frequency rate is 33. Let assume that a character is coded with 4 sub-symbols with 16-bit like $_{X4+X3+X2+X1.}$ Here, low subscript is the sequence number of combination. If the code consists of $_{04+03+02+01}$ and $8_4+4_3+2_2+1_1$, scrambling is occurred 4-time and 3-time as follows respectively.

( ) = 0000 0000 0000 0000$_{bin}$ (10)
( ) = 1000 0100 0010 0001$_{bin}$

In these two cases, scrambling occurs more than two times in one code. However, in case of $_{54+23+32+11}$ and $_{54+63+32+21,}$ scrambling is not occurred.

( ) ( ) (11)
( ) ( )

Let the number of scrambling characters be 16ea. The characters are coded sequentially from the leftmost blank in 0*th* row in Table 2. Let suppose that a character $C_N(0)$ is coded in the leftmost coding position $_{00hex}$ *(lu)* and probability of the character, *P (0)* is 0.3. And then, 16- characters are coded by symbols in row 0 to 15 with same condition as the leftmost one and even probability. In this case, according to (5) and (6), the average amount of the scrambling frequency Sfr_{av} *is* calculated as follows,

([( ) ( )+ [∑ ( ) ( )+ (12)
[( ) ( )+ [( ) ( )+
[( ) ( ) [( ) ( )+
( ) [( ) +

However, when characters are coded in 1*th* and 9*th* in row with equal condition in 0*th* in row, average amount of the scrambling frequency, Sfr_{av}(1*th*) and Sfr_{av}(9*th*) are as following,

( ) ( ) [( )+ (13)

( ) ( ) [( )+

If character of probability 0.3 in symbol *Ss (10)* and *Ss (90)* is assigned in unscrambling symbol, average amount of scrambling frequency rate for characters in the rows becomes 0.658 and 0.094 respectively.

5 Conclusion

This paper develops the minimum-scrambling character source coding method for the 8-bit code based on upper sub- symbol 4-bit and lower sub-symbol 4-bit. The codes made by the method will be optimum source codes in the line coder which uses the HDB-3 scrambling method. When the character is coded by using the developed method of the minimum-scrambling source coding, the scrambling frequency in the line coding is drastically minimized.
This paper simulates the developed method with Latin UTF-8 code table but this method can be applied all kinds of source coding system which affects data transmission signal. The simulation result shows that average amount of scrambling frequency rate is decreased 0.0 from minimum 0.43 to maximum 0.96. This result indicates that from the viewpoint of cost-effectiveness, operational and data transmission efficiency in the line coder is greatly improved. Besides such effects,

operation flexibility of device also improves. The reason is that the scrambling is performed in the physical layer with hardware-based but the minimum-scrambling source coding method is treated in the presentation layer with software-based same as to an existing source coding layer. Hereafter, the character source coding system such as Spanish, Chines, Korean and so on would be developed on the basis of the developed minimum scrambling method. To do that and prove the effectiveness, feasibility, reality and so on of the developed method in this paper, the similar source coding systems such as UNICODE, UTF-8 coding system and so on will also be studied.

References

1. Information and Communication Systems Character Code Structure and Extension Techniques, ISO/IEC 2022 ECMA-35 6th Ed standards-1994
2. Behrouz A. Forouzan, “4.1 Digital to Digital conversion,” Data communications and networking, 4th ed., McGraw Hill New York, 2007, pp.101.
3. Murat Torlak, “Digital Transmission (line coding),” in Telecom. Switching & Transmission, the University of Texas at Dallas. Available: http://www.utdallas.edu/~torlak/courses/ee4367/lectures/CodingI.pdf
4. Behrouz A. Forouzan, “4.1 Digital to Digital conversion.” Data communications and Networking, 4th ed., McGraw Hill New York, 2007, pp.118.
5. Physical/ electrical characteristics of hierarchical digital interface, ITU-T Rec. G.703 standards-1998.
6. Test Method for Telecommunication Terminal Equipment, TTA TTAS. KO-05.0028/R1 standards-2004
7. David A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes,” in Proc. I.R.E, 1952, pp. 1908-1101.
8. David A. Huffman, “A Method for the Construction of Minimum-Redundancy Codes,” in Proc. I.R.E, 1952, pp. 1909.
9. Information and Communication Systems Character Code Structure and Extension Techniques, ISO/IEC 2022 ECMA-35 6th Ed.-1994
10. American National Standard for Information Systems-Coded Character Sets 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII), ANSI X3.41986.
11. Unicode/Versions, wikibooks. Accessed May 09, 2014. Available: http://en.wikibooks.org/wiki/Unicode/Versions#Unicode_1.1. Apr 2014
12. UTF-8 Usage Statistics, BuiltWith® Pty Ltd. Accessed May 09, 2014. Available: http://trends.builtwith.com/encoding/ UTF-8
13. Unicode Wikipedia. Accessed May 09, 2014. http://en.wikipedia.org/wiki/Unicode_Transformation_Format. Available: #Unicode_ Transformation_Format_and_Universal_Character_Set.
14. Wan-Pyo Hong, “Coding Rule of Characters by 2 bytes with 4x4 bits to Improve the Transmission Efficiency in Data Communications”, Journal of Korea Navigation Institute, Vol. 15, No. 5, pp.749-756, Oct 2011.
15. 8-Bit Single-Byte Coded Graphic Character Sets: Latin Alphabet No. 5 ISO/IEC 2022 ECMA-128 2nd Ed.standards-1999
16. Information technology - 8-bit single-byte coded graphic character set -- Part 16: Latin alphabet No. 10, ISO/IEC FCD 8859-16 standards-2000.