Rate-optimal Binary Locally Repairable Codes with Joint Information Locality

Jung-Hyun Kim, Mi-Young Nam, and Hong-Yeop Song Yonsei University, Korea (jh.kim06, my.nam, hysong@yonsei.ac.kr)

> 2015 / 10 / 12 2015 IEEE Information Theory Workshop







Prior Work

BLRC with Joint Inform. Locality

Summary & Conclusion





Distributed Storage System (DSS)



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





Locally repairable code (LRC)
 Codes with good (small) locality

Locality

- **Symbol locality** : # of symbols required to repair a failed symbol
- (Code) locality : the maximum value of symbol locality







Locally repairable code (LRC)
 Codes with good (small) locality

Locality

- **Symbol locality** : # of symbols required to repair a failed symbol
- (Code) locality : the maximum value of symbol locality







Locally repairable code (LRC)
 Codes with good (small) locality

Locality

- **Symbol locality** : # of symbols required to repair a failed symbol
- (Code) locality : the maximum value of symbol locality







Locally repairable code (LRC)
 Codes with good (small) locality

Locality (Generalized definition)

• *l*-locality (r_{ℓ}) : locality for ℓ symbols repair

* 1-locality (r_1) is the same with "code locality" in the previous definition



A. S. Rawat, A. Mazumdar, and S. Vishwanath, "Cooperative local repair in distributed storage," arXiv Preprint arXiv:1409.3900, 2014.

Jung-Hyun Kim, Mi-Young Nam, Ki-Hyeon Park, and Hong-Yeop Song, "New Binary Locally Repairable Codes with Joint Locality and Average Locality," under revision, IEEE Trans. on Inf. Theory. 7







Prior Work

BLRC with Joint Inform. Locality

Summary & Conclusion





(Binary) Simplex codes

Prior Work

 $r_1 = 2$ (VERY GOOD) $R = \frac{k}{2^{k}-1}$ (VERY LOW)

Only better code is repetition code ($r_1 = 1$), but its code rate is extremely low.





(Binary) Simplex codes

 $r_1 = 2$ (VERY GOOD) $R = \frac{k}{2^{k}-1}$ (VERY LOW)

Only better code is repetition code $(r_1 = 1)$, but its code rate is extremely low. **Q1**: Can we improve the code rate maintaining the locality?

 $(r_1 = 2)$





(Binary) Simplex codes

$$r_1 = 2$$
 (VERY GOOD)
 $R = \frac{k}{2^{k}-1}$ (VERY LOW)

Only better code is repetition code $(r_1 = 1)$, but its code rate is extremely low. **Q1**: Can we improve the code rate maintaining the locality? $(r_1 = 2)$



Jung-Hyun Kim, Mi-Young Nam, Ki-Hyeon Park, and Hong-Yeop Song, "New Binary Locally Repairable Codes with Joint Locality and Average Locality," under revision, IEEE Trans. on Inf. Theory.





Q2: What about multiple failure patterns? (every *l*-locality)









Q2: What about multiple failure patterns? (every *l*-locality)

Which one is better? C_1 ? or C_2 ?









Q2: What about multiple failure patterns? (every *l*-locality)

Which one is better? C_1 ? or C_2 ?









Q2: What about multiple failure patterns? (every *l*-locality)

Which one is better? C_1 ? or C_2 ?

 $r_1 = 2$ > $r_1 = 3$ $r_2 = 5$ < $r_2 = 4$









Q2: What about multiple failure patterns? (every *l*-locality)

Which one is better? C_1 ? or C_2 ?

 $r_1 = 2$ > $r_1 = 3$ $r_2 = 5$ < $r_2 = 4$ $(r_1, r_2) = (2, 5)$ $(r_1, r_2) = (3, 4)$









Q2: What about multiple failure patterns? (every *l*-locality)

Which one is better? C_1 ? or C_2 ?

Joint locality

$$(r_1, r_2, r_3) = (2, 5, 5)$$
 $(r_1, r_2, r_3) = (3, 4, 5)$













































Main results



Joint locality							
Code (dimension k)	Code rate	(r ₁ , r ₂)	Another metric?				
Simplex code	$\frac{k}{2^k - 1}$	(2,3)	?	increasing			
Complete graph code	$\frac{2}{k+1}$	(2,3)	?	code rate			
Complete multipartite graph code (<i>p</i> -partite)	$\frac{2}{k - \frac{k}{p} + 2}$	(2,4)	?				
New code?	?	(2,4)	?				







Prior Work

BLRC with Joint Inform. Locality

Summary & Conclusion



- Joint Information Locality
 - a set of numbers of symbols for repairing various erasure patterns of information symbols









Joint Information Locality

 a set of numbers of symbols for repairing various erasure patterns of information symbols

Can we design rate-optimal codes with joint inform. locality (2,3) or (2,4)?









Joint Information Locality

- a set of numbers of symbols for repairing various erasure patterns of information symbols
- Can we design rate-optimal codes with joint inform. locality (2,3) or (2,4)?



We begin with a **simple graph**.

an unweighted, undirected, connected graph containing no loops or multiple edges





BLRC with Joint Inform. Locality



simple graph

$$k = \#v$$

$$n = \#v + \#e \longrightarrow \text{If } \#e \uparrow, \text{ then } \frac{k}{n} \downarrow$$

Vertex : inform. symbol







edge : parity symbol









Simple graph-based code construction

Minimum distance

obtained straightforwardly

We found this expression.

$$d = \min_{S \subseteq V} [|Cut(S, S^{c})| + |S|]$$

where V is the set of all the vertices.















Simple graph-based code construction

Lemma 1. Always $(r_1)_{info} = 2$



Node failure (Information symbol)

Repair set



Simple graph-based code construction

Lemma 1. Always $(r_1)_{info} = 2$



Node failure (Information symbol)

Repair set

Lemma 2. If every vertex pair is in 2-hop distance, $(r_2)_{info} = 3$





2-hop distance vs. higher rate





2-hop distance vs. higher rate

of edges ↑ Too many edges \Rightarrow low rate Too few edges \Rightarrow 2-pop

of edges \downarrow





2-hop distance vs. higher rate

of edges ↑ Too many edges ⇒ low rate

of edges \downarrow Too few edges \Rightarrow 2-hop



2-hop Low rate High rate <mark>3-hop</mark>

BLRC with Joint Inform. Locality



Lemma 3. $(r_1, r_2)_{info} = (2, 3)$, if and only if any vertex pair is in either of triangle, quadrangle, or pentagon.









• Rate-optimal code with joint information locality $(r_1, r_2)_{info} = (2, 3)$



For every positive integer $k \ge 5$, the code construction is possible.















Theorm 1.
 $(r_1, r_2)_{info} = (2, 3)$
Rate-optimalAny vertex pair should be in
either of \Box or \bigcirc , and no more.
The graph should contain at
least one \bigcirc .



 $(r_1, r_2)_{info} = (2, 3)$ Not rate-optimal









Not rate-optimal

 $(r_1, r_2)_{info} = (2, 3)$ $(r_1, r_2)_{info} = (2, 3)$ **Rate-optimal**



43



For every positive integer $k \ge 3$, the code construction is possible. When k = 5, $(r_1, r_2)_{info} = (2, 3)$ since it is also a crown code.









Ring code

Theorm 2. $(r_1, r_2)_{info} = (2, 4)$ Rate-optimal



The graph should be single cycle structure



Ring code

Theorm 2. $(r_1, r_2)_{info} = (2, 4)$ Rate-optimal



The graph should be single cycle structure



 $(r_1, r_2)_{info} = (2, 4)$ Not rate-optimal



Ring code









Prior Work

BLRC with Joint Inform. Locality

Summary & Conclusion







Average locality

	V Joint inform. locality V				
Code (dimension k)	Code rate	(r_1, r_2)	$\overline{r_2}$	$(r_1, r_2)_{info}$	$(\overline{r_2})_{info}$
Simplex code	$\frac{k}{2^k - 1}$	(2,3)	3	(2,3)	3
Complete graph code	$\frac{2}{k+1}$	(2,3)	3	(2,3)	3
Complete multipartite graph code (<i>p</i> -partite)	$\frac{2}{k - \frac{k}{p} + 2}$	(2,4)	$3 + \frac{2\binom{k/p}{2}^2 \binom{p}{2}}{\binom{n}{2}}$	(2,3)	3
Crown code	$\frac{k}{3k-5}$	(2,4)	$3 + \frac{2k^2 - 4k - 10}{9k^2 - 33k + 30}$	(2,3)	3
Ring code	$\frac{1}{2}$	(2,4)	$4 - \frac{7}{2k-1}$	(2,4)	$4 - \frac{4}{k-1}$





- The rate of Crown/Ring codes gives a global lower bound, since it is Rate-optimal within a framework of codes based on simple graph. How good is it?
- LRC construction not based on simple graph
- Binary LRC with joint inform. locality (r_1, r_2, r_3, r_4)
- **Non-binary LRC** construction with the same *G* for either Crown or Ring code