논문 16-41-12-01 The Journal of Korean Institute of Communications and Information Sciences '16-12 Vol.41 No.12
https://doi.org/10.7840/kics.2016.41.12.1671

두 개의 다른 부분접속수 요건을 가진 부분접속 복구 부호

김 건 우◆, 이 정 우°

Locally Repairable Codes with Two Different Locality Requirements

Geonu Kim◆, Jungwoo Lee°

요 약

부분접속 복구 부호(Locally Repairable Code)는 분산 저장 시스템(Distributed Storage System)의 효율적인 노드 복구(repair)를 위한 부호로서, 부분접속수(locality), 즉 복구 과정에서 사용되는 노드의 개수를 작게 함으로써 복구의 효율성을 높이는 것을 목적으로 한다. 본 논문에서는 각 노드의 부분접속수가 서로 다른 값으로 규정되는 상황을 다룬다. 다중 부분접속수에 대한 기존의 연구 결과를 (r,δ)-부분접속수의 경우로 확장하여, 서로 다른 두 부분접속수로 규정되는 부호의 최소 거리 상계 및 이를 달성하는 최적 부호의 설계를 제시한다. 제안되는 상계는 기존의 연구와 달리 다중 부분접속수의 개수가 두 개로 제한되지만, 부호의 부분접속수가 정확하게 주어지지 않고 상한으로만 주어지는 보다 일반적인 경우에 직접 적용 가능하다.

Key Words : Distributed Storage, Locality, LRC, Singleton Bound, Gabidulin Code

ABSTRACT

Locally repairable codes (LRCs) constitute an important class of codes for distributed storage, where repair efficiency is a key metric of system performance. In LRCs, efficient repair is achieved by small locality—number of nodes participating in the repair process. In this paper, we focus on situations where different locality is required for different nodes. We present a non-trivial extension of the recent results on multiple (or unequal) localities to the (r,δ)-locality case. A new Singleton-type minimum distance upper bound is derived and an optimal code construction is provided. While the result is limited to the case of only two different localities, it should be noted that it can be directly applied to the more general case where the localities are specified not exactly but by upper limits.

Ⅰ. 서 론

빅 데이터(Big Data) 시대를 맞이하여 많은 양의 데이터를 안정적이고 효율적으로 저장 및 관리하는 것이 중요해지고 있으며, 기존 데이터 저장 방식의 한계로 분산 저장 시스템(Distributed Storage System)

※ 본 연구는 한국연구재단 중견연구자지원사업(NRF-2015R1A2A1A15052493), 미래창조과학부 정보통신·방송연구개발사업(IITP-2016-B0717-16-0023), 산업통상자원부 산업기술혁신사업(10051928), 방위사업청 국방생체모방자율로봇특화센터(UD130070ID), BK21플러스 창의정보기술 인재양성사업단 및 뉴미디어통신공동연구소 지원으로 수행되었습니다.

◆ First Author : Department of Electrical and Computer Engineering, INMAC, Seoul National University, bdkim@wspl.snu.ac.kr, 학생회원

° Corresponding Author : Department of Electrical and Computer Engineering, INMAC, Seoul National University, junglee@snu.ac.kr, 종신회원

논문번호 : KICS2016-09-260, Received September 18, 2016; Revised November 10, 2016; Accepted November 10, 2016

의 활용이 크게 주목받고 있다[1-3]. 분산 저장 시스템의 중요한 장점은, 데이터를 물리적으로 분리된 여러 노드(node)에 분산 저장함으로써, 저장 장치의 고장, 네트워크 불안정 등의 이유로 발생하는 순간적 혹은 영구적 저장 데이터 소실(erasure)의 범위를 국부적으로 제한할 수 있다는 것이다. 따라서 적절한 부호화를 통하여 전체 저장 데이터의 안정성과 신뢰성을 크게 향상 시키는 것이 가능하다. 이러한 데이터의 부호화를 위해 간단한 반복 부호(repetition code)가 많이 사용되고 있으나, 낮은 저장 효율 문제를 극복하기 위해 일반적인 소실 부호(erasure code)의 적용이 중요하다[2]. RS(Reed Solomon) 부호를 비롯한 MDS (Maximum Distance Seperable) 부호들은 일정한 소실 복호(decoding) 능력 하에서 저장 효율을 최대화한다는 측면에서 분산 저장 시스템에 일견 매우 적합해 보이지만, 소실 노드 복구(repair)에 있어 중대한 문제를 보인다. 분산 저장 시스템에 있어 노드 소실이 누적되면, 당장 복호 가능성에 문제가 없더라도 데이터의 안정성이 점점 훼손되기 때문에 적절한 시점에서 부호의 소실된 노드를 복구하는 것이 필수적이며, 이러한 복구 과정의 효율성이 전체 시스템 성능의 중요한 척도가 된다[3]. 반복 부호에서는 단순히 소실 노드와 데이터를 공유하는 하나의 노드로부터 내용물을 복사하여 복구를 수행할 수 있는데 반해, MDS 부호의 경우에는 단 하나의 소실 노드를 복구하기 위해서도 일반적으로 전체 부호를 복호하는 과정을 거쳐야 한다. 이를 위해 부호 차원(dimension) 만큼의 데이터를 네트워크로부터 읽어 들여야 하는데, 이러한 복구에 수반되는 네트워크 사용량이 전체 네트워크를 포화시켜버릴 수 있을 정도로 비효율적임이 알려져 있다[2].

위의 논의와 같이 반복 부호와 MDS 부호는 저장 및 복구 효율에서 있어 서로 다른 양 극단의 특성을 가지고 있어, 이들의 장단점을 적절히 조합한 새로운 부호의 필요성 하에 분산 저장 시스템을 위한 부호에 대한 연구가 활발히 이루어져왔으며, 대표적으로 복구 대역폭(repair bandwidth)에 초점을 맞춘 재생 부호(Regenerating Code)[3,4]와 부분접속수(locality), 즉 소실 노드의 복구를 위해 접속해야하는 노드의 개수에 초점을 둔 부분접속 복구 부호(locally repairable code)[2,5,6]를 들 수 있다. 이중 부분접속 복구 부호는 소실 노드의 복구를 위해 필요한 노드들, 즉 복구 집합(repair set)을 네트워크의 지리적 분포 혹은 계층적 구조를 고려하여 접속이 용이한 노드들에 대응시킴으로써 보다 실질적인 복구 효율을 높일 수 있다는 장점을 지닌다[2].

선형(linear) 부분접속 복구 부호에서 어떤 노드가 특정 부분접속수 요건(requirement) r을 만족시키기 위해서는 해당 노드와 복구 집합으로 천공(puncture)된 부호의 최소 거리가 2 이상이어야 하는데, 이를 δ 이상이 되도록 부분접속수 개념을 확장한 것을 (r,δ)-부분접속수라고 한다[7,8]. 이러한 (r,δ)-부분접속 복구 부호에서는 다중 노드 소실 상황에서 보다 효과적인 복구가 가능하다. 가령, 전통적인 최적(optimal) 부분접속 복구 부호에서는 2개 이상의 노드가 소실될 경우 일반적으로 노드 복구 작업의 수행이 불가능하고 비효율적인 전체 부호의 복호를 수행해야 하는데 반해, (r,δ)-부분접속 복구 부호에서는 최소한 두 번의 단일 노드 복구 작업을 통해 원래의 부호를 복원하거나 단일 노드 복구 작업을 위한 네트워크 사용 한 번으로 두 소실 노드의 복구를 수행할 수 있다.

한편, 부분접속 복구 부호 최소 거리의 이론적 한계 및 이를 달성하는 최적 부호의 설계 문제는, 모든 노드들이 하나의 부분접속수 요건을 만족시키는 조건하에서 규명되었는데[5], 최근 각 노드의 부분접속수가 다르게 규정되는 경우에 대한 연구가 이루어졌다[9,10]. 이러한 연구 결과는 분산 저장 시스템 또는 데이터에 어떤 비균질성이 존재할 경우에 적용할 수 있다는 점에서 유용하다. 가령 몇몇 노드에 저장된 데이터가 핫 데이터(hot data)에 해당되는 경우, 이들 노드의 부분접속수 요건을 더 작게 함으로써 핫데이터의 복구 시간을 줄일 수 있고, 또한 다중 가용성(availability)을 고려한 시나리오에서 핫데이터 다운로드 속도를 향상시킬 수 있다. 이들 연구에서는 최소 거리 상계 및 최적 부호의 설계 문제를 모두 다루었으나, 각 노드의 부분접속수가 정확히 주어져야 한다는 한계가 있다. 기존의 문제 설정에서는 부분접속수가 요건, 즉 어떤 값 이하로 규정되는데, 부분접속수는 작을수록 유리하다는 점에서 기존과 같이 이를 포괄하여 문제를 다룰 필요가 있다.

본 논문에서는 기존의 연구[9,10]를 두 가지 측면에서 확장한 결과를 소개한다. 즉, 확장된 (r,δ)-부분접속수 개념을 사용하며, (r,δ)-부분접속수가 요건으로 규정되는 경우를 다룬다. 이러한 문제 설정 하에서 최소 거리의 상계를 규명하고, 이를 달성하는 최적 부호를 제시함으로써 상계가 엄격(tight)함을 보인다.

본 논문의 연구 결과는 부분적으로 학술대회[16]에 발표된 바 있다.

Ⅱ. 배경 지식

2.1 표기법

본 논문에서는 아래와 같은 표기법을 사용한다.

1) 정수 i에 대해 $[i]=\{1,...,i\}$이다.
2) 길이가 n인 벡터 $\boldsymbol{v}$에 대해 $\boldsymbol{v}=(v_1,...,v_n)$로 나타낸다.
3) 크기가 $k\times n$인 행렬 G에 대해 $G=(g_{i,j})_{i\in[k],j\in[n]}$으로 표기한다.
4) 집합 S_1, S_2에 대해 $S_1\sqcup S_2$는 두 집합의 합집합이며, 동시에 두 집합이 서로소(disjoint)임을 함의한다.
5) 길이가 n이고 생성 행렬(generator matrix)이 G인 부호 C와 심볼(symbol) 지표(index) 집합 $T\subset[n]$에 대해, C_T는 T로 천공된 천공 부호이며 G_T는 그에 해당되는 생성 행렬이다. 또한, $\mathrm{rank}(T)$는 G_T의 계수(rank)를 의미한다.

2.2 부호 최소 거리

부호의 최소 거리에 대해 아래의 보조 정리가 성립함이 잘 알려져 있다[8].

보조정리 1. $[n,k,d]$ 선형 부호의 심볼 지표 집합 $T\subset[n]$에 대해 $\mathrm{rank}(T)\le k-1$ 일 경우, 최소 거리 d는 다음의 상계를 만족한다.

$$d\le n-|T| \tag{1}$$

이때, $|T|$ 가 최대이면 등호가 성립한다.

보조정리 1로부터 아래의 보다 유용한 보조정리를 얻을 수 있다[11]. 본 논문에서는 이를 이용하여 최소 거리 상계를 증명한다.

보조정리 2. $[n,k,d]$ 선형 부호 및 $\mathrm{rank}(T)\le k-1$ 인 심볼 지표 집합 $T\subset[n]$에 대해 잉여(redundant) 심볼의 개수를 γ, 즉 $\gamma=|T|-\mathrm{rank}(T)$ 라고 하면, 최소 거리 d는 다음의 상계를 만족한다.

$$d\le n-k+1-\gamma \tag{2}$$

증명. 집합 T를 확대하여 $\mathrm{rank}(T')=k-1$ 인 집합 T'을 만들 수 있다. T'에서 γ개의 잉여 심볼을 제외한 집합 T''에 대해 $\mathrm{rank}(T'')=k-1$이므로 $|T''|\ge k-1$ 이다. 따라서 T'에 보조정리 1을 적용하면 다음과 같다.

$$\begin{aligned} d &\le n-|T'|=n-|T''|-\gamma \\ &\le n-k+1-\gamma \end{aligned} \tag{3}$$

보조정리 1에 대한 따름정리로서 다음의 보조정리는 자명하며, 이를 이용해 설계한 부호의 최소 거리를 분석할 수 있다.

보조정리 3. $[n,k,d]$ 선형 부호에서 $|T|=\tau$ 인 임의의 심볼 지표 집합 $T\subset[n]$에 대해 $\mathrm{rank}(T)=k$가 만족할 경우, 최소 거리 d는 다음의 하계를 만족한다.

$$d\ge n-\tau+1 \tag{4}$$

비고 1. 조건 $\mathrm{rank}(T)=k$ 및 $\mathrm{rank}(T)\le k-1$는 각각 T에 해당되는 심볼로 부호를 복호할 수 있음과 없음을 등가적으로 의미한다.

2.3 (r,δ)-부분접속수

본 논문에서는 혼선을 방지하기 위해서 (r,δ)-부분접속수[7] 대신 $\mathrm{Loc}_\delta(\cdot)$ 라는 표기를 사용하며, 이의 명확한 정의는 아래와 같다.

정의 1. $[n,k,d]$ 선형 부호 C에서, $i\in[n]$에 대해 i번째 심볼을 포함하여 길이가 $r+\delta-1$, 최소 거리가 δ인 천공 부호가 존재한다고 하자. 즉, 어떤 집합 $S_i\subset[n]$와 정수 r이 존재하여 아래 조건을 만족한다.

1) $i\in S_i$.
2) $|S_i|=r+\delta-1$.
3) $d(C_{S_i})=\delta$.

이때, r이 위 조건을 만족하는 최소일 때 i번째 심볼에 대하여 $\mathrm{Loc}_\delta(i)=r$이다. 또한 부호 C에 대하여 $\mathrm{Loc}_\delta(C)=\max\{\mathrm{Loc}_\delta(i)|i\in[n]\}$로 정의한다.

비고 2. 싱글톤 한계식(Singleton bound)에 의해 $\mathrm{rank}(S_i)\le r$이다.

$2\le\delta\le d$인 조건에서 $\mathrm{Loc}_\delta(i)$ 는 잘 정의된다. 최소 거리가 d인 부호 C에서 i를 제외한 임의의 심볼을 하나씩 천공해나가면 매번 최소 거리는 최대 1만큼 감소하는데, 최종적으로 최소 거리는 1이 되어야하

므로 중간 과정에서 $d(C_{S_i})=\delta$인 S_i의 존재가 보장되기 때문이다. 그중에서 $|S_i|$ 가 최소일 때, $\mathrm{Loc}_\delta(i)=|S_i|-\delta-1$ 이다.

정의 2. 선형 부호 C에서 $i\in[n]$에 대해 $\mathrm{Loc}_\delta(i)\le r$ 인 경우 i번째 심볼이 (r,δ)-부분접속수 요건을 만족한다고 한다. 또한 $\mathrm{Loc}_\delta(C)\le r$ 인 경우 부호 C가 (r,δ)-부분접속수 요건을 만족한다고 정의한다.

정의 2에서 (r,δ)-부분접속수 요건을 만족한다는 것은 기존 논문[7]에서 (r,δ)-부분접속수를 갖는다는 표현과 등가적으로 같음을 보일 수 있다.

(r,δ)-부분접속수 요건을 만족하는 $[n,k,d]$ 선형 부호 C의 최소 거리는 아래와 같은 상계를 만족한다[7].

$$d\le n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1) \tag{5}$$

또한 위 상계의 등호 조건을 만족하는 여러 최적 부호들이 존재함이 알려져 있다[7,8,12,13].

2.4 Gabidulin 부호

본 논문에서 제시하는 최적 부호는 Gabidulin 부호를 기반으로 한다[10,12-14]. 본 항에서는 Gabidulin 부호 및 이를 응용한 중요 성질에 대해서 다룬다.

확대체(extension field) F_{q^t}는 기저체(base field) F_q에 대해 벡터 공간(vector space) 구조를 갖기 때문에, 각 원소를 F_q 상에서 길이 t인 벡터로 나타낼 수 있다. 따라서 임의의 벡터 $\boldsymbol{v}\in F_{q^t}^n$는 등가적인 행렬 $V\in F_q^{t\times n}$으로 표현 할 수 있고, 벡터 $\boldsymbol{v}$의 계수를 $\mathrm{rank}(\boldsymbol{v})=\mathrm{rank}(V)$로 정의할 수 있다. 또한 벡터 $\boldsymbol{u},\boldsymbol{v}\in F_{q^t}^n$에 대하여 아래와 같이 계수 거리(rank distance)를 정의할 수 있다.

$$d_R(\boldsymbol{u},\boldsymbol{v})=\mathrm{rank}(\boldsymbol{u}-\boldsymbol{v})=\mathrm{rank}(U-V) \tag{6}$$

해밍 거리(Hamming distance) $d_H(\boldsymbol{u},\boldsymbol{v})$에 대해서 $d_R(\boldsymbol{u},\boldsymbol{v})\le d_H(\boldsymbol{u},\boldsymbol{v})$임을 쉽게 알 수 있는데, 따라서 부호의 최소 계수 거리도 싱글톤 한계식을 만족하며, 이의 등호 조건을 만족시키는 부호를 MRD (Maximum Rank Distance) 부호라고 부른다.

Gabidulin 부호는 대표적인 MRD 부호로 RS(Reed Solomon) 부호를 비롯한 여러 대수적 부호(algebraic code)와 마찬가지로 다항식 값매김(evaluation)으로 부호어(codeword)를 생성한다. 구체적으로 $[n,k]_{q^t}$ Gabidulin 부호어의 생성은 아래와 같다.

1) 메시지 벡터 $\boldsymbol{a}=(a_1,...,a_k)\in F_{q^t}^k$에 대해 F_{q^t}상에서의 선형 다항식(linearized polynomial)[15]를 아래와 같이 만든다.

$$f(x)=\sum_{i=1}^{k}a_i x^{q^{i-1}} \tag{7}$$

2) F_q 상에서 선형 독립(linearly independent)인 n개의 점, $\{x_1,...,x_n\}\subset F_{q^t}$에서 다항식 $f(x)$를 값매김하여 부호어를 얻는다. 즉, 부호어 $\boldsymbol{c}=(f(x_1),...,f(x_n))$이고, 등가적으로 $\{x_1,...,x_n\}\subset F_q^t$라고 봤을 때, $\mathrm{rank}(\{x_1,...,x_n\})=n$이다.

선형 다항식은 F_{q^t}에서 F_{q^t}로 가는 F_q-선형 변환(F_q-linear transformation)으로, 임의의 $a,b\in F_q$와 $x,y\in F_{q^t}$에 대해 아래의 관계가 성립한다[15].

$$f(ax+by)=af(x)+bf(y) \tag{8}$$

Gabidulin 부호가 MDS인 이유는 선형 독립인 임의의 점 k개에서 값매김을 알면, 수식 (8)을 통해 q^k의 서로 다른 점에서 값매김을 구할 수 있고, 보간법(interpolation)을 통해 차수가 q^{k-1} 이하인 다항식 $f(x)$를 구할 수 있기 때문이다. 더군다나 k개의 값매김 점이 부호어 생성에 사용된 값매김 점과 달라도 되는데, 본 논문에서는 부호에 부분접속 복구 성질을 부여하기 위해 Gabidulin 부호어를 분할하여 각 부분을 MDS 부호화시킴으로써 그와 같은 상황이 발생한다. 이러한 부호에서 복호 가능성의 판별은, 원래 Gabidulin 부호에서와 마찬가지로 소실되지 않고 남아 있는 값매김 점들의 잔여 계수(remaining rank)를 F_q 상에서 구하여 k와 비교함으로써 가능하다. 이때, 심볼 소실에 의한 값매김 점 계수의 소실을 계수 소실(rank erasure)이라고 한다. 아래의 보조정리[13]는 본 논문에서 제시하는 부호의 계수 소실 혹은 잔여 계수를 분석하는데 이용된다.

보조정리 4. F_{q^t} 상에서의 선형 다항식 $f(\cdot)$를 F_q 상에서 선형 독립인 k개의 점에서 값매김하여 이를

원소로 하는 벡터 $\boldsymbol{u}$를 $[n,k]_q$ MDS 부호로 부호화한 부호어 $\boldsymbol{v}$에서 각 심볼은 $f(\cdot)$의 값매김에 해당하며, 그 값매김 점들은 원래 k개의 점들에 의해 생성되는 부분공간(subspace)에 속한다. 또한 임의의 부호어 심볼 s개에 해당되는 값매김 점들의 계수는 $\min(s,k)$이다.

증명. $\boldsymbol{u}=(f(x_1),...,f(x_k))$, $\mathrm{rank}(\{x_1,...,x_k\})=k$이고, $[n,k]_q$ MDS 부호의 생성 행렬을 $G=(g_{i,j})_{i\in[k],j\in[n]}$라고 하면 $\boldsymbol{v}=\boldsymbol{u}G$이다. 일반성을 잃지 않고, 부호어 $\boldsymbol{v}$에서 임의의 심볼 s개를 $\{v_1,...,v_s\}$로 두면 아래와 같다.

$$\begin{aligned}(v_1,...,v_s)&=\boldsymbol{u}G_{[s]}\\&=(\sum_{i=1}^{k}g_{i,1}f(x_i),...,\sum_{i=1}^{k}g_{i,s}f(x_i))\\&=(f(\sum_{i=1}^{k}g_{i,1}x_i),...,f(\sum_{i=1}^{k}g_{i,s}x_i))\end{aligned}\quad(9)$$

따라서 $j\in[s]$에 대하여 v_j는 $y_j=\sum_{i=1}^{k}g_{i,j}x_i$에서 다항식 $f(\cdot)$를 값매김한 값이다. 또한 이와 같은 새로운 값매김 점들을 아래와 같이 표현할 수 있다.

$$\begin{aligned}(y_1,...,y_s)&=(\sum_{i=1}^{k}g_{i,1}x_i,...,\sum_{i=1}^{k}g_{i,s}x_i)\\&=(x_1,...,x_k)\,G_{[s]}\end{aligned}\quad(10)$$

수식 (10)에서 $\mathrm{rank}(\{x_1,...,x_k\})=k$인 점과 MDS 생성 행렬 G의 성질을 이용하여 심볼 s개에 해당하는 값매김 점들의 계수는 아래와 같다.

$$\mathrm{rank}(\{y_1,...,y_s\})=\mathrm{rank}(G_{[s]})=\min(s,k)\quad(11)$$

Ⅲ. 두 개의 부분접속수로 규정되는 부분접속 복구 부호

본 논문의 연구 대상인 부호는 아래와 같이 정의된다.

정의 3. $N_1\sqcup N_2=[n]$인 두 집합 N_1과 N_2에 대해 $|N_1|=n_1$, $|N_2|=n_2$이고, 두 정수 r_1, r_2에 대해 $r_1<r_2$라고 했을 때, $[n,k,d]$ 선형 부호 C에서 N_1에 해당되는 심볼들이 (r_1,δ)-부분접속수 요건을 만족하고, N_2에 해당되는 심볼들이 (r_2,δ)-부분접속수 요건을 만족할 경우, 부호 C는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족한다고 정의한다.

(r_1,δ)-부분접속수 요건의 만족은 (r_2,δ)-부분접속수 요건의 만족을 함의한다. 따라서 (r_1,δ)-부분접속수 요건을 만족하는 부호는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하므로, 수식 (5)의 등호 조건을 만족시키는 최적 부호는 아래와 같은 최소 거리 d를 가지면서 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족한다.

$$d=d_1:=n-k+1-\left(\left\lceil\frac{k}{r_1}\right\rceil-1\right)(\delta-1)\quad(12)$$

한편 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수를 만족하는 임의의 부호는 (r_2,δ)-부분접속수 요건을 만족해야 하므로, 마찬가지로 수식 (5)에 의해 부호의 최소 거리 d는 다음의 상계를 만족해야 한다.

$$d\le d_2:=n-k+1-\left(\left\lceil\frac{k}{r_2}\right\rceil-1\right)(\delta-1)\quad(13)$$

$d_1\le d_2$이고 등호는 일반적으로 성립하지 않음에 유의했을 때, 본 연구에서 풀고자하는 문제를 다음과 같은 질문으로 표현할 수 있다. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하면서 최소 거리가 d_1보다 큰 부호는 존재하는가? 또한 이러한 부호의 최소 거리에 대해 d_2보다 엄격한 상계를 구할 수 있는가? 본 논문의 3절과 4절에서 각각 그러함을 보이도록 한다.

예 1. $((n_1=12,r_1=2),(n_2=12,r_2=4),\delta=3)$-부분접속수 요건을 만족하는 $[n=24,k=10,d]$ 선형 부호에 대해서 수식 (12)와 (13)에 의해 $d_1=7$이고, $d_2=11$이다. 즉, 기존의 알려진 (r_1,δ)-부분접속수 요건을 만족시키는 최적 부호를 통해 최소거리 $d=d_1=7$을 달성할 수 있으나 이는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건에 대한 최적 최소거리가 아닐 개연성이 크며, 부호의 최소거리는 상계 $d\le d_2=11$를 만족해야 한다. 3절에서 부호의 최소거리는 상계 $d\le 9$를 만족함을 보이고, 4절에서 상계의 등호조건을 만족하는, 즉 $d=9$인 최적 부호의

생성을 보이도록 한다. 한편, 이러한 최적 부호와 동일한 $[n=24,k=10,d=9]$ 선형 부호가 (r,δ)-부분접속수 요건을 만족한다고 하면 수식 (5)에 의하여 $r=3$임을 알 수 있다. 즉, 본 논문에서 제시하는 최적 부호를 사용함으로써, 최소 거리는 동일하게 유지하면서 일부 심볼의 부분접속수를 $r_2=4$로 높이는 대신 상대적으로 중요한 심볼의 부분접속수를 $r_1=2$로 낮추는 것이 가능하며, 이를 통해 핫데이터의 복구 효율 및 다운로드 속도를 개선할 수 있다.

Ⅳ. 최소 거리 상계

본 절에서는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 부호의 최소 거리에 대해 엄격한 상계를 구한다. 이를 위해 우선 몇 가지 보조 파라미터를 정의하여 상계를 표현하고, 추후에 보조 파라미터를 표현식에서 제거하는 단계적 접근 방법을 취한다.

정의 4. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수를 만족하는 부호 및 이에 대응되는 집합 N_1, N_2에 대해서, 아래와 같이 정의한다.

1) $\Delta N=\{i\in N_2|\mathrm{Loc}_\delta(i)\le r_1\}$.
2) $\widehat{N_1}=N_1\sqcup\Delta N$, $\widehat{N_2}-\Delta N$.
3) $\widehat{n_1}=|\widehat{N_1}|$ $\widehat{n_2}=|\widehat{N_2}|$.

추가적으로, 표현의 간략화를 위해 아래와 같이 정의한다.

정의 5. n_1에 대해 p_1, q_1, m_1은 아래와 같다.

1) p_1과 q_1는 $n_1=p_1(r_1+\delta-1)+q_1$이고 $0\le q_1\le r_1+\delta-2$인 정수.
2) $m_1=\dfrac{n_1}{r_1+\delta-1}=p_1+\dfrac{q1}{r_1+\delta-1}$.

$\widehat{n_1}$에 대해도 마찬가지로 $\widehat{p_1}$, $\widehat{q_1}$, $\widehat{m_1}$을 정의한다.

비고 3. 정의에 의해 $\widehat{n_1}\ge n_1$이고, 또한 $\widehat{m_1}\ge m_1$이다.

첫 번째 단계로 $\widehat{n_1}$과 $\mathrm{rank}(\widehat{N_1})$를 알고 있다고 가정하여 다음의 보조정리에서 최소 거리 상계를 구한다.

표 1. 알고리즘-1 (보조정리 5와 6에서 사용)
Table 1. Algorithm-1 (Used in Lemma 5 and 6)

1: Let $Q_0=\varnothing$, $l=0$
2: **while** $\mathrm{rank}(Q_l)<\mathrm{rank}(\widehat{N_1})$ **do**
3: Pick any $i\in\widehat{N_1}-Q_l$ such that $\mathrm{rank}(Q_l\cup S_i)>\mathrm{rank}(Q_l)$
4: $l=l+1$
5: $Q_l=Q_{l-1}\cup S_i$
6: **end while**
7: $l_{end}=l$

표 2. 알고리즘-2 (보조정리 5에서 사용)
Table 2. Algorithm-2 (Used in Lemma 5)

1: Let $Q_0=\widehat{N_1}$, $l=0$
2: **while** $\mathrm{rank}(Q_l)<k$ **do**
3: Pick any $i\in\widehat{N_2}-Q_l$ such that $\mathrm{rank}(Q_l\cup S_i)>\mathrm{rank}(Q_l)$
4: $l=l+1$
5: $Q_l=Q_{l-1}\cup S_i$
6: **end while**
7: $l_{end}=l$

보조정리 5. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 $[n,k,d]$ 선형 부호의 최소 거리는 다음의 상계를 만족한다.

1) $\mathrm{rank}(\widehat{N_1})=k$인 경우.

$$d\le n-k+1-\left(\left\lceil\frac{k}{r_1}\right\rceil-1\right)(\delta-1) \tag{14}$$

2) $\mathrm{rank}(\widehat{N_1})<k$인 경우.

$$d\le n-k+1-(\widehat{n_1}-\mathrm{rank}(\widehat{N_1}))-\left(\left\lceil\frac{k-\mathrm{rank}(\widehat{N_1})}{r_2}\right\rceil-1\right)(\delta-1) \tag{15}$$

증명. 각각 표-1의 알고리즘-1과 표-2의 알고리즘-2를 수행하여 $\mathrm{rank}(T)\le k-1$인 집합 $T\subset[n]$을 만들고 잉여 심볼의 개수 γ를 분석한 후 보조정리 2를 이용하여 최소 거리의 상계를 구한다.

1) 표-1의 알고리즘-1에서 S_i는 i번째 심볼이

(r_1,δ)-부분접속수 요건을 만족하게 하는 천공 부호에 해당되는 심볼 지표 집합이다. 이 때, $\widehat{N}_1$의 정의에 의해 $S_i \subset \widehat{N}_1$이므로 $Q_l \subset \widehat{N}_1$이고 $\mathrm{rank}(Q_l) \le \mathrm{rank}(\widehat{N}_1)$ 이다. 또한 2행의 조건에 의해 3행에서 항상 적절한 i를 선택할 수 있다. 알고리즘의 반복문은 l_{end}번 수행되는데, 비고 2에 의해 l_{end}에 대해서 다음이 성립한다.

$$l_{end} \ge \left\lceil \frac{\mathrm{rank}(Q_{l_{end}})}{r_1} \right\rceil = \left\lceil \frac{k}{r_1} \right\rceil \tag{16}$$

한편, l번째 반복문 수행에 있어 C_{S_i}에서 임의로 택한 $\delta-1$개의 심볼은 잉여이고, C_{S_i}에서 잉여인 심볼은 당연히 C_{Q_l}에서도 잉여이므로 다음이 성립한다.

$$|Q_l|-|Q_{l-1}| \ge \mathrm{rank}(Q_l)-\mathrm{rank}(Q_{l-1})+\delta-1 \tag{17}$$

이제 $l_0 = \left\lceil \frac{k}{r_1} \right\rceil - 1 \le l_{end}-1$에 대해 $T = Q_{l_0}$으로 놓으면, $\mathrm{rank}(T) \le k-1$ 임이 자명하다. 또한 T에서 잉여 심볼의 개수 γ는 수식 (17)을 이용하여 아래의 부등식을 만족함을 보일 수 있다.

$$\begin{aligned}\gamma &= |T|-\mathrm{rank}(T)\\ &= \sum_{l=1}^{l_0}(|Q_l|-|Q_{l-1}|)\\ &\quad -\sum_{l=1}^{l_0}(\mathrm{rank}(Q_l)-\mathrm{rank}(Q_{l-1}))\\ &\ge l_0(\delta-1)\\ &= \left(\left\lceil \frac{k}{r_1} \right\rceil -1\right)(\delta-1)\end{aligned} \tag{18}$$

따라서 보조정리 2에 의해 수식 (14)가 성립한다.

2) 표-2의 알고리즘-2를 수행하면 수식 (16)에서와 비슷하게 아래가 성립한다.

$$l_{end} \ge \left\lceil \frac{k-\mathrm{rank}(\widehat{N}_1)}{r_2} \right\rceil \tag{19}$$

$l_0 = \left\lceil \frac{k-\mathrm{rank}(\widehat{N}_1)}{r_2} \right\rceil - 1 \le l_{end}-1$에 대해 $T = Q_{l_0}$으로 놓으면, $\mathrm{rank}(T) \le k-1$ 이고, T에 해당하는 심볼 중 잉여 심볼의 개수 γ는 수식 (18)에서와 비슷하게 다음을 만족한다.

$$\begin{aligned}\gamma &= |T|-\mathrm{rank}(T)\\ &= \sum_{l=1}^{l_0}(|Q_l|-|Q_{l-1}|)+|Q_0|\\ &\quad -\sum_{l=1}^{l_0}(\mathrm{rank}(Q_l)-\mathrm{rank}(Q_{l-1}))\\ &\quad -\mathrm{rank}(Q_0)\\ &\ge |\widehat{N}_1|-\mathrm{rank}(\widehat{N}_1)+l_0(\delta-1)\\ &= \widehat{n_1}-\mathrm{rank}(\widehat{N}_1)\\ &\quad +\left(\left\lceil \frac{k-\mathrm{rank}(\widehat{N}_1)}{r_2} \right\rceil -1\right)(\delta-1)\end{aligned} \tag{20}$$

마찬가지로 보조정리 2에 의해 수식 (15)가 성립한다. 아래 보조정리에서는 $\mathrm{rank}(\widehat{N}_1)$의 상계를 구한다.

보조정리 6. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 선형 부호에서 $\mathrm{rank}(\widehat{N}_1)$은 아래의 상계를 만족한다.

1) $0 \le \widehat{q_1} \le \delta-2$인 경우.

$$\mathrm{rank}(\widehat{N}_1) \le \lfloor \widehat{m_1} \rfloor r_1 \tag{21}$$

2) $\delta-1 \le \widehat{q_1} \le r_1+\delta-2$인 경우.

$$\mathrm{rank}(\widehat{N}_1) \le \widehat{n_1} - \lceil \widehat{m_1} \rceil (\delta-1) \tag{22}$$

증명. 표-1의 알고리즘-1에서 매 반복문 수행마다 집합 Q_l에 누적되는 심볼의 개수에 대해 보조정리 5에서와 비슷하게 아래의 부등식을 얻을 수 있다.

$$\begin{aligned}\widehat{n_1} &\ge |Q_{l_{end}}|\\ &= \sum_{l=1}^{l_{end}}(|Q_l|-|Q_{l-1}|)\\ &\ge \sum_{l=1}^{l_{end}}(\mathrm{rank}(Q_l)-\mathrm{rank}(Q_{l-1}))\\ &\quad +l_{end}(\delta-1)\\ &\ge \mathrm{rank}(\widehat{N}_1)+\left\lceil \frac{\mathrm{rank}(\widehat{N}_1)}{r_1} \right\rceil (\delta-1)\end{aligned} \tag{23}$$

1) $\mathrm{rank}(\widehat{N}_1) \ge \widehat{p_1} r_1 + 1$ 임을 가정하면, 수식 (23)에 이를 대입하여 아래와 같은 모순을 보일 수

있다.

$$\begin{aligned}\widehat{n_1} &\geq \widehat{p_1}r_1+1+(\widehat{p_1}+1)(\delta-1)\\ &=\widehat{p_1}(r_1+\delta-1)+\delta\\ &>\widehat{p_1}(r_1+\delta-1)+\widehat{q_1}\\ &=\widehat{n_1}\end{aligned} \tag{24}$$

따라서 수식 (21)이 성립한다.

2) $\mathrm{rank}(\widehat{N_1}) \geq \widehat{p_1}r_1+\widehat{q_1}-(\delta-1)+1$ 임을 가정하면 $\mathrm{rank}(\widehat{N_1}) \geq \widehat{p_1}r_1+1$ 이고, 각각 수식 (23)에 대입하면 아래와 같이 모순이다.

$$\begin{aligned}\widehat{n_1} &\geq \widehat{p_1}r_1+\widehat{q_1}-(\delta-1)+1+(\widehat{p_1}+1)(\delta-1)\\ &=\widehat{p_1}(r_1+\delta-1)+\widehat{q_1}+1\\ &>\widehat{n_1}\end{aligned} \tag{25}$$

따라서
$\mathrm{rank}(\widehat{N_1}) \leq \widehat{p_1}r_1+\widehat{q_1}-(\delta-1)=\widehat{n_1}-(\widehat{p_1}+1)(\delta-1)$
이 성립하여 수식 (22)를 얻을 수 있다.

보조정리 5의 최소 거리 상계에 보조정리 6에서 구한 $\mathrm{rank}(\widehat{N_1})$의 상계를 적용하여 표현식에서 $\mathrm{rank}(\widehat{N_1})$이 제거된 최소 거리 상계를 구할 수 있다. 단, 해당 상계를 적용하기 위해서는 여전히 $\widehat{n_1}$이 주어져야 한다.

보조정리 7. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 $[n,k,d]$ 선형 부호 C에서 $\widehat{k_1^*}$는 아래와 같다고 하자.

$$\widehat{k_1^*}=\begin{cases}\lfloor \widehat{m_1} \rfloor r_1 & if\ \ 0 \leq \widehat{q_1} \leq \delta-2\\ \widehat{n_1}-\lceil \widehat{m_1} \rceil (\delta-1) & if\ \ \delta-1 \leq \widehat{q_1} \leq r_1+\delta-2\end{cases} \tag{26}$$

이때, 부호 C의 최소 거리는 다음의 상계를 만족한다.

1) $\widehat{k_1^*} \geq k$인 경우.

$$d \leq n-k+1-\left(\left\lceil \frac{k}{r_1} \right\rceil -1\right)(\delta-1) \tag{27}$$

2) $\widehat{k_1^*} < k$인 경우.

$$d \leq n-k+1-(\widehat{n_1}-\widehat{k_1^*}) -\left(\left\lceil \frac{k-\widehat{k_1^*}}{r_2} \right\rceil -1\right)(\delta-1) \tag{28}$$

증명. 보조정리 5와 6을 이용하여 아래와 같이 각각 증명할 수 있다.

1) $\mathrm{rank}(\widehat{N_1})=k$ 인 경우 보조정리 5에 의해 자명하다. $\mathrm{rank}(\widehat{N_1}) \leq k-1$ 인 경우, 보조정리 5의 수식 (15)에서 마지막 항이 음이 아닌 점과 보조정리 6을 사용하면 아래와 같다.

$$\begin{aligned}d &\leq n-k+1-(\widehat{n_1}-\mathrm{rank}(\widehat{N_1}))\\ &\leq n-k+1-(\widehat{n_1}-\widehat{k_1^*})\end{aligned} \tag{29}$$

이제 $0 \leq \widehat{q_1} \leq \delta-2$인 경우에 수식 (29)의 마지막 항에 대한 하한을 아래와 같이 구할 수 있다.

$$\begin{aligned}\widehat{n_1}-\widehat{k_1^*} &= \widehat{n_1}-\lfloor \widehat{m_1} \rfloor r_1\\ &\geq \widehat{n_1}-\widehat{m_1}r_1=\widehat{m_1}(\delta-1)\\ &\geq \lfloor \widehat{m_1} \rfloor (\delta-1)\\ &=\frac{\widehat{k_1^*}}{r_1}(\delta-1)\end{aligned} \tag{30}$$

마찬가지로 $\delta-1 \leq \widehat{q_1} \leq r_1+\delta-2$인 경우에 대하여 아래와 같다.

$$\begin{aligned}\widehat{n_1}-\widehat{k_1^*} &= \lceil \widehat{m_1} \rceil (\delta-1)\\ &\geq \frac{\widehat{m_1}r_1}{r_1}(\delta-1)\\ &=\frac{\widehat{n_1}-\widehat{m_1}(\delta-1)}{r_1}(\delta-1)\\ &\geq \frac{\widehat{n_1}-\lceil \widehat{m_1} \rceil (\delta-1)}{r_1}(\delta-1)\\ &=\frac{\widehat{k_1^*}}{r_1}(\delta-1)\end{aligned} \tag{31}$$

한편, 다음의 부등식이 성립한다.

$$\frac{\widehat{k_1^*}}{r_1} \geq \frac{k}{r_1} > \left\lceil \frac{k}{r_1} \right\rceil -1 \tag{32}$$

따라서 수식 (30), (31), (32)를 수식 (29)에 대입하여 수식 (27)을 얻을 수 있다.

2) 보조정리 6에 의하여 $\mathrm{rank}(\widehat{N}_1) \le \hat{k}_1^* \le k-1$ 이므로 보조정리 5의 수식 (15)가 성립하고, 보조정리 6을 적용하면 수식 (28)을 얻을 수 있다.

본 절에서 구하는 최종적인 최소 거리 상계를 다음의 정리에 나타내었다. 보조정리 7을 기반으로 비고 3에서 명기한 $\widehat{n_1}$에 대한 조건에 의하여 상계의 표현식에서 $\widehat{n_1}$이 제거된다. 따라서 아래의 상계는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건에서 직접적으로 주어지는 파라미터만으로 표현된다.

정리 1. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 $[n,k,d]$ 선형 부호의 최소 거리는 다음의 상계를 만족한다.

1) $\lfloor m_1 \rfloor r_1 \ge k$인 경우.

$$d \le n-k+1-\left(\left\lceil \frac{k}{r_1} \right\rceil -1 \right)(\delta-1) \tag{33}$$

2) $\lfloor m_1 \rfloor r_1 < k$인 경우.

$$d \le n-k+1 -\left(\lfloor m_1 \rfloor + \left\lceil \frac{k-\lfloor m_1 \rfloor r_1}{r_2} \right\rceil -1 \right)(\delta-1) \tag{34}$$

증명. 아래의 부등식을 통해 수식 (33)의 상한이 수식 (34)의 상한보다 작거나 같음을 알 수 있다.

$$\begin{aligned} \left\lceil \frac{k}{r_1} \right\rceil &= \lfloor m_1 \rfloor + \left\lceil \frac{k-\lfloor m_1 \rfloor r_1}{r_1} \right\rceil \\ &\ge \lfloor m_1 \rfloor + \left\lceil \frac{k-\lfloor m_1 \rfloor r_1}{r_2} \right\rceil \end{aligned} \tag{35}$$

따라서 보조정리 7에서 정의된 $\hat{k}_1^*$에 대해 $\hat{k}_1^* \ge k$인 경우, 수식 (33)과 동일한 수식 (27)이 성립하므로 증명은 자명하다. $\hat{k}_1^* \le k-1$인 경우는 수식 (28)이 성립하며, 이로부터 각 경우에 대해 아래와 같이 증명할 수 있다.

1) 수식 (28)의 마지막 항이 음이 아니므로 다음이 성립한다.

$$d \le n-k+1-(\widehat{n_1}-\hat{k}_1^*) \tag{36}$$

수식 (36)의 마지막 항은, $0 \le \hat{q}_1 \le \delta-2$인 경우와 $\delta-1 \le \hat{q}_1 \le r_1+\delta-2$인 경우에 대해 각각 아래와 같다.

$$\begin{aligned} \widehat{n_1}-\hat{k}_1^* &= \widehat{n_1}-\lfloor \widehat{m_1} \rfloor r_1 \\ &\ge \widehat{n_1}-\widehat{m_1}r_1 \\ &= \widehat{m_1}(\delta-1) \end{aligned} \tag{37}$$

$$\begin{aligned} \widehat{n_1}-\hat{k}_1^* &= \lceil \widehat{m_1} \rceil (\delta-1) \\ &\ge \widehat{m_1}(\delta-1) \end{aligned} \tag{38}$$

수식 (37)과 (38)에 비고 3과 $\lfloor m_1 \rfloor r_1 \ge k$인 조건을 적용하면 다음의 부등식을 얻을 수 있으며, 따라서 수식 (33)이 성립한다.

$$\begin{aligned} \widehat{n_1}-\hat{k}_1^* &\ge \widehat{m_1}(\delta-1) \ge \lfloor m_1 \rfloor (\delta-1) \\ &\ge \frac{k}{r_1}(\delta-1) \\ &\ge \left(\left\lceil \frac{k}{r_1} \right\rceil -1 \right)(\delta-1) \end{aligned} \tag{39}$$

2) $0 \le \hat{q}_1 \le \delta-2$인 경우, $\hat{k}_1^* = \lfloor \widehat{m_1} \rfloor r_1$이므로 아래의 부등식을 얻을 수 있다.

$$\begin{aligned} \widehat{n_1}-\hat{k}_1^* &= \widehat{n_1}-\lfloor \widehat{m_1} \rfloor r_1 \\ &\ge \widehat{n_1}-\widehat{m_1}r_1 = \widehat{m_1}(\delta-1) \\ &\ge \lfloor \widehat{m_1} \rfloor (\delta-1) \end{aligned} \tag{40}$$

각각 수식 (28)에 대입하면 다음과 같다.

$$d \le n-k+1 -\left(\lfloor \widehat{m_1} \rfloor + \left\lceil \frac{k-\lfloor \widehat{m_1} \rfloor r_1}{r_2} \right\rceil -1 \right)(\delta-1) \tag{41}$$

또한, $\delta-1 \le \hat{q}_1 \le r_1+\delta-2$인 경우에는 $\hat{k}_1^* = \widehat{n_1}-\lceil \widehat{m_1} \rceil r_1$이고, 다음이 성립한다.

$$\begin{aligned} \hat{k}_1^* &= \widehat{n_1}-\lceil \widehat{m_1} \rceil (\delta-1) \\ &\le \widehat{n_1}-\widehat{m_1}(\delta-1) = \widehat{m_1}r_1 \\ &\le \lceil \widehat{m_1} \rceil r_1 \end{aligned} \tag{42}$$

마찬가지로 각각 수식 (28)에 대입하면 아래와 같다.

$$d \le n-k+1 - \left(\lceil \widehat{m_1} \rceil + \left\lceil \frac{k-\lceil \widehat{m_1} \rceil r_1}{r_2} \right\rceil - 1 \right)(\delta-1) \quad (43)$$

마지막으로, $a \ge b$인 두 정수 a와 b에 대해 아래와 같은 부등식이 성립하므로 수식 (34)의 상한은 수식 (41)과 (43)의 상한보다 크거나 같다.

$$\begin{aligned} a+\left\lceil \frac{k-ar_1}{r_2} \right\rceil &= b+\left\lceil \frac{k-br_1}{r_2}+(a-b)\left(1-\frac{r_1}{r_2}\right) \right\rceil \\ &\ge b+\left\lceil \frac{k-br_1}{r_2} \right\rceil \end{aligned} \quad (44)$$

예 2. 예 1의 $((n_1=12,r_1=2),(n_2=12,r_2=4),\delta=3)$-부분접속수 요건을 만족하는 $[n=24,k=10,d]$ 선형 부호는 정리 1에 의하여 $d \le 9$를 만족한다. 이는 수식 (13)에 의한 기존 상계 $d \le 11$보다 엄격하다.

n_1이 특별한 조건을 만족시키는 경우, 정리 1의 상계보다 일반적으로 조금 더 엄격한 상계를 구할 수 있다. 이를 다음의 정리에 나타내었다.

정리 2. $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 $[n,k,d]$ 선형 부호 C에 대해 $q_1=0$이거나 $\delta-1 \le q_1 \le r_1+\delta-2$인 경우 C의 최소 거리는 다음의 상계를 만족한다.

1) $\lceil m_1 \rceil r_1 \ge k$인 경우.

$$d \le n-k+1-\left(\left\lceil \frac{k}{r_1} \right\rceil - 1 \right)(\delta-1) \quad (45)$$

2) $\lceil m_1 \rceil r_1 < k$인 경우.

$$d \le n-k+1 - \left(\lceil m_1 \rceil + \left\lceil \frac{k-\lceil m_1 \rceil r_1}{r_2} \right\rceil - 1 \right)(\delta-1) \quad (46)$$

증명. 정리 1의 증명에서와 마찬가지로 $\widehat{k_1^*} \ge k$인 경우의 증명은 자명하고, $\widehat{k_1^*} \le k-1$인 경우에 대해서만 증명하면 된다.

1) 정리 1에서와 동일하게 수식 (36)이 성립한다. $0 \le \widehat{q_1} \le \delta-2$인 경우, $\widehat{m_1} \ge \lceil m_1 \rceil$가 성립하므로 수식 (37)에 의해 $\widehat{n_1}-\widehat{k_1^*} \ge \widehat{m_1}(\delta-1) \ge \lceil m_1 \rceil (\delta-1)$이 다. $\delta-1 \le \widehat{q_1} \le r_1+\delta-2$인 경우에도 비고 3에 의해 마찬가지로 $\widehat{n_1}-\widehat{k_1^*}=\lceil \widehat{m_1} \rceil (\delta-1) \ge \lceil m_1 \rceil (\delta-1)$이다. 이를 수식 (36)에 대입하고, 수식 (39)에서와 비슷하게 $\lceil m_1 \rceil r_1 \ge k$ 조건을 적용하면 수식 (45)를 얻을 수 있다.
2) 정리 1의 증명 과정과 동일하게 수식 (41)과 (43)을 얻는다. 특별히 $0 \le \widehat{q_1} \le \delta-2$인 경우에 $\lfloor \widehat{m_1} \rfloor \ge \lceil m_1 \rceil$임에 유의하여, 수식 (44)와 같은 전개를 거치면 수식 (46)이 성립함을 알 수 있다.

$\delta=2$인 경우, n_1에 대한 조건은 항상 충족된다는 점에서 정리 2는 매우 유용하다. 즉, $\delta=2$인 기본적인 부분접속 복구 부호에서는 정리 1의 상계보다 엄격한 정리 2의 상계가 항상 적용된다. 이 경우 정리 2의 상계는 기존에 알려진 상계[9]와 파리미터 범위 조건의 차이를 제외하면 동일한데, 파라미터 범위를 나눠서 따져보면 두 상계가 완전히 동일함을 보일 수 있다. 단, 기존의 상계는 부분접속수가 요건이 아닌 정확한 값으로 주어진 전제하에서 증명되었다는 점에서 제한적이다.

Ⅴ. 최적 부호의 설계

앞서 구한 최소 거리 상계의 등호 조건을 달성하는 최적 부호의 제시를 통해, 상계가 엄격함을 보인다. 본 절에서 제시하는 최적 부호는 기존의 연구들[10,12-14]과 동일하게 Gabidulin 부호화한 부호어를 분할하여 각 부분을 MDS 부호화하는 두 단계를 거친다. 이를 다음의 정의에 형식적으로 나타내었다.

정의 6. $1 \le r_1 < r_2$, $\delta \ge 2$, $m_1, m_2 \ge 0$인 정수 r_1, r_2, δ, m_1, m_2에 대해 $n_1=m_1(r_1+\delta-1)$, $n_2=m_2(r_2+\delta-1)$, $n=n_1+n_2$이고, 정수 k와 t는 $k \le m_1r_1+m_2r_2 \le t$을 만족할 때, 다음과 같이 $[n,k,d]_{q^t}$ 선형 부호를 만든다.

1) F_{q^t} 상에서 k개의 메시지 심볼로 길이 $m_1r_1+m_2r_2$의 Gabidulin 부호어를 생성한다.
2) Gabidulin 부호어를 크기 r_1인 부분접속 그룹(local group) m_1개와 크기 r_2인 부분접속 그룹 m_2개로 분할한다.
3) 각 부분접속 그룹을 $[r_1+\delta-1,r_1,\delta]_q$, $[r_2+\delta-1,r_2,\delta]_q$ MDS 부호로 부호화한다.

정의 6에서 제시한 부호를 적용하기 위해서는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건 파라미터가 $n_1|(r_1+\delta-1)$ 및 $n_2|(r_2+\delta-1)$, 즉 m_1과 m_2가 정수인 조건을 만족해야 한다. 또한 생성할 부호의 파라미터 $[n,k]_{q^t}$에 대해 $k \le m_1r_1+m_2r_2$가 만족해야 하는데, 이는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하는 모든 부호가 공통적으로 만족해야 되는 조건임을 보조정리 6을 응용하여 보일 수 있다. 마지막으로 부호화에 사용되는 유한체(field) F_{q^t}에 대하여 $t \ge m_1r_1+m_2r_2$이고, q는 정의 6의 3)에서 MDS 부호가 존재할 만큼 충분히 커야 한다.

예 3. 예 1 및 2의 $((n_1=12,r_1=2),(n_2=12,r_2=4),\delta=3)$-부분접속수 요건을 만족하는 $[n=24,k=10,d]$ 선형 부호를 그림 1과 같이 생성할 수 있다. 본 예에서는 MDS 부호로 RS(Reed Solomon) 부호를 사용하였다. Gabidulin 및 RS 부호 모두 선형 부호로서, 다항식 값매김 또는 생성행렬을 사용할 수 있다.

아래의 정리에서 정의 6에서 만든 부호가 정리 1과 2에 대한 최적 부호임을 보인다.

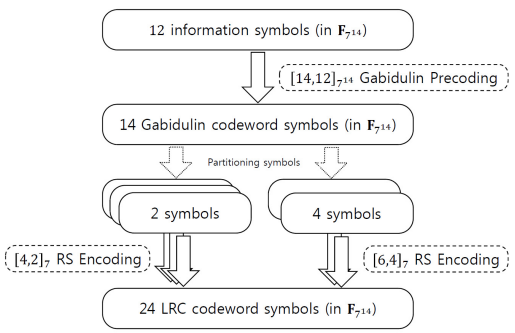

그림 1. 예 3에 따른 부호 생성
Fig. 1. Code construction according to Example 3

정리 3. 정의 6에서 생성한 부호 C는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족하고 정리 1과 2의 최소 거리 상계 등호 조건을 만족한다.

증명. 정의에 의해 각 심볼에 해당되는 천공 부호의 길이는 $r_1+\delta-1$ 혹은 $r_2+\delta-1$이고, 각각 $[r_1+\delta-1,r_1,\delta]_q$, $[r_2+\delta-1,r_2,\delta]_q$ 부호의 부분부호(subcode)로서 최소 거리가 δ 이상이므로 부호 C는 $((n_1,r_1),(n_2,r_2),\delta)$-부분접속수 요건을 만족한다.

최소 거리의 증명에 앞서, 주어진 파라미터 조건에서 $q_1=0$이므로 부호 C는 정리 2의 부호 조건에 부합한다. 따라서 정리 1과 2의 최소 거리 상계가 모두 유효하고, 더군다나 두 상계의 크기는 동일하다.

최소 거리 상계의 등호 조건 증명을 위해 보조정리 3과 비고 1을 이용하여 부호 C가 동일한 크기의 하계를 만족함을 보인다. 즉, 1) $m_1r_1 \ge k$, 2) $m_1r_1 < k$인 경우에 대해 각각 아래와 같은 크기를 갖는 임의의 심볼 집합 J로부터 항상 복호가 가능함을 보인다.

$$|J| = k+\left(\left\lceil \frac{k}{r_1} \right\rceil -1\right)(\delta-1) \tag{47}$$

$$|J| = k+\left(m_1+\left\lceil \frac{k-m_1r_1}{r_2} \right\rceil -1\right)(\delta-1) \tag{48}$$

한편 부호 C에서 임의의 소실 패턴(erasure pattern)에 대하여 잔여 심볼에 해당되는 값매김 점들이 생성하는 부분공간은 부분접속 그룹 단위로 잔여 값매김 점들이 생성하는 부분공간의 직합(direct sum)이다. 이는 이들이 속하는 각 부분접속 그룹 전체 단위의 부분공간들이 자명한 교집합을 이루기 때문이다. 따라서 임의의 소실 패턴에 해당되는 잔여 계수를 구할 때, 부분접속 그룹 별로 잔여 계수를 따로 구하여 합산하면 된다.

또한 고정된 소실 심볼 개수 하에서 잔여 심볼들이 최대한 작은 크기의 부분접속 그룹에 위치하는 경우에 계수 소실이 최대화, 즉 잔여 계수가 최소화된다. 이는 잔여 심볼의 분포가 그러하지 않은 경우에, 임의의 잔여 심볼을 큰 크기의 부분접속 그룹에서 작은 크기의 부분접속 그룹으로 이동시키면 보조정리 4와 앞 문단의 내용에 의해 잔여 계수가 줄거나 동일하기 때문이다.

1) $m_1r_1 \ge k$인 경우. 아래 수식과 $0 \le z \le r_1-1$이 만족하도록 정수 w와 z를 정의한다.

$$k-1=wr_1+z \tag{49}$$

이를 바탕으로 크기가 아래와 같은 임의의 심볼 집합 J를 상정한다.

$$\begin{aligned}|J| &= w(r_1+\delta-1)+z+1\\ &= k+w(\delta-1)\end{aligned} \tag{50}$$

이러한 집합들 중, 계수 소실이 최대화되는 집합 J는 MDS 부호화된 크기가 $r_1+\delta-1$인 부분접속 그룹 w개의 전체 심볼들과 추가적인 $z+1$개의 심볼을 포함한다. 이러한 잔여 심볼의 분포는 수식 (49)와 조건 $m_1r_1 \geq k$에 근거한 아래의 부등식과 $z+1 \leq r_1$인 점에 의해 유효함을 알 수 있다.

$$w=\left\lfloor \frac{k-1}{r_1} \right\rfloor = \left\lceil \frac{k}{r_1} \right\rceil -1 < \frac{k}{r_1} \leq m_1 \tag{51}$$

따라서 보조정리 4와 앞서 언급한 직합 성질을 이용하여 잔여 계수를 구하면 $wr_1+z+1=k$이므로 이보다 잔여 계수가 크거나 같은 임의의 J로부터 복호가 가능함을 알 수 있고, 수식 (51)을 수식 (50)에 대입하면 수식 (47)을 얻을 수 있다.

2) $m_1r_1 < k$인 경우. 아래 수식과 $0 \leq z \leq r_2-1$이 만족하도록 정수 w와 z를 정의한다.

$$k-1-m_1r_1=wr_2+z \tag{52}$$

잔여 심볼 집합 J의 크기는 다음과 같다.

$$\begin{aligned}|J| &= n_1+w(r_2+\delta-1)+z+1\\ &= k+(m_1+w)(\delta-1)\end{aligned} \tag{53}$$

계수 소실이 최대화되는 집합 J는 MDS 부호화된 크기가 $r_1+\delta-1$인 부분접속 그룹 m_1개와 크기가 $r_2+\delta-1$인 부분접속 그룹 w개의 전체 심볼들, 그리고 크기가 $r_2+\delta-1$인 부분접속 그룹 하나에 속하는 $z+1$개의 심볼을 포함한다. 아래의 부등식과 $z+1 \leq r_2$인 점에 의해 이러한 잔여 심볼의 분포는 유효함을 알 수 있다. 아래의 부등식은 수식 (52)와 정의 6의 부호 파라미터 조건에 근거한다.

$$\begin{aligned}w &= \frac{k-m_1r_1}{r_2}-\frac{z+1}{r_2}\\ &\leq \frac{(m_1r_1+m_2r_2)-m_1r_1}{r_2}-\frac{z+1}{r_2}\\ &< m_2\end{aligned} \tag{54}$$

앞에서와 마찬가지로 잔여 계수는 $m_1r_1+wr_2+z+1=k$이므로 임의의 J로부터 복호가 가능하고, 수식 (52)에 근거한 아래의 수식을 수식 (53)에 대입하면 수식 (48)을 얻을 수 있다.

$$w=\left\lfloor \frac{k-m_1r_1-1}{r_2} \right\rfloor = \left\lceil \frac{k-m_1r_1}{r_2} \right\rceil -1 \tag{55}$$

예 4. 정리 3에 의해, 예 3에서 생성한 부호의 최소 거리는 $d=9$를 만족한다. 이는 예 1에서 보인 기존 방식의 부호 $d=7$ 보다 큰 값이다. 또한 부호가 예 2에서 구한 상계 $d \leq 9$의 등호 조건을 만족시키므로 정리 1에 대한 최적 부호임을 알 수 있다. 이는 정리 2에 대해서도 마찬가지이다.

Ⅵ. 결 론

본 논문에서는 (r,δ)-부분접속 복구 부호의 부분접속수 요건이 두 개의 심볼 그룹에 대해서 다르게 주어지는 경우에 대해 최소 거리 상계를 구하고 이의 등호 조건을 달성하는 최적 부호를 제시하였다. 추가적인 연구로, 부분접속수 요건의 개수를 임의의 개수로 확장하는 것을 생각해볼 수 있다.

References

[1] J.-C. Park, "Improving data availability by data partitioning and partial overlapping on multiple cloud storages," *J. KICS*, vol. 36, no. 12, pp. 1498-1508, Dec. 2011.

[2] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: novel erasure codes for big data," in *Proc. Int. Conf. Very Large Data Bases*, pp. 325-336, Trento, Italy, Aug. 2013.

[3] 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.

[4] J. S. Park, J.-H. Kim, K.-H. Park, and H.-Y Song, "Average repair read cost of linear repairable code ensembles," *J. KICS*, vol. 39, no. 11, pp. 723-729, Nov. 2014.

[5] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6925-6934, Nov. 2012.

[6] J.-H. Kim, M.-Y. Nam, K.-H. Park, and H.-Y Song, "Construction of $(2^k-1+k, k, 2^{k-1}+1)$ codes attaining Griesmer bound and its locality," *J. KICS*, vol. 40, no. 3, pp. 491-496, Mar. 2015.

[7] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property," in *IEEE ISIT*, pp. 2776-2780, Cambridge, MA, Jul. 2012.

[8] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with local regeneration and erasure correction," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4637-4660, Aug. 2014.

[9] A. Zeh and E. Yaakobi, "Bounds and constructions of codes with multiple localities," in *2016 IEEE ISIT*, pp. 640-644, Barcelona, Spain, Jul. 2016.

[10] S. Kadhe and A. Sprintson, "Codes with unequal locality," in *2016 IEEE ISIT*, pp. 435-439, Barcelona, Spain, Jul. 2016.

[11] M. Kuijper and D. Napp, *Erasure codes with simplex locality*, *CoRR*, vol. abs/1403.2779, 2014, [Online]. Available: http://arxiv.org/abs/1403.2779

[12] N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," in *IEEEISIT,* pp. 1819 - 1823, Istanbul, Turkey, Jul. 2013.

[13] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 212 - 236, Jan. 2014.

[14] E. M. Gabidulin, "Theory of codes with maximum rank distance," *Problems of Inf. Transmission*, vol. 21, no. 1, pp. 3-16, Jan. 1985.

[15] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*, North-Holland Publishing Company, 1977.

[16] G. Kim and J. Lee, "A study on (r,δ)-LRC with two localities," in *Proc. KICS Int. Conf. Commun.*, pp. 465-466, Jeju, Korea, Jun. 2016.

김 건 우 (Geonu Kim)

2004년 8월 : 한국과학기술원 전기 및 전자공학과 학사

2007년 2월 : 한국과학기술원 전기 및 전자공학과 석사

2013년 3월~현재 : 서울대학교 전기·정보공학부 박사과정

<관심분야> 부분접속복구 부호, 부호이론, 통신공학

이 정 우 (Jungwoo Lee)

1988년 : 서울대학교 전기공학 학사

1990년 : Princeton대학교 전기공학 석사

1994년 : Princeton대학교 전기공학 박사

2002년~현재 : 서울대학교 전기·정보공학부 교수

<관심분야> 무선통신, 분산저장시스템 부호, 머신러닝