ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 18-04-43
278
j.inst.Korean.electr.electron.eng.Vol.22,No.4,1163~1167,December 2018
http://dx.doi.org/10.7471/ikeee.2018.22.4.1163

정규 허프만 코드의 선택적 암호화

Selective Encryption of Canonical Huffman code

박 상 호*★

Sang-ho Park*★

Abstract

The selective encryption scheme for canonical Huffman codes using the inversion of bit values is proposed. The symbols are divided into blocks of a certain size, and each symbol in the block is compressed by canonical Huffman coding. Blocks are determined to be sent in the original code or encrypted form. The encryption block inverts the values of the whole bits, and bits of block that do not encrypt are not inverted. Those compressed data are transmitted with the encryption information. It is possible to decrypt the compressed data on the receiving side using the encryption information and compressed data.

요 약

비트 값의 반전을 이용하여 정규 허프만 코드의 선택적인 암호화 방법을 제안하였다. 심벌들을 일정한 크기의 블록으로 나누어 블록안의 각 심벌들을 정규 허프만 코딩으로 압축한다. 블록별로 원 코드로 보낼 것인지 암호화하여 보낼것인지 결정하고 암호화 블록은 전제 비트들의 값을 반전시키고 암호화하지 않는 블록들은 원 코딩 데이터를 암호화 정보와 함께 전송한다. 수신측에서 압축된 데이터를 암호화 정보를 이용하여 해독가능하다.

Key words : Entropy coding, Huffman coding, Canonical tree, Selective Encoding, Encryption

Ⅰ. 서론

데이터를 저장하거나 전송하기 위하여 데이터의 크기를 줄이는 압축 단계를 거치게 된다. 압축방법에는 무손실압축과 손실압축이 있다[1]. 손실압축은 압축비가 높으면 높을수록 원 데이터와 복원된 데이터 간에 오차가 크다. 그러나 손실 압축한 데이터는 청각, 시각 등의 감각기관으로 품질의 열화를 알아낼 수 없을 정도이므로 오디오, 영상, 비디오 데이터의 압축방법으로 채택되고 있다. 무손실 압축방식은 데이터를 무손실로 압축하기 위하여 사용되고 있지만, 데이터의 크기가 큰 오디오, 영상, 비디오와 같은 멀티미디어 데이터들은 최대한 많이 압축하기 위하여 손실 압축한 데이터를 무손실 압축방법으로 한 번 더 압축하는 데 사용되고 있다. 허프만 코딩[2]은 무손실 압축방식으로 압축 알고리즘이 단순하며 우수한 성능을 나타내므로 문서 압축, 무손실 영상 압축 등 광범위한 분야에서 압축 알고리즘으로 사용하고 있다.

멀티미디어 데이터는 정보를 보호하기 위하여 암호화하기도 하고 저작권을 위하여 워터마크 기법이 사용된다. 멀티미디어 데이터는 크기가 크므로

* Dept. of Information and Communication Engineering, Andong National University
★ Corresponding author
E-mail : spark@anu.ac.kr, Tel : +82-54-820-5748
※ This work was supported by a Research Grant of Andong National University.
Manuscript received Dec. 9, 2018; revised Dec. 11, 2018; accepted Dec. 13, 2018

데이터 전체를 암호화하기도 하지만 데이터 중 일부만 선택적으로 암호화하기도 한다. 본 논문에서는 심벌들을 블록으로 나누어 정규 허프만 코딩으로 압축된 데이터를 블록 별로 선택적 암호화하는 방법을 제안하였다. 각 코드는 지정된 위치의 비트 값을 반전시켜 암호화 정보가 없는 사람은 디코딩이 가능 하지 않도록 하였다.

II. 정규 허프만 코드의 암호화

1. 정규 허프만 트리

허프만 코딩은 엔트로피 코딩방식이며 무손실 압축에 사용되고 있다. 심벌의 발생횟수가 크면 클수록 작은 수의 비트를 할당하고 확률이 낮으면 작을수록 긴 비트를 할당하는 가변 길이 코딩방식이다. 허프만 코딩의 예를 표 1에 나타내었으며 허프만 코드는 그림 1과 같이 트리 형태로 표현한다.

그림 1의 허프만 트리는 같은 레벨에 있는 심벌들은 왼쪽에서 오른쪽으로 갈수록 코드의 크기가 커진다. 심벌 C, D, E는 트리의 같은 레벨에 있는데 코드가 각각 1100, 1101, 1110으로 오른쪽으로 올림차순으로 정렬되어 있다. 이러한 형태로 각 레벨의 코드가 정렬된 허프만 코드를 정규 허프만 코드라 한다[3]. 정규 허프만 코드는 허프만 코드보다 트리를 구성하기 위한 정보를 적제 전송할 수 있고 그 정보를 이용하여 트리를 구성하는 알고리즘을 단순화할 수 있어 효율적으로 디코딩할 수 있다[4]. 허프만 코딩하여 얻은 허프만 트리가 정규 허프만 트리 구조가 아닌 경우에는 같은 레벨에 있는 심벌들의 코드를 올림차순으로 정렬하여 재구성하면 정규 허프만 트리 구조가 된다. 본 논문에서는 정규 허브만 트리의 정점을 활용하기 위하여 허프만 코드가 정규 허프만 코드로 재정렬된 코드의 암호화 방법에 대해 논의한다.

Table 1. Huffman codes for seven symbols.
표 1. 7개의 심볼을 위한 허프만 코드

Symbol	Probability	code
A	48%	0
B	31%	10
C	7%	1100
D	6%	1101
E	5%	1110
F	2%	11110
G	1%	11111

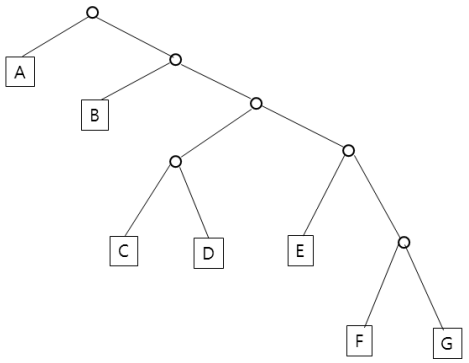


Fig. 1. Huffman tree for seven symbols.
그림 1. 7개의 심벌을 위한 허프만 트리

2. 정규 허프만 트리의 선택적 암호화

압축된 디지털 데이터는 정보의 유출, 변조 등을 방지하기 위하여 데이터를 암호화하고 디지털 변조한 후 전송한다. 암호화는 파일의 모든 데이터를 암호화하기도 하지만 데이터 중 일부만 암호화하기도 한다. 최근 허프만 코드의 선택적 암호화 방법이 제안되었다[5]. [5]에서는 허프만 트리의 특정 레벨에서 트리의 좌우 위치를 바꿈으로 코드의 값을 변화시켜 암호화하였다. 예를 들면 그림 1의 허프만 트리에서 레벨 3에서 트리의 좌우 위치를 바꾸면 그림 2와 같이 되고 암호화한 코드는 표 2와 같다.

표 2에서 암호화된 코드를 보면 코드의 세 번째 비트 값이 반전되었다. 심벌 A와 B는 각각 1-bit 코드와 2-bit 코드이므로 반전된 비트가 없다. 레벨 3 이하에 있는 심벌들은 코드 중 세 번째 비트가 반전되었음을 알 수 있다. 허프만 트리의 좌우 위치를 바꾸어 암호화하는 과정은 반전하기를 원하는 비트를 결정한 후 그 비트를 반전하는 레벨 이하의 좌우 위치를 바꾼다. 송신 측에서 표 1의 허프만 트리와 암호화 한 블록의 위치 그리고 반전한 레벨의 위치를 송신한다. 수신 측에서는 암호화한 블록을 암호화한 레벨의 비트를 반전하여 암호를 해독하여 코드를 재구성한 후 허프만 테이블을 참조하여 코드를 심벌로 변환한다.

Table 2. Encrypted Huffman codes for seven symbols based on algorithm in [5].
표 2. [5]의 알고리즘으로 암호화한 7개의 심볼을 위한 허프만 코드

Symbol	Probability	Huffman code	encrypted code
A	48%	0	0
B	31%	10	10
C	7%	1100	1110
D	6%	1101	1111
E	5%	1110	1100
F	2%	11110	11010
G	1%	11111	11011

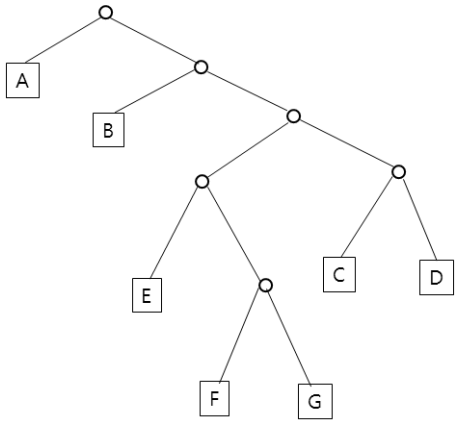


Fig. 2. Encrypted Huffman tree for seven symbols based on algorithm in [5].
그림 2. [5]의 알고리즘으로 7개의 심볼을 위한 암호화한 허프만 트리

본 논문에서는 [5]의 암호화 기법을 확장하여 다양한 암호화 방법을 제안한다. 그림 2에서 세 번째 레벨에서 좌우의 위치를 바꾸는 대신 레벨 3에서 180도 회전하면 그림 3과 같은 트리로 재구성되며 암호화한 코드는 표 3과 같다. 표 3에서 허프만 코드와 암호화된 코드를 보면 코드의 세 번째 비트 이하 피트들의 값이 반전되었다. 심벌 A와 B는 각각 1-bit 코드와 2-bit 코드이므로 반전된 비트가 없다. 암호화하는 레벨인 레벨 3 이하에 있는 심벌들은 코드 중 세 번째 비트부터 끝까지 반전되어 있고 오른쪽으로 내림차순으로 정렬되어 있음을 알 수 있다. 암호화된 코드는 암호화하는 레벨 이하의 비트가 반전되고 같은 레벨의 코드들은 오름차순에서 내림차순으로 바뀌게 되고 트리의 형태는 여전히 정규 허프만 트리 구조를 유지하게 된다. 제안한 암호화 방법은 레벨 1에서부터 암호화하면 허프만 코드의 모든 비트 값이 반전되며 레벨 k에서부터 암호화하면 k비트부터 이후의 모든 비트가 반전되어 암호화된다. 정규 허프만 코드는 디코딩이 구조가 간단해지며 디코딩 시간을 단축할 수 있으므로[4] 본 논문에서는 정규 허프만 트리 구조에 대한 암호화 방법을 제안한다.

Table 3. Encrypted Huffman code for seven symbols based on proposed method.
표 3. 제안한 방법으로 암호화한 7개의 심볼을 위한 허프만 코드

Symbol	Probability	Huffman code	encrypted code
A	48%	0	0
B	31%	10	10
C	7%	1100	1111
D	6%	1101	1110
E	5%	1110	1101
F	2%	11110	11001
G	1%	11111	11000

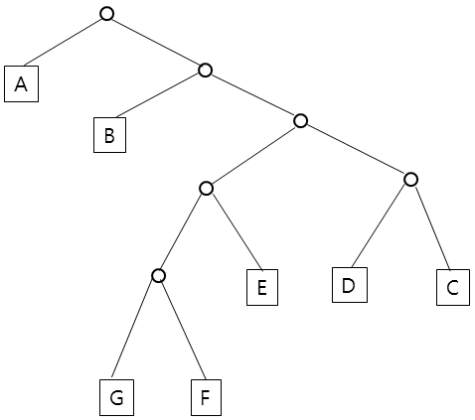


Fig. 3. Encrypted Huffman tree for seven symbols based on proposed method.
그림 3. 제안한 방법으로 암호화한 7개의 심볼을 위한 허프만 트리

표 1의 심벌들로부터 스트링 $S = ABDFCA$ 이면 허프만 코드 $H = 01011011111011000$ 이다. 만약 레벨 1에서 반전시키면 허프만 코드는 전체 비트가 반전되어 $H = 10100100000100111$ 로 암호화되어 수신측에서는 17-bit 스트링 중 처음 14비트는 $BBABAAAABA$ 로 복호화되고 끝에 있는 3-bit는 알 수 없는 심벌로 복호가 되지 않는다. 만약 레벨 2에

서 반전시키면 각 심벌의 코드가 두 번째 비트부터 반전되므로 허프만 코드는 $H = 01110101000110110$ 으로 암호화되고 수신 측에서는 $AEBBAADB$ 로 모든 비트가 해석되나 원래의 정보를 알아낼 수 없게 된다.

허프만 트리에서 반전하는 레벨 수의 비트 위치의 값이 반전하게 되는데 두 번 이상 반전하여 암호화하는 경우 반전하는 위치가 이전 값을 반전하여 암호화된 코드를 만든다. 예를 들면 그림 1의 정규 허프만 트리에서 코드가 $A = 0$, $B = 10$, $C = 1100$, $D = 1101$, $E = 1110$, $F = 11110$, $G = 11111$ 로 인코딩되어 있다. [5]의 방법으로 레벨 1에서 반전하면 암호화된 코드는 $A = 1$, $B = 00$, $C = 0100$, $D = 0101$, $E = 0110$, $F = 01110$, $G = 01111$ 으로 첫 번째 비트만 반전되어 있다. 레벨 3에서 한번더 암호화하면 $A = 1$, $B = 00$, $C = 0110$, $D = 0111$, $E = 0100$, $F = 01010$, $G = 01011$ 으로 이전에 암호화한 코드에서 세 번째 비트 값만 반전되어 있음을 볼 수 있다. 정규 허프만 트리의 하나 이상의 레벨에서 비트값을 반전 시킬 수 있다. 본 논문에서 제안한 방법으로 레벨 1에서 반전하면 암호화된 코드는 $A = 1$, $B = 01$, $C = 0011$, $D = 0010$, $E = 0001$, $F = 00001$, $G = 00000$ 으로 첫 번째 위치부터 마지막 위치까지 비트 값이 반전되어 있다. 이 코드를 레벨 3에서 한번더 암호화하면 $A = 1$, $B = 01$, $C = 0000$, $D = 0001$, $E = 0010$, $F = 00110$, $G = 00111$ 으로 이전에 암호화한 코드에서 세 번째 이하 모든 비트 값이 반전되어 있음을 볼 수 있다. 이는 첫 번째 비트와 두 번째 비트만을 반전하여 암호화한 것과 같다. 같은 방법으로 레벨 2에서 반전시키고 레벨 4에서 또 반전시키면 원 코드의 2번째 비트부터 세 번째 비트까지 반전되어 원하는 비트 블록만 반전시켜 암호화할 수 있다. [5]의 방법은 특정 위치의 비트를 반전시키는 방법으로 암호화하는 데 효과적이며 제안한 방법은 블록 단위로 반전시켜 암호화하는데 효율적임을 알 수 있다.

3. 코드의 인코딩과 디코딩

선택적 암호화 방법으로 허프만 코드를 인코딩하는 방법들은 아래와 같다.

① k 개의 심벌들을 하나의 블록으로 생각한다. 전체 심벌 블록은 n 개로 가정한다.
② 각 블록들은 허프만 코딩하기 전에 암호화할 것인가 결정하고 암호화하는 경우에는 반전 횟수i 와 반전하는 레벨 l 을 결정한다. 블록 내의 심벌들은 결정한 방법대로 암호화 없이 코딩하거나 암호화하여 인코딩한다.
③ ②의 과정을 n 개의 블록에 대해 반복한다.

암호화된 코드들은 허프만 테이블과 암호화 정보와 함께 전송된다. 암호화 정보는 블록의 크기 k, 블록의 수 n, 그리고 각 블록마다 블록의 암호화 방법 i와 l들로 구성되어 있다. 레벨 1에서 한 번만 반전하는 경우 암호화 방법 i와 l들은 필요 없고 반전한 블록의 위치만 전송하면 되므로 암호화 정보 파일의 크기가 작아지며 수신 측에서 디코딩이 단순해진다. 암호화 정보는 보안을 유지하기 위하여 데이터 전송경로와 다른 좀 더 보안이 보장되는 경로로 보낸다.

수신 측에서 데이터를 수신한 데이터 즉 선택적 암호화 방법으로 암호화된 코드를 디코딩하는 방법들은 아래와 같다.

① 첫 번째 블록의 암호화 정보에 따라 블록 내 심벌들을 허프만 테이블에 따라 디코딩한다. 수신된 블록이 암호화된 경우 수신된 허프만 테이블에서 암호화한 위치의 비트 값을 반전하여 새로운 테이블 값을 얻을 수 있다.
② 두 번째 블록부터 마지막 블록까지 ①의 방법을 반복한다.

레벨 1에서 한 번만 반전된 경우 블록 내의 모든 비트가 반전되어 있으므로 블록 내의 모든 비트 값을 반전시킴으로 원 코드를 얻을 수 있어 디코딩 과정이 간단하고 디코딩 시간이 최소화되므로 약한 보안이 요구되고 빠른 디코딩이 중요한 경우 사용할 수 있고, 두 번 이상 비트를 반전한 경우 오버헤드가 커지고 디코딩 과정이 복잡해져서 디코딩 시간이 길어지므로 디코딩 시간보다 강한 보안이 요구되는 경우에 사용할 수 있다.

III. 결론

디지털 데이터를 무손실 압축방식인 허프만 코딩

으로 압축한 파일을 선택적으로 암호화하는 방법을 제안하였다. 정규 허프만 트리는 디코딩 과정이 단순하여 디코딩 시간을 단축할 수 있으므로 정규 허프만 코드에 대한 암호화 방법과 전송과정 해독 방법을 제안하였다.

심벌들을 일정한 크기의 블록으로 나누고 각 블록의 심벌들을 정규 허프만 코딩 한 다음 암호화할 것인지 결정하여 비트 값을 반전하여 암호화한다. 블록들의 암호화 패턴은 매번 달라지며 패턴정보는 송신측과 수신측이 보유하여 암호화 정보를 가진 사람만 암호를 해독할 수 있게 된다. 정규 허프만 트리에서 트리를 회전시키는 레벨은 코드에서 비트 값이 변하기 시작하는 위치이며 그 위치에서 마지막 비트까지 값이 변하므로 암호화가 단순하나 강력한 암호화가 가능하다. 정규 허프만 트리를 회전하여 만든 암호화는 코드의 일정 부분을 블록으로 반전 시키는데 효과적이며 정규허프만 트리를 좌우 이동하여 만든 암호화는 코드의 한 특정 비트를 암호화 하는데 효과적이었다. 수신측에서 암호해독을 위하여 블록의 크기, 암호화 패턴이 필요하며 한 코드에서 2회 이상 반전시켜 암호화 하였으면 트리의 회전 위치에 관한 정보도 필요하게 된다.

References

[1] K. Sayood, *Introduction to Data Compression, 3rd ed.* Morgan Kaufmann, 2006.

[2] D. Huffman, “A method for the construction of minimum redundancy codes,” *Proc. of the IRE*, vol.40, no.90, pp.1098-1101, 1952. DOI:10.1109/JRPROC.1952.273898

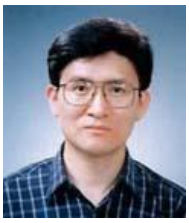
[3] E. S. Schwartz and B. Kallick, “Generating a canonical prefix encoding,” *Communications of the ACM*, vol.7, no.3, pp.166-169, 1964. DOI:10.1145/363958.363991

[4] S. Park, “Efficient Huffman decoding using canonical Huffman tree,” *J. of KSCI*, vol.12, no.4, pp.111-117, 2007.

[5] S. Park, “A study on selective encryption of Huffman codes,” *J. Information and Security*, vol.7, no.2, pp.57-63, 2007.

BIOGRAPHY

Sangho Park (Member)



1979 : BS degree in Electronics Engineering, Kyungpook National University.
1981 : MS degree in Electronics Engineering, Yeungnam University.
1989 : MS degree in Computer Engineering, Syracuse University.
1995 : PhD degree in Electrical and Computer Engineering, State University of New York at Buffalo.
1996-current : Professor, Andong Nation University.