http://dx.doi.org/10.7840/kics.2016.41.10.1176

최적의 이진 부분접속 복구 부호 생성법

남 미 영◆, 송 홍 엽°

Constructions for Optimal Binary Locally Repairable Codes

Mi-Young Nam◆, Hong-Yeop Song°

요 약

본 논문에서는 패리티 검사 행렬에 기반하여 부분접속수가 2인 이진 부분접속 복구 부호의 생성법을 제안한다. 제안하는 부호는 항상 최소거리 6을 갖는다. 이 부호는 부호의 길이와 최소거리가 주어졌을 때 부분접속수가 2인 부호가 가질 수 있는 최대의 차원을 갖는다는 관점에서 최적이다.

Key Words : locality, locally repairable codes

ABSTRACT

We propose some binary locally repairable codes with locality 2 uising a parity-check matrix. The minimum distance of the proposed codes is 6. The proposed codes are optimal in the sense of achieving the upper bound of dimension for given length, minimum distance, and locality.

Ⅰ. 서 론

분산 저장 시스템은 대용량의 데이터를 네트워크로 연결된 다수의 저장노드에 분산하여 저장하는 시스템으로, 노드 장애에 대응하여 데이터를 안정적으로 저장하기 위해서 다양한 소실 부호가 사용된다[1].

노드의 장애와 복구는 빈번하게 발생하기 때문에 효율적인 노드 복구를 가능하도록 하는 소실 부호에 관한 연구가 활발히 이루어지고 있다[2].

복구 과정의 효율성을 판단할 수 있는 중요한 척도 중 하나로 부분접속수 (locality)가 사용된다[3]. 부분접속수는 임의의 노드를 복구하기 위해 접속해야 하는 최소 노드의 수를 의미한다. 부호의 파라미터가 주어졌을 때, 부호가 가질 수 있는 최소 거리의 이론적인 상한계가 제시되었고[3], 이러한 한계식을 달성하는 최소거리를 갖는 최적의 부분접속 복구 부호를 생성하는 방법에 관한 연구가 발표되었다[4].

부호의 사용에 따른 연산 복잡도를 줄이고 실제 시스템에 효과적인 적용을 위해 작은 크기의 유한체에서의 부분접속 복구 부호의 생성이 많은 주목을 받았다[5]. 특히, 이진 부분접속 복구 부호에 관한 연구가 최근 들어 활발히 이루어지고 있다[6,7].

$(n,k,r)_q$부분접속 복구 부호는 크기 q인 유한체에서 정의되는 길이 n, 차원 k, 그리고 부분접속수가 r인 부호라고 하자. 최소거리가 d인 $(n,k,r)_q$부분접속 복구 부호의 차원 k의 유한체의 크기를 고려한 상한계는 다음과 같이 알려져 있다[8].

$$k \le min_{\ t \in Z_+}[tr + k_{opt}^{(q)}(n-t(r+1),d)], \qquad (1)$$

여기서 $k_{opt}^{(q)}(n,d)$는 길이 n이고 최소거리가 d인 부호가 가질 수 있는 최대의 차원의 크기이다.

본 논문에서는 항상 최소거리가 6인 이진 부분접속 복구 부호의 생성법을 소개한다. 제안하는 부호는 식 (1)의 상한을 달성함으로써 최적이다.

Ⅱ. 부분접속 복구 부호의 패리티 검사 행렬

$(n,k,r)_2$ 부호 C의 부호어가 $(c_1,c_2,...,c_n)$이라 하자. i번째 심볼 c_i의 부분접속수가 r 이하라면 C의 패리티 검사 행렬의 행 공간 (row space)에 i번째 열의 값이 1이고 무게가 $r+1$이하인 행 벡터가 항상 존재해야 함을 의미한다.

이러한 사실로부터 다음과 같은 형태의 패리티 검사 행렬을 갖는 부호가 부분접속수 r인 부분접속 복구 부호가 됨을 쉽게 알 수 있다.

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. 2013R1A1A2062061)
◆ First Author : School of Electrical and Electronic Engineering, Yonsei University, my.nam@yonsei.ac.kr, 학생회원
° Corresponding Author : School of Electrical and Electronic Engineering, Yonsei University, hysong@yonsei.ac.kr, 종신회원
논문번호 : KICS2016-09-249, Received September 9, 2016; Revised September 27, 2016; Accepted September 27, 2016

$$H=\left[\frac{H_L}{H_G}\right]=\begin{bmatrix}1_{r+1}\,0_{r+1}\,...\,0_{r+1}\\0_{r+1}\,1_{r+1}\,...\,0_{r+1}\\\vdots\quad\vdots\quad\ddots\quad\vdots\\0_{r+1}\,0_{r+1}\,...\,1_{r+1}\\H_G^1\quad H_G^2\quad ...\quad H_G^s\end{bmatrix}, \tag{2}$$

여기서 1_{r+1}과 0_{r+1}은 모두 길이가 $r+1$인 행벡터로 각각 원소가 모두 1인 벡터와 원소가 모두 0인 벡터를 나타낸다. 그리고 H_G^i, $i=1,2,...,s$,는 임의의 정수 l에 대해, 크기가 $l\times(r+1)$인 H_G의 i번째 부행렬을 나타낸다.

이러한 행렬을 패리티 검사 행렬로 갖는 부호는 $((r+1)s, rs+l, r)_2$ 부호가 된다. 위의 행렬 H에서 상위 부행렬 H_L에 의해 부호의 부분접속수가 r임이 보장된다.

부호의 패리티 검사 행렬 H의 임의의 $\delta-1$개 이하의 열벡터가 항상 선형독립이고 어떤 δ개의 열벡터가 선형독립이 아니라면 이 부호의 최소거리는 $d=\delta$라는 사실은 잘 알려져 있다. 이 사실로부터 패리티 검사 행렬이 상위 부행렬 H_L로만 이루어진 경우, 이 부호는 최소거리가 $d=2$인 부호가 된다. 하위 부행렬 H_G를 설계하는 방법에 따라 이 부호의 최소거리가 결정된다.

하나의 부행렬 H_G^i, $i=1,2,...,s$,의 모든 s개의 열벡터가 서로 다르게 설계함으로써 최소거리가 $d=4$인 $(n,k,r)_2$부호가 제안되었다[6,7].

Ⅲ. 제안하는 이진 부분접속 복구 부호의 생성

이 장에서는 Ⅱ장의 (2)와 같은 구조를 갖는 패리티 검사 행렬을 통해 부분접속수가 $r=2$인 이진 부분접속 복구 부호를 생성하는 새로운 방법을 제안한다. 이 부호의 패리티 검사 행렬 H는 다음과 같다.

$$H=\begin{bmatrix}111\,000\,...\,000\\000\,111\,...\,000\\\vdots\quad\vdots\quad\ddots\quad\vdots\\000\,000\,...\,111\\H_G^1\;H_G^2\;...\;H_G^s\end{bmatrix} \tag{3}$$

정리 1. s는 $s\geq 3$인 정수라고 하자. 식 (3)의 행렬 H의 하위 부행렬 H_G가 다음 조건을 만족하면 H를 패리티 검사 행렬로 하는 부호 C의 최소거리는 $d\geq 6$이다.

① H_G의 모든 열벡터가 서로 다르다.

② 임의의 서로 다른 두 부행렬 H_G^i와 H_G^j에 대해, 네 개의 열벡터

표 1. 이진 선형 부호의 최대 차원 $k_{opt}^{(2)}(n,6)$
Table 1. The maximum possible dimension $k_{opt}^{(2)}(n,6)$ of a binary linear $(n,k)_2$ code with $d=6$

	$n=9$	$n=12$
$k_{opt}^{(2)}(n,6)$	2	4

$h_{i,a}, h_{i,b}, h_{j,c}, h_{j,d}, i\neq j\in\{1,2,...,s\}$,
$1\leq a\neq b, c\neq d\leq 3$, 가 선형독립이다.

증명. 패리티 검사 행렬 H로부터 선택된 임의의 5개 이하의 열벡터가 항상 선형독립임을 보임으로써 정리 1을 증명할 수 있다. 이때 (3)에서 주어진 H의 구조에 의해 홀수 개의 열벡터의 선형조합은 항상 영이 아닌 벡터가 되므로, 선형독립이다. 따라서 짝수 개의 열벡터 집합이 항상 선형독립임을 보이면 된다. ①의 특성으로 인해 임의의 두 개의 열벡터는 항상 선형독립이다. 이는 최소거리가 $d\geq 4$ 임을 보장한다.

다음으로 임의의 네 개의 열벡터의 집합을 고려하자. 하나의 부행렬 H_G^i가 포함되는 세 개의 열벡터를 하나의 그룹 G_i라고 정의하자. 두 개의 서로 다른 그룹 G_i와 G_j로부터 각각 두 개의 열벡터를 선택하는 경우만 고려하면 된다. ②의 특성으로 인해 이렇게 선택된 네 개의 열벡터가 항상 독립임이 보장된다. 이는 최소거리가 $d\geq 6$임을 보장한다.

따라서 부호 C의 최소거리는 항상 $d\geq 6$이다. ■

정리 1에서 주어진 두 개의 조건을 만족하는 행렬을 다음과 같이 생성할 수 있다.

$$H_1=\begin{bmatrix}111000000\\000111000\\000000111\\000011011\\000101101\\011000011\\101000101\end{bmatrix},\ H_2=\begin{bmatrix}111000000000\\000111000000\\000000111000\\000000000111\\000011011011\\000101101101\\011000011110\\101000101011\end{bmatrix},$$
$$H_3=\begin{bmatrix}111000000000000\\000111000000000\\000000111000000\\000000000111000\\000000000000111\\000011011011011\\000101101101101\\011000011110101\\101000101011110\end{bmatrix} \tag{4}$$

위의 식 (4)에 주어진 세 개의 행렬 H_1, H_2, H_3를 패리티 검사 행렬로 갖는 부호를 각각 C_1, C_2, C_3라고 하자.

$(9,2,2)_2$ 부호 C_1, $(12,4,2)_2$부호 C_2, 그리고 $(15,6,2)_2$부호 C_3는 모두 부분접속수가 2이고 최소거리가 6인 이진 부분접속 복구 부호이다.

식 (1)의 상한계를 이용하여 생성된 부호가 최적임을 확인하기 위해, 표 1의 $k_{opt}^{(2)}(n,d)$를 이용해 세 부호가 최적임을 확인하였다. 표 1은 [9]를 참조한 것이다. 부호 C_1은 부분접속수 특성을 고려하지 않더라도 주어진 길이 $n=12$와 최소거리 $d=6$을 갖는 모든 이진 선형 부호가 가질 수 있는 최대 가능한 차원 $k_{opt}^{(2)}(9,6)=2$를 달성하는 부호이므로 최적이다. 부호 C_2를 식 (1)에 대입할 경우, $t=1$일 때 (1)의 우변이 최소가 되고 이 때의 값이 4이고 C_2의 차원과 같다. 따라서 부호 C_2역시 최적이다. 부호 C_3을 식 (1)에 대입하면 $t=1$일 때 (1)의 우변이 최소가 되고, 부호 C_3의 차원이 이때의 값 6과 같다. 따라서 부호 C_3역시 최적이다.

제안하는 부호의 장점을 알아보기 위해 MTTDL (Mean Time To Data Loss)[7]을 이용하였다. [7]의 Proposition 5를 이용해, 기존의 각 부호의 MTTDL을 구한다. 부호의 안정성을 평가할 수 있는 척도로, 3회 반복부호의 MTTDL과의 비교값인 $\zeta=\frac{\text{부분접속 복구부호의 } MTTDL}{\text{3회 반복부호의 } MTTDL}$를 이용하였다. MTTDL 및 ζ를 구하는데 필요한 파라미터 및 설정은 [7]을 따랐다. 그림 1은 기존의 최소거리 4인 부호[7]와 본 논문에서 제안하는 최소거리가 6인 부호의 ζ의 그래프로, 부호의 길이 n이 증가함에 따라 감소하는 형태를 보인다. 기준선으로 3회 반복부호의 ζ인 1을 검정색 실선으로 함께 표시하였다. 파란색 점선은 부분접속수가 2이고 최소거리가 4인 부호[7]의 ζ 값이고 표식 ○를 갖는 붉은 실선이 제안하는 부호의 ζ이다. 부호의 길이 n이 14 이상이 되면 [7]의 부호는 3회 반복부호에 비해 MTTDL이 더 작아지는 반면, 제안하는 부호는 103 이하의 부호길이 까지는 3회 반복부호보다 큰 MTTDL 값을 갖는다.

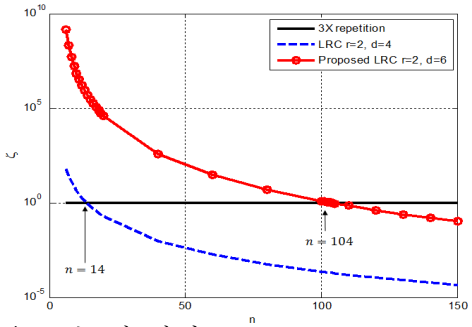

그림 1. 부호 간 ζ의 비교
Fig. 1. The values of ζ versus n of the proposed code and the existing LRC with $d=4$

Ⅳ. 결 론

본 논문에서는 패리티 검사 행렬에 기반한 부분접속 복구 부호의 생성법에 대해 소개하였다. 생성된 부호는 항상 최소거리가 6으로, 기존에 알려진 부호에 비해 높은 안정성을 보장할 수 있다. 또한 생성된 부호는 이진부호이므로 복잡도가 낮아 실제 시스템에 적합하다.

References

[1] J.-H. Kim, J. S. Park, K.-H. Park, M.-Y. Nam, and H.-Y. Song, "Trends of regenerating codes for next-generation cloud storage systems," *Inf. Commun. Mag.*, vol. 31, no. 2, pp. 125-131, Feb. 2014.

[2] M.-Y. Nam, J.-H. Kim, and H.-Y. Song, "Locally repairable fractional repetition codes," *J. KICS*, vol. 40, no. 09, Sept. 2015.

[3] P. Gopalan, H. Cheng, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, Nov. 2012.

[4] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4661-4676, 2014.

[8] V. R. Cadambe and A. Mazumdar, "Bounds on the size of locally recoverable codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 5787-9448, Sept. 2015.

[5] A. Zeh and E. Yaakobi, "Optimal linear and cyclic locally repairable codes over small fields," Online available at http://arxiv.org/pdf/1502.06809.pdf

[6] M. Shahabinejad, M. Khabbazian, and M. Ardakani, "A class of binary locally repairable codes," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3182-3193, Aug. 2016.

[7] J. Hao, S.-T. Xia, and B. Chen, "Some results on optimal locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 440-444, Jun. 2016.

[9] M. Grassl, *Bounds on the minimum distance of linear codes and quantum codes*, Online available at http://www.codetables.de.