
데이터 전송 효율을 고려한 3x4비트 1 바이트 문자 부호화 규칙에 관한 연구

홍완표*

Study on the coding rule of Characters by 1 byte of 3x4 bits with consideration for the transmission efficiency of data communications

Wan-Pyo Hong*

요 약

본 논문은 컴퓨터 등 정보기기에 입력되는 문자나 기호의 부호화 규칙을 제시하였다. 문자나 부호를 부호화할 때 본 논문에서 제시하는 규칙을 적용할 경우, 데이터를 전송로에 전송할 때 회선부호화과정에서 발생하는 스크램블링의 발생으로 인한 전송효율이 떨어지는 것을 감소시킬 수 있다. 본 논문에서는 3비트열 x 4비트행의 1바이트 부호화에 대한 것을 고찰하였다. 이를 위해 대표적인 3비트 열 x 4비트 행의 1바이트 부호체계인 ASCII 부호체계를 적용하였다.

ABSTRACT

This paper propose the rule of coding for the characters and symbol, etc which are used in computer, information devices, etc. When they use the rule of coding, they may improve the efficiency of transmission in data communications by reducing the number of scrambling during the line coding in the coder in the transmitter. The paper considered the codes of one byte of 3bits column x 4bits rows. ASCII code is one of the representative code for the codes of one byte of 3bits column x 4bits rows. Therefore, this paper applied ASCII code to study the coding rule of the characters and symbols, etc.

키워드

source coding, line coding, transmission efficiency, ASCII, HDB-3
원천부호화, 회선부호화, 전송효율, 아스키코드, 고밀도2진부호화-3

1. 서 론

자판기(keyboard) 등 정보입력기로부터 컴퓨터 등 정보기기에 입력되는 문자나 기호는 정해진 부호로 부호화되어 정보기기에 저장되거나 프린터 등을 통하여 출력된다. 정보기기에 입력되는 문자나 기호를 2진 비트로 구성된 일정한 패턴으로 변환하는 것을 원천

부호화라고 한다.

ASCII 또는 EBDIC 부호는 각각의 문자나 기호들을 2진 비트로 표현한 대표적인 문자, 기호 부호이다. 이렇게 부호화되어 정보기기에 저장되거나 입력되는 데이터는 전송장치를 통해 인터넷에 전송된다. 컴퓨터 등 정보기기로부터 인터넷에 전송되기 위해 전송장치에 입력되는 정보의 2진 부호는 원거리 전송에 적합

* 한세대학교 정보통신공학과 (wphong@hansei.ac.kr)

접수일자 : 2011. 07. 02

심사(수정)일자 : 2011. 08. 10

게재확정일자 : 2011. 08. 12

한 신호로 부호화된다. 이것을 회선부호화라고 한다. 전송장치에 입력된 2진 데이터를 회선부호화하는 과정에서 입력되는 2진 비트열에 “0”이 연속되어 입력될 경우, 수신기에서 각 비트에 대한 동기를 맞추지 못하여 데이터를 복호화하지 못하는 경우가 발생한다. 이러한 문제를 해결하기 위해 입력되는 데이터에 일정 갯수의 연속된 “0”이 입력될 경우 이 연속된 “0”의 비트열을 “0”의 비트가 연속되지 않도록 정해진 비트 패턴으로 바꾸어 준다. 이러한 과정을 스크램블링이라고 한다. 그러므로 정보기기에 저장된 문자나 기호 등을 표현하는 2진부호에 “0”의 연속이 많이 발생하는 부호가 있으면, 회선 부호화 과정에서 스크램블링이 많이 발생하게 된다. 결국 이것은 데이터의 전송효율을 떨어트리게 되는 요인이 된다.

회선부호화 방식으로 AMI방식을 사용하는 경우에 사용되는 대표적인 스크램블링 방식은 B8ZS와 HDB-3방식이 있다. 전자는 미국의 표준방식이고 후자는 ITU-T[1]와 우리나라의 표준방식[2]이다.

본 논문은 국내 표준인 HDB-3 스크램블링 방식을 사용하는 경우에 스크램블링이 가능한 최소화되도록 하여 전송효율을 떨어뜨리지 않도록 하는 문자나 기호 등의 원천부호화 규칙을 제시하였다. 현재의 원천부호화방법은 기본적으로 3비트열 x 4비트행의 1바이트 또는 4비트열 x 4비트행의 1바이트로 부터 문자나 기호를 부호화한다. ASCII 부호와 EBDIC 부호는 이것들에 대한 대표적인 부호체계이다.

본 논문에서는 3비트열 x 4비트행의 1바이트로 부호화하는 ASCII 부호를 기준으로 하여 문자나 기호 등의 부호화규칙을 제시하였다.

본 논문에서는 문자나 기호 등의 원천부호화과정에서 각각의 부호 그리고 부호와 부호간을 조합할 때 연속된 “0”의 비트가 4개이상 발생하는 부호체계를 찾아내는 규칙을 제시하였다. 즉 이를 통하여 문자나 기호를 부호화 할 때 연속된 4개이상의 “0”이 발생하는 부호를 사용하지 않도록 하는 것이다. 즉, 본 논문에서 제시한 규칙을 적용할 경우 회선부호화과정에서 발생하는 스크램블링을 최소화할 수 있고 이를 통하여 전송효율을 향상시킬 수 있도록 하였다.

II. ASCII 부호체계

ASCII(American Standard Code for Information Interchange)는 영어의 알파벳순으로 문자나 기호 등을 부호화한 체계이다[3]. ASCII 부호는 컴퓨터, 통신 장비 및 기타 문자를 처리하는 장치에서 사용된다. 현재의 대부분의 문자 부호화 체계는 ASCII 부호체계를 근간으로 하고 있다. 다만 현재의 부호체계는 ASCII 부호체계보다 상당히 많은 문자 부호를 가지고 있다. ASCII부호체계는 전신부호로부터 시작되었다. 이 부호체계의 최초의 상업적 사용은 벨 데이터서비스에 의해 시작된 7비트 전신인쇄기 부호였다. 이 부호체계에 대한 최초 표준은 1963년에 공표되었다 [4][5]. 가장 최근에 수정보완 된 것은 1986년에 수정된 것이다[6]. 이 수정 보완된 부호체계는 초기 전신부호와 비교하여 알파벳 순으로 되었고 전신인쇄기에서 사용하였던 것 보다 더 많은 장치들에 대한 특징을 추가하게 되었다. 이 ASCII부호 체계는 2007년 10월까지 웹에서 일반적으로 사용된 문자 부호계였다 [7]. 현재는 웹상에서의 부호체계는 일반적으로 UTF-8 부호체계를 [8][9][10] 사용하고 있다. ASCII 부호 체계는 128개의 문자를 정의하고 있다. 이중 33개는 현재 대부분 사용하고 있지 않은 비인쇄 제어문자들이다. 94개의 문자만 인쇄 가능한 문자이다. 이중에서 문자간의 간격을 나타내는 “스페이스”문자는 가시적으로 표현되지 않는 문자로 취급된다[11].

ASCII 부호 체계내에서 0x20-0x7E는 인쇄문자, 표현문자, 숫자, 기능부호와 몇 가지 기호에 대한 부호계이다. 부호 0x20는 스페이스(space)문자로서 키보드(keyboard) 상에서 스페이스 바(bar)에 의해 입력되는 문자간의 간격을 표시하는 문자이다. 유니코드(unicode)에서는 U+2420 “SP”에 해당하는 문자이다 [12]. 0x00 -0x1F의 부호계는 제어문자로서 인쇄되지 않는 문자의 부호계이다. 표1은 ASCII 부호체계이다.

III. 문자 부호화 규칙

1. 부호화 규칙 조건

표 1. ASCII 부호체계
Table 1. ASCII code table

					0	0	0	0	1	1	1	1	1
					0	0	1	1	0	0	1	0	1
Bits	b ₄	b ₃	b ₂	b ₁	Column	0	1	2	3	4	5	6	7
0	0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	9	HT	EM)	9	I	Y	i	y
1	0	1	0	10	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	11	VT	ESC	+	;	K	[k	{
1	1	0	0	12	12	FF	FC	,	<	L	\	l	
1	1	0	1	13	13	CR	GS	-	=	M]	m	}
1	1	1	0	14	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	15	SI	US	/	?	O	_	o	DEL

본 논문에서는 각 문자나 기호를 3비트열 x 4비트 열의 1바이트로 부호화하는 ASCII 부호 체계와 HDB-3 스크램블링 방식[13]을 기준으로 하여 문자부호화 규칙을 연구하였다. 본 논문에서 제시한 문자부호화 규칙의 원리는 문자나 기호의 원천부호를 구성할 때 그리고 각 문자나 기호의 부호가 조합될 때, 연속된 4개 이상의 "0"이 발생되지 않도록 하였다.

2. 3비트 열 x 4비트 행 조합 규칙

표1에서 각 문자와 기호 등은 7비트의 1바이트로 부호화된다[14]. 1바이트 중에서 b7~b5의 3비트를 상위비트 열(열)이라 하고, b4~b1의 4비트를 하위비트 열(행)이라고 한다. 표1에서 상위 비트열 b7~b5이 "000"인 경우에 하위비트열 b4~b1과 조합되는 형태를 재구성해 보면 다음 표2와 같다. 표2에서와 같이 하위비트열 b4~b1의 4개의 비트는 상위비트열 b7~b5의 3개 비트열과 조합된다.

표 2. b7~b5 비트열 중 "000"의 b4~b1 비트열과의 부호조합

Table 2. Character code composition of "000" of b7~b5

구분	b7~b5	b4~b1
NUL	000	0000
SOH	000	0001
STX	000	0010
FTX	000	0011
EOT	000	0100
ENQ	000	0101
ACK	000	0110
BEL	000	0111
BS	000	1000
HT	000	1001
LF	000	1010
VT	000	1011

FF	000	1100
CR	000	1101
SO	000	1110
SI	000	1111

표2에서 상위비트열 b7~b5가 “000”인 경우, 하위비트열 b4~b1의 최상위비트인 b4에 해당되는 비트가 이진 비트“0”이 되는 부호는 통신용 문자나 기호 그리고 자주 사용되는 문자나 기호에는 부여하지 않도록 하는 것이다.

이렇게 각 문자나 기호 등에 부호를 부여할 경우, 전체 문자와 기호를 나타낼 수 있는 부호수가 적어진다. 그러므로 사용하고자 하는 문자와 기호 등 중에서 사용빈도수가 높은, 즉 전송빈도수가 많은 문자나 기호 등에 연속 4개의 이진 비트“0”이 발생되지 않는 부호를 배정하여야 한다.

2.1 3비트 열 x 4비트 행 조합 규칙

표3은 3비트 열 x 4비트 행의 단위조합 비트열에 대한 각 단위조합별 조합이 가능한 것과 조합이 제한되는 조건을 보여주는 것이다. 즉, 3비트로 구성된 8개의 상위비트열과 4비트로 구성된 16개의 하위비트열간의 조합 조건에 대한 것이다.

표3에서와 같이 상위비트열이 이진비트“000”인 경우에 하위비트열의 16진수 0,1,2,3,4,5,6,7과 조합이 제한된다. 이것은 4개의 “0”비트조합이 일어나는 것을

피하여야 하기 때문이다.

이러한 조건으로 상위비트열이 001, 011, 101, 111인 경우에는 하위비트열 16진수 0과만 조합이 제한된다. 상위비트열이 010, 110인 경우에는 하위비트열 16진수 0,1과 조합이 제한된다. 상위비트열이 100인 경우에는 하위비트열 16진수가 0, 1, 2, 3인 경우와 조합이 제한된다. 이 표3에서 알 수 있듯이 하위비트열, b4~b1과 가장 잘 부합되는 1순위 상위비트열, b7~b5은 001, 011, 101, 111이다. 그 다음 2순위로는 010, 110이다. 다음 3, 4순위는 각각 100과 000이다. 즉 3, 4순위는 부호조합에서 가장 적합지 않은 비트열이다. 그러므로 사용빈도가 높은 문자나 기호 등은 1순위 비트열을 이용하여 부호화하도록 하는 것이 바람직하다.

이 표3에 의하면 총128개의 부호중에서 20개의 부호가 조합이 제한된다. 이 부호의 조합이 제한되는 20개의 부호는 통신로상에 전송되는 빈도가 낮은 문자나 기호 등에 배정하는 것이 전송회선부호기 및 복호기의 운용효율을 제고시킬 수 있게 된다.

2.2 2바이트간 조합규칙

표4는 2바이트간의 조합규칙에 대한 것이다. 바이트간의 조합은 첫 번째 출력된 바이트의 최상위 3비트와 두 번째 출력된 바이트의 최하위 4비트가 연결되어 조합된다.

첫 번째 출력되는 한 개의 문자를 표현하는 부호를 나타내는 한 바이트의 최상위 비트열, b7~b5가

표 3. 문자부호화 규칙 ; 3비트 열 x 4비트 행
Table 3. Rule of Characters Coding ; 3-bit column x 4-bit rows

상위비트 열 (2진수, 3비트)	하위 비트 행(16진수, 4비트)	
	부 여 제 한	부 여 가 능
000	0, 1, 2, 3, 4, 5, 6, 7	8, 9, A, B, C, D, E, F
001	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
010	0, 1	2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
011	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
100	0, 1, 2, 3	4, 5, 6, 7, 8, 9, A, B, C, D, E, F
101	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
110	0, 1	2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
111	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

표 4. 바이트 (7비트)간 조합규칙 ; 첫 번째 바이트, b7-b5 기준
Table 4. Compositon Rule between Bytes (7bits)

두 번째 바이트, b4-b1	첫 번째 바이트, b7-b5	첫 번째 바이트, b7-b5 기준		
		두 번째 바이트, b4-b1		
		부 호 부 여 제 한		부 호 부 여 가 능
		조건 1	조건 2	조건
0000	000	0, 2, 4, 6, 8, A, C, E	0	1, 3, 5, 7, 9, B, D, F
0001	001	0, 4, 8, C	0	1, 2, 3, 5, 6, 7, 9, A, B, D, E, F
0010	010	0, 8	0	1, 2, 3, 4, 5, 6, 7, 9, A, B, C, D, E, F
0011	011	0, 8	0	1, 2, 3, 4, 5, 6, 7, 9, A, B, C, D, E, F
0100	100	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
0101	101	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
0110	110	0	0	1, 2, 3, 4, 5, 6, 8, 9, A, B, C, D, E, 7, F
0111	111	0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, C, D, E, B, F
1000	-	-	-	-
1001	-	-	-	-
1010	-	-	-	-
1011	-	-	-	-
1100	-	-	-	-
1101	-	-	-	-
1110	-	-	-	-
1111	-	-	-	-
	-	-	-	-

001이라고 하자. 이때 첫 번째 바이트에 이어서 출력되는 두 번째 바이트의 최상위 비트열 b4~b1의 비트열이 0100이라고 하자.

이 경우에 첫 번째 바이트와 두 번째 바이트는 0100 001과 같이 조합된다. 이 두 바이트의 연결 결과, 이진비트 “0”이 연속으로 4개가 발생하게 되었다. 이결과는 문자부호과정에서 이진비트 “0”이 연속으로 4개 나오지 않도록 하는 규칙에 위배되는 것이다. 그러므로 한 개의 문자나 기호 등의 한 바이트 조합에 대한 것과 함께 한 문자나 기호 등을 구성하는 바이트간, 즉 한 문자나 기호 등간의 비트열 조합에 대한 조건도 고려하여야 함을 알 수 있다.

표4에 바이트간의 연결에 대한 조합제한과 조합가능 조건을 보여 주고 있다. 이 표4의 분석결과에 의하면 첫 번째 출력되는 바이트의 최상위 3개의 비트열이 “000”인 경우, 두 번째 출력되는 바이트의 최하위 4개의 비트열 16진수 중에서 0, 2, 4, 6, 8, A, C, E와 조합이 제한된다.

첫 번째 바이트의 최상위 비트열이 “001”인 경우, 두 번째 바이트의 최하위 비트 열 16진수는 0, 4, 8, C

와 조합이 제한된다. 첫번째 바이트의 최상위 비트열이 “010” “011”인 경우, 두 번째 바이트의 최하위 비트열, 16진수 0, 8과 조합이 제한된다. 첫 번째 바이트의 최상위 비트열이 “100”, “101”, “110”, “111”인 경우, 두 번째 바이트의 최하위 비트 열 16진수 0과 조합이 제한된다. 표3과 같이 이 표4에서 알 수 있듯이 두 번째 바이트 b4-b1가 가장 잘 부합되는 첫 번째 바이트 b7-b5는 1순위가 100 101 110 111이다. 2순위는 010 011이고 3, 4번 순위는 001과 000이다. 표3과 표4를 통합하여 볼 때 가장 잘 부합되는 3비트열은 1순위가 101, 111이고 2순위는 010, 011, 110이다. 다음 3, 4순위는 각각 001, 100과 000이다. 즉, 이 문자 부호화 규칙에 의거하여 통신용으로 사용되는 문자나 기호중에서 사용빈도가 높은 문자나 기호 등에 조합이 가능한 부호를 부여하는 것이다.

IV. 결 론

정보 기기 내에서 문자를 어떠한 규칙에 의해 부호

화하여 전송하는가에 따라 전송장치의 운용효율에 영향을 줄 수 있다. 본 논문은 컴퓨터 등 정보기기로부터 생성되는 문자나 기호 등에 대한 부호를 전송로를 통하여 원거리에 전송할 때 전송효율을 높일 수 있는 문자, 기호 등에 대한 부호화 규칙을 제시하였다. 현재 ITU-T[1]와 국내[2]의 회선부호화 (Line coding) 방식은 HDB-3 스크램블 방식[13]을 채택하고 있다. 이 방식은 입력데이터의 회선부호화 과정에서 연속되는 "0"의 비트가 4개 이상 발생되지 않도록 하는 방식이다. 이 방식을 채택한 목적은 AMI회선부호화를 적용한 이후에 연속"0"비트로 인한 동기의 소실을 막기 위한 것이다. 따라서 문자를 "0"의 연속이 4개 이상 발생되도록 부호화할 경우 데이터의 전송효율을 감소시키게 된다. 본 논문에서는 문자를 부호화 할 때 사용빈도가 높은 문자의 경우에 "0"의 연속이 4개 이상 발생 되는 부호를 갖지 않도록 하는 문자부호화 규칙을 제시하였다.

향후 본 연구에서 제시한 문자, 기호 등의 부호화 규칙을 기존의 ASCII 부호, EBDIC 부호체계, unicode 부호체계 등에 적용하여 데이터 전송효율에 미치는 영향을 입증할 필요가 있다.

감사의 글

이 논문은 2011년도 한세대학교 교내 학술연구비지원에 의해 연구되었음

참고 문헌

[1] ITU-T Recommendation G.703, "Physical/electrical characteristics of hierarchical digital interfaces", pp. 24-41, Oct., 1998.
 [2] TTA Standard, "Test Method for Telecommunication Terminal Equipment" TTAS. KO-05.0028/R1 pp306-451, Revised on 23 Dec., 2004
 [3] http://en.wikipedia.org/wiki/ASCII#_History
 [4] a b Mary Brandel, "1963: The Debut of ASCII: CNN".
 [5] American Standards Association, "American Code (July 6, 1999). for Information Interchange", ASA X3.4- 1963, 17 Jun., 1963
 [6] American National Standards Institute, "American National Standard for Information Systems-

Coded Character Sets 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII)", ANSI X3.4-1986, Inc., 26 Mar., 1986.

[7] abc Internet Assigned Numbers Authority (May 14, 2007). "Character Sets". Accessed 14 Apr., 2008.
 [8] Dubost, Karl (May, 2008), "UTF-8 Growth On The Web". W3C Blog. World Wide Web Consortium.<http://www.w3.org/QA/2008/05/utf8-web-growth.html>.
 [9] Davis, Mark (May, 2008). "Moving to Unicode 5.1". Official Google Blog. Google. <http://googleblog.blogspot.com/2008/05/moving-to-unicode-51.html>.
 [10] Davis, Mark (Jan., 2010). "Unicode nearing 50% of the web". Official Google Blog. Google. <http://googleblog.blogspot.com/2010/01/unicode-nearing-50-of-web.html>.
 [11] a b Mackenzie, p.223.
 [12] http://en.wikibooks.org/wiki/Unicode/Character_reference/2000-2FFF
 [13] Behrouz A. Forouzan, "Data communications" McGraw Hill Korea, pp.132-134. Jan., 2008.
 [14] 산업자원부 기술표준원. "정보 교환용 부호계 (한글 및 한자) 부속서 4. 7비트 한글 낱자 부호계" KS X 1001 : 2004, 12.

저자 소개

홍완표(Wan-Pyo Hong)



1991년 서울과학기술대학교 전자공학과 졸업(공학사)

1994년 연세대학교대학교 공학대학원 산업공학과 졸업(공학석사)

1999년 광운대학교 대학원 전자

1990년 전기통신기술사합격

1991년 정보통신부 5급특별채용고시합격 본부 통신정책실, 전파방송관리국, 정보화기획실

1997년 삼성전자(주) 통신사업부 전송영업그룹장

1999년 광운대학교 연구전담교수

2000년 한국정보통신기술협회장

2002년 한세대학교 IT학부 정보통신공학전공 교수
한세대학교 정보통신연구소장

※ 관심분야 : 위성통신방송/문자코딩/통신정책