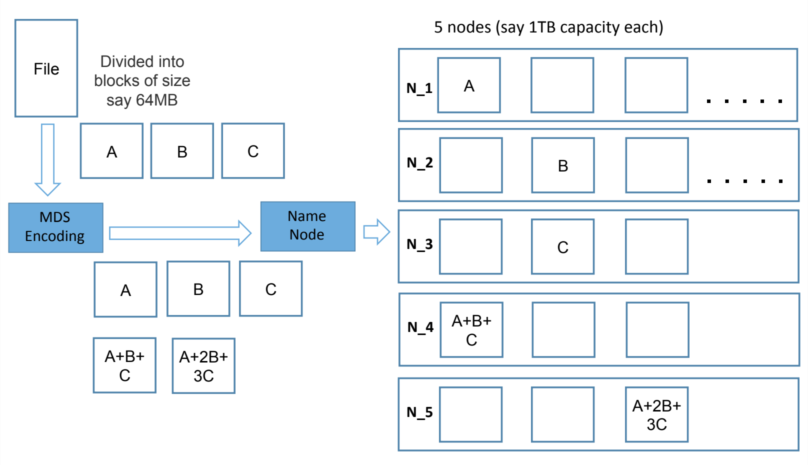
Efficient Repair of Reed-Solomon Codes and Tamo-Barg Codes

Lalitha Vadlamani
IIIT Hyderabad

CNI Seminar Series, IISc
November 26, 2024

Distributed Storage System with Erasure Coding

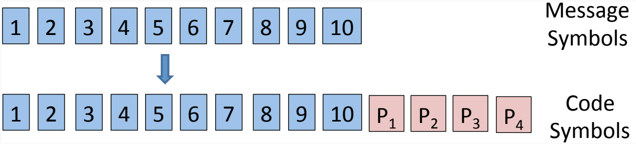


- The blocks obtained after encoding placed in different nodes
- Encoding is done by dividing 64MB blocks into symbols of size 8 bits each

Failures in DSS

- Node is considered a failure domain
- Each encoded block is placed in a different failure domain (in this case different node)
- Permanent Failures: Data is lost because of hardware failure
- Temporary Failures: Power Outage, Software Upgrade. Data is temporarily unavailable but needs efficient recovery if there is a request for such data

MDS Codes in DSS



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

Single Node Failures in DSS

- 98% of failures are single node failures

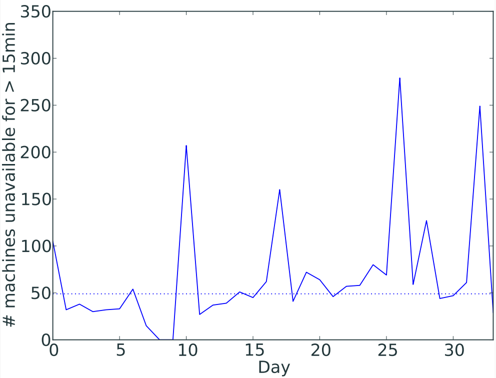


Image Courtesy: K. V. Rashmi, et al. "A solution to the network challenges of data recovery in erasure-coded distributed storage systems: A study on the Facebook warehouse cluster." HotStorage 2013.

Metrics of Interest in Repair

For a given storage overhead,

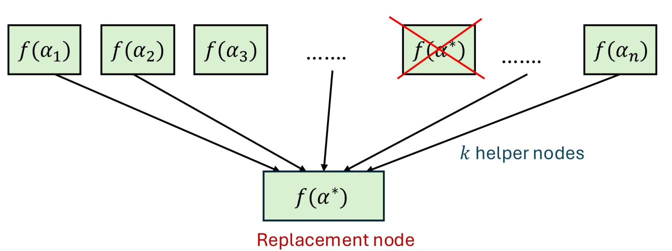
- Maximize the reliability wrt worst case failures. Ensured by
  - "k out of n" property
  - Maximizing $d_{\min}$
- Minimize the repair bandwidth in case of single node failures (Regenerating Codes)
- Minimize the number of nodes contacted in case of single node failures (Locally Repairable Codes)

Repairing Reed-Solomon Codes

Reed-Solomon Codes

- Let $\underline{m} = [m_0, \dots m_{k-1}]$ be message vector over finite field $\mathbb{F}_q$
- Form the message 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
- Codeword corresponding to $\underline{m}$ is $\underline{c} = [f(\alpha_1), \dots, f(\alpha_n)]$
- This code can tolerate $n - k$ erasures ($k - 1$ degree polynomial can be uniquely determined by evaluations at k points)
- Minimum distance of RS code is $n - k + 1$

Naive Repair of Reed-Solomon Codes



- If a node $f(\alpha^*)$ is erased, k of the remaining $n-1$ nodes are downloaded to obtain $f(x)$ and subsequently $f(\alpha^*)$.

For Better Repair Bandwidth [SPDC14]

- Code symbols from the finite field treated as vectors over a subfield
- Helper nodes send symbols from the subfield by performing vector linear operations
- In [SPDC14], improvements from (5,3) and (6,4) RS codes were shown

[SPDC14] Shanmugam, K., Papailiopoulos, D.S., Dimakis, A.G. and Caire, G., "A repair framework for scalar MDS codes," *IEEE Journal on Selected Areas in Communications*, May 2014.

Dual codes of Reed-Solomon Codes

Dual of a Reed-Solomon code is a Generalized Reed-Solomon code (GRS) code

- GRS Code: For some non-zero elements $v_1, v_2, \ldots, v_n \in \mathbb{F}_q$ and message polynomial $f(x) = \sum_{i=0}^{k-1} m_i x^i$, the codeword corresponding to $\underline{m}$ is $\underline{c} = [v_1 f(\alpha_1), \ldots, v_n f(\alpha_n)]$.
- For an $[n, k]$ Reed-Solomon code, the dual code is an $[n, n-k]$ GRS code with $d_{min} = k + 1$.

Trace of a field element

Let $\mathbb{F} = \mathbb{F}_{q^l}$ and $\mathbb{B} = \mathbb{F}_q$. The trace polynomial is defined as,

$$\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}(\alpha) = \alpha + \alpha^q + \alpha^{q^2} + \cdots + \alpha^{q^{l-1}}$$

- Trace of an element takes values from a field $\mathbb{F}$ and maps it to a subfield $\mathbb{B}$.
- Trace is a $\mathbb{B}$-linear

  $$\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}(b_1\alpha + b_2\beta) = b_1\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}(\alpha) + b_2\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}(\beta),$$

  $\alpha, \beta \in \mathbb{F}$ and $b_1, b_2 \in \mathbb{B}$.
- Every $\mathbb{B}$-linear function is $Tr(\gamma\alpha), \alpha \in \mathbb{F}$, γ fixed element in $\mathbb{F}$.

Trace Repair Framework [GW17]

- An erased node $f(\alpha^*)$ can be recovered from the equation

$$f(\alpha^*) = \sum_{j=1}^{l} \mathsf{Tr}_{\mathbb{F}/\mathbb{B}}(u_j f(\alpha^*)) v_j$$

  where $u_1, u_2, \ldots u_l$ is a basis of $\mathbb{F}$ over $\mathbb{B}$ and $v_1, v_2, \ldots v_l$ is the dual-basis.

- Say $A = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. $f(x)$ is the message polynomial of RS code and $g(x)$ is the message polynomial of its dual code, $\sum_{\alpha \in A} f(\alpha) g(\alpha) = 0$.

- Applying trace,

$$\mathsf{Tr}_{\mathbb{F}/\mathbb{B}}(g_j(\alpha^*) f(\alpha^*)) = - \sum_{\alpha \in A \setminus \{\alpha^*\}} \mathsf{Tr}_{\mathbb{F}/\mathbb{B}}(g_j(\alpha) f(\alpha))$$

Guruswami-Wootters Scheme [GW17]

- Choice of g_j: If $f(\alpha^*)$ has been erased, choose $\forall j \in [l]$

$$g_j(x) = \frac{\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}\left(u_j\left(x - \alpha^*\right)\right)}{x - \alpha^*},$$

  where $u_1, u_2, \ldots u_l$ forms a basis of $\mathbb{F}$ over $\mathbb{B}$.

- $g_j(\alpha^*) = u_j$ and

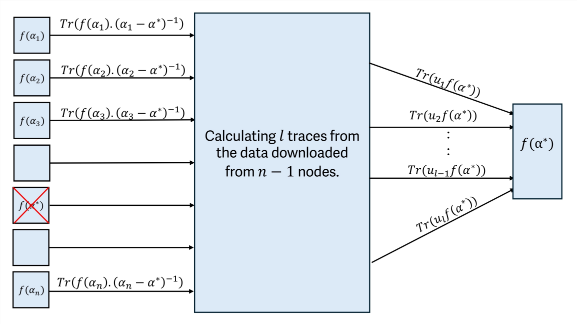
$$\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}\left(u_j f(\alpha^*)\right) = -\sum_{\alpha \in A \setminus \{\alpha^*\}} \mathrm{Tr}_{\mathbb{F}/\mathbb{B}}(u_j(\alpha - \alpha^*))\mathrm{Tr}_{\mathbb{F}/\mathbb{B}}\left(\frac{f(\alpha)}{\alpha - \alpha^*}\right).$$

- Dimension of span given by

$$\dim_{\mathbb{B}}(< g_1(\alpha), g_2(\alpha), \ldots g_l(\alpha) >) = \begin{cases} l & \alpha = \alpha^* \\ 1 & \alpha \neq \alpha* \end{cases}$$

Guruswami-Wootters Scheme [GW17]

- The l traces required for the repair can be obtained by downloading 1 symbol from each of the remaining $n-1$ nodes.
- The repair bandwidth of this framework is $(n-1)\log_2 q$ bits.



[GW17] Guruswami, V. and Wootters, M., "Repairing Reed-Solomon codes," *IEEE Transactions on Information Theory*, Sept. 2017.

Constraints on Parameters

- $n \leq q^l$
- All the l polynomials can act as check polynomials if $n - k \geq q^{l-1}$.
- Repair bandwidth is optimal if $n = q^l$ and $n - k = q^{l-1}$

Dau-Milenkovic Scheme [DM17]

- Linearized Polynomial: A monic polynomial of the form

$$L(x) = \sum_{i=0}^{d} \ell_i x^{q^i},$$

where $\ell_i \in \mathbb{F}$. Trace is an example and so is $L_W(x)$ below.

- Choice of g_j: If $f(\alpha^*)$ has been erased, choose $\forall j \in [l]$

$$g_j(x) = \frac{L_W\left(u_j\left(x - \alpha^*\right)\right)}{x - \alpha^*},$$

where W is a subspace of dimension s over $\mathbb{B}$. $L_W(x) = \prod_{w \in W}(x - w)$.

- Dimension of span given by

$$\dim_{\mathbb{B}}(< g_1(\alpha), g_2(\alpha), \ldots g_l(\alpha) >) = \begin{cases} l & \alpha = \alpha^* \\ \leq l - s & \alpha \neq \alpha* \end{cases}$$

[DM17] H. Dau and O. Milenkovic, "Optimal Repair Schemes for Some Families of Full-Length Reed-Solomon Codes," in 2017 *IEEE International Symposium on Information Theory.*

Constraints on Parameters

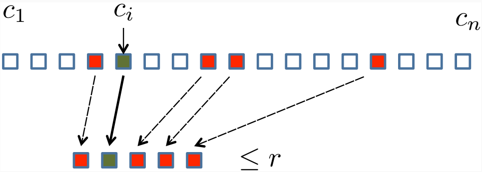
- $n \leq q^l$
- All the l polynomials can act as check polynomials if $n - k \geq q^s$
- Repair bandwidth is optimal if $n = q^l$ and $n - k = q^s$

Repairing Locally Recoverable Codes

Locality Parameter [GHYS12]

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 Locality Tradeoff [GHYS12]

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

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

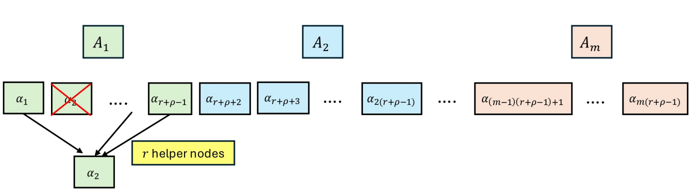
[GHSY12] Gopalan, P., Huang, C., Simitci, H. and Yekhanin, S., "On the locality of codeword symbols," *IEEE Transactions on Information theory*, 2012.

(r, ρ) locality

- ith symbol in an (n, k, d) code is said to have (r, ρ) locality if there exists a punctured subcode $\mathbb{C}_i$ with support containing i,
    - whose length is at most $r + \rho - 1$
    - whose minimum distance is atleast ρ

- A code in which all the symbols have (r, ρ) locality is said to be an (n, k, r, ρ) LRC.

Tamo-Barg Codes [TB14]



- $g(x)$ is of degree $r+\rho-1$.
- Encoding polynomial is $f(x) = \sum_{i,j} a_{ij} x^i g(x)^j$
- $A_1, A_2, \ldots, A_m$ form a partition such that
  $g(\alpha_j) = c_i \quad \forall \alpha_j \in A_i, \quad i.e., \quad j \in [(i-1)(r+\rho-1)+1, i(r+\rho-1)].$
- g can be picked to be polynomial of additive or multiplicative cosets of a subgroup

[TB14] Itzhak Tamo and Alexander Barg, "A family of optimal locally recoverable codes, *TIT*, Jul 2014.

Tamo-Barg Codes with Local Repair

- The construction yields an (n, k, r, ρ) LRC with m disjoint $RS_{\mathbb{F}}(r + \rho - 1, \rho)$ local codes.
- Objective: Minimise repair bandwidth required to repair a single erasure.
- Two schemes in which the evaluation points are chosen differently.
  - In one scheme, the evaluation points are picked from cosets of additive subgroup.
  - In the other scheme, the evaluation points are picked as elements of prime degree over a field.

Tamo-Barg Codes based on Additive Cosets

- Let $B = \{\alpha_1, \alpha_2, \ldots, \alpha_{r+\rho-1}\} = \mathbb{F}_{q^a}$ and $\{\beta_1 + B, \beta_2 + B, \ldots, \beta_m + B\}$ are additive cosets of B in $\mathbb{F}_{q^l}$
where $a \mid l$ and $m \leq q^{l-a}$.
- Let $A_i = \{\alpha_1 + \beta_i, \alpha_2 + \beta_i, \ldots, \alpha_{r+\rho-1} + \beta_i\} \subset \mathbb{F}_{q^l}$ for all $i \in [m]$.
- Let W be an s dimensional $\mathbb{F}_q$ subspace of $\mathbb{F}_{q^a}$.
- Define $g_{ij}(x) = \frac{L_W(u_j(x-(\alpha^*+\beta_i)))}{x-(\alpha^*+\beta_i)}, \forall j \in [a]$ to repair $f(\alpha^* + \beta_i)$.

Sasanka, U. S. S., and V. Lalitha, "Tamo-Barg Codes with Efficient Local Repair," ITW 2022

Tamo-Barg Codes based on Additive Cosets

- We need l traces for the repair framework but $\{g_{ij}(x), j \in [a]\}$ are only a polynomials.
- Let $\{\gamma_1, \gamma_2, \ldots \gamma_{\frac{l}{a}}\}$ be a basis of $\mathbb{F}_{q^l}$ over $\mathbb{F}_{q^a}$.
- The l polynomials are $\{\gamma_1 g_{ij}(x), \gamma_2 g_{ij}(x), \ldots \gamma_{\frac{l}{a}} g_{ij}(x)\}$ for some $i \in [m]$ and $\forall j \in [a]$.
- The bandwidth required in this scheme is $\frac{l}{a}((r + \rho - 1) - 1)(a - s)$.

Revisiting Reed-Solomon Codes

Cut-set Bound for Repair Bandwidth

- Cut-set Bound: For any $[n, k, l]$ MDS code where l is the sub-packetization, the repair bandwidth for a single erasure is given by

$$b \geq \frac{dl}{d-k+1},$$

  where d, such that $k < d < n$, are number of helper nodes

- Bound above corresponds to the Minimum Storage Regeneration (MSR) point of storage-bandwidth tradeoff

- To achieve the cutset bound, require different sub-fields over which trace is computed for different failed nodes

Optimal RS Code Achieving Cut-set Bound [TYB19]

- Let $s = d - k + 1$. Let $p_1, p_2, \ldots, p_n$ be the smallest distinct primes satisying $p_j \equiv 1 \bmod s$ for all $i = 1, 2, \ldots, n$.

- Let $\mathbb{F}_p$ be a field of prime order. Let $A = \alpha_1, \alpha_2, \ldots, \alpha_n$ be the evaluation set. Choose α_i to be an element of degree p_i over $\mathbb{F}_p$, i.e.,

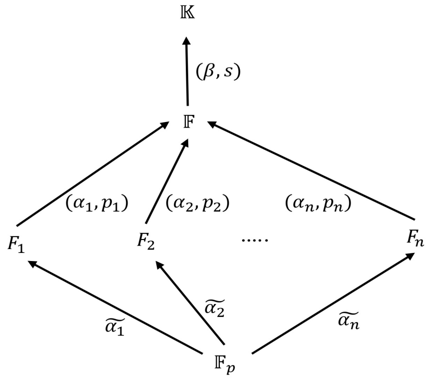
$$[\mathbb{F}_p(\alpha_i) : \mathbb{F}_p] = p_i,$$

  where $\mathbb{F}_p(\alpha_i)$ denotes the field obtained by adjoining α_i to $\mathbb{F}_p$.

- Define $\mathbb{F} = \mathbb{F}_p(\alpha_1, \alpha_2, \ldots, \alpha_n)$ and so $[\mathbb{F} : \mathbb{F}_p] = \prod_{i=1}^{n} p_i$.

- Define $\mathbb{K} = \mathbb{F}(\beta)$ where β is an element of degree s over $\mathbb{F}$.

Optimal RS Code Achieving Cut-set Bound [TYB19]



- RS code defined over the field $\mathbb{K}$
- Sub-packetization $O(n^n)$

- Repair of a failed node corresponding α_i occurs over field $F_i = \mathbb{F}_p(\alpha_j : j \in [n] \text{ and } j \neq i)$ by using the trace $\mathrm{Tr}_{\mathbb{K}/F_i}$.
- Check polynomials can be chosen and repair process is similar to the trace repair framework.

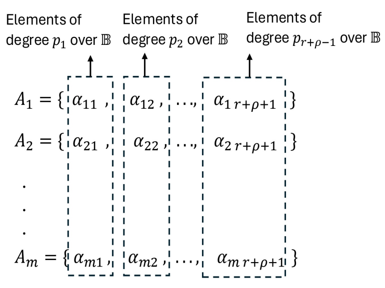
[TYB19] I. Tamo, M. Ye, and A. Barg, "The Repair Problem for Reed–Solomon Codes: Optimal Repair of Single and Multiple Erasures with Almost optimal node size," *IEEE Trans. on Inf. Theory*, May 2019.

Tamo-Barg Codes with Optimal Local Repair

- Goal: Design an (n, k, r, ρ) LRC which achieves the cut-set bound for single node repair within the local group.
- Since the local codes are RS codes, the MSR construction in [TYB19] can be used.
- Node failure can happen in any of the m RS codes. So all of them must be MSR codes.

Tamo-Barg Codes with Optimal Local Repair

- Extend [TYB19] so that the j^{th} element of each local group is chosen to be a distinct primitive element of the (same) extension field of degree p_j over the base field.



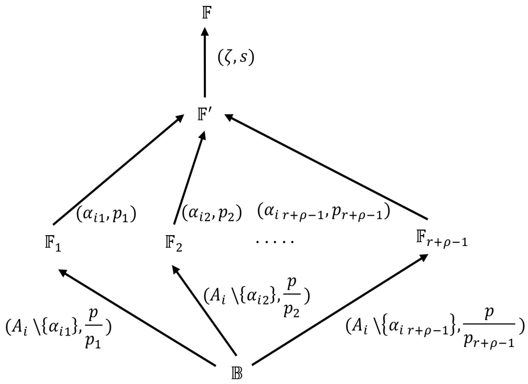
- Let $\mathbb{B} = \mathbb{F}_q$. Choose α_{ij} $\forall i \in [m],\ j \in [r+\rho-1]$ such that $[\mathbb{F}_q(\alpha_{ij}) : \mathbb{F}_q] = p_j$.

Sasanka, U. S. S., and V. Lalitha, "Tamo-Barg Codes with Efficient Local Repair," ITW 2022

Tamo-Barg Codes with Optimal Local Repair

- Pick different generators for the same extension field
- The number of primitive elements in a finite field $\mathbb{F}_p$ is given by $\phi(p-1)$, where $\phi(x)$ is the Euler's totient function.
- Constraint $m < min\{\phi(q^{p_1}-1), \phi(q^{p_2}-1), \ldots, \phi(q^{p_{r+\rho-1}}-1)\}$.

Tamo-Barg Codes with Optimal Local Repair



- Let $P = \prod_{i=1}^{n} p_i$. The code is defined on $\mathbb{F} = \mathbb{F}_{q^l}$, where $l = sP$. The repair for the erased node corresponding to α_{ij} is done over the field $\mathbb{F}_j$.
- Each of the local RS codes is an MSR code and the repair bandwidth of the LRC code is $\frac{dl}{d-k+1}$.

Comparison of Our Schemes

Scheme	Repair bandwidth	Code length and restrictions	Achieving Cut-set Bound
Additive cosets	$\frac{l}{a}(n'-1)(a-s)$	$n' = q^a$; $q^s \leq n'-k'$, $a \mid l$	No
Optimal repair	$\frac{l(n'-1)}{n'-k'}$	$(n')^{(n')} \approx l$	Yes

- $n' = r + \rho - 1$ and $k' = r$ are the length and dimension of the local RS code.

Thanks!