Basics of Coding Theory and Introduction to Codes with Locality

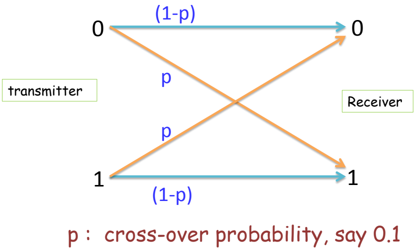
Lalitha Vadlamani
IIIT Hyderabad

Trivandrum School of Communications, Coding and Networking

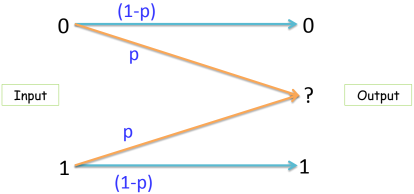
January 27-30, 2017

How Coding Theory Started?

Channel Models



(a) Binary Symmetric Channel



(b) Binary Erasure Channel

Pre-Shannon Reliable Communication

- If you send the bits as it is, the probability of error is 0.1. How to reduce it to 0.03?

  - Repeat every bit thrice, rate = 1/3

    0 → 0 0 0
    1 → 1 1 1

  - Decoding Scheme – Use majority rule

  - Probability of error – say 0 is transmitted
    - 0 0 0, 0 0 1, 0 1 0, 1 0 0 ✓
    - 0 1 1, 1 0 1, 1 1 0, 1 1 1 ✗ Prob. = 0.028

Rate of the code goes to zero if we require vanishingly small probability of error.

Shannon's 1948 Paper

Theorem

Any rate $R < C = 1 - H_2(p)$ is achievable for the binary symmetric channel with probability of error asymptotically going to zero. Conversely, if probability of error asymptotically goes to zero, then $R <= C$.

- Shannon showed achievability of the capacity of BSC using random codes
- These codes require long block length
- Not practical because of above two reasons

History of Hamming Code

- Richard Hamming worked at Bell Telephone Laboratories
- Goal was to correct errors in a error-prone punched card reader
- Hamming code can correct a single bit error
- It can detect all single bit and two bit errors

Binary Block Codes

Binary Field

- $\mathbb{F}_2 = \{0, 1\}$, addition $+$, multiplication .

+	0	1
0	0	1
1	1	0

.	0	1
0	0	0
1	0	1

- Closed and commutative w.r.t addition, multiplication
- Associative
- Additive identity exists, additive inverse exists for all elements
- Multiplicative identity exists, multiplicative inverse exists for all non-zero elements

Binary Field

- $\mathbb{F}_2 = \{0, 1\}$, addition $+$, multiplication $.$

+	0	1
0	0	1
1	1	0

.	0	1
0	0	0
1	0	1

- Closed and commutative w.r.t addition, multiplication
- Associative
- Additive identity exists, additive inverse exists for all elements
- Multiplicative identity exists, multiplicative inverse exists for all non-zero elements

Any property missing?

Binary Block Codes

Definition

A binary block code of length n is any subset of $\mathbb{F}_2^n$

- The elements of the code are called codewords
- Binary block code since the channels have binary input and binary output

Hamming Weight, Hamming Distance

Hamming Weight

The Hamming weight $w_H(\underline{x})$ of a vector $\underline{x} \in \mathbb{F}_2^n$ is the number of nonzero components in $\underline{x}$.

Hamming Distance

The Hamming distance $d_H(\underline{x}, \underline{y})$ between two vectors $\underline{x}, \underline{y} \in \mathbb{F}_2^n$ is defined as

$$d_H(\underline{x}, \underline{y}) = w_H(\underline{x} + \underline{y})$$

Properties of Hamming Distance

- (Positivity) $d_H(\underline{x}, \underline{y}) \geq 0$ with equality if $\underline{x} = \underline{y}$.
- (Symmetry) $d_H(\underline{x}, \underline{y}) = d_H(\underline{y}, \underline{x})$.
- (Triangle Inequality) $d_H(\underline{x}, \underline{z}) \leq d_H(\underline{x}, \underline{y}) + d_H(\underline{y}, \underline{z})$.

Parameters of a Code

- Block length of the code n
- Size of a code $\mathcal{C}$, which is the number of codewords in the code $|\mathcal{C}|$.
- Rate R of $\mathcal{C}$, $R = \frac{\log_2 |\mathcal{C}|}{n}$.
- Minimum distance

$$d_{\min}(\mathcal{C}) = \min\{d_H(\underline{x}, \underline{y}) | \underline{x}, \underline{y} \in \mathcal{C}, \underline{x} \neq \underline{y}\}$$

Repetition Code

$$\mathcal{C} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

- Block length $n = 7$
- Size of the code $|\mathcal{C}| = 2$
- Rate $R = \frac{1}{7}$
- $d_{\text{min}} = 7$

Single Parity Check Code

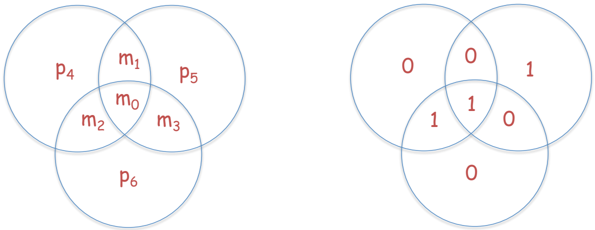
$$\mathcal{C} = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_7 \end{bmatrix} \text{such that} \sum_{i=1}^{7} x_i = 0 \right\}$$

- Block length $n = 7$
- Size of the code $|\mathcal{C}| = 2^6$
- Rate $R = \frac{6}{7}$
- $d_{\min} = 2$

$$d_H([0\ 0\ 0\ 0\ 0\ 0\ 0]^t, [1\ 1\ 0\ 0\ 0\ 0\ 0]^t)$$

Hamming Code

$$
\begin{aligned}
p_4 &= m_0 + m_1 + m_2 \\
p_5 &= m_0 + m_1 + m_3 \\
p_6 &= m_0 + m_2 + m_3
\end{aligned}
$$



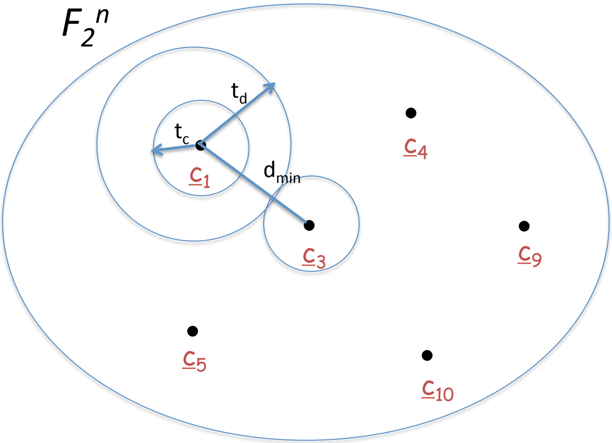
- Block length $n = 7$
- Size of the code $|\mathcal{C}| = 2^4 = 16$
- Rate $R = \frac{4}{7}$
- $d_{\min} = 3$ (Will be calculated later)

Question

- Consider Hamming code. To each codeword of the Hamming code, we add a bit which is the sum of all the bits in the codeword.
- What is the block length, size and rate of the new code?
- Make an observation about Hamming weights of the codewords in the new code.

Error Detection and Error Correction

Error Detection and Error Correction



- Any combination of $d_{\min} - 1$ errors can be detected.
- Any combination of $\lfloor \frac{d_{\min}-1}{2} \rfloor$ errors can be corrected.

(t_c, t_d) Code

Definition

A (t_c, t_d) code, $t_c \leq t_d$ is a code in which

(a) any combination of $\leq t_c$ errors can be detected and corrected and

(b) any combination of t errors, $t_c < t \leq t_d$ can be detected as an uncorrectable error

- Error detection corresponds to the case $t_c = 0$.
- Error correction corresponds to the case $t_c = t_d$

Minimum Distance Bound of (t_c, t_d) Code

Theorem

A binary block code $\mathcal{C}$ is a (t_c, t_d) code iff

$$d_{\min}(\mathcal{C}) \geq t_c + t_d + 1$$

Example

Repetition Code: $d_{\min} = 7$
$(t_c = 0, t_d = 6), (t_c = 1, t_d = 5), (t_c = 2, t_d = 4), (t_c = 3, t_d = 3)$

Simple Parity Code: $d_{\min} = 2$
$(t_c = 0, t_d = 1)$

Hamming Code: $d_{\min} = 3$
$(t_c = 0, t_d = 2), (t_c = 1, t_d = 1)$

Proof of 'If' Part

- Assume that $d_{\min} \geq t_c + t_d + 1$.
- Define for any vector $\underline{a} \in \mathbb{F}_2^n$

$$B(\underline{a}, r) = \{\underline{z} \in \mathbb{F}_2^n \mid d_H(\underline{a}, \underline{z}) \leq r\}$$

- Decoding algorithm: If $B(\underline{y}, t_c)$ contains a codeword $\underline{x}$, then declare $\underline{x}$ to be transmitted codeword. If not, declare that uncorrectable number of errors have occurred.

Proof of 'If' Part

- If one codeword lies in $B(\underline{y}, t_c)$ and another lies in $B(\underline{y}, t_d)$

$$\begin{aligned} d_H(\underline{y}, \underline{x}_1) \leq t_c, d_H(\underline{y}, \underline{x}_2) \leq t_d \\ \Rightarrow d_H(\underline{x}_1, \underline{x}_2) \leq d_H(\underline{y}, \underline{x}_1) + d_H(\underline{y}, \underline{x}_2) (\text{Triangle Inequality}) \\ \leq t_c + t_d < t_c + t_d + 1 \end{aligned}$$

- Two codewords cannot lie in $B(\underline{y}, t_c)$
- If no codeword lies in $B(\underline{y}, t_c)$, then its not a correctable error and hence declared uncorrectable

Proof of 'Only If' Part

- Suppose $d_{\min} = t_c + t_d - \ell$, $\ell \geq 0$
- There exists a pair $(\underline{x}_1, \underline{x}_2)$ in C such that

$$d_H(\underline{x}_1, \underline{x}_2) = t_c + t_d - \ell$$

- Pick a vector $\underline{y}$ such that $d_H(\underline{y}, \underline{x}_1) = t_d$ and $d_H(\underline{y}, \underline{x}_2) = t_c - \ell$,
- $\underline{y}$ can be a uncorrectable error if $\underline{x}_1$ was transmitted and can be a correctable error if $\underline{x}_2$ was transmitted. This can't be resolved.

Question

- Decoding algorithm presented above: If $B(\underline{y}, t_c)$ contains a codeword $\underline{x}$, then declare $\underline{x}$ to be transmitted codeword. If not, declare that uncorrectable number of errors have occurred. is termed as "bounded distance decoding"

- If bounded distance decoding is used, then we cannot achieve capacity. Make an intuitive guess why?

Linear Block Codes

Why Linear Codes?

- Complexity of encoding of an arbitrary block code - A lookup table of the size of the code
- Complexity of encoding of a linear code - Matrix multiplication of the message vector with a matrix of the order of logarithm of the size of the code
- Construction of codes becomes simpler since there is rich structure in the code

Vector Space, Subspace of a Vector Space

Vector Space

A vector space $(V, +, \mathbb{F}, .)$ is a set of a vectors, a field $\mathbb{F}$ of scalars and two operations: vector addition denoted by $+$ and scalar multiplication denoted by . which satisfy the following properties:

(i) Closure, commutativity, associative, additive identity and additive inverse for $(V, +)$

(ii) Closed under scalar multiplication

(iii) Multiplication with identity is the vector itself

(iv) Associativity under scalar multiplication

(v) Distributive under scalar multiplication

Subspace

A subspace of a vector space $(V, +, \mathbb{F}, .)$ is a subset W of V such that $(W, +, \mathbb{F}, .)$ is also a vector space.

Basis, Dimension of a Vector Space

Basis

A basis of a vector space $(V, +, \mathbb{F}, .)$ is a collection $\{\underline{\alpha}_1, \underline{\alpha}_2, \ldots\}$ such that

(a) the set is linearly independent.

(b) the set spans the vector space V.

Dimension

The dimension k of a *finite dimensional vector space* $(V, +, \mathbb{F}, .)$ is the number of elements in any basis for V.

Examples of Vector Spaces

For each vector space below, write a basis, dimension and identify one subspace

- $(\mathbb{R}^n, +, \mathbb{R}, .)$
- $(\mathbb{F}_2^n, +, \mathbb{F}_2, .)$
- $(\mathbb{F}[x], +, \mathbb{F}, .)$
- $(\mathbb{R}^{m\times n}, +, \mathbb{R}, .)$

Linear Block Codes

Definition

A linear code of block length n is any subspace of $\mathbb{F}_2^n$ (of the vector space $(\mathbb{F}_2^n, +, \mathbb{F}_2, .)$

Example

Hamming code is the set of all codewords such that

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \\ p_4 \\ p_5 \\ p_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

If $\underline{x}, \underline{y} \in \mathcal{C}$, then $H\underline{x} = 0, H\underline{y} = 0 \Rightarrow H(\underline{x} + \underline{y}) = 0 \Rightarrow \underline{x} + \underline{y} \in \mathcal{C}$.

Minimum Distance of a Linear Block Code

Theorem

The minimum distance $d_{\min}$ of a linear block code $\mathcal{C}$ is equal to the minimum Hamming weight $w_{\min}$ of a nonzero codeword.

Proof.

Let $\underline{c}$ have $w_H(\underline{c}) = w_{\min}$. Then,

$$d_H(\underline{c}, \underline{0}) = w_{\min} \quad \Rightarrow \quad d_{\min} \leq w_{\min}$$

Let $\underline{c}_1, \underline{c}_2 \in \mathcal{C}$ such that $d_H(\underline{c}_1, \underline{c}_2) = d_{\min}$. Then,

$$w_H(\underline{c}_1 + \underline{c}_2) = d_{\min} \quad \Rightarrow \quad w_{\min} \leq d_{\min}$$

The second step follows because $\underline{c}_1 + \underline{c}_2 \in \mathcal{C}$. □

Dimension of a Linear Code

Dimension

The dimension of a linear code $\mathcal{C}$ of block length n is its dimension as a subspace of the vector space $(\mathbb{F}_2^n, +, \mathbb{F}_2, .)$.

- An $[n, k, d]$ code denotes a block code of length n, dimension k and minimum distance d.
- An $[n, k]$ code denotes a code of block length n and dimension k

Generator Matrix of a Linear Block Code

Generator Matrix

Let $\mathcal{C}$ be an (n, k) code. Then any $k \times n$ matrix G whose rows form a basis for $\mathcal{C}$ is called a generator matrix of $\mathcal{C}$.

- Map from message vectors to codewords in terms of G

$$\underline{c}^t = \underline{m}^t G_{k \times n}$$

- A code can in general have more than one generator matrix.

Examples of Generator Matrices

- Single Parity Check Code:

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- Hamming Code:

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Systematic Generator Matrix

Definition

A generator matrix G is said to be a systematic generator matrix for an (n, k) code if it can be expressed in the form

$$G = [I_k \mid P]$$

- A code has systematic generator matrix iff first k columns of G are full rank

Why Systematic Generator Matrix?

- Map between message vector and codeword

$$\begin{aligned}\underline{c}^t &= \underline{m}^t G_{k\times n}\\ &= \underline{m}^t [I_k \mid P]\\ &= [\underline{m}^t \mid \underline{m}^t P]\end{aligned}$$

  Message symbols are explicitly present in the codewords

- Every linear code $\mathcal{C}$ is equivalent (upto permutation of coordinates) to a second linear code $\mathcal{C}'$ which has a systematic generator matrix

Question

An (n, k) binary linear code has been given. Are there ways to construct new codes from this code?

Recall that we constructed a new code from Hamming code in an earlier example.

Recap

- Binary block codes - block length, size of a code, rate of a code, minimum distance of a code
- Error Detection and error correction capability of a code
- Linear block codes - dimension of a code, minimum distance, generator matrix, systematic generator matrix

Dual Code, Parity Check Matrix

Definition

Let $\mathcal{C}$ be an (n, k) code. The dual $\mathcal{C}^\perp$ of $\mathcal{C}$ is defined by

$$\mathcal{C}^\perp = \{\underline{y} \in \mathbb{F}_2^n \mid \underline{x}^t \underline{y} = 0 \text{ for all } \underline{x} \in \mathcal{C}\}$$

Example

If $\mathcal{C}$ is the repetition code, the dual of this code is the single parity check code

Definition

A parity check matrix for a linear code $\mathcal{C}$ is any basis of the dual code $\mathcal{C}^\perp$. Rank of parity check matrix is $n - k$ by rank nullity theorem.

Example

Parity check matrix of the repetition code is

$$H = [I_6 \mid \underline{1}]$$

Bound on Minimum Distance

Singleton Bound

Theorem

Upper bound on minimum distance of a code is given by

$$d_{\min} \leq n - k + 1$$

Proof.

Let H be the parity check matrix of code $\mathcal{C}$.

$s = d_{min} - 1$ is the largest integer such that any s columns of H are linearly independent.

Since $s <= \text{rank}(H) = n - k$, the bound follows.

$\square$

Minimum Distance of Hamming Code

- $s = d_{min} - 1$ is the largest integer such that any s columns of H are linearly independent.

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- Check that $s = 2$ for parity check matrix above.

Maximum Distance Separable Codes

Definition

A code whose d_{min} achieves the Singleton bound with equality is called a Maximum Distance Separable (MDS) code.

Example

$(n, 1, n)$ repetition code, $d_{\text{min}} = n = n - 1 + 1$.

$(n, (n-1), 2)$ single parity check code $d_{\text{min}} = 2 = n - (n-1) + 1$.

Other MDS Codes over Binary Field

- Repetition code and the single parity check code are the only possible families of binary MDS codes
- Proof: Try constructing the parity check matrix of a binary $(n, n-2), n \geq 4$ code

Reed Solomon Codes

Non-binary Alphabets

- Binary codes were designed for binary symmetric channel which introduces i.i.d. errors in the bits
- Suppose there are burst errors in the symbols, then we can think of non-binary alphabets (alphabets of size 2^m) where we treat m bits as one symbol.
- To define codes over alphabets of size 2^m, finite field arithmetic is needed.

Finite Fields

- We have already seen $\mathbb{F}_2$. The structure of $\mathbb{F}_p$ is similar.
- To imagine the structure of $\mathbb{F}_{2^m}$, consider the following analogy:
- The set of complex numbers $\mathbb{C}$ is obtained from the set of real numbers $\mathbb{R}$ by considering the polynomials $\mathbb{R}[x]$ and taking modulo a polynomial x^2+1. x^2+1 is irreducible over $\mathbb{R}$.
- The finite field $\mathbb{F}_{2^4}$ is obtained from the field $\mathbb{F}_2$ by considering the polynomials $\mathbb{F}[x]$ and taking modulo a polynomial x^4+x+1. x^4+x+1 is irreducible over $\mathbb{F}_2$.

Theorem

Let $\mathbb{F}$ be a finite field and $\mathbb{F}[x]$ denote the ring of polynomials with coefficients from $\mathbb{F}$. Consider $f[x] \in \mathbb{F}_2[x]$ be a polynomial of degree k, then $f[x]$ has at most k roots in $\mathbb{F}$.

Reed-Solomon Codes

- Let $\underline{m}^t = [m_0, \ldots m_{k-1}]$ be message vector over finite field $\mathbb{F}_q$
- Form the polynomial $f(x) = \sum_{i=0}^{k-1} m_i x^i$
- Pick $\alpha_i \in \mathbb{F}_q, 1 \leq i \leq n$ all distinct, assuming $q \geq n$.
- Codeword corresponding to $\underline{m}^t$ is $\underline{c}^t = [f(\alpha_1), \ldots, f(\alpha_n)]$.

Minimum Distance of Reed-Solomon Codes

- This code can tolerate $n - k$ erasures ($k - 1$ degree polynomial can be uniquely determined by evaluations at k points). This implies that $d_{\min} \geq n - k + 1$.
- By Singleton bound, $d_{\min} \leq n - k + 1$.
- Thus, minimum distance of RS code is $n - k + 1$
- Reed Solomon codes are MDS codes

Generator Matrix of Reed-Solomon Code

$$f(\alpha_j) = \sum_{i=0}^{k-1} m_i \alpha_j^i$$

$$\underline{c}^t = [f(\alpha_1), \ldots, f(\alpha_n)] = \underline{m}^t \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \ldots & \alpha_n^{k-1} \end{bmatrix}$$

- The above generator matrix is called Vandermonde matrix
- Any $k \times k$ submatrix of the above matrix is full rank.

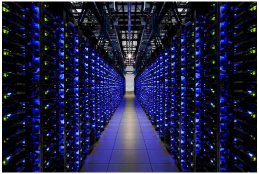
$$\underline{y}^t \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_k \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \ldots & \alpha_k^{k-1} \end{bmatrix} = 0 \Rightarrow \underline{y} = \underline{0}$$

Applications of Reed-Solomon Codes

- CD/DVD
- RAID systems
- Bar Code Scanning

Codes with Locality

Distributed Storage Systems



Servers in a Google data center

- DSS with hundreds of nodes
- Petabytes of data adding to data center everyday
- Node failures (modeled as erasures) are a norm

High Level Objective

- Data should not be lost at any cost
- Efficient utilization of storage space and network resources

Some parameters of interest

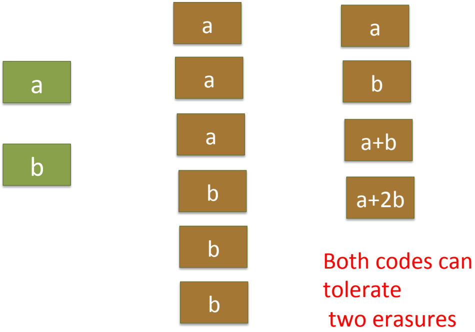
- High Resiliency
- Low Storage Overhead
- Efficient node repair

Pic courtesy:

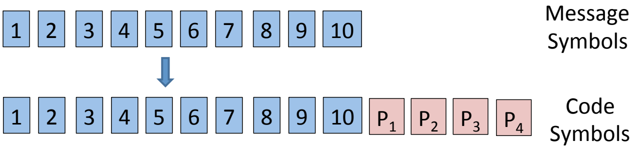
`http://webodysseum.com/technologyscience/visit-the-googles-data-centers/`

Conventional Distributed Storage Systems

- Replication and Reed Solomon codes are commonly used
- For same erasure tolerance, Reed Solomon codes have less storage overhead than replication

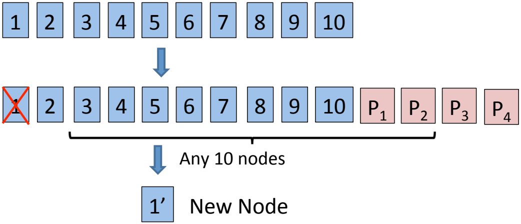


Reed-Solomon Codes in DSS



- $[14, 10]$ Reed-Solomon code - storage overhead $1.4x$
- Can recover data by connecting to any 10 nodes
- Used in Facebook for "cold" storage

RS Codes inefficient for Node Repair



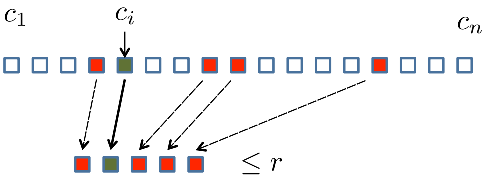
For node repair, the known strategy is:

- Connect to any 10 nodes
- Download 10 code symbols
- Reconstruct entire data file and then reconstruct data stored in the node

Locality Parameter

Setting:

- Linear code $\mathcal{C}$ with parameters $[n, k, d_{\min}]$
- Code symbol c_i has locality r



- Consider a code in systematic form. The code is said to have information locality r if all the message symbols in the code have locality r

Storage vs Repair Locality Tradeoff

Theorem

For $[n, k, d_{\mathrm{min}}]$ code with information locality r

$$d_{min} \leq \underbrace{n-k+1}_{\substack{\text{Singleton} \\ \text{bound}}} - \underbrace{\left(\left\lceil \frac{k}{r} \right\rceil - 1\right)}_{\substack{\text{Term due to} \\ \text{locality constraint}}}$$

P. Gopalan, C. Huang, S. Yekhanin, H. Simitci, "On the Locality of Codeword Symbols," *IEEE Trans. Inform. Th.*, Nov. 2012. 2014 ComSoc/IT Joint Paper Award

Main Lemma

Lemma

Any $n - d_{\min} + 1$ columns of generator matrix G have rank k. Thus, if a set of T columns of G has rank $\leq k - 1$, then

$$|T| \leq n - d_{\min}$$

Proof.

- Let S be set of $n - d_{\min} + 1$ columns of G which have rank say $k - 1$. G_1 is submatrix of G corresponding to columns in S.
- G_1 can be row reduced to give all zero vector in one row.
- If we do the same row reduction to G, we will end up in a vector (that corresponding to all zero vector in G_1), that has support in $\leq d_{\min} - 1$ columns. Contradicts the fact that $d_{\min}$ is the minimum distance of the code.

□

Sketch of Proof of the Theorem

- To find a upper bound on d_{min}, we will find a lower bound on the size of T by applying the locality constraint.
- There are sets of columns (of size $\leq r+1$) of generator matrix G which are linearly dependent. These sets we will call them "local groups"
- Construct T by accumulating local groups and try to reach the rank $k-1$.
- For each local group we are adding, we get one linearly dependent column which doesn't add rank.
- Since the number of local groups is at least $\frac{k}{r}$, to accumulate rank of $k-1$ by adding local groups, we will need a support of at least $k-1+(\frac{k}{r}-1)$. Thus, $|T| \geq k-1+(\frac{k}{r}-1)$.

Pyramid Code Construction via Example

- Given generator matrix G of a systematic $[11, 8, 4]$ MDS code:

$$G = \begin{bmatrix} 1 & & & & g_{11} & g_{12} & g_{13} \\ & 1 & & & g_{21} & g_{22} & g_{23} \\ & & \ddots & & \vdots & \vdots & \vdots \\ & & & 1 & g_{81} & g_{82} & g_{83} \end{bmatrix}$$

- Split first parity column, and then rearrange columns:

$$G' = \left[\begin{array}{ccccc|ccccc|cc} 1 & & & & g_{11} & & & & & & g_{12} & g_{13} \\ & 1 & & & g_{21} & & & & & & g_{22} & g_{23} \\ & & 1 & & g_{31} & & & & & & g_{32} & g_{33} \\ & & & 1 & g_{41} & & & & & & g_{42} & g_{43} \\ \hline & & & & & 1 & & & & g_{51} & g_{52} & g_{53} \\ & & & & & & 1 & & & g_{61} & g_{62} & g_{63} \\ & & & & & & & 1 & & g_{71} & g_{72} & g_{73} \\ & & & & & & & & 1 & g_{81} & g_{82} & g_{83} \end{array}\right]$$

Optimality of Pyramid Code Construction

- The new $[12, 8, ?]$ code has two $[5, 4, 2]$ local codes.
- Minimum distance of code generated by G' is at least that generated by G. Thus $d_{\min} \geq 4$.
- Applying the bound on minimum distance,

$$\begin{aligned} d_{\min} &\leq n - k - \frac{k}{r} + 2 \\ &= 12 - 8 - \frac{8}{4} + 2 = 4 \end{aligned}$$

- Thus, $d_{\min} = 4$ and the pyramid code constructed is optimal

References

- NPTEL Lectures Notes of a course on Error Correcting Codes taught by Prof. P. Vijay Kumar
- Shu Lin, Daniel J.Costello, Error Control Coding - Second Edition, Pearson Education Inc. Pearson Prentice Hall, 2004.
- W. Cary Huffman, and Vera Pless, Fundamentals of Error-Correcting Codes, Cambridge University Press, 2010.
- Ron M.Roth, Introduction to Coding Theory, Cambridge University Press, 2006.

Thanks!

Email: lalitha.v@iiit.ac.in