

회선부호의 스크램블링을 고려한 새로운 한국표준 한글글자마디부호에 관한 연구

박요셉* · 홍완표**

Considering the scrambling code of the line Study on the New Korea joint protection Standard Hangul character

Yo-Seph Park* · Wan-Pyo Hong**

요 약

본 논문은 정보통신용 표준 부호인(정보 교환용 부호계 KS X 1001 : 2004)에 정의되어 있는 한글낱자 부호집합에 대하여, Date link 계층에서 AMI/HDB-3 스크램블링 측면에서 효율적인 데이터 전송을 위한 새로운 부호집합 체계를 제시하였다. 기존 부호집합 체계와 상호비교를 위하여 (4x4) 비트 원천부호화 규칙과 한글 빈도통계(국립국어원)를 적용한 결과 약44 %의 데이터 전송 효율이 개선시킬 수 있음을 나타냈다.

ABSTRACT

This paper, information communication code standard(KS X 1001, confirmation in 2004), as defined in Hangul Character Code Hangul AMI/HDB-3 the code set for the new system Hangul consonant and vowel tables presented. 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(The National Institute of The Korea Language) and statistics showed that 44% of the data processing efficiency is improved.

키워드

Source Coding, AMI, HDB-3, Character Code
한글 낱자, 한글 코드, 스크램블링, 회선 부호화

1. 서 론

정보통신장비에서 처리되어 통신망으로 전송되는 데이터는 2단계의 부호화 처리과정을 거친다. 1번째 부호화 과정은 OSI-7 Layer의 6계층인 Presentation 계층에서 처리된다[1]. 이 과정을 원천부호화 과정이라 한다. ASCII [1-3], EBCDIC [4], UNICODE [5] 및 KS X 1001등이 위의 처리과정에 해당된다. 2번째

부호화 과정은 1계층인 Physical 계층에서 처리되는 데, 이과정이 회선부호화 이다. 회선부호화 방식은 RZ, NR 및 AMI등이 있다[2-3]. 특별히 AMI방식은 장거리 전송에 뛰어난 방식이다[6]. 회선부호화방식으로 AMI방식을 적용 할 경우 0 비트 연속발생으로

1) <http://en.wikipedia.org/wiki/EBCDIC>

<http://en.wikipedia.org/wiki/Unicode>

* 한세대학교 IT융합학과(squirt01@naver.com) · Received : Oct 21, 2015, Revised : Dec 13, 2015, Accepted : Dec 24, 2015
** 교신저자 : 한세대학교 정보통신공학과 · Corresponding Author : Wan-Pyo Hong
Dept. of Information and Communication Technology, Hansei University
Email : wphong@hansei.ac.kr
· 접수 일 : 2015. 10. 21
· 수정완료일 : 2015. 12. 13
· 게재확정일 : 2015. 12. 24

인한 동기 상실을 방지하기 위하여 스크램블링 기술 방식을 적용한다. 스크램블링의 대표적인 방식은 HDB-3방식[6] 및 B8ZS방식[7]이 있다. 본 논문에서는 연속 4개 이상의 0비트 전송을 방지하기 위하여 HDB-3방식을 사용하였다. 따라서 원천부호화 처리과정중 연속 4개이상의 0비트의 전송을 방지하기 위하여 HDB-3방식을 사용하였다[4-5]. 그러므로 원천부호화 과정중 0비트 연속4개 이상 존재하는 비트조합이 다수 발생 할 경우 회선부호화 과정에서 스크램블링이 많이 발생되어, 데이터의 전송효율에 영향을 주게 된다. 또한 OSI-7 Layer의 2계층인 Data Link계층에서 HDLC 프로토콜을 사용할 때, 프레임내의 문자 비트열중에 FLAG 구성 비트열과 유사한 비트열이 존재 할 경우, FLAG비트열로 오인되는 것을 사전에 방지하기 위하여 위해 비트 또는 문자 스텀핑을 한다. 즉 문자 비트열에 이러한 비트열이 증가하면 통신규약 프레임체계를 변경시키게 되며, 데이터의 전송율도 저하 된다[6].

본 논문에서는 이러한 것을 토대로 하여 KS X 1001 한글부호체계에 대한 원천부호화 위배율이 한글 사용 빈도율을 기준으로 하여 분석하였다. 현재 KS X 1001의 한글부호체계에는 총 2350개의 한글부호가 있으며, 이 한글부호체계에 대하여 분석하였다. 본 논문은 참고문헌[8] 에서 제시하고 있는 원천부호화 규칙을 적용하여 분석하였다. 이를 통하여 원천부호화 과정에서 얻어지는 개선효과를 정량적으로 산출 하였다.

II. 문자의 원천부호화 규칙 및 한글 유니코드체계

2.1 문자의 원천부호화 규칙

표 1에서 제시하고 있는 문자의 원천부호화 규칙이다[6-10]. 이 규칙은 OSI 2계층인 Data Link 계층에 적용되는 HDLC 프로토콜과 1계층인 Physical 계층에서 수행되는 AMI 회선부호화방식과 HDB-3 스크램블링 기술을 적용한 것이다.

표 1. 문자 부호화 규칙 : 4 비트 X 4비트 [8]

Table 1. Characters coding rule : 4 bit x 4 bit

HEXA	high bit line	low bit line	
		no mix	mix
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	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	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	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	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	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	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	1,2,3,4,5,6,7

2.2 문자의 원천부호화 규칙 위배 문자부호

표 2는 KS X 1001의 한글부호체계를 표 1의 원천

부호화 규칙에 적용하여, 원천부호화 규칙간의 관계를 나타내는 것이다. 표2에서 는 원천부호화 규칙에 1회 위배되는 문자부호이다. 표2에서 는 원천부호화 규칙에 2회 위배되는 문자부호이다. 즉, 표2에서 보듯이 원천부호화 규칙에 위배되는 문자부호의 수는 546개이다. 이와같이 원천부호화 규칙에 1회 위배되는 문자는 505개, 2회 위배되는 문자는 41개, 총 546개다.

표 2. 원천부호화 규칙

Table 2. Source coding rule

code	char acter	code	char acter	code	chara cter	code	char acter
B0A1	가	B1A3	갓	C0E2	잡	BFD1	얌
B0A2	각	B1A7	팠	C0E3	갓	BFD9	얌
B0A3	간	B1A8	팠	C0E4	갓	BFDA	얌
B0A4	갈	B1A9	팠	C0E5	장	BFDD	익
B0A5	갈	B1B1	팠	C0E6	갓	BFE2	잇
B0A6	갈	B1B6	팠	C0E7	재	BFE3	윙
B0A8	감	B1BF	쿨	C0E8	재	BFF1	윙
B0A9	갈	B1B6	극	C0E9	젠	BFFB	윙
B0AA	갈	B1BA	쿨	C0EB	잼	BFFD	윙
B0AB	갓	B1BC	켓	C0EC	잼	C4A4	친
B0AC	갓	B1BE	극	C0ED	갓	C4BE	캉
B0AD	강	B1E3	극	C0EF	쟁	C4D4	킴
B0AE	갓	B1E7	깁	C0F0	자	C4D5	킴
B0AF	갓	B1EC	깁	C0F7	재	C4D8	경
B0B0	갈	B1F3	깁	C0FA	저	C4E4	관
B0B1	갈	B1FB	깁	C0FB	적	C4E6	관
B0B2	강	B2A6	깁	C0FC	진	C4F2	콧
B0B3	개	B2A7	깁	C0FD	절	C4F3	콧
B0B4	객	B2B3	깁	C0FE	젊	C4F8	퀵
B0B5	객	B2B5	깁	C1A1	집	C5A2	퀵
B0B6	객	B2B6	깁	C1A2	집	C5A4	퀵
B0BB	갱	B2BA	깁	C1A3	젓	C5A6	칸
B0BC	가	B2BB	갓	C1A4	정	C5A8	콧
B0BF	갈	B2CC	팠	C1A5	젓	C5CC	탕
B0C2	개	B2CD	팠	C1A6	제	C5D1	땀
B0C3	객	B2D3	편	C1A7	적	C5DF	탱
B0C4	객	B2D4	편	C1A8	젠	C5E2	테
B0C5	거	B2D6	편	C1A9	젤	C5EC	툰
B0C6	격	B2D7	편	C1AC	젓	C5ED	타
B0C7	건	B2D8	표	C1AE	저	C5EE	판
B0C8	걸	B2E4	편	C1AF	전	C5F1	땀
B0C9	걸	B2E8	편	C1B3	젓	C5F3	땀

code	char acter	code	char acter	code	chara cter	code	char acter
B0CB	검	B2EB	팠	C1B6	조	C5FB	툰
B0CC	겁	B2EC	팠	C1B7	족	C6A6	툰
B0CD	것	B2ED	팠	C1B8	존	C6AA	툰
B0CE	것	B2F3	꾸	C1B9	졸	C6B3	툰
B0CF	경	B2F9	꿈	C1BB	좁	C6B8	툰
B0D1	걸	B3B1	낚	C1BC	좁	C6B9	툰
B0D2	걸	B3CF	낚	C1AE	중	C6BA	툰
B0D3	경	B3DB	넌	C1AF	좁	C6BB	툰
B0D4	게	B3E5	넌	C1C0	좁	C6C9	괘
B0D5	겐	B3EA	넌	C1C1	중	C6D5	팠
B0D6	객	B3EF	늑	C1C2	좌	C6D9	파
B0DA	갓	B3FB	넌	C1CB	죄	C6F4	팠
B0DC	겨	B3FD	넌	C1CC	권	C6F5	팠
B0DD	격	B3FE	넌	C1D2	조	C6F6	팠
B0DE	귀	B4A4	논	C1D6	주	C6FC	팠
B0DF	견	B4A6	늑	C1D7	죽	C7A2	팠
B0E1	걸	B4A7	늑	C1D8	준	C7A3	피
B0E2	검	B4B4	넌	C1D9	줄	C7A4	편
B0E3	겁	B4B9	넌	C1DC	좁	C7A7	팠
B0E5	갓	B4BB	늑	C1DD	좁	C7AD	팠
B0E6	경	B4BE	늑	C1DE	좁	C7B1	팠
B0E7	걸	B4C5	늑	C1DF	중	C7B5	팠
B0E8	게	B4C8	늑	C1E0	취	C7B7	편
B0ED	고	B4CD	넌	C1E1	취	C7B8	편
B0EE	곡	B4E6	닷	C1E3	취	C7B9	편
B0EF	근	B4F7	땀	C1E5	권	C7BA	팠
B0F0	곧	B4F9	땀	C1E6	퀵	C7BC	팠
B0F1	곧	B5AC	땀	C1EA	쥬	C7BE	팠
B0F4	곧	B5BA	늑	C1EE	즈	C7BF	팠
B0F5	곧	B5C4	달	C1EF	죽	C7C6	팠
B0F6	곧	B5D9	탕	C1F0	준	C7E7	협
B0F7	곳	B5DB	땀	C1F1	줄	C7FE	헛
B0F8	공	B5DC	땀	C1F2	좁	C8A7	훈
B0F9	곳	B5DD	땀	C1F3	좁	C8B5	헛
BOFA	과	B5E4	땀	C1F5	중	C8CF	힘
BOFB	곽	B6A8	땀	C1F6	지	C8D2	헛
BOFC	관	B6AC	땀	C1F7	직	C8DA	힘
BOFD	괄	B6B4	땀	C1F8	진	C8DF	훈
B1C0	굽	B6B6	땀	C1FA	질	C8E3	훈
B1C1	굽	B6C5	땀	C1FB	깁	C8E8	훈
B1C2	곳	B6C6	땀	C1FC	깁	C8EE	훈
B1C3	궁	B6CD	땀	C1FD	깁	C8F4	힘
B1E0	궁	B6D3	땀	C1FE	깃	C8F6	헛

code	char acter	code	char acter	code	chara cter	code	char acter
B1F0	깎	B6D6	뚝	C2A1	징	B1A2	깎
B2B0	꺾	B6D8	뻬	C2A2	깃	B1AC	꺾
B2C0	꼭	B6DE	뵁	C2A3	길	B1AD	퀸
B2C4	꼴	B6EF	뵁	C2A4	깊	B1AE	퀸
B2D0	꽤	B6F1	뵁	C2A5	짜	B1AF	퀸
B2E0	꿏	B6FC	랏	C2A6	작	B1B4	꾼
B2E1	꿏	B7AC	란	C2A7	잔	B1B5	꾼
B2F0	꿏	B7CC	렘	C2A8	짱	B1B7	꿏
B3A0	꺾	B7D6	와	C2A9	짱	B1C9	꺾
B3B0	낱	B7D7	환	C2AA	짱	B1D2	꺾
B3C0	넙	B7D8	황	C2AB	짱	B1DA	꺾
B3C1	넙	B7D9	렘	C2AC	짱	B1E5	꺾
B3C2	넙	B7DB	린	C2AF	짱	B1F7	꺾
B3C3	넙	B7DC	릴	C2B0	째	B1FD	꺾
B3D0	넙	B7DD	림	C2B7	째	B1FE	꺾
B3E0	넙	B7DF	릿	C2B8	쟁	B2A2	꺾
B3E1	넙	B7E3	롤	C2B9	짜	B2B4	꺾
B3F0	넙	B7E4	롭	C2BC	찌	B2B7	꺾
B4B0	넙	B7E5	롯	C2BD	찍	B2B9	꺾
B4C0	느	B7F1	웨	C2BE	전	B2BD	꺾
B4C1	느	B7F5	윌	C2BF	쩍	B2BE	꺾
B4C2	느	B7F6	윌	C2C0	쩍	B2D5	꺾
B4C3	느	B7F7	윌	C2C1	쩍	B2E9	꺾
B4D0	넙	B7F8	윌	C2C4	쟁	B2EA	꺾
B4E0	넙	B7FE	윌	C2C7	찌	B2F2	꺾
B5A0	넙	B8AB	윌	C2C9	쪼	B2F5	꺾
B5A1	넙	B8AC	윌	C2CA	쪼	B2FC	꺾
B5C0	넙	B8BD	윌	C2CC	쪼	B3AE	넙
B5D0	넙	B8CF	야	C2CD	쪼	B3C5	넙
B5E0	넙	B8D2	망	C2D0	쪼	B3C8	넙
B5F0	넙	B8D7	뵁	C2D1	쪼	B3D2	넙
B6A1	넙	B8EF	메	C2D3	쪼	B3DC	넙
B6B0	넙	B8F5	물	C2D8	쪼	B3DF	넙
B6C3	넙	B8FA	과	C2DE	쪼	B3E3	넙
B6E0	넙	B8FB	관	C2DF	쪼	B3F1	넙
B6E1	넙	B8FC	랏	C2E1	쪼	B3F7	넙
B7A1	래	B8FD	망	C2E5	쪼	B3F8	넙
B7B0	래	B9A2	윌	C2EA	쪼	B3FC	넙
B7C1	래	B9A3	윌	C2EB	쪼	B4A3	넙
B7C2	래	B9A4	윌	C2EC	쪼	B4A5	넙
B7C3	래	B9A5	윌	C2ED	쪼	B4A8	넙
B7D0	론	B9A7	문	C2EE	찌	B4AC	넙
B7E1	료	B9A8	물	C2EF	찍	B4B1	넙

code	char acter	code	char acter	code	chara cter	code	char acter
B7F0	뤌	B9A9	몹	C2F0	찢	B4B7	넙
B8B0	뤌	B9AA	몹	C2F1	찢	B4B8	넙
B8C0	뤌	B9B3	몹	C2F2	찢	B4BC	넙
B8C1	뤌	B9BE	몹	C2F3	찢	B4C7	넙
B8C2	뤌	B9C6	몹	C2F4	찢	B4CE	넙
B8C3	뤌	B9D1	몹	C2F5	찢	B4D3	넙
B8E0	뤌	B9DC	몹	C2F6	찢	B4D8	넙
B8E1	뤌	B9F2	바	C2F7	차	B4E8	넙
B8F0	뤌	B9F3	박	C2F8	차	B4E9	넙
B9B0	물	B9F4	반	C2F9	찬	B4F4	다
B9C0	물	B9F5	밥	C2FA	찬	B4FB	넙
B9C2	물	B9FB	벌	C2FB	찰	B5AA	넙
B9D0	물	BAA8	뵁	C2FC	참	B5AB	넙
B9E0	물	BAA8	뵁	C2FD	참	B5AF	넙
B9F0	물	BAB6	베	C2FE	차	B5B2	넙
BAA1	물	BAB7	벤	C3A1	차	B5B3	넙
BAB0	물	BAC7	빅	C3A2	차	B5B4	넙
BAC0	물	BACD	본	C3A3	차	B5BB	넙
BAC1	물	BAD4	봄	C3A4	차	B5CC	넙
BAC3	물	BADA	봄	C3A5	차	B5CD	넙
BAD0	물	BADB	뷔	C3A6	차	B5D8	넙
BBA1	물	BADC	빌	C3A7	차	B5E2	넙
BBB0	물	BADD	뤌	C3A8	차	B5E3	넙
BBC0	물	BADE	웨	C3AA	차	B5EA	넙
BBC1	물	BAE2	윌	C3AB	차	B5EF	넙
BBC2	물	BAE3	윌	C3AC	차	B5F6	넙
BBF0	물	BAE8	윌	C3B3	차	B6A2	넙
BCC0	물	BAE9	윌	C3B4	차	B6A3	넙
BCC1	물	BAEE	윌	C3B5	차	B6A6	넙
BCC2	물	BAEF	윌	C3B6	차	B6AD	넙
BCD0	물	BAF5	윌	C3B7	차	B6B5	넙
BCF0	물	BBA2	뵁	C3B8	차	B6B7	넙
BDA1	물	BBA4	뵁	C3B9	차	B6BA	넙
BDB0	물	BBAE	뵁	C3BC	차	B6BD	넙
BDC0	물	BBB2	뵁	C3BD	차	B6C4	넙
BDC1	물	BBB3	뵁	C3BE	차	B6C9	넙
BDC2	물	BBBA	뵁	C3BF	차	B6CC	넙
BDC3	물	BBC6	뵁	C3C2	차	B6CE	넙
BDE1	물	BBCE	뵁	C3C4	차	B6DD	넙
BDF0	물	BBD6	뵁	C3C6	차	B6E5	넙
BEC0	물	BBD9	뵁	C3C7	차	B6E9	넙
BEC1	물	BBE3	뵁	C3CA	차	B6EB	넙
BEC2	물	BBE4	뵁	C3CB	차	B6F2	넙

code	char acter	code	char acter	code	chara cter	code	char acter
BEC3	씹	BBEB	산	C3CC	촌	B6FD	랏
BED0	압	BCA9	샌	C3CD	출	B7AD	랏
BEE0	약	BCAA	셀	C3CE	춤	B7BE	랩
BEE1	안	BCAB	샘	C3D0	춧	B7CB	렌
BEF0	언	BCAC	생	C3D1	충	B7CD	렛
BFA1	에	BCC6	석	C3D4	찰	B7DE	웁
BFB0	염	BCCA	섭	C3D6	취	B7E2	론
BFC0	오	BCCB	섯	C3DD	초	B7F3	뢰
BFC1	옥	BCD1	성	C3DF	추	B7F4	뢴
BFC2	온	BCD7	숨	C3E0	축	B8C8	멜
BFC3	올	BCDF	찬	C3E1	춘	B8D1	말
BFD0	알	BCE3	센	C3E2	출	B8D9	몹
COA2	웬	BCE4	셀	C3E3	춤	B8E3	몹
COA3	웬	BCE5	셈	C3E4	춤	B8E5	몹
COA5	웬	BCE6	셋	C3E5	춧	B8EB	몹
COA7	위	BCE7	셋	C3E6	춤	B9B1	몹
COA9	윈	BCEB	침	C3E7	취	B9B4	몹
COAA	윌	BCEC	침	C3E8	취	B9B7	몹
COAD	윌	BCF2	숨	C3E9	취	B9B8	몹
COAE	윌	BDA5	셋	C3EB	취	B9C4	몹
COAF	유	BDA7	셋	C3F2	취	B9C5	몹
COB0	윳	BDAA	셈	C3F7	츠	B9C9	몹
COB1	윤	BDAB	셈	C3F8	츠	B9CA	몹
COB2	울	BDB5	숙	C3FB	춤	B9CB	몹
COB6	움	BDDD	쌌	C3FE	춤	B9ED	몹
COB8	으	BDE5	쌌	C4A1	치	B9EF	몹
COB9	옥	BDEC	셀	C4B0	갑	B9F9	몹
COBA	은	BDED	센	C4C0	격	BAA6	몹
COBB	을	BDF3	숨	C4C1	컨	BAA9	몹
COBC	을	BDF4	숨	C4C2	컨	BAC4	몹
COBD	음	BDF5	숨	C4C3	컨	BAC5	몹
COBE	읍	BDF9	싼	C4E0	콧	BAC8	몹
COC0	응	BDFA	쌌	C4E1	콩	BACA	몹
COC7	의	BDFC	쌌	C4F0	쿨	BACC	몹
COC8	윈	BEA2	셈	C5B0	키	BAD2	몹
COC9	이	BEA3	셈	C5C0	닷	BAD6	몹
COCD	익	BEA4	쏘	C5C1	당	BAE5	몹
COCE	인	BEA8	쓸	C5C2	대	BAE7	몹
COCF	일	BEA9	숨	C5C3	택	BAEB	몹
COD0	임	BEAE	췌	C5D0	털	BAF7	몹
COD3	임	BEAF	췌	C6A1	퇴	BBA3	몹
COD4	입	BEB1	쌌	C6B0	튼	BBA8	몹
COD5	잇	BEB6	쌌	C6C0	팀	BBB1	몹

code	char acter	code	char acter	code	chara cter	code	char acter
C0D6	있	BEB7	쌌	C6C3	팅	BBBE	몹
C0D7	잉	BEBB	션	C6D0	패	BBBF	몹
C0D8	잇	BEBC	셀	C7B0	폼	BBC5	몹
C0D9	잎	BEBD	쌌	C7C1	프	BBC8	몹
C0DA	자	BEEC	앨	C7C2	픈	BBC9	몹
C0DB	작	BEED	얹	C7C3	플	BBD7	몹
C0DC	잔	BEF5	얹	C7D0	학	BBD8	몹
C0DD	잖	BFAE	얹	C7E0	행	BBDE	몹
C0DF	잘	BFBD	얹	C8B0	활	BBE5	몹
C0E1	잡	BFBF	얹	C8F0	흘	BBE6	몹

2.3 한글 부호의 사용 빈도율

표 3은 KS X 1001의 한글부호체계의 일부 빈도율이다.

표 3. 사용빈도율
Table 3. Frequency rate

char acter	Freque ncy rate	char acter	Freque ncy rate	char acter	Freque ncy rate	char acter	Freque ncy rate
가	1.7013	뻬	0.0001	셀	0.0002	췌	0.0003
각	0.2037	뻬	0.0349	셋	0	췌	0
간	0.2473	뻬	0.0068	소	0.3281	췌	0
간	0.0014	뻬	0.0000	속	0.1687	췌	0
갈	0.0530	뻬	0.0003	슈	0.0000	췌	0.0000
감	0.0002	뻬	0.0003	손	0.0657	췌	0.0004
값	0	뻬	0	솔	0.0090	췌	0.0135
감	0.1291	뻬	0	숨	0	췌	0.0010
갑	0.0230	뻬	0.0001	숨	0.0029	췌	0.0000
값	0.0095	뻬	0.0002	습	0.0002	췌	0.0000
갓	0.0023	뻬	0.0001	숫	0.0049	췌	0.0127
갓	0.0364	뻬	0	송	0.0414	췌	0.0134
강	0.1169	뻬	0	솔	0.0010	췌	0.0006
갓	0.0395	뻬	0.1096	쇠	0.0000	췌	0.0028
갓	0.0004	뻬	0.0104	속	0.0000	췌	0.0010
갈	0.1660	뻬	0.0000	싼	0	췌	0.0003
값	0.0017	뻬	0.0015	살	0	췌	0.0000
강	0.0008	뻬	0.0043	상	0	췌	0.0009
개	0.2265	뻬	0.0001	쇄	0.0055	췌	0.0021
객	0.0220	뻬	0	셀	0	췌	0.0005

character	Frequency rate	character	Frequency rate	character	Frequency rate	character	Frequency rate
갠	0.0003	팔	0.0000	쌔	0	차	0.1604
겔	0.0007	꽤	0.0002	셋	0	착	0.0267
갸	0.0001	피	0	쌌	0	찬	0.0288
갹	0.0001	핀	0.0055	쇠	0.0077	참	0.0085
갻	0.0326	빨	0.0028	즉	0.0209	홀	0.0002
되	0.3418	뽀	0.0001	즌	0.0025	화	0.3277
된	0.1850	뽀	0.0035	즐	0.0124	확	0.0574
될	0.0554	빠	0.0001	즘	0.0161	환	0.0608
뒸	0.0041	뻑	0.0000	줍	0.0013	활	0.0805
뒹	0.0064	뻐	0.0005	증	0.0644	핫	0.0002
뒻	0.0001	뻐	0	지	1.5071	황	0.0491
도	0.0000	뽀	0.0001	직	0.1318	쾌	0.0002
두	0.1889	뻐	0.0001	진	0.2448	쾌	0.0002
득	0.0051	뽀	0.9497	진	0.0001	퀵	0
둔	0.0113	사	0.0043	질	0.1089	햇	0.0005
둘	0.0282	삭	0.0002	줍	0.0006	행	0.0000
둠	0.0054	삿	0.1837	짐	0.0170	회	0.2484
둑	0.0012	산	0	집	0.1269	획	0.0251
뒸	0.0004	살	0.1045	짓	0.0211	흰	0.0000
등	0.0130	살	0.0000	징	0.0211	흰	0.0000
뒤	0.0019	삿	0.0205	짓	0.0024	휼	0
뒸	0.0007	쌌	0.0598	질	0.0026	휼	0.0000
웨	0.0001	삼	0.0016	짚	0.0028	휼	0.0007
헝	0	삿	0.0011	짜	0.0214	횡	0.0031
뒤	0.0515	삿	0.0014	짜	0.0153	효	0.0325
된	0	삿	0.4435	잔	0.0007	흔	0
될	0	상	0.0004	짤	0.0002	홀	0
뒸	0	살	0.0887	짤	0.0013	흡	0
뒸	0.0099	새	0.0471	짤	0.0059	훗	0
뒸	0.0007	색	0.0013	짤	0.0004	후	0.1023
두	0.0009	센	0.0006	짤	0.0004	후	0.0002
둔	0	셀	0.0021	짤	0.0000	훈	0.0161
둘	0.0000	샘	0.0000	짤	0.0001	홀	0.0052
둑	0.0000	쌔	0.0006	짤	0.0026	홀	0.0011
둑	0	셋	0.0000	째	0	흠	0.0025
드	0.1360	쌔	0.3423	째	0.0315	훗	0.0007
득	0.0293	생	0.0037	쩍	0.0001	흥	0.0001
든	0.1066	샤	0.0001	쩍	0.0000	휘	0.0001
든	0.0168	삭	0.0001	쩍	0	흰	0.0008
들	0.7826	산	0.0002	쩍	0.0000	흰	0.0074
듬	0.0001	살	0.0003	쩍	0	휼	0
듬	0.0090	삼	0.0002	쩍	0.0001	횡	0.0000

character	Frequency rate	character	Frequency rate	character	Frequency rate	character	Frequency rate
듬	0.0045	삿	0.0003	찢	0.0040	휘	0.0019
뒸	0.0601	삿	0.0001	쟁	0.0004	획	0
등	0.1963	새	0	짜	0.0003	흰	0.0000
디	0.0000	센	0	잔	0	흰	0.0000
디	0.0747	셀	0	짤	0	횡	0.0003
딕	0.0003	쌔	0	째	0.0000	휘	0.0149

표 3은 원천부호화 규칙에 위배되는 한글 부호에 대한 위배율이다. 1번 위배되는 한글부호와 2번 위배되는 한글 부호로 구분하였다.

표 4. 원천부호화 규칙 위배 문자의 사용 빈도율
Table 4. Frequency rate of use of violated characters with Source coding rule

character	Frequency rate	character	Frequency rate	character	Frequency rate	character	Frequency rate
원천부호화 규칙 1번 위배							
각	0.2037	땀	0.0042	작	0.1729	쭈	0.0028
간	0.2473	떠	0.0390	잔	0.0259	쭈	0.0003
간	0.0014	땀	0.0002	잖	0.0198	쯔	0.0004
갈	0.0530	땀	0.0040	잘	0.0676	쯔	0.0135
감	0.0002	땀	0.0021	잡	0.0543	쯔	0.0010
감	0.1291	래	0.1632	жат	0.0025	쯔	0.0000
갑	0.0230	력	0.0044	жат	0.0011	찌	0.0127
값	0.0095	려	0.2485	장	0.4271	찍	0.0134
жат	0.0023	력	0.1228	жат	0.0014	쩍	0.0028
갔	0.0364	런	0.0567	재	0.1720	쩍	0.0010
강	0.1169	론	0.0866	재	0.0007	쩍	0.0003
गत	0.0395	료	0.0520	젠	0.0003	쩍	0.0009
गत	0.0004	뤼	0.0004	젼	0.0002	쩍	0.0021
값	0.0017	린	0.0886	젼	0.0005	쩍	0.0005
값	0.0008	맛	0.0126	젼	0.0005	차	0.1604
개	0.2265	망	0.0566	쟁	0.0379	착	0.0267
객	0.0220	맛	0.0482	재	0.0006	찬	0.0288
갠	0.0003	말	0.0129	저	0.1624	참	0.0085
갸	0.0007	멘	0.0039	적	0.5126	찰	0.0322
갱	0.0030	멜	0.0015	친	0.4459	참	0.0624
가	0.0014	모	0.2792	절	0.0776	참	0.0005

char acte r	Freque ncy rate	char acte r	Freque ncy rate	char acter	Freque ncy rate	char acter	Freque ncy rate
갈	0.0008	물	0.2240	젊	0.0145	찾	0.0009
겔	0.0000	뮌	0.0005	접	0.0389	창	0.0600
거	0.3526	뮤	0.0048	젓	0.0027	찾	0.0415
걱	0.0097	밀	0.0355	정	0.6133	채	0.0518
건	0.1462	밖	0.0278	젓	0.0070	책	0.0705
겉	0.0059	뱅	0.0017	제	0.4918	첸	0.0003
겉	0.0761	병	0.0013	젝	0.0011	첸	0.0002
검	0.0373	별	0.0635	젠	0.0064	첸	0.0007
겉	0.0133	봉	0.0300	젤	0.0016	첸	0.0002
것	0.6582	봐	0.0212	젯	0.0009	첸	0.0003
겉	0.0011	봤	0.0088	저	0.0728	쟁	0.0043
경	0.0003	분	0.2094	전	0.0005	차	0.0001
겉	0.0032	빨	0.0195	젯	0.0460	처	0.1418
겉	0.0003	뺨	0.0002	조	0.2824	척	0.0174
겉	0.0004	뺨	0.0041	죽	0.0609	천	0.1085
게	0.6419	뺨	0.0004	존	0.0468	철	0.0580
겐	0.0049	뺨	0.0004	졸	0.0110	첩	0.0056
겉	0.0012	삽	0.0016	좁	0.0388	첩	0.0035
겉	0.0974	셍	0.0080	좁	0.0048	첩	0.0169
겨	0.0548	셍	0.0005	중	0.0963	칭	0.0795
격	0.0664	셍	0.0050	죇	0.0005	체	0.1827
겉	0.0069	셍	0.0002	죄	0.0177	첸	0.0004
건	0.0404	손	0.0002	권	0.0001	첸	0.0007
겉	0.0041	숏	0.0007	조	0.0316	처	0.0470
겉	0.0045	쉽	0.0005	주	0.4946	첸	0.0133
겉	0.0113	습	0.2085	죽	0.0436	체	0.0013
경	0.2923	숫	0.0068	준	0.0736	초	0.0746
겉	0.0047	승	0.0516	줄	0.0561	축	0.0110
계	0.2666	시	0.7027	좁	0.0053	촌	0.0152
고	1.7232	써	0.0394	좁	0.0016	출	0.0002
곡	0.0222	쑤	0.0004	죇	0.0005	춤	0.0006
곤	0.0237	쑤	0.0102	중	0.2583	춤	0.0406
골	0.0306	쑤	0.0003	귀	0.0074	촬	0.0031
꿍	0.0002	쑤	0.0002	권	0.0011	최	0.0792
곰	0.0077	쑤	0.0019	질	0.0002	쇼	0.0002
곰	0.0082	압	0.0184	쑤	0.0005	추	0.0808
곳	0.0516	약	0.0836	즈	0.0167	출	0.0902
공	0.2402	안	0.0041	죽	0.0209	춤	0.0085
곳	0.0003	언	0.0807	즐	0.0124	춤	0.0007
과	0.4925	에	1.9332	좁	0.0161	춇	0.0003
곽	0.0030	염	0.0219	좁	0.0013	춤	0.0375
관	0.2477	오	0.2969	증	0.0644	취	0.0066
괄	0.0028	옥	0.0149	지	1.5071	첸	0.0011

char acte r	Freque ncy rate	char acte r	Freque ncy rate	char acter	Freque ncy rate	char acter	Freque ncy rate
굼	0.0003	온	0.0774	직	0.1318	첸	0.0006
굼	0.0029	올	0.0816	진	0.2448	취	0.0367
굿	0.0019	왈	0.0008	질	0.1089	취	0.0003
궁	0.0167	웬	0.0036	좁	0.0006	츠	0.0109
궁	0.0039	웰	0.0005	짐	0.0170	츄	0.0271
깁	0.0023	웹	0.0014	집	0.1269	츄	0.0013
겉	0.0003	위	0.2927	깃	0.0211	츄	0.0222
꼭	0.0154	윈	0.0017	깃	0.0024	치	0.2470
꿀	0.0053	윌	0.0014	질	0.0026	캡	0.0004
꽤	0.0002	윌	0.0026	짚	0.0028	컱	0.0014
꿇	0.0005	윌	0.0004	짜	0.0214	컨	0.0061
꿇	0.0005	유	0.2163	짜	0.0153	컨	0.0010
꿇	0.0005	윤	0.0238	짠	0.0007	컬	0.0034
꿇	0.0368	울	0.0235	짹	0.0002	콧	0.0019
납	0.0027	응	0.0141	짹	0.0013	콩	0.0069
넙	0.0014	으	0.7484	짹	0.0059	쿨	0.0007
넙	0.0013	으	0.0025	짹	0.0004	키	0.0654
넙	0.0138	은	1.1964	짹	0.0004	탯	0.0012
넙	0.0075	을	1.8327	짹	0.0026	탕	0.0145
넙	0.0113	음	0.0009	짹	0.0040	태	0.1138
너	0.0769	음	0.2253	쟁	0.0004	택	0.0325
넙	0.0087	읍	0.0042	짜	0.0003	털	0.0103
넙	0.0146	의	1.9403	짜	0.0080	퇴	0.0137
넙	0.0002	윈	0.0001	쩍	0.0060	튀	0.0001
느	0.1018	이	3.5495	쩍	0.0015	튼	0.0097
늑	0.0020	익	0.0259	쩍	0.0045	팁	0.0115
느	2.6196	인	0.7244	쟁	0.0010	팅	0.0049
늘	0.0763	일	0.5178	쩌	0.0002	패	0.0237
닉	0.0021	임	0.0953	쪼	0.0033	폼	0.0725
넙	0.0027	입	0.1716	쪽	0.0563	프	0.0744
넙	0.0010	잇	0.0051	쪼	0.0005	픈	0.0054
넙	0.0047	있	0.8777	쪼	0.0012	플	0.0140
넙	0.0003	잉	0.0040	쪼	0.0060	학	0.2479
넙	0.0113	잇	0.0084	쪽	0.0003	행	0.1581
넙	0.0009	잎	0.0064	쪼	0.0006	활	0.0805
넙	0.0747	자	0.8494	쪼	0.0019	흘	0.0022
원천부호화 규칙 2번 위배							
가	1.7013	입	0.0199	첸	0.0030	첸	0.0006
갈	0.1660	잡	0.0402	즌	0.0025	첸	0.0011
개	0.0029	자	0.0008	징	0.0211	첸	0.0003
겐	0.0002	짐	0.1191	째	0.0315	춇	0.0010
겉	0.1415	죇	0.0007	쩍	0.0005	춇	0.0472
곧	0.0180	중	0.0669	쩍	0.0003	춘	0.0154

character	Frequency rate	character	Frequency rate	character	Frequency rate	character	Frequency rate
육	0.0544	좌	0.0144	중	0.0004		
응	0.0319	쥐	0.0062	쫄	0.0003		

전송 효율의 개선에 대한 정량적 계산은 표4의 원천부호화 규칙에 위배되는 한글부호들에 대한 사용빈도율을 사용하였다.

표 5. 개선율

Table 5. Improvement rate

사용빈도율 총합계	99.99%
원천부호화 1번 위배 빈도율	39.3650%
원천부호화 2번 위배 빈도율	2.5095%
1회 위배 글자수	627
1회 위배 글자중 빈도율 0%인 글자수	122
2회 위배 글자수	50
2회 위배 글자중 빈도율 0%인 글자수	9
교체 필요한 글자수	428
비위배 영역내 빈도율 0%인 글자수	283
사용 빈도율 0.0001% 글자수	272

1. $\Sigma(\text{원천부호화 규칙 1번 위배율}) + \Sigma(\text{원천부호화 규칙 2번 위배율}) * 2 - 0.0001 * 272$

(원천부호화 규칙 1번 위배율: 한글글자마다부호 원천부호화 1번 위배율

원천부호화 규칙 2번 위배율: 한글글자마다부호 원천부호화 2번 위배율)

$= 39.3650\% + 2.5095\% * 2 - 0.0272\%$

$= 44.3568\%$

위와같이 원천부호화 규칙에 위배되는 한글부호를 빈도율이 낮은 한글부호로 대체할 경우 회선부호화과정에서 발생하는 스크램블링 처리율을 개선시키게 되어 데이터의 전송효율을 재고시키는 효과를 얻게 된다.

III. 결 론

본 논문에서는 Unicode에 규정되어 있는 한글부호 체계와 원천부호화 규칙간의 관계를 연구하였다. 본 연구는 Datelink 계층의 HDLC 통신규약 내 Flag 비트열과 문자 비트열간의 에러를 방지하는 기능과 AMI/HDB-3 회선부호화와 스크램블링방식을 적용할 때에 대하여 연구하였다.

연구를 위한 분석대상 한글 유니코드 문자부호는 KS X 1001의 한글부호체계에 대하여 연구하였다.

연구결과 2350개의 한글부호중 원천부호화 규칙에 위배되는 한글부호는 428개로 나타났다. 이 428개의 한글부호에 대한 사용 빈도율은 44.38%였다. 이 원천부호화 규칙에 위배되는 사용빈도가 높은 한글부호를 사용빈도가 상대적으로 낮은 문자부호에 적용 할 경우, 사용 빈도율에 해당하는 정도의 회선부호기 내의 스크램블링 효율을 약 44% 정도 개선시킬 수 있는 것으로 나타났다.

References

- [1] American Standards Association, *American Code(July 6, 1999). for Information Interchange, ASA X3.4-1963*, Washington, D.C., USA 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., Washington, D.C., USA , Mar. 1986.
- [3] RFC 20, *ASCII format for Network Interchange*, New York, USA, Oct. 1969 (<http://tools.ietf.org/html/rfc20>).
- [4] N. B. A. Forouzan, *Data communication*. New York, USA, McGraw Hill, 2007.
- [5] ITU-T Recommendation G.703, "Physical/

electrical characteristics of hierarchical digital interfaces," Oct. 1998, pp. 24-41. Technical Report

- [6] 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 Institue*, vol. 15, no. 5, 2011, Technical Report
- [7] 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. 227-240.
- [8] 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. 753-760.
- [9] W. Hong, "A Study on the Hangeul Syllables of Unicode System considering Data Transmission Efficiency," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 1, 2015, pp. 39-46.
- [10] W. Hong, "An Analysis on the Korean Language for Optimum Transmission of Hangeul Code," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 1, 2015, pp. 33-38.



홍완표(Wan-Pyo Hong)

1991년 서울과학기술대학교 전자공학과 졸업(공학사)
1994년 연세대학교 공학대학원 산업공학과 졸업(공학석사)

- 1999년 광운대학교 대학원 전자공학과 졸업 (공학박사)
- 1990년 정보통신기술사 합격
- 1991년 정보통신부 5급 특별채용고시합격 본부 통신정책실, 전파방송관리국, 정보화기획실
- 1997년 삼성전자(주) 통신사업부 전송영업그룹장
- 1999년 광운대학교 연구전담교수
- 2000년 한국정보통신기술협회장
- 2002년 한세대학교 IT학부 정보통신공학 교수
한세대학교 정보통신연구소장
- ※ 관심분야 : 위성통신방송, 문자코딩, 통신정책

저자 소개



박요셉(Yo-Seph Park)

2010년 한세대학교 정보통신학과 졸업(공학사)
2015년 한세대학교 대학원 IT융합학과 졸업(공학석사)

※ 관심분야 : 문자코딩, 데이터통신, 빅데이터

