New Bounds on the $(n, k, d)$ Storage Systems with Exact Repair

Soheil Mohajer  
Department of Electrical and Computer Engineering  
University of Minnesota  
Minneapolis, MN 55404  
Email: soheil@umn.edu

Ravi Tandon  
Discovery Analytics Center  
Department of Computer Science  
Virginia Tech, Blacksburg, VA 24060  
Email: tandonr@vt.edu

Abstract—The exact-repair problem for distributed storage systems is considered. Characterizing the optimal storage-vs-repair bandwidth tradeoff for such systems remains an open problem for more than four storage nodes. A new family of information theoretic bounds is provided for the storage-vs-repair bandwidth tradeoff for all $(n, k, d)$ systems. The proposed bound readily recovers Tian’s result for the $(4, 3, 3)$ system, and hence suffices for exact characterization for this system. In addition, the bound improves upon the existing bounds for the $(5, 4, 4)$ system. More generally, it is shown that this bound characterizes the optimal boundary of the exact repair tradeoff for all distributed storage systems, with $(n, k, d) = (n, n-1, n-1)$ when $\beta \leq 2\alpha/k$.

Index Terms—Distributed storage system, Exact repair, New outer bounds

I. INTRODUCTION

Contemporary distributed storage systems store massive amounts of data over a set of distributed nodes. Besides the traditional goals of achieving reliability by introducing redundancy, new aspects such as efficient repair of failed storage nodes are becoming increasingly important. To address these issues, the concept of regenerating codes for distributed storage systems (DSS) was introduced by Dimakis et al. [1]. A DSS consists of $n$ storage nodes each with a storage capacity of $\alpha$ units, such that the entire file of size $F$ can be recovered by accessing any $k < n$ nodes. This is called as the reconstruction property of the DSS. Whenever a node fails, $d$ nodes (where $k \leq d \leq n-1$) participate in the repair process by sending $\beta$ units of data each. This procedure is termed as the regeneration of a failed node and $\beta$ is referred to as the per-node repair bandwidth. In [1], it was shown that the maximum amount of data, $F$, that any regenerating code can store satisfies

$$F \leq \sum_{i=0}^{k-1} \min(\alpha, (d-i)\beta).$$

Thus, in order to store data of size $F$, there exists a fundamental tradeoff between $\alpha$ (storage) and $d\beta$ (total repair bandwidth). It was also shown in [1] that the above tradeoff is achievable for functional repair, which does not require the contents of the repaired node to be the same as the original node. In contrast to functional repair, exact repair requires that the contents of the failed node must match with those stored in the original node. Exact repair is a practically appealing property specially when it is desirable that the stored contents remain intact over time. Furthermore, the file recovery process is also easier in this case as the reconstruction procedure need not change whenever a failed node is replaced. While characterizing the storage-vs-bandwidth tradeoff for the case of exact repair remains a challenging open problem in general, two extreme points of this tradeoff namely, the minimum storage regenerating case (MSR) and the minimum bandwidth regenerating (MBR) case have been studied extensively (see [3], [4] and references therein). Other notable works on code constructions beyond MSR and MBR points include [8], [9].

Tian has recently characterized the exact repair tradeoff for the $(4, 3, 3)$-DSS [5]. This result, which is based on a novel computer-aided approach showed that functional and exact repair problems are fundamentally different. Despite its originality, the solution involved solving an optimization problem with a large number of variables and constraints. More importantly, the number of variables/constraints grow (at least) exponentially, and hence, it is not clear whether this approach can be generalized for larger system parameters. Moreover, such a computer-aided approach does not necessarily lead to intuition and insights which could be used to understand the exact repair problem for a general set of parameters. Notably, Sasidharan et al. in [7] brought some intuition in this regard and presented a simpler proof for the $(4, 3, 3)$ problem and also presented new bounds for the $(5, 4, 4)$-DSS.

In this paper, we present a new and general approach for obtaining information theoretic upper bounds on $F$ for the exact repair problem. This approach is used to develop a family of bounds which hold for any $(n, k, d)$-DSS. Using these bounds, together with the code constructions in [9], we characterize the partial boundary of the optimal exact repair tradeoff for $(n, n-1, n-1)$-DSS in the regime when $\beta \leq 2\alpha/k$. We also show that the proposed bound yields a new and simple proof for the $(4, 3, 3)$-DSS. For the $(5, 4, 4)$-DSS, our bounds improve upon the ones obtain in [7].

II. PROBLEM STATEMENT AND RESULT

Notation: We use $[i : j] = \{i, i+1, \ldots, j\}$ to denote the set of positive integers between (and including) $i$ and $j$. If $i = 1$, we drop it, and simply use $[j]$ to denote set $\{1, 2, \ldots, j\}$, hence $\{n\} = \{1, 2, \ldots, n\}$ denotes the set of all node indices. We use $W_i$ to denote the content stored in node $i$, and extend
this definition to \( W_A = \{ W_i : i \in A \} \) for any \( A \subseteq [n] \). In the rest of this paper, unless otherwise mentioned, we focus on a subset of the nodes indexed by \( N = \{ 1, 2, \ldots, d + 1 \} \). Note that any upper bound for this sub-system of \( (d + 1) \) nodes holds for the original system with \( n \) nodes as well.

In this sub-system, the repair data from \( i \) to \( j \) is denoted by \( S^j_i \). Note that since \( |N| = d + 1 \), there is a unique way of choosing \( d \) helper nodes to repair any failed node within \( N \). Therefore, the dependence of \( S^j_i \) on the remaining \( (d - 1) \) helper nodes, that is \( N \setminus \{ i, j \} \), is clear due to their uniqueness and is hence dropped from the notation for simplicity. We also set \( S^i_i \), to be a dummy variable with zero entropy, for consistency. Moreover, \( S^i_A = \{ S^j_i : i \in A, j \in B \} \).

Next we describe the exact repair problem and the associated constraints. An exact repair distributed storage system with parameters \( (n, k, d) \) and \( (\alpha, \beta) \) is defined as follows. A DSS consists of \( n \) storage devices, each with capacity \( \alpha \), which is used to store some Data in a distributed fashion, such that the following properties hold:

- **MDS Property (Data recovery):** Data can be recovered from the content of any \( k \) nodes: \( H(\text{Data}|W_A) = 0 \) for any \( A \subseteq N \) satisfying \( |A| \geq k \).
- **Repairability Requirements:** The content of any failed node can be exactly recovered (repaired) by receiving no more that \( \beta \) units of repair data from any other \( d \) nodes, that is, \( H(W_i|S^A) = 0 \) for any \( A \subseteq N \setminus \{ i \} \), with \( |A| \geq d \), where \( H(S^j_i) \leq \beta \) and \( H(S^i_A) = 0 \).

We next present the main result of this paper which is a new set of lower bounds on the exact repair tradeoff for the \((n, k, d)\) distributed storage system.

**Theorem 1.** The exact repair capacity of an \((n, k, d)\) distributed storage system with per node storage \( \alpha \) and total repair bandwidth \( d \beta \) is upper bounded by a family of bounds, namely,

\[
3F \leq (3k - 2m)\alpha + \frac{m(2d - 2k + m + 1)}{2} \beta + (d - k + 1) \min(\alpha, k\beta),
\]

for \( m = 0, 1, \ldots, k \).

The following corollary is an immediate consequence of this theorem together with the code construction in [9].

**Corollary 1.** The exact repair capacity of an \((n, k, d)\) DSS for \( \beta \leq 2\alpha/k \) is given by

\[
F \leq \min \left\{ \frac{k + 1}{3} \alpha + \frac{k(k + 1)}{6} \beta, \frac{k(k + 1)}{2} \beta \right\}. \tag{2}
\]

**Proof of Corollary 1.** The first bound in the minimum above follows by setting \( d = m = k \) in Theorem 1, while the second one, i.e., \( F \leq k(k + 1)\beta/2 \) is simply the cut-set bound. Moreover, achievability of the MBR point \((\alpha, \beta) = \left( \frac{2F}{k+1}, \frac{2F}{k(k+1)} \right)\) is given in [10].

Finally, the other extreme point of this region is \((\alpha, \beta) = \left( \frac{3F}{2(k+1)}, \frac{3F}{k(k+1)} \right)\), which is shown to be achievable by the code construction in [9, Theorem 1 for \( k = 2 \)] for every \( k \).

**Remark 1.** For the \((4, 3, 3)\)-DSS, setting \( m = d = k = 3 \), we obtain \( 3F \leq 4\alpha + 6\beta \), which is precisely the new converse bound obtained by Tian [5] through a novel computer aided approach. This bound together with the cut-set bound and the achievability in [5] suffices to characterize the exact repair tradeoff for \((4, 3, 3)\)-DSS.

**Remark 2.** For the \((5, 4, 4)\)-DSS, Theorem 1 leads to the following set of new bounds which improve upon the cut-set bound:

1) \( 3F \leq 5\alpha + 10\beta \) (setting \( m = 4 \))
2) \( 3F \leq 7\alpha + 6\beta \) (setting \( m = 3 \))

It is interesting to note that the bound \( 3F \leq 7\alpha + 6\beta \) was also obtained in [7] through a different set of arguments. Moreover, the other bound \( 3F \leq 5\alpha + 10\beta \) (corresponding\(^1\) to \( m = 4 \)) gives the optimal characterization of the exact repair tradeoff for the regime in which \( \beta \leq 2\alpha/k \) (see Corollary 1). These bounds together with the cut-set bound and the best known code-constructs are shown in Fig. 1.

![Fig. 1. Existing and new results for (5, 4, 4) DSS.](image)

### III. PROOF OF THEOREM 1

For the sake of brevity and simplicity, we focus on the tradeoff of the symmetric\(^2\) exact-repair regeneration codes for DSSs, in which the information-theoretical quantities are invariant under any relabeling of the nodes. We adopt the notation in [5] in order to formally define this symmetry:

**Definition 1.** A permutation \( \pi \) is given by a one-to-one mapping \( \pi : [n] \rightarrow [n] \). We denote the set of all permutations by \( \Pi \).

Then a symmetric DSS can be defined as the following.

\(^1\)This specific bound for \((5, 4, 4)\) first appeared in our prior work [11]. In contrast to [11], the bounding technique of this paper is general and applicable to any \((n, k, d)\) DSS.

\(^2\)Note that the symmetry assumption is made without any loss in generality, as any asymmetric code can be symmetrized by augmenting its \( n! \) copies, each copy corresponding to a permutation of the node labels. The resulting symmetric code and the original asymmetric code achieve the same \((F, \alpha, \beta)\) up to the scaling factor of \( n! \).
For any pair of disjoint sets $P = \{1, 2, \ldots, m\}$ and $Q = \{m + 1, m + 2, \ldots, k\}$. Then, since $|P \cup Q| = k$, the entire data can be recovered from disks in $P \cup Q$. Hence,

\[
F = H(\text{Data}) = H(W_{P \cup Q}) = H(W_Q) + H(W_P|W_Q)
\]

\[
= H(W_Q) + \sum_{i=1}^{m} H(W_i|W_{[i-1]}|W_Q)
\]

\[
= H(W_Q) + \sum_{i=1}^{m} [H(W_i|W_Q) - I(W_i; W_{[i-1]}|W_Q)]
\]

\[
\leq H(W_Q) + \sum_{i=1}^{m} I(W_i; W_{[i-1]}|W_Q) - \sum_{i=1}^{m} I(W_i; W_{[i-1]}|W_Q)
\]

\[
\leq (k - m)\alpha + m\alpha - \sum_{i=1}^{m} I(W_i; W_{[i-1]}|W_Q).
\]

Next, note that the repair data $S_{i-1}^i$ (which is sent by node $i$) is a function of $W_i$. Similarly, $S_{i-1}^i$ is a function of $W_{[i-1]}$. Hence we can write

\[
\sum_{i=1}^{m} I(W_i; W_{[i-1]}|W_Q) \geq \sum_{i=1}^{m} I(S_{i-1}^i, S_{[i-1]}^i|W_Q)
\]

\[
= \sum_{i=1}^{m} [H(S_{i-1}^i|W_Q) + H(S_{[i-1]}^i|W_Q) - H(S_{i-1}^i, S_{[i-1]}^i|W_Q)]
\]

\[
= \sum_{i=1}^{m} H(S_{i-1}^i|W_Q) + \sum_{i=1}^{m} H(S_{[i-1]}^i|W_Q) - \sum_{i=1}^{m} H(S_{i-1}^i, S_{[i-1]}^i|W_Q),
\]

which together with (6) implies that

\[
F \leq k\alpha - |B|\alpha.
\]

Our next goal is lower bounding $\text{Term}_1$ and $\text{Term}_2$, and upper bounding $\text{Term}_3$.

### A. LowerBounding $\text{Term}_1$

First we use Lemma 1 for $B = Q$ to get

\[
\sum_{i=1}^{m} H(S_{i-1}^i|W_Q) + (d + 1 - k)H(S_{k+1}^i|W_Q) \geq F - |Q|\alpha.
\]

Hence,

\[
\text{Term}_1 = \sum_{i=1}^{m} H(S_{i-1}^i|W_Q)
\]

\[
\geq F - (k - m)\alpha - (d + 1 - k)H(S_{k+1}^i|W_Q).
\]
that is where the last equality follows from the symmetry property,

\[
H \left( S_{i-1}^{[i-1]} | W_Q \right) \geq \frac{[i-1]}{d-|Q|} H(S_N^i | W_Q)
\]

where in the last equality we used the symmetry property implying that \( H(S_N^i | W_Q) = H(S_N^i | W_Q) \). Summing up (10) for \( i = 1, 2, \ldots, m \), we get

\[
\text{Term}_2 = \sum_{i=1}^{m} H \left( S_{i-1}^{[i-1]} | W_Q \right) \geq \sum_{i=1}^{m} \frac{i-1}{d-(k-m)} H(S_N^i | W_Q)
\]

C. Upper Bounding Term3

Finally, in order to bound Term3 in (7), we apply Lemma 3 with \( C = Q \) and \( (A, B) = ([i-1], \{i\}) \) (so that \( A \cup B = \{i\} \)) to get

\[
H \left( S_N^i | W_Q \right) + H \left( S_{i-1}^{[i-1]} | W_Q \right)
\]

where

\[
\text{Term}_3 = \sum_{i=1}^{m} H \left( S_{i-1}^{[i-1]} | W_Q \right) \leq \sum_{i=1}^{m} H \left( S_N^i | W_Q \right).
\]

Recall that \( P = [m] \), and having the repair data from every node in \( N \) to \( P \) (i.e., \( S_N^i \)), we can recover the contents for every node in \( P \) (i.e., \( W_P \)). Thus we have

\[
H \left( S_N^m | W_Q \right) \geq H(W_P | W_Q) = H(W_P | Q | W_Q)
\]

\[
\geq H(\text{Data} | W_B) = H(\text{Data}) - H(W_B) \geq F - (k-m) \alpha.
\]

Hence (13) together with (14) imply

\[
\text{Term}_3 = \sum_{i=1}^{m} H \left( S_{i-1}^{[i-1]} | W_Q \right) \leq \sum_{i=1}^{m} H \left( S_N^i | W_Q \right) - H(\text{Data}) + H(W_B) \geq F - (k-m) \alpha.
\]

D. Upper Bounding F

Next, we plug (9), (11) and (15) in (8), to get

\[
F \leq k\alpha - \text{Term}_1 - \text{Term}_2 + \text{Term}_3
\]

\[
\leq k\alpha - \left[ F - (k-m) \alpha - (d+1-k)H \left( S_{k+1}^| W_Q \right) \right]
\]

Finally, in order to bound Term2 in (7), we use Lemma 2 to lower bound each individual conditional entropy in the summation in Term2. Evaluating Lemma 2 for \( A = [i-1] \) and \( B = Q \), we get

\[
H \left( S_{[i-1]}^{[i-1]} | W_Q \right) \geq \frac{[i-1]}{d-|Q|} H(S_N^i | W_Q)
\]

where

\[
\sum_{i=1}^{m} H \left( S_{i-1}^{[i-1]} | W_Q \right) \leq \sum_{i=1}^{m} H \left( S_N^i | W_Q \right) - [F - (k-m) \alpha] = mH \left( S_N^i | W_Q \right) + (k-m) \alpha - F;
\]

where the last equality follows from the symmetry property, that is \( H \left( S_N^i | W_Q \right) = H \left( S_N^i | W_Q \right) \) for \( i = 1, 2, \ldots, m \).
where in (a) we used the fact that $S_B^{[m]} = S_{[m+1:k]}^{[m]}$ is a function of $W_B$. On the other hand, using the symmetry property in Definition 2, we have

$$\begin{align*}
(d-k+1)H\left(S_{k+1}^{[k]} | W_B\right) &= \sum_{i=k+1}^{d+1} H\left(S_i^{[k]} | W_B\right) \\
&\geq H\left(\{S_{k+1}^{[k]} : i \in [k+1 : d+1]\} | W_B\right) \\
&= H\left(\{S_{k+1}^{[k+1 : d+1]} : i \in [k]\} | W_B\right).
\end{align*}$$

Hence, adding (19) and (20) we get

$$\sum_{i=1}^{m} H\left(S_{i}^{[i-1]} | W_B\right) + (d-k+1)H\left(S_{k+1}^{[k]} | W_B\right) \geq H(\text{Data} | W_B) - H(W_B) \geq F - |B|\alpha.$$ 

This completes the proof of Lemma 1. \hfill \square

**Proof of Lemma 2.** Let $C = N \setminus (B \cup \{i\})$. It is clear that $A \subseteq C$, and $|C \cup B| = |N| - 1 = d$. Now consider any arbitrary $A' \subseteq C$ with $|A'| = |A|$. The symmetry property of the code (Definition 2) implies that $H(S_A^i | W_B) = H(S_{A'}^i | W_B)$. Hence, we can write

$$\begin{align*}
\frac{1}{|A|} H(S_A^i | W_B) &= \frac{1}{|A|} \left| \frac{C}{|A|}\right| \sum_{A' \subseteq C} \frac{A'}{|A'|} H(S_{A'}^i | W_B) \\
&= \frac{1}{|A|} \left| \frac{C}{|A|}\right| \sum_{A' \subseteq C} H(S_{A'}^i | W_B) \\
&\geq \frac{1}{|C|} \left| \frac{C}{|C|}\right| \sum_{C' \subseteq C} H(S_{C'}^i | W_B) \\
&= \frac{1}{|C|} H(S_{C}^i | W_B),
\end{align*}$$

where in (*) we used the conditional version of Han’s inequality [6]. Next note that, given $W_B$, all the repair data outgoing from nodes in $B$ are determined. Moreover, $S_C^i$ is just a dummy variable with zero entropy. Hence,

$$H(S_C^i | W_B) = H(S_C^i, S_B^i, S_A^i | W_B) = H(S_A^i | W_B).$$

Substituting (23) in (22), and incorporating $|C| = d - |B|$, we get the desired bound. \hfill \square

**Proof of Lemma 3.** First note that $S_A^A$ includes $S_A^B$ since $B \subseteq N$. Moreover, $S_A^A$ provides all repair data required to repair nodes in $A$. Hence $W_A$ can be reconstructed from $S_A^A$, from which the outgoing repair data $S_A^B$ can be found. Hence

$$H(S_A^A | W_C) = H(S_A^A, S_B^B, S_A^A | W_C),$$

and similarly

$$H(S_A^B | W_C) = H(S_A^A, S_B^B, S_A^B | W_C).$$

Therefore, using the inequality

$$H(X, Y | T) + H(X, Z | T) \geq H(X, Y, Z | T) + H(X | T),$$

we get

$$H(S_A^A | W_C) + H(S_A^B | W_C) \geq H(S_A^A, S_B^B, S_A^A | W_C) + H(S_A^A, S_B^B, S_A^B | W_C) \geq H(S_A^A, S_B^B, S_A^A | W_C) + H(S_A^A, S_B^B, S_A^B | W_C),$$

which implies the desired inequality. \hfill \square

**REFERENCES**


