

AMI/HDB-3 회선부호화 및 HDLC FLAG를 고려한 KS X 1001 정보 교환용 로마문자 부호체계고찰

홍완표*

Consideration of Roman Character in KS X 1001 Code System for Information Interchange considered AMI/HDB-3 and HDLC FLAG

Wan-Pyo Hong*

요 약

데이터를 원거리에 전송하는 회선부호화 방식으로는 AMI방식을 사용한다. AMI방식의 단점은 원천부호에 일정한 개수 이상의 0의 비트가 존재할 경우 비트동기를 상실하게 되는 것이다. 이 단점을 보완하기 위해 스크램블링 기술을 적용한다. HDB-3 스크램블링 방식은 원천부호에 비트 0이 네 개 이상 연속하여 있을 때 이것을 인위적으로 변환시킨다. 그러므로 원천 부호 중에 연속하여 네 개 이상의 0의 비트로 구성된 부호가 많을 경우, 회선부호화 과정 중에서 데이터전송효율을 감소시키게 된다. 한편 HDLC 통신규약에서는 플래그 비트열과 유사한 비트열의 오인을 방지위해 비트 또는 문자 스템핑 방식을 사용한다. 본 논문은 이러한 관점에서 KS X 1001에 포함된 로마문자용 부호집합을 분석대상으로 하였다. 이러한 분석결과를 토대로 하여 데이터의 전송효율을 제고시키는 최적의 로마문자 원천부호체계를 제시하였다. 본 연구에는 문자의 (4 x 4)비트 원천 부호화 규칙과 영어 알파벳의 사용빈도 통계를 적용하였다. 연구결과 본 논문에서 제시하는 로마문자용 부호집합체계를 적용할 경우에 약134%의 데이터처리 효율을 제고시키는 것으로 나타났다.

ABSTRACT

Datacommunications transmit the source codes that are coded in information devices, such as computer to the transmission line by means of the line coded signal. AMI method is applied to the line coding method to transmit the signal for long distance. The disadvantage of the AMI method is to lose the bit synchronization when consecutive binary bit '0' over 4ea is coming into line coder. The scrambling technique is used to overcome the problem. The HDB-3 scrambling method is used in Korea standard which standard in ITU-T. When the HDB-3 technology is used, the method should convert the consecutive bit '0' over 4ea to certain bits format. As a result, when there are many such kind of '0' bit stream in source codes, data transmission efficiency will be decreased to treat in line coder, etc. This paper is directed to study the Roman character code system in KS X 1001, Korea standard for information exchange code in datacommunication systems. Based on the study result, this paper proposed the maximum optimized Roman character code system. In the study, Character coding rule for 4 x 4bits and the statistical data for roman character using frequency were considered to simulate. The paper shows the result that when the proposed new roman character coding system is applied to use, the data transmission efficiency could be increased to about 134% compared to existing code system.

키워드

Roman Character, Source Code, Line Coding, HDB-3, HDLC
로마문자, 원천부호, 회선부호화, HDB-3, HDLC

* 교신저자(corresponding author) : 한세대학교 정보통신공학과(wphong@hansei.ac.kr)
접수일자 : 2013. 05. 06

심사(수정)일자 : 2013. 06. 20

게재 확정일자 : 2013. 07. 23

1. 서론

컴퓨터 등 정보 기기의 키보드 등 문자입력 자판에서 정보를 처리하기 위해 입력되는 문자 등의 정보를 부호화하는 것을 원천부호화 (source coding)라고 한다. 대표적인 원천부호체계로는 ASCII [1][2][3] EBCDIC[4] Unicode[5]등이 있다. 이 부호체계 중에서 ASCII와 EBCDIC 부호체계는 하나의 문자가 각각 7비트, 8비트로 부호화된다. 유니코드의 BMP에서는 한 개의 문자나 기호를 32비트로 부호화하고 있다.

데이터통신에서는 컴퓨터 등 정보기기에서 생성된 문자 등의 원천부호를 전송에 적합하도록 변환하여 전송한다. 이것을 회선부호화(line coding)라고 한다. 회선부호화 방식에는 NRZ, RZ, Manchester 및 AMI 방식 등이 있다 [6].

데이터통신에서는 이들 회선부호화 방식 중에서는 장거리 전송에 적합한 AMI방식을 주로 사용하고 있다. 이 AMI방식은 다른 방식에 비하여 교류적인 신호특성 및 점유주파수대역폭 측면에서 유리하다.

그러나 이 방식은 원천부호에 일정한 개수이상의 비트 0이 존재할 경우, 수신측에서 비트에 대한 동기를 상실하는 단점이 있다. 따라서 이러한 비트 0의 연속으로 인한 문제를 극복하기 위해 연속되는 비트 0을 정해진 규칙에 따라 스크램블링(scrambling)한다. 연속되는 비트 0이 4개일 때 스크램블링하는 방식이 HDB(High Density Bipolar 3-zero)방식이다[6]. 연속되는 비트 0이 8개일 때 스크램블링하는 방식이 B8ZS(Binary 8 Zero Sequence)방식이다[7]. 전자는 ITU-T 국제표준방식으로 우리나라에서 적용하고 있는 표준방식이다[8]. 후자는 미국표준방식이다.

따라서 원천부호를 구성하는 비트의 조합은 회선부호기의 운용에 영향을 미치게 되고 결과적으로 이것은 데이터의 전송효율에 영향을 주게 된다.

우리나라의 정보교환용 부호계 (한글 및 한자) (KS X 1001)는 정보처리 및 데이터를 사용하는 시스템에서 정보교환에 사용하는 부호계의 표현 형식을 규정하고 있다[10]. 이 표준에는 한글날자, 한글글자마디와 함께 로마문자에 대한 부호체계가 규정되어 있다[12]. 본 논문은 참고문헌 [13]에서 제시하고 있는 원천부호화 규칙을 적용하여 분석하였다.

이 분석결과를 토대로 하여 HDLC 프로토콜 및

AMI회선부호화 방식에 적합한 로마문자용 부호집합 체계를 제시하였다. 또한 본 논문에서 제시하는 로마 문자부호집합을 적용할 경우 데이터 통신의 전송효율에 주는 영향을 정량적으로 산출하였다.

II. 문자의 원천부호화규칙

표 1은 참고문헌 [13]에서 제시하고 있는 문자의 원천부호화 규칙이다. 이 규칙은 OSI의 2계층인 데이터링크 계층에 적용되는 HDLC프로토콜과 AMI회선 부호방식을 사용할 때 HDB-3의 스크램블링을 적용한 경우에 대한 것이다. 즉 HDLC 프로토콜의 FLAG 비트 01111110와 정보기기에서 생성되는 원천부호에 연속 네 개의 0비트가 발생한 경우, 이에 대한 대처를 고려한 것이다. 상위비트열은 8비트 1바이트를 구성하는 상위4비트를 말한다. 예를 들어 표1에서 @기호의 경우 부호는 A3C0이다. 이 부호는 16진수로 나타낸 것이다. 그러므로 A를 3에 대한 상위비트열이고 3은 C에 대한 상위비트열이고 C는 0에 대한 상위비트열이 된다. 그러므로 표2에서 상위비트열 16진수 0은 그 자체에 연속된 0이 네 개가 존재하므로 어떤 하위비트열과도 조합될 수 없다. 상위비트열 16진수 1의 경우는 하위비트열 16진수 0과 F를 제외한 모든 비트열과 조합이 가능하다. F가조합이 제한되는 것은 16진수 1과 F가 조합될 경우 00011111이 되어 HDLC flag 비트열과 혼동을 일으키기 때문이다. 이러한 규칙으로 하위비트열들을 조합시켜 나갈 수 있다. 이렇게 함으로서 연속하여 비트 0이 네 개이상 조합되는 것과 비트 1이 다섯 개이상 조합되는 부호와 그렇지 않은 부호를 선별해 낼 수 있다.

표 1. 문자 부호화 규칙 ; 4 비트 x 4비트[13]
Table 1. Characters coding rule ; 4-bit x 4-bit

HEXA	upper bits	lower bits	
		Unavailable connection	Available connection
0	0000	0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F	X
1	0001	0,F	1,2,3,4,5,6,7,8,9,A,B,C,D,E
2	0010	0,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F

3	0011	0,E,F	1,2,3,4,5,6,7,8,9,A,B,C,D,
4	0100	0,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
5	0101	0,F	1,2,3,4,5,6,7,8,9,A,B,C,D,E
6	0110	0,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
7	0111	0,C,D,E,F	1,2,3,4,5,6,7,8,9,A,B
8	1000	0,1,2,3,4,5,6,7	8,9,A,B,C,D,E,F
9	1001	0,F	1,2,3,4,5,6,7,8,9,A,B,C,D,E
A	1010	0,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
B	1011	0,E,F	1,2,3,4,5,6,7,8,9,A,B,C,D
C	1100	0,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
D	1101	0,F	1,2,3,4,5,6,7,8,9,A,B,C,D,E
E	1110	0,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
F	1111	0,8,9,A,B,C,D,E,F	1,2,3,4,5,6,7

* 하위비트 행의 숫자는 16진수임

II. 문자원천부호화규칙과 로마문자용 부호집합체계

2.1 로마문자용 부호집합 체계



표 2는 [12]에 규정된 로마문자용 부호집합이다. 이 표에서 보듯이 행은 A3XX로 구성되어 있고 열은 0~15까지로 구성되어 있다. 이 숫자들은 16진수이다. A3A0과 A3FF는 비어있다. 이 부호체계는 ASC11부호체계와 동일한 부호배열체계로 구성되어 있다. 다만 A3A0, A3FF, A3DC부호만 다른데 ASCII부호체계에서는 각각 NUL, DEL 및 \로 되어 있다. 또한 이 부호체계는 2바이트 16비트로 구성되어 있는데 ASCII 부호는 7비트로 구성되어 있다. Unicode에서는 ASCII 부호가 8비트로 구성되어 있다. 이 표1의 로마문자용 부호체계의 숫자와 대소 알파벳 문자부호는 순서대로 배열되어 있음을 알 수 있다. 일부 유사한 부분도 있지만 특별히 자판(keyboard)의 배열을 고려하여 배열된 것도 아니다. 또한 이 부호체계는 데이터통신의 회선부호화를 염두에 두고 부호화된 것도 아님을 알 수 있다. 한편 모르스부호 등과 같이 사용빈도에 의한 통

계적 부호체계로 배열된 것도 아님을 알 수 있다.

표 2. 로마문자용 부호집합[12]
Table 2. Roman characters code system

	A3A	A3B	A3C	A3D	A3E	A3F
0		0	@	P	`	p
1	!	1	A	Q	a	q
2	"	2	B	R	b	r
3	#	3	C	S	c	s
4	\$	4	D	T	d	t
5	%	5	E	U	e	u
6	&	6	F	V	f	v
7	'	7	G	W	g	w
8	(8	H	X	h	x
9)	9	I	Y	i	y
A	*	:	J	Z	j	z
B	+	;	K	[k	{
C	,	<	L	W	l	
D	-	=	M]	m	}
E	.	>	N	^	n	~
F	/	?	O	_	o	

2.2 원천부호화규칙 적용 로마문자용부호집합

표 3은 표 1의 원천부호화 규칙을 표 2에 적용한 것을 보여 주는 것이다. 이 표 1의 원천부호화규칙에서 상위비트열 16진수 A에는 16진수 0과 1을 제외한 모든 비트열이 조합이 가능하다. 그러므로 표 2에서 보듯이 최상위비트열 다음의 16진수 3은 조합이 가능하다. 다음으로 16진수 3은 16진수 0, E와 F가 조합이 제한된다. 표 3에서 이 이에 해당된다. 그러므로 소문자 알파벳을 포함한 일부 기호가 규칙에 위배되고 있음을 알 수 있다. 다음으로 3번째 비트열 A는 0,1, B는 0,E,F, C는 0,1,2,3, D는 0,F, E는 0,1, F는 0,8,9,A,B,C,D,E,F가 조합이 제한된다. 이 표에서 는 원천부호화 규칙에 두 번이상 위배되는 부분을 나타낸다. 즉 알파벳의 경우, 대문자 A, B, C와 P가 해당된다. 알파벳 소문자의 경우에는 모두 원천부호화 규칙에 위배되는 부호열에 배치되어 있음을 알 수 있다. 특히 a, x, y, z는 두 번 중복하여 위배되

는 것으로 나타났다.

결과적으로 이 표 3에서 알 수 있듯이 원천부호화 규칙에 위배되지 않는 부호는 총 53개 된다. 알파벳 대소문자가 총52개 임을 고려하여, 원천부호화에 부합되는 부호열에 알파벳 대소문자 모두를 배열하였다. 그리고 기호중에서 가장 많이 사용되는 “.”을 나머지 하나에 부여하였다. 숫자는 A3E2~A3E8까지 순서대로 부여하였다. 기타 기호는 문자입력자판의 좌측 상단으로부터 우측으로 순서대로 부여하였다. 다만 사용빈도가 높은 “.”는 원천부호화 규칙에 위배되지 않는 A3BE에 부여하였고 사용빈도가 상대적으로 높은 “,” 과 “?”는 “=”과 “+”와 대치하여 부여하였다.

표 3. 원천부호화규칙을 고려한 로마문자용 부호집합
Table 3. New roman character code system by characters coding rule

A3A	A3B	A3C	A3D	A3E	A3F
	,	#	&	-	
~	O	\$	m	-	\
A	P	%	n	0	{
B	Q	^	o	1	}
C	R	a	p	2	[
D	S	b	q	3]
E	T	c	r	4	:
F	U	d	s	5	;
G	V	e	t	6	“
H	W	f	u	7	’
I	X	g	v	8	<
J	Y	h	w	9	>
K	Z	i	x	(/
L	.	j	y)	+
M	!	k	z	,	=
N	@	l	*	?	

2.3 로마문자용 부호집합체계 개선효과

가. 개선효과 분석용 알파벳 사용빈도통계

이상에서 문자의 원천부호화 규칙을 적용한 로마문자용 부호집합체계의 문제점과 개선체계를 제시하였다. 다음은 이러한 새로 제시한 로마문자용 부호집합체계가 기존의 부호집합체계와 비교하여 데이터의 전송효율측면에서 어느 정도 개선되었는지 분석하였다. 이 분석에 대한 정량적 결과를 얻기 위해 표 4의 영어알파벳 사용빈도 통계자료를 사용하였다.

이 사용빈도에 대한 통계자료는 모르스 부호, 옥스퍼드 사전 및 실용단어의 사용빈도를 적용한 것이다. 모르스 부호는 사용빈도가 높은 알파벳에 부호구성이 가장 간단한 부호를 부여하였다. 모르스는 단순 평문 문서에서 사용빈도수가 높은 알파벳을 추출하였다. 따라서 실제 다양한 단어가 사용되는 문서에서의 알파벳 사용빈도수 통계는 이와는 다르게 나타날 수 있다.

옥스퍼드 사전에 의한 사용빈도는 컨사이스 옥스퍼드사전 (Concise Oxford Dictionary)의 단어 목록 글자로부터 추출한 것이다.

실생활사용단어의 사용빈도는 영국, 미국 및 호주의 영자신문, 잡지, 서적, TV, 라디오 및 실생활회화에서 사용하는 영어에서 가장 많이 사용되는 500단어에서 추출한 것이다.

표 4. “라틴어 사용빈도에 대한 모르스 부호, 옥스퍼드 사전, 실용500단어 및 전체평균”

Table 4. “Usage frequency of morse code, oxford dictionary, practical english and total average in latin”

Morse code		Oxford Dictionary		Practical 500words		Average	
E	17.83	E	11.16	E	11.7	E	13.57
T	13.37	A	8.50	A	7.91	A	9.43
A	11.89	R	7.58	O	7.87	T	9.25
I	11.89	I	7.54	T	7.43	O	8.97
N	11.89	O	7.16	R	7.33	N	8.33
O	11.89	T	6.95	N	6.45	I	8.25
S	11.89	N	6.65	S	5.52	R	8.04
H	9.51	S	5.74	I	5.32	S	7.71
R	9.21	L	5.49	L	5.18	H	5.72
D	6.54	C	4.54	H	4.64	L	5.54
L	5.94	U	3.63	D	4.01	D	4.64
U	5.05	D	3.38	U	3.57	C	4.09
C	4.46	P	3.17	C	3.27	U	4.08
M	4.46	M	3.01	W	2.93	M	3.45

F	3.71	H	3.00	M	2.88	P	2.69
W	2.97	G	2.47	G	2.59	F	2.61
Y	2.97	B	2.07	P	2.39	G	2.53
G	2.53	F	1.81	F	2.30	W	2.40
P	2.53	Y	1.78	Y	2.00	Y	2.25
B	2.38	W	1.29	B	1.86	B	2.10
V	1.78	K	1.10	V	1.22	V	1.34
K	1.19	V	1.01	K	1.17	K	1.15
Q	0.74	X	0.29	X	0.20	X	0.36
J	0.59	Z	0.27	J	0.10	Q	0.35
X	0.59	J	0.20	Q	0.10	J	0.30
Z	0.30	Q	0.20	Z	0.05	Z	0.21

나. 개선효과 분석 방법

이러한 새로 제시한 로마문자용 부호집합체계가 기존의 부호집합체계와 비교한 데이터의 전송효율측면에서의 분석은 세 가지 점을 고려하였다. 이 분석은 정량적 결과를 얻을 수 있는 사용빈도에 대한 통계자료가 있는 알파벳 부호에 대하여서만 이루어졌다.

첫째는 OSI의 데이터링크 계층에서 수행되는 HDLC프로토콜의 Flag비트열에 대한 stuffing 측면에서 분석하였다. 두 번째는 OSI의 물리계층에서 수행되는 회선부호화과정의 HDB-3 스크램블링측면에서 분석하였다. 세 번째는 이 두 개의 분석결과를 종합하여 전체적인 개선효과를 추출하였다.

나. 개선효과 분석 및 결과

1) 데이터링크 계층 개선효과분석

Violated code in Source code	Characters
A3BE	>
A3BF	?
A3DF	-
A3E0~A3EF	a ~ o
A3F0~A3FF	p ~ z { \ }
A3F8~A3FF	x y z { \ }

0 사용빈도율 (개선 율) ;

$$\sum_a^z(\text{사용빈도율}) + \sum_x^z(\text{사용빈도율}) \quad (1)$$

$$= 100 + (0.36 + 2.25 + 0.21)$$

$$= 102.82(\%)$$

2) 물리계층 개선효과분석

Violated code in Source code	Characters
A3AO	None
A3A1	!
A3A0-3	@ A B C
A3D0	P
A3B0	0
A3E0-1	` a
A3F0	p

$$\sum_A^C(\text{사용빈도율}) + P + a + p \quad (2)$$

$$= (9.43 + 2.1 + 4.09 + 2.69 + 9.43 + 2.25)$$

$$= 31.44(\%)$$

3) 전체 개선도

$$(1) + (2) = 134.26(\%)$$

III. 결론

본 논문에서는 우리나라 정보 교환용 부호 표준(KS X 1001; 2009확인)에 규정되어 있는 로마문자용 부호집합에 대하여 데이터 링크 계층의 HDLC Flag

의 비트 또는 문자 stuffing, 물리계층의 AMI/HDB-3 스크램블링 측면에서의 새로운 로마문자용 부호집합 체계를 제시하였다.

이를 위해 (4 x 4) 비트열에 대한 문자의 원천부호화 규칙과 알파벳의 사용빈도에 대한 통계를 사용하였다.

연구결과 새로운 로마문자용 부호집합체계를 적용할 경우 기존 부호 집합 체계 대비 약 134%의 개선 효과가 있는 것으로 나타났다. 따라서 본 연구결과에서 제시한 로마문자용 부호집합체계를 사용할 경우, 알파벳의 전송에 있어서는 데이터링크계층에서의 HDLC 프로토콜에서의 Flag로 인한 문자부호열에 대한 비트 또는 문자 스테핑과 물리계층에서의 HDB-3 스크램블링이 완전히 발생되지 않는 것으로 나타나 이 문자부호체계를 적용할 경우에 문자의 전송효율개선에 큰 영향을 줄 것으로 전망된다. 향후 추가 연구 사항으로는 AMI회선부호화과정에서 직류성분을 완전히 제거할 수 있는 원천부호체계를 제시하는 것이다. 또한 본 논문에서 제시한 새로운 로마문자용 부호집합체계를 UTF-8 등의 부호체계로 변환할 때 어떠한 영향이 있는가를 연구할 필요가 있다.

감사의 글

이 논문은 2013년도 한세대학교 교내 학술연구비 지원에 의해 연구되었음

참고 문헌

[1] American Standards Association, "American Code (July 6, 1999). for Information Interchange", ASA X3.4-1963, 17 June 1963.
 [2] 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 March 1986.
 [3] RFC 20 "ASCII format for Network Interchange" October 1969 (<http://tools.ietf.org/html/rfc20>)
 [4] <http://en.wikipedia.org/wiki/EBCDIC>

[5] <http://en.wikipedia.org/wiki/Unicode>
 [6] Behrouz A. Forouzan, "Data communications" McGraw Hill Korea, p. 1031. 2007.
 [7] ITU-T Recommendation G.703, "Physical/ electrical characteristics of hierarchical digital interfaces", pp. 24-41, Oct. 1998.
 [8] TTA Standard, "Test Method for Telecommunication Terminal Equipment" TTAS. KO-05.0028/R1, pp. 306-451, Revised on 23 Dec. 2004.
 [9] Young-Oh Han, "A study on motion prediction and subband coding of moving pictuers using GRNN", The Journal of The Korea Institute of Electronic Communication Sciences, Vol. 5, No. 3, pp. 255-265, 2010.
 [10] Ministry of Trade, Industry & Energy, Korean Agency for Technology and Standards" CODE FOR INFORMATION INTERCHANGE(HANGUL AND HANJA) Annex 3. Supplemental Code(2 bit combination type, KS X 1001 : Dec. 2004. (Confirmed 2009)
 [11] Jin-woo Yu, Hyung-kyu Im, Se-won Park, "Constructions of the special sign pattern matrices that allow normality", The Journal of The Korea Institute of Electronic Communication Sciences, Vol. 6, No. 2, pp. 193-198, 2011.
 [12] Ministry of Trade, Industry & Energy, Korean Agency for Technology and Standards" CODE FOR INFORMATION INTERCHANGE(HANGUL AND HANJA) p. 9, Table 1. row 3 : Roman characters code
 [13] Wan-Pyo Hong, "Coding Rule of Characters by 2 bytes with 4x4 bits to Improve the Transmission Efficiency in Data Communications", The Journal of Korea Navigation Institute, Vol. 15, No. 5, 2011.
 [14] Chang-young Lee, "Improvement of the Linear Predictive Coding with Windowed Autocorrelation", The Journal of The Korea Institute of Electronic Communication Sciences, Vol. 6, No. 2, pp. 186-190, 2011.

저자 소개



홍완표(Wan-Pyo Hong)

1991년 서울과학기술대학교 전자공학
학과 졸업(공학사)

1994년 연세대학교대학교 공학대학
원 산업공학과 졸업(공학석사)

1999년 광운대학교 대학원 전자공학과 졸업(공학박사)

1990년 전기통신기술사합격

1991년 정보통신부 5급특별채용고시합격 본부 통신
정책실, 전파방송관리국, 정보화기획실

1997년 삼성전자(주) 통신사업부 전송영업그룹장

1999년 광운대학교 연구전담교수

2000년 한국정보통신기술협회장

2002년 한세대학교 IT학부 정보통신공학전공 교수

한세대학교 정보통신연구소장

※ 관심분야 : 위성통신방송, 문자코딩, 통신정책