논문 18-43-01-03 The Journal of Korean Institute of Communications and Information Sciences '18-01 Vol.43 No.01
https://doi.org/10.7840/kics.2018.43.1.16

차 커버링 배열을 이용한 복제 기반 분산 저장 기술

김 하 성•, 박 호 성°

Replication-Based Distributed Storage Technique Using Difference Covering Arrays

Haseong Kim•, Hosung Park°

요 약

본 논문에서 차 커버링 배열을 이용하여 분할 반복 부호를 설계하는 방법을 제안한다. 제안하는 부호는 두 개의 노드가 공통된 파일 조각을 2개까지 공유할 수 있도록 하는 특징이 있다. 모의실험을 통하여 제안하는 부호가 기존 부호와 유사한 파일 크기를 지원함을 보인다.

Key Words : Distributed storage code, fractional repetition code, difference covering array, replication degree

ABSTRACT

In this paper, we propose a construction method of fractional repetition codes using difference covering arrays. The proposed codes have the property that two nodes are allowed to have up to 2 file fragments in common. It is shown via simulation that the proposed code supports a similar file size to the conventional code.

Ⅰ. 서 론

모바일, IoT 등의 애플리케이션에서 데이터 발생량이 기하급수적으로 증가함에 따라 데이터센터 역시 그 규모 및 개수가 빠르게 증가하고 있다. 효율적으로 데이터를 저장하기 위한 기술은 매우 중요해지고 있는데, 현재 핫 데이터 (hot data)를 위한 복제 기반 분산 저장 기술 및 콜드 데이터 (cold data)를 위한 소실부호 (erasure code) 기반 분산 저장 기술[1,2]로 나뉘어서 발전하고 있다.

복제 기반 분산 저장 기술은 현재 HDFS[3] 같은 파일시스템에서 이미 사용되고 있는 방법이며, 랜덤한 방법 대신 수학적 구조를 바탕으로 복제된 파일 조각들을 분산시키는 방법은 분할 반복 부호를 통하여 구현될 수 있으며 현재까지 다양한 분할 반복 부호가 제안되었다[4,5]. 기존의 분할 반복 부호는 두 개의 노드가 최대 1개까지의 파일 조각을 공유하도록 설계되었다.

본 논문에서는 차 커버링 배열[6]을 이용하여 반복 차수가 3인 새로운 분할 반복 부호의 설계 방법을 제안한다. 제안하는 부호는 기존 분할 반복 부호와는 달리 두 개의 노드가 최대 2개까지의 파일 조각을 공유하도록 허락한다. 이 특징을 효율적으로 활용할 수 있는 예로써 이종 네트워크로 구성된 분산 저장 시스템에서의 데이터 복구와 데이터의 무결성을 확인하기 위한 메시지 인증 등이 있다. 모의실험을 통해 제안하는 부호의 지원하는 파일 크기가 기존 부호에 비하여 작지 않음을 보인다.

Ⅱ. 분할 반복 부호

분할 반복 부호는 저장하려는 파일을 여러 조각들로 분할하고, 각각을 수 회 복제한 후, 노드들에 일정한 규칙에 따라 분산시킨다. 분할 반복 부호는 최소 대역폭 재생성 부호의 변형으로서 복구를 위한 네트워크 대역폭을 최소한으로 이용할 뿐 아니라 디스크의 입출력 및 연산을 최소로 하는 특징을 가지고 있다. 현재까지 그래프 이론, 디자인 이론 등을 이용하여 다양한 분할 반복 부호들이 제안되었고, 각 설계 방법마다 부호의 파라미터들이 다르므로 보다 다양한 부호 파라미터를 위해서 새로운 설계 방법들이 필요하다.

본 논문에서 데이터의 각 심볼들은 F_q의 원소이고, 파일은 외각 부호화를 거친 후 (생략가능) θ개의 심볼로 이루어지며, 데이터를 저장하기 위한 노드의 수는

※ 본 연구는 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이며 (No. NRF-2015R1D1A1A01060941), 또한 2015년도 전남대학교 학술연구비 지원에 의하여 연구되었음
◆ First Author : Department of LG Smart Hybrid Engineering, Chonnam National University, nivell@nate.com, 학생회원
° Corresponding Author : School of Electronics and Computer Engineering, Chonnam National University, hpark1@jnu.ac.kr, 종신회원
논문번호 : KICS2017-11-352, Received November 20, 2017; Revised December 25, 2017; Accepted December 27, 2017

n이라고 가정하자. (n,α,ρ) 분할 반복 부호 C는 다음의 조건을 만족시키는 $\{0,1,...,\theta-1\}$의 n개의 부분집합 $N_0,N_1,...,N_{n-1}$으로 정의된다.

- $n\alpha=\theta\rho$;
- $|N_i|=\alpha,\ i=0,...,n-1$;
- 심볼 j는 N_i들에 정확히 ρ번 포함된다.

위와 같이 정의된 분할 반복 부호 C에 의해 노드 i는 N_i에 있는 심볼들을 저장한다. 파라미터 ρ는 각 심볼에 대한 반복 차수라고 불린다.

분할 반복 부호를 사용한 분산 저장 장치에서 하나의 노드 i가 손실되었을 때 다른 노드들로부터 α개의 심볼을 다운로드 받으면 그 노드는 복구될 수 있다. 그리고 사용자가 임의의 k개의 노드에 접속해서 다운로드 받을 수 있는 최소의 심볼 수 $M(k)$는 다음과 같이 표현된다.

$$M(k)=\min_{\substack{I\subset\{0,...,n-1\}\\|I|=k}}|\cup_{i\in I}N_i|$$

주어진 k에 대하여 $M(k)$는 분할 반복 부호가 저장을 보장하는 심볼의 개수를 의미하며, 저장하고자 하는 파일의 크기와 일치한다.

분할 반복 부호에서 $k\le\alpha$에 대하여 $M(k)$에 대한 하나의 상계는 다음과 같으며 MBR 상계라 불린다.

$$M(k)\le k\alpha-\binom{k}{2}$$

이 MBR 상계를 달성하기 위하여 기존 분할 반복 부호는 임의의 두 개의 노드에 대하여 최대 1개의 공통의 파일 조각을 가질 수 있도록 제한하였다.

Ⅲ. 차 커버링 배열을 이용한 분할 반복 부호 설계

아벨리안 군 $(G,+)$에 대한 차수 t인 차 행렬 $DM(l,t;\lambda)$은 $0\le j\ne j'\le t-1$에 대하여 다음의 차 멀티셋 $\Delta_{j,j'}$이 G의 모든 원소들을 동일하게 λ번 가지고 있게 되는 $\lambda l\times t$ 행렬 $Q=[q(i,j)]$, $q(i,j)\in G$, 으로 정의된다.

$$\Delta_{j,j'}=\{q(i,j)-q(i,j')|0\le i\le\lambda l-1\}$$

아벨리안 군 $(G,+)$에 대한 차수 t인 차 커버링 배열 $DCA(t,\eta;l)$은 $0\le j\ne j'\le t-1$에 대하여 다음의 차 멀티셋 $\Delta_{j,j'}$이 G의 모든 원소들을 적어도 한 번 가지고 있게 되는 $\eta\times t$ 행렬 $Q=[q(i,j)]$, $q(i,j)\in G$, 으로 정의된다.

$$\Delta_{j,j'}=\{q(i,j)-q(i,j')|0\le i\le\eta-1\}$$

일반성을 잃지 않고 $DM(l,t;\lambda)$과 $DCA(t,\eta;l)$의 마지막 행과 마지막 열의 모든 원소들을 0으로 놓기로 하자.

본 논문에서 분할 반복 부호를 설계하기 위해 $DCA(4,2m+1;2m)$를 이용하는데, $m=73$을 제외하면 모든 $m\ge3$에 대하여 존재하며 그 설계 방법은 [6]을 참조한다. 해당 차 커버링 배열에서 마지막 열을 제외한 임의의 두 개의 열의 차 멀티셋은 아래와 같이 나타난다.

$$\Delta_{j,j'}=\{0,1,...,m-1,m,m,m+1,...,2m\}$$

예를 들어 $DCA(4,7;6)$은 아래와 같으며, 열 간의 차에서 항상 3이 두 번 나타나게 됨을 알 수 있다.

$$Q^T=\begin{bmatrix}0&1&2&3&4&5&0\\1&3&5&0&2&4&0\\3&0&4&1&5&2&0\\0&0&0&0&0&0&0\end{bmatrix}$$

이제부터 $DCA(4,2m+1;2m)$으로부터 새로운 $(n=6m,\alpha=2m,\rho=3)$ 분할 반복 부호를 제안한다. $DCA(4,2m+1;2m)$의 마지막 행과 마지막 열을 제외한 모든 원소를 이용하여 $0\le i\le6m-1$에 대하여 다음과 같이 설계한다.

$$N_i=\{2mj+(2m-q(j,\lfloor i/2m\rfloor)+i)\bmod 2m \mid 0\le j\le 2m-1\}$$

제안하는 분할 반복 부호가 지원하는 데이터 심볼 수는 $\theta=4m^2$이며 다음의 특징을 가진다.

성질1) $6m$개의 노드들은 $2m$개씩 3개의 그룹으로 나누어지고 각 그룹 내의 노드들은 공통된 심볼을 저장하지 않는다. 즉,

$$\bigcup_{i=0}^{2m-1}N_i=\bigcup_{i=2m}^{4m-1}N_i=\bigcup_{i=4m}^{6m-1}N_i=\{0,...,4m^2-1\}$$

성질2) 하나의 노드는 다른 그룹에 있는 $2m$개의 노드들 중 한 노드와는 공통된 심볼을 가지지 않고 또 다른 하나와는 2개의 공통된 심볼을 가지며 나머지와는 1개의 공통된 심볼을 가진다. 그 노드는 나머지 다른 그룹에 대해서도 마찬가지 성질을 가진다.

성질1)은 하나의 파일을 저장할 때 노드들을 세 개의 그룹으로 나누기 적절하며, 하나의 그룹에 속한 노드들만 접속해도 전체 파일을 읽을 수 있는 장점이 있다. 성질2)는 기존 분할 반복 부호와는 달리 제안하는 부호가 처음으로 가지게 되는 특징이다. $k\le\alpha$에 대

하여 파일 크기가 기존 부호에 비하여 작을 수 있는데 파일 크기는 다음 장에서 모의실험을 통하여 보이기로 한다. 성질2)가 효율적으로 적용될 수 있는 예는 다음과 같다.

분산 저장 시스템에서 기존 분할 반복 부호가 사용되는 경우 하나의 손실 노드를 복구하기 위해 α개의 다른 노드들로부터 하나의 파일 조각을 전송받아야만 한다. 만일 분산 저장 시스템에서 링크 대역폭에 심한 불균형이 있거나 이종 네트워크로 구성되었을 경우 손실 노드를 복구하기 위해 대역폭이 큰 링크를 통해서는 2개의 파일 조각을 전송받고 대역폭이 매우 작은 링크를 통해서는 파일 조각을 전송받지 않는 것이 더 효율적이다. 제안하는 분할 반복 부호는 이러한 상황에 적합할 수 있는 성질을 가진다.

또한 분산 저장 시스템에서 각 노드들의 데이터 무결성을 확인하기 위한 절차가 필요한 경우 각 노드마다 메시지 인증 데이터를 생성하여 저장하게 된다. 이러한 경우 메시지를 인증하기 위해서 다른 노드들을 이용하게 되는데 중복 되는 데이터를 통해서 메시지 인증이 가능하다. 제안하는 분할 반복 부호는 각 노드에 대하여 2개의 파일 조각이 중복되는 다른 노드들이 2개씩 존재하므로 강력한 메시지 인증 알고리즘을 구현하는데 적합하다.

Ⅳ. 기존 부호와의 비교

비교를 위해 차 행렬로부터 생성된 분할 반복 부호를 제시한다. 이 부호는 상호 직교 라틴 방진으로부터 설계된 부호[4]와 동일하며, 횡단 디자인 (transversal design)으로부터 설계된 부호와도 등가이다. 이 부호는 제안하는 부호의 성질1)을 만족하지만 성질2)는 만족하지 않는다. 이 부호의 $M(k)$는 작은 k에 대해 최적임이 알려져 있으며, 모의실험을 통해 제안하는 부호와 $M(k)$를 비교한다.

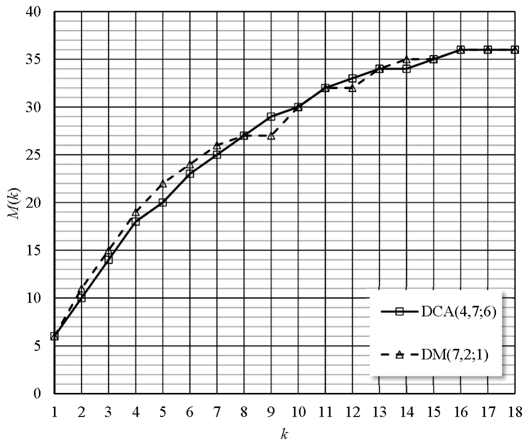

그림 1. $DCA(4,7;6)$와 $DM(7,2;1)$으로부터 생성된 $(18,6,3)$ 분할 반복 부호들의 k에 따른 파일 크기 비교
Fig. 1. File size comparison of FR codes constructed from $DCA(4,7;6)$ and $DM(7,2;1)$ with respect to k

제안하는 부호로서 $DCA(4,7;6)$으로부터 $(18,6,3)$ 분할 반복 부호를 설계하였다. 비교를 위해 $DM(7,2;1)$을 이용하여 동일한 파라미터의 부호를 설계하였다. 그림 1에서 볼 수 있듯이 작은 k에 대하여 DM 부호가 더 큰 파일 크기를 지원하지만 $k \geq 8$에 대하여 제안하는 부호가 지원하는 파일 크기가 DM 대비 동일하거나 약간 큰 경향이 있다. 노드들이 2개의 심볼을 공통으로 가질 수 있도록 허용함에 따라 작은 $M(k)$가 예상되었으나 거의 대등한 값을 가짐을 확인하였다. 참고로 복잡도를 고려해볼 때 실제 시스템에서 외각 부호를 사용하지 않거나 고부호율의 외각 부호를 사용하게 되므로 큰 k에 대해 좋은 $M(k)$를 가지는 것은 중요하다.

Ⅴ. 결 론

본 논문에서 $DCA(4,2m+1;2m)$를 이용하여 $(6m,2m,3)$ 분할 반복 부호를 설계하는 방법을 제안하였다. 제안하는 부호는 기존 부호와는 달리 2개의 노드들이 최대 2개의 파일 조각을 공유할 수 있게 허용하였다. 이러한 성질은 이종 네트워크로 구성된 분산 저장 시스템 및 데이터 무결성을 위한 메시지 인증 알고리즘 구현에서 장점을 보일 수 있다. 모의실험을 통하여 제안하는 부호가 기존 차 행렬로부터 설계된 부호와 비슷한 파일 크기를 지원함을 보였다.

References

[1] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. IEEE Symp. Mass Storage Syst. Technol.*, pp. 1-5, Incline Village, NV, USA, May 2010.

[2] O. Olmez and A. Ramamoorthy, "Fractional repetition codes with flexible repair from combinatorial designs," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1565-1591, Apr. 2016.

[3] H. Park and C.-S. Kim, "Distributed storage

codes with multiple replication degrees using relative difference families," *J. KICS*, vol. 41, no. 12, pp. 1768-1770, Dec. 2016.

[4] D. S. Papailiopoulos and A. G. Dimakis, "Locally repairable codes," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5843-5855, Oct. 2014.

[5] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539-4551, Sept. 2010.

[6] F. Demirkale, D. Donovan, J. Hall, A. Khodkar, and A. Rao, "Difference covering arrays and pseudo-orthogonal Latin squares," *Graphs and Combinatorics*, vol. 32, pp. 1353-1374, 2016.