

AMI/HDB-3 회선부호화 및 HDLC FLAG를 고려한 KS X 1001 정보교환용 한글날자 부호체계 개선연구

우제택* · 홍완표**

A Study on the Hangul Character Code System for KS X 1001 Information
Interchange considering AMI/HDB-3 Line Encoding and HDLC Flag

Je-Teak Woo* · Wan-Pyo Hong**

요 약

스크램블링 기술을 적용한 AMI/HDB-3 방식은 장거리 데이터전송 회선부호화에 주로 사용된다. 본 논문은 정보통신용 부호 표준(KS X 1001 ; 2014 확인)에 규정되어 있는 한글날자, 한글고어날자용 부호집합에 대하여 데이터 링크 계층에서 HDLC Flag의 비트 또는 문자 스테핑과 물리계층의 AMI/HDB-3 스크램블링 측면에서 데이터 전송효율을 높이는 새로운 한글날자용 부호집합 체계를 제시하였다. 기존 부호집합 체계와 비교를 위해 (4x4) 비트 원천부호화 규칙과 한글날자 사용빈도 통계를 적용한 결과, 약 22.01%의 데이터 처리효율이 향상되는 것으로 나타났다.

ABSTRACT

AMI / HDB-3 method used a scrambling technique is used primarily for long distance data transmission line encoding. In this paper, information communication code standard (KS X 1001; 2014 confirmation), as defined in Hangul Character Code HDLC Flag bit or character stuffing at the data link layer and physical layer with respect to the code set for Hangul AMI / HDB-3 the code set for the new system to increase the data transmission efficiency Hangul consonant and vowel tables presented in terms of scrambling. The result of the existing system and the code set (4x4) bit source coding rules for comparing the frequency of use Hangul consonant and vowel tables and statistics showed that about 22.01% of the data processing efficiency is improved.

키워드

Source Coding, Scrambling, HDB-3, HDLC, Character Code
통신규약, 스크램블링, 한글날자, 데이터통신, 한글코드

1. 서 론

데이터 문자정보 및 숫자 등의 정보를 부호화하여 컴퓨터, 통신에서 효율을 향상시키는 방법을 원천부호

화(source coding)라고한다. 대표적인 원천부호체계로는 ASCII, EBCDIC, Unicode등이 있다. 이 부호체계 중에서 ASCII와 EBCDIC 부호체계는 하나의 문자가 각각 7비트, 8비트로 부호화된다. 유니코드의 BMP에

* 대전대학교 IT경영공학과(jtwo@dju.ac.kr)

** 교신저자(corresponding author) : 한세대학교 정보통신공학과(wpohng@hansei.ac.kr)

접수일자 : 2014. 11. 05

심사(수정)일자 : 2014. 12. 15

게재확정일자 : 2015. 01. 12

서는 한 개의 문자나 기호를 32비트로 부호화하고 있다[7]. 네트워크 상에서 통신은 컴퓨터 등 정보기기에 서 생성된 문자 등의 원천부호를 전송에 적합하도록 변환하여 전송한다. 이것을 회선부호화(line coding)라고 한다. 회선부호화 방식에는 NRZ, RZ, Manchester 및 AMI방식 등이 있다[3]. 이들 회선부호화 방식 중에서 데이터통신의 장거리 전송에 적합한 AMI방식을 주로 사용하고 있다. 이 AMI방식은 다른 방식에 비하여 교류적인 신호특성 및 점유주파수대역폭 측면에서 유리하다.

그러나 이 방식은 원천부호에 일정한 개수이상의비트 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[3]. 연속되는 비트 0이 8개일 때 스크램블링하는 방식이 B8ZS(Binary 8 Zero Sequence)방식이다. 전자는 ITU-T 국제표준방식으로 우리나라에서, 후자는 미국 표준방식으로 미국에서 적용하고 있는 표준방식이다. 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. 이 표준에는 한글날자, 한글글자마디와 함께 로마문자에 대한 부호체계가 규정되어 있다. 본 논문은 참고문헌 [5]에서 제시하고 있는 원천부호화 규칙으로 적용하여 분석하였다. 이 분석결과를 토대로 하여 HDLC 프로토콜 및 AMI회선부호화 방식에 적합한 한글날자용 부호집합체계를 제시하였다. 또한 본 논문에서 제시하는 한글날자용 부호집합을 적용할 경우 데이터 통신의 전송효율에 주는 영향을 정량적으로 산출하였다.

II. 문자의 원천부호화 규칙

2.1 부호화연구 범위와 조건

본 논문에서는 정보기내에서 문자를 2진 부호화

하는 규칙을 논하는데 위에서 언급한 HDB-3 회선(line) 부호화 방식을 기준으로 원천 부호화하는 규칙으로 다음과 같은 범위와 조건을 적용하였다. 첫째, 1개의 조합단위를 구성하는 4비트 열 또는 단위조합 4,4비트열과 바이트 8,8비트열간 및 바이트의 조합 16,16비트열의 조합에 있어서 4개 이상의 연속된 "0" 비트열이 발생하는 조합을 찾아낸다. 둘째, 1바이트는 4비트 행 x 4비트 열=8 비트로 한다. 셋째, 단위조합간의 결합은 4,4행x4,4행=16비트 또는 4,4,4행 x4열=16비트로 한다. 넷째, 문자부호의 단위바이트 또는 바이트의 조합에 있어서 "0"이 연속하여 4개 이상 발생하는 조합을 찾아 제시한다.[7]

2.2 4 x 4비트 단위조합 규칙

표 1은 참고문헌[4]에서 제시한 HDB-3 스크램블링에 적용하는 4비트 x 4비트 단위조합에 대한 각 비트열 별 조합이 가능한 것과 제한되는 조합조건을 보여주는 것이다[4],[6].

이러한 조건으로 볼 때 16진수 0의 경우가 조합제한 개수가 8개로 가장 많고 16진수 7이 5개, 16진수 4와 C가 각각 4개, 16진수 2, 6, A, E가 각각 2개이다. 기타 1, 3, 5, 9, B, D, F는 16진수 0을 제외한 모든 비트열과 조합이 가능하다. 결과적으로 이러한 방식으로 조합할 경우 조합이 가능한 16x16=총256개에서 52개의 경우에는 조합이 되었을 때 회선부호화 과정에서 스크램블링이 발생되게 됨을 나타내고 있다[6]. 즉, 52개의 부호는 사용빈도가 낮은 문자부호나 디스플레이 장치로 출력되지 않는 비표현 문자부호에 부여한다.

표 1 원천 부호화 규칙 ; 4 x 4비트(HDB-3방식)
Table 1. Source coding rule ; 4 x 4-bits(HDB-3)

HEXA	high rank bits	Low Rank bits	
		Combination limited	Possible Combination
0	0000	0,1,2,3,4,5,6,7,8,9, A,B,C,D,E,F	X
1	0001	0	1,2,3,4,5,6,7,8,9,A,B, C,D,E,F
2	0010	0,1	2,3,4,5,6,7,8,9,A,B,C, D,E,F
3	0011	0	1,2,3,4,5,6,7,8,9,A,B, C,D,E,F

4	O100	O,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
5	O101	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
6	O110	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
7	O111	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
8	1000	O,1,2,3,4,5,6,7	8,9,A,B,C,D,E,F
9	1001	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
A	1010	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
B	1011	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
C	1100	O,1,2,3	4,5,6,7,8,9,A,B,C,D,E,F
D	1101	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
E	1110	O,1	2,3,4,5,6,7,8,9,A,B,C,D,E,F
F	1111	O	1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

표 2는 이러한 원칙을 적용한 4비트 x 4비트 단위 조합에 대한 각 비트열 별 조합이 가능한 것과 조합이 제한되는 조합조건을 보여 주는 것이다[4]. 상위비트열이 16진수 2, 4, 6, 8, A, C, E인 경우 하위 모든 비트열과 조합이 가능하다. 상위비트열이 16진수 1, 5, 9, D의 경우에는 하위 비트열 16진수 F를 제외한 모든 조합이 가능하다. 상위비트열이 16진수 3과 B의 경우에는 하위 비트열 16진수 E, F를 제외한 모든 조합이 가능하다. 상위 비트열이 16진수 7의 경우에는 하위 비트열 16진수 C, D, E, F외의 모든 조합이 가능하다. 상위비트열이 16진수 F인 경우에는 하위비트열 16진수 8, 9, A, B, C, D, E, F는 조합이 제한되고 O, 1, 2, 3, 4, 5, 6, 7은 조합이 가능하다.

표 2. 원천부호화 규칙 ; 4 x 4비트(HDLC 규격)
Table 2. Source coding rule ; 4 x 4-bit(HDLC)

HEXA	High bits Row	Low bits line	
		Limited Combination	Possible Combination
O	OOOO		O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

1	0001	F	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E
2	0010	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
3	0011	E,F	O,1,2,3,4,5,6,7,8,9,A,B,C,D
4	0100	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
5	0101	F	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E
6	0110	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
7	0111	C,D,E,F	O,1,2,3,4,5,6,7,8,9,A,B
8	1000	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
9	1001	F	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E
A	1010	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
B	1011	E,F	O,1,2,3,4,5,6,7,8,9,A,B,C,D
C	1100	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
D	1101	F	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E
E	1110	-	O,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
F	1111	8,9,A,B,C,D,E,F	O,1,2,3,4,5,6,7

* 하위비트열의 숫자는 16진수임

표 2와 같이 4비트 행 X 4비트 열의 조합으로 8비트열 바이트가 구성된다. 그 다음단계로 8비트열 각 바이트간의 조합이 고려되어야 한다. .

III. HDB-3, HDLC 문자의 원천부호화 규칙과 한글날자용 부호집합체계

3.1 한글날자용 부호집합 체계

표 3에 규정된 KS X 1001 한글날자용 부호집합이다. 이 표에서 보듯이 행은 0~15 까지 구성되어 있고 열은 A4XX 로 구성되어있다. 이 숫자들은 16진수이다. A4A0 과 A4FF는 비어있다. 이 부호체계는 유니코드 호환용 부호체계와 동일한 부호배열체계로 구성되어있다[2]. 부호체계 배열에서 같이 한글부호순으로 자모를 배열되어 있음을 알 수 있다. 특별히 한

표 3. KS X 1001 호환용 한글날자 부호와 원천부호화 규칙
Table 3. KS X 1001 hangul compatibility jamo and source coding rule

U+	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A4A		ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ
A4B	ㅇ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ
A4C	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ
A4D	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ
A4E	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ
A4F	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㆁ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ	ㄷ	ㄹ

 : Violating Code of Table 1 (HDB-3)
  : Violating Code of Table 2 (HDLC)

글자판(Keyboard)의 배열을 고려한 방식이 아님을 알 수 있다. 또한 이 부호체계는 데이터통신의 회선부호화를 염두에 두고 효율성을 고려한 부호화가 된 것도 아님을 알 수 있다. 물론 한글날자 자모의 사용빈도수를 고려한 배열 또한 아니며 단순 ㄱ-ㄴ순서와 자음, 모음, 고어받침순이라고 말할 수 있다.

개선효과를 얻을 수 있다. 빈도수가 집중된 자모들이 전체 88개 한글날자에서 차지하는 비율은 19.3%이지만 사용 빈도율은 전체의 80.9%에 달한다. 이 표 4에서 빈도율은 초성, 중성 및 종성의 각 성내에서의 상대비율이다. 총 빈도율은 총 날자 빈도수 21,036,022를 한 상대비율이다.

3.2 원천부호화 규칙 적용 한글날자용 부호집합

가. 개선효과 분석용 한글날자 사용빈도통계

표 3은 표 1의 원천부호화 규칙을 적용하여 보여 주는 것이다. 이 표 1의 원천부호화 규칙에서 상위비트열 16진수 A에는 16진수 0과 1을 제외한 모든 비트열이 조합이 가능하다. 그러므로 표 3에서 보듯이 최상위비트열 다음으로 16진수 4는 조합이 가능하다. 다음으로 16진수 3는 16진수 0, 1, 2, 3 조합이 제한된다. 표 3에서 격자무늬가 이에 해당된다. 다음으로 3번째 비트열 A는 0,1, B는 0, C는 0,1,2,3, D는 0, E는 0,1, F는 0이 조합이 제한된다.

이 표 2에서 빗살무늬 부호는 원천부호화 규칙의 HDLC를 나타내는 것으로 ㅎ, ㅏ, 를 비롯하여 한글날자에 해당되는 위배되는 부호열에 배치되어 있음을 알 수 있다[1]. 결과적으로 이 표 3에서 알 수 있듯이 원천부호화규칙에 위배되지 않는 부호는 총 88개 중 73개가 된다. 원천부호화규칙에 위배된 부호조합들은 근접된 빈도수가 매우낮은 부호와 배열을 교체하여

나. 개선효과 분석 방법

이러한 새로 제시한 한글날자용 부호집합체계가 기존방식과 비교한 전송효율측면에서 세 가지 점을 고려하였다. 이 분석은 정량적 결과를 얻을 수 있는 사용빈도에 대한 한글날자 부호에 대하여서만 이루어졌다. 첫째는 OSI의 7 Layer 중 데이터링크 계층에서 수행되는 HDLC프로토콜의 Flag비트열에 대한 stuffing 측면에서 분석하였다[7]. 두 번째는 OSI의 물리 계층에서 수행되는 회선부호화 과정의 HDB-3 스크램블링 측면에서 분석하였다[7]. 세 번째는 이 두 개의 분석결과를 종합하여 전체적인 개선효과를 추출하였다[7].

1) 제 1 방안

한글날자용 부호집합체계가 기존방식과 비교 전송효율측면에서 위배된 문자체계와 가장 인접하고 빈도율이 낮은 부호를 대체하는 방안으로 기존배열에서 최소변경된 방식으로 가장적합하다. 전송효율은 22.01% 증가이다.

표 4. 한글날자(자모) 사용빈도수, 빈도율 및 총 빈도율
Table 4. Hangeul jamo using rate, using frequency rate, total using frequency rate

consonant	hexa	Using Frequency	Using Frequency rate (%)	Total Using Frequency rate (%)	Vowel	hexa	Using Frequency	Using Frequency rate (%)	Total Using Frequency rate (%)
ㄴ	A4A4	1924983	39.46	9.15	ㅏ	A4BF	1926007	23.76	9.16
ㅇ	A4B7	2722792	38.95	12.94	ㅣ	A4D3	1355527	16.72	6.44
ㄹ	A4A9	1387102	26.03	6.59	ㅡ	A4D1	1151976	14.21	5.48
ㄱ	A4A1	1582576	23.34	7.53	ㅓ	A4C7	862073	10.63	4.1
ㅅ	A4B5	871039	11.35	4.14	ㅗ	A4CC	592419	7.31	2.82
ㅇ	A4B1	688203	10.85	3.29	ㅋ	A4C5	422490	5.21	2.01
ㄷ	A4A7	819320	9.49	3.9	ㅕ	A4C0	409768	5.05	1.95
ㅈ	A4B8	765682	8.86	3.64	ㅖ	A4C4	393695	4.86	1.87
ㅎ	A4BE	662247	7.82	3.15	ㅑ	A4C3	942819	2.22	0.85
ㄴ	A4B2	492758	7.35	2.35	ㅓ	A4D2	179818	2.22	0.85
ㅆ	A4B6	248370	5.73	1.18	ㅓ	A4C8	162656	2.01	0.77
ㅊ	A4BA	210210	2.51	1	ㅓ	A4CA	99794	1.23	0.47
ㅌ	A4BC	132020	1.82	0.63	ㅓ	A4CB	94576	1.17	0.45
ㅍ	A4BD	115924	1.54	0.55	ㅓ	A4C1	63728	0.79	0.3
ㅊ	A4A2	71065	0.88	0.34	ㅓ	A4CD	57373	0.71	0.27
ㅊ	A4A8	72604	0.82	0.35	ㅓ	A4CF	47506	0.59	0.23
ㄴㅎ	A4A6	30376	0.75	0.14	ㅓ	A4D0	47135	0.58	0.22
ㄴㅅ	A4B4	25868	0.64	0.12	ㅓ	A4C6	44394	0.55	0.21
ㅋ	A4BB	48399	0.55	0.23	ㅓ	A4C9	11215	0.14	0.05
ㅆ	A4B9	20434	0.23	0.1	ㅓ	A4C2	3769	0.05	0.02
ㄹㄱ	A4AA	7330	0.18	0.03	ㅓ	A4CE	2035	0.03	0.01
ㅃ	A4B3	16037	0.18	0.08					
ㄹㅇ	A4AB	4344	0.11	0.02					
ㄴㅈ	A4A5	2790	0.07	0.01					
ㄹㅎ	A4B0	3023	0.07	0.01					
ㄱㅅ	A4A3	357	0.01	0					
ㄹㄴ	A4AC	2183	0.01	0					
ㄹㅅ	A4AD	5	0	0					
ㄹㅌ	A4AE	131	0	0					
ㄹㅍ	A4AF	78	0	0					
계		12,928,250		61.47	계		8,870,773		38.53

Total Using Frequency : 21,036.022

2) 제 2 방안

한글날자용 부호집합체계가 기존방식과 비교 전송 효율측면에서 위배된 문자체계와 가장 빈도율이 낮은 부호만을 선별하여 대체하는 방안으로 기존배열표와 상당한 배열차이를 보여 소팅(sorting)시 별도 프로그램을 요구하는 문제점이 있다. 하지만 전송효율은

23.19%로 가장 좋다.

3) 제 2 방안

한글날자용 부호집합체계가 기존방식과 비교 전송 효율측면에서 위배된 문자체계를 블록단위로 빈도율이 낮은부호와 대체하는 방안으로 기존배열표에서 블

록단위로 변경된 방식이다. 전송효율이 3.77% 개선으로 다른안들에 비해 상대적으로 낮고 소팅할 때 문제가 발생할 수 있다.

다. 개선효과 분석 및 결과

KS X 1001 한글날자용 부호집합체계 개선 방안 3가지중 한글날자용 부호집합체계가 기존방식과 비교 전송효율측면에서 위배된 문자체계과 가장 인접하고 빈도율이 낮은 부호를 대체하여 전송효율을 높일수 있는 개선 1안을 제시한다.

표 5. 개선 방안별 비교표
Table 5. Efficiency improvement by comparison

Code	1st/rate	2nd/rate	3rd/rate
HDLC	11.13%	12.31%	12.31%
HDB-3	10.88%	10.88%	-8.54%
Total	22.01%	23.19%	3.77%

1) 제 1 방안 개선효과 분석

1.1) 데이터링크 계층 개선효과 분석
(HDLC의 Flag 비트열에 대한 stuffing)

Violated code in Source code	Characters
A4BE	ㅎ
A4BF	ㅏ

0 사용빈도율 (개선율) ;

$$\sum_{\Gamma}^{\Sigma} + \sum_{\Gamma}^{-1} (\text{사용빈도율}) \quad (1)$$

$$= (3.15 - 0.63) + (9.16 - 0.55)$$

$$= 11.13 (\%)$$

1.2) 물리계층 개선효과 분석
(HDB-3 원천부호화 규칙위배)

Violated code in Source code	Characters
A4A0~A4A1	ㄱ
A4B0	ㄴ
A4C0~A4C4	ㄷ, ㅌ, ㅊ, ㅋ, ㆁ
A4D0	ㅍ

0 사용빈도율 (개선율) ;

$$\sum_{\Gamma}^{\Sigma} + \sum_{\Gamma}^{-1} (\text{사용빈도율}) \quad (2)$$

$$= 7.53 + 0.01 + (1.95 + 0.30 + 0.02 + 0.85 + 0.22) = 10.88 (\%)$$

1.3) 전체 개선도

$$(1) + (2) = 11.13 \% + 10.88 \%$$

$$= 22.01 (\%)$$

2) 제 2 방안 개선효과 분석

2.1) 데이터링크 계층 개선효과 분석
(HDLC의 Flag 비트열에 대한 stuffing)

Violated code in Source code	Characters
A4BE	ㅎ
A4BF	ㅏ

0 사용빈도율 (개선율) ;

$$\sum_{\Gamma}^{\Sigma} + \sum_{\Gamma}^{-1} (\text{사용빈도율}) \quad (1)$$

$$= (3.15) + (9.16)$$

$$= 12.31 (\%)$$

2.2) 물리계층 개선효과 분석
(HDB-3 원천부호화 규칙위배)

Violated code in Source code	Characters
A4A0~A4A1	ㄱ
A4B0	ㄴ
A4C0~A4C4	ㄷ, ㅌ, ㅊ, ㅋ, ㆁ
A4D0	ㅍ

0 사용빈도율 (개선율) ;

$$\sum_{\Gamma}^{\Sigma} + \sum_{\Gamma}^{-1} (\text{사용빈도율}) \quad (2)$$

$$= 7.53 + 0.01 + (1.95 + 0.30 + 0.02 + 0.85 + 0.22) = 10.88 (\%)$$

2.3) 전체 개선도

$$(1) + (2) = 12.31 \% + 10.88 \% = 23.19 (\%)$$

3) 제 3 방안 개선효과 분석

3.1) 데이터링크 계층 개선효과 분석

(HDLC의 Flag 비트열에 대한 stuffing)

Violated code in Source code	Characters
A4BE	ㅎ
A4BF	ㅏ

0 사용빈도율 (개선율) ;

$$\sum_{\neg}^{\times} + \sum_{\vdash}^{-1} (\text{사용빈도율}) \tag{1}$$

$$= (3.15) + (9.16) = 12.31 (\%)$$

3.2) 물리계층 개선효과 분석

(HDB-3 원천부호화 규칙위배)

Violated code in Source code	Characters
A4A0~A4A1	ㄱ
A4B0	ㄹ
A4C0~A4C4	ㅅ, ㅆ, ㅈ, ㅊ, ㅋ
A4D0	ㅍ

0 사용빈도율 (개선율) ;

$$\sum_{\neg}^{\times} + \sum_{\vdash}^{-1} (\text{사용빈도율}) \tag{2}$$

$$= - 7.53 - 0.01 + (0.30 - 1.95 + 0.02 + 0.85 - 0.22) = -8.54 (\%)$$

3.3) 전체 개선도

$$(1) + (2) = 12.31 \% - 8.54 \% = 3.77 (\%)$$

V. 연구결과

본 논문에서는 대한민국 정보통신용 부호 표준 (KS X 1001; 2014 확인)에 규정되어 있는 한글날자,한글고어날자용 부호집합에 대하여 데이터 링크 계층의 HDLC Flag의 비트 또는 문자 stuffing, 물리계층의 AMI/HDB-3 스크램블링 측면에서의 새로운 한글날자용 부호집합 체계를 제시하였다.이를위해 (4x4) 비트열에 대한 문자의 원천부호화 규칙과 한글날자의 사용빈도에 대한 통계[8]를 사용하였다. 연구결과 새로운 한글날자, 한글고어날자용 부호집합체계를 적용할 경우 새로운 3가지 방안을 제시하였다. 그 중 전송효율이 가장 높은 2 방안의 경우 소팅(sorting)상의 불편함이 있으므로 인접부호별 대체로 전송효율과 소팅측면에서 우수한 개선 1방안을 제시한다. 이 방안은 기존 부호 집합 체계 대비 22.01%의 개선효과가 있는 것으로 나타났다. 따라서 본 연구결과에서 제시한 한글날자, 한글고어날자용 부호집합 체계를 사용할 경우, 한글날자의 전송에 있어서는 데이터링크 계층에서의 HDLC 프로토콜에서의 Flag로 인한 문자부호열에 대한 비트 또는 문자 스티핑과 물리계층에서의 HDB-3 스크램블링이 완전히 발생되지 않는 것으로 나타나 이 문자부호체계를 적용할 경우에 문자의 전송효율개선에 큰 영향을 줄 것으로 전망된다. 향후 추가 연구사항으로는 AMI방식의 회선부호화과정에서 직류성분을 완전히 제거할 수 있는 원천부호체계를 제시하여 국제표준으로써 사용이 가능하다. 또한 본 논문에서 제시한 새로운 한글날자 부호집합체계를 UTF-8 등의 부호체계로 변환할 때 나타내어지는 효율적 변화가 있는가를 연구할 가치가 있다.

References

[1] W. Hong, "Composition Rule of Character Codes to Efficiently Transmit the Character Code in HDLC(High-level Data Link Control) Protocol," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 7, no. 4, 2012, pp. 753-759.

[2] W. Hong, "Compatibility of UTF-8 Encoding System to HDB-3 Scrambling Method," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 8, no. 2, 2013, pp. 227-238.

[3] B. A. Forouzan, *Data Communications and Netwo-*

orking, Fourth Edition. New York : McGraw Hill, May. 2007.

- [4] W. Hong, "Coding Rule of Characters by 2 Bytes with 4x4 Bits to Improve the Transmission Efficiency in Data Communications," *The J. of Korea Navigation Institute*, vol. 15, no. 5, 2011. pp. 745-751.
- [5] C. Lee, "Improvement of the Linear Predictive Coding with Windowed Autocorrelation," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 6, no. 2, 2011, pp. 186-190.
- [6] W. Hong, "Study on the coding rule of characters by 1 byte of 3x4 bits with consideration for the transmission efficiency of data communications," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 6, no. 4, 2011, pp. 499-503.
- [7] W. Hong, "Consideration of Roman Character in KS X 1001 Code System for Information Interchange considered AMI/HDB-3 and HDLC FLAG," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 8, no. 7, 2013, pp. 1017-1024.
- [8] H. Kim, "The research No II frequently used modern korean languages," *National Institute of Korean, Ministry of Culture and Tourism*, vol. 1 no. 2, 2005, pp. 10-13.

저자 소개



우제택(Je-Teak Woo)

1993년 중앙대학교 기계공학과 졸업(공학사)
 2001년 서강대학교 정보통신대학원 정보통신공학과 졸업(공학석사)
 2007년 서강대학교 대학원 전자공학과 수료(공학박사)
 2001년 정보통신기술사합격 65회
 1993년 삼성SDS(주) 그룹공채34기입사
 테크윈IS실, IT컨설팅실, 컨설팅본부
 2012년 대전대학교 공과대학 IT경영공학과 전임교수
 2013년 대전대학교 공과대학 IT경영공학과 조교수
 2014년 한국엔지니어링협회 정보화위원
 ※ 관심분야 : 데이터통신,빅데이터 분석, 문자코딩



홍완표(Wan-Pyo Hong)

1991년 서울과학기술대학교 전자공학과(공학사)
 1994년 연세대학교 공학대학원 전자공학전공(공학석사)
 1999년 광운대학교 대학원 전자공학과(공학박사)
 1990년 전기통신기술사합격
 1991년 정보통신부 5급특별채용고시합격 본부 통신정책실, 전파방송관리국, 정보화기획실
 1997년 삼성전자(주) 통신사업부 전송영업그룹장
 1999년 광운대학교 연구진담교수
 2000년 한국정보통신기술사협회장
 2002년 한세대학교 정보통신공학과 교수
 2014년 USC 동북아언어문화학과 방문학자
 ※ 관심분야 : 위성통신방송, 문자코딩, 통신정책