

Towards Optimal Secure Distributed Storage Systems with Exact Repair

Ravi Tandon[†], SaiDhiraj Amuru^{*}, T. Charles Clancy^{*,†} and R. Michael Buehrer^{*}

^{*}Bradley Department of Electrical and Computer Engineering

[†]Hume Center for National Security and Technology,

Virginia Tech, Blacksburg, VA USA

Email: {tandonr, adhiraj, tcc, rbuehrer}@vt.edu

Abstract

Distributed storage systems in the presence of a wiretapper are considered. A distributed storage system (DSS) is parameterized by three parameters (n, k, d) , in which a file stored across n distributed nodes, can be recovered from any k out of n nodes. This is called as the reconstruction property of a DSS. If a node fails, any d out of $(n - 1)$ nodes help in the repair of the failed node so that the regeneration property of the DSS is preserved. For such a (n, k, d) -DSS, two types of wiretapping scenarios are investigated: (a) Type-I (node) adversary which can wiretap the data stored on any $l < k$ nodes; and a more severe (b) Type-II (repair data) adversary which can wiretap the contents of the repair data that is used to repair a set of l failed nodes over time. The focus of this work is on the practically relevant setting of *exact* repair regeneration in which the repair process must replace a failed node by its exact replica. We make new progress on several non-trivial instances of this problem which prior to this work have been open.

The main contribution of this paper is the optimal characterization of the secure storage-vs-exact-repair-bandwidth tradeoff region of a (n, k, d) -DSS, with $n \leq 4$ and any $l < k$ in the presence of both Type-I and Type-II adversaries. While the problem remains open for a general (n, k, d) -DSS with $n > 4$, we present extensions of these results to a $(n, n - 1, n - 1)$ -DSS, in presence of a Type-II adversary that can observe the repair data of any $l = (n - 2)$ nodes. The key technical contribution of this work is in developing novel information theoretic converse proofs for the Type-II adversarial scenario. From our results, we show that in the presence of Type-II attacks, the *only efficient* point in the storage-vs-exact-repair-bandwidth tradeoff is the MBR (minimum bandwidth regenerating) point. This is in sharp contrast to the case of a Type-I attack in which the storage-vs-exact-repair-bandwidth tradeoff allows a spectrum of operating points beyond the MBR point.

I. INTRODUCTION

A continuous rise in the volume of data managed across various systems, calls for new storage mechanisms that maintain this data reliably. Distributed storage is the default technique for storing data in all new generation applications. The data from a file is stored in a decentralized manner on several un-reliable nodes/disks that when collectively used are capable of recovering the entire file. To ensure robustness to disk failures, the simplest scheme is to replicate and store the data across several disks. For instance, the Google File System (GFS) and the Hadoop Distributed File System (HDFS) store 3 copies of the data across several nodes [1]. While replication is robust to failures, it is not a scalable strategy to store large volumes of data. As an alternative, erasure codes (for instance, Reed-Solomon codes) have been used by Facebook, OceanStore, RAID-6 [2] and others to introduce redundancy into the storage system. However, to repair a failed node, the entire file must be downloaded from the remaining alive nodes. Thus, the repair process for such codes can be excessively bandwidth intensive. The concept of regenerating codes for distributed storage was introduced in the seminal work by Dimakis *et al.* [3]. A typical distributed storage system (DSS) consists of n storage nodes each with a storage capacity of α (symbols or units of data) such that the entire file of size \mathcal{B} can be recovered by accessing any $k \leq n$ nodes. This is called as the reconstruction property of the DSS. Whenever a node fails, d nodes (where $k \leq d \leq n$) participate in the repair process by sending β units of data each. This procedure is termed as the regeneration of a failed node and β is often referred to as the per-node repair bandwidth.

In [3], the authors show that the in order to store a file of size \mathcal{B} , the parameters of a DSS must necessarily satisfy

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

Thus, in order to store a file of size \mathcal{B} , there exists a fundamental tradeoff between α (storage) and $d\beta$ (total repair bandwidth). Furthermore, it was shown that the reconstruction-regeneration requirements of a DSS can be equivalently formulated as a multicasting problem over an appropriately defined graph. This revelation along with the celebrated result of network coding for multicasting over graphs [4] were used to show that this tradeoff is indeed achievable. However, this tradeoff is in general achievable only for the functional-repair case. In functional repair, a failed node is replaced by a new node such that the resulting DSS has the same reconstruction and regeneration capabilities as before. In particular, the content of the repaired node may not necessarily be identical to the failed node even though the desirable properties of the DSS are preserved.

In contrast to functional repair, exact repair regeneration requires the repair process to replace a failed node with an identical new node. Exact repair is appealing for many practical applications where the data has to be stored intact. The file recovery

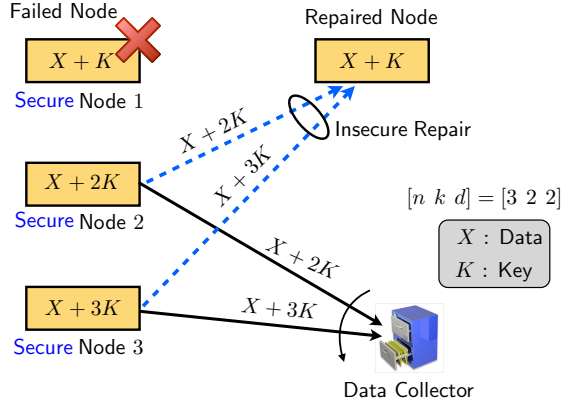


Fig. 1. $(3, 2, 2)$ -DSS with Type-I (node) security.

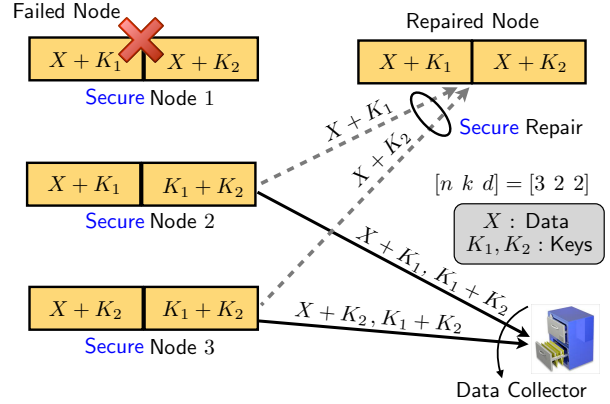


Fig. 2. $(3, 2, 2)$ -DSS with Type-II (repair) security.

process is also easier in this case compared to the functional repair scenario since the file 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 (depending on whether α or β is minimized first) have been studied extensively. They are the minimum storage regenerating (MSR) and the minimum bandwidth regenerating (MBR) points for which the explicit exact-repair regenerating codes have been developed (see [5], [6] and references therein). Beyond these results, Tian [7] has recently characterized the optimal exact-repair tradeoff for the $(4, 3, 3)$ -DSS where it has been shown that there is a gap between the optimal tradeoffs for functional repair and exact repair.

Sensitive data such as personal, health and financial records are increasingly being stored in a DSS. Securing such data from adversaries/eavesdroppers is necessary to ensure data secrecy for the users. Hence a DSS should be secure apart from satisfying the reconstruction and regenerating requirements. Two types of eavesdropping attacks can potentially occur in a DSS, (a) Type-I attack, in which the eavesdropper can wiretap the storage contents of l nodes in the DSS and (b) Type-II attack, in which the eavesdropper wiretaps the contents of the repair data (and thereby the storage content as well) of l nodes in the DSS. Throughout this paper, we assume that $l < k$ since k is the minimum number of nodes required to reconstruct the entire file of size \mathcal{B} . Else, if $l \geq k$, the eavesdropper can recover the file by using the reconstruction property of the DSS. Encryption and decryption i.e., cryptographic approaches to ensure data security have been deemed to offer less secrecy compared to other information/coding theoretic approaches [11]. Further, handling secure key management issues in a DSS are more complicated compared to developing information theoretically secure codes that offer the desired level of secrecy.

The concept of security in a DSS against Type-II attacks was introduced in [9] where, linear codes that achieve the optimal tradeoff region in the bandwidth limited regime were proposed. In the context of DSS, secure exact repair regenerating codes are beneficial as the functional repair process may reveal additional information such as coding coefficients that are used in the process of regenerating a functionally equivalent node [8]. Optimal exact repair codes that are secure against eavesdropping have been explored for the MSR and MBR points in [10]-[12]. The codes developed in [10] achieve the MBR point for all (n, k, d) configurations with any $l < k$. The MSR code in [10] was shown to be optimal for Type-I attacks while it was shown to be optimal for Type-II attacks with $l = 1$ in [11]. The maximum file size \mathcal{B} that can be securely stored using linear MSR codes with exact repair was studied in [12] and proved to be optimal for $d = n - 1$.

As an example, consider the $(3, 2, 2)$ -DSS and consider the setting in which the wiretapper can only read the contents of any $l = 1$ node (Type-I adversary). For this setting, a secure exact repair code for this problem is shown in Fig. 1. When the 1st node fails, the other nodes send their data contents to enable its repair. Using these two data symbols, the 1st node can recover its initial data contents (since two symbols can be recovered using two linearly independent combinations) thus satisfying the exact repair requirements. Since the eavesdropper is unaware of the secure key K , it cannot recover the data symbol X by wiretapping on any single node in the DSS.

Now, consider the same $(3, 2, 2)$ -DSS in the presence of a more powerful Type-II adversary that can observe the repair data of any $l = 1$ node. It is clear that the code in Figure 1 will leak out the entire data X during the repair of node 1, as the wiretapper can use $X + 2K, X + 3K$, (2 linear combinations of 2 symbols) to recover X . Thus, a different secure exact repair DSS that can handle a Type-II adversary is shown in Fig. 2. It is seen that the storage capacity should be increased to $\alpha = 2$ in this case compared to the Type-I attack (where $\alpha = 1$). Since the eavesdropper is not aware of the keys K_1 and K_2 , it cannot get any information about the message symbol X even if it observes the repair data of any one node.

From these schemes, the secure repair scheme in Figure 2 requires higher storage per-node ($\alpha = 2$) compared to the secure node scheme in Figure 1 ($\alpha = 1$ per-node). Thus, a natural question arises: does there exist a DSS with a smaller storage per

node $\alpha < 2$ that can store a file of size $\mathcal{B} = 1$, with repair bandwidth $\beta = 1$ while still preserving the security of the repair process? Or, equivalently, is there any other more efficient tradeoff pair (α', β') than $(\alpha, \beta) = (2, 1)$ that can store a file of size $\mathcal{B} = 1$ while ensuring secure exact repair?

In this paper, we answer this question in the negative through an information theoretic converse proof that shows that if secure and exact repair requirements are imposed, then the storage per-node cannot be smaller than $\alpha = 2$ and hence the scheme in Figure 2 is optimal. Equivalently, our result shows that the *only efficient* point in the (α, β) tradeoff for this example is the MBR point which corresponds to $\alpha = d\beta$. This result is extended and generalized for any $(n, n-1, n-1)$ -DSS and Type-I/Type-II adversaries for $l = (k-1) = (n-2)$.

We next summarize the main contributions of this paper.

- 1) We characterize and prove the optimal exact repair region that is achievable under Type-I and Type-II attacks for the (n, k, d) -DSS with $n \leq 4$ and $l < k$. Specifically, we look at $(3, 2, 2)$ -DSS with $l = 1$, $(4, 2, 3)$ -DSS with $l = 1$ and $(4, 3, 3)$ -DSS with $l = 1, 2$. The main contributions of this paper are the novel converse proofs that characterize the storage-vs-bandwidth tradeoff region against Type-II attacks for these distributed storage systems.
- 2) For the $(4, 3, 3)$ -DSS with $l = 1$ and Type-I attack, we leverage upon the characterization of the exact repair region obtained in the absence of an eavesdropper in [7]. Specifically, we introduce the additional security constraint in order to obtain the secure storage-vs-bandwidth tradeoff region in this case. Further, a novel code construction for this storage and bandwidth efficient region of the tradeoff curve is also introduced. This code construction also achieves the optimal exact repair region for $(4, 3, 3)$ in the absence of any eavesdroppers and is shown to be read efficient (in terms of the number of disk reads required to repair/recover a file [13]) when compared to the code given in [7].
- 3) These results are extended and proved for a general $(n, n-1, n-1)$ -DSS when any $l = n-2$ nodes are compromised in the DSS.

The rest of this paper is organized as follows. In Section II we describe the system and the eavesdropper's model. The main theorems that describe the optimal achievable storage-bandwidth tradeoff regions are given in Section III and are proved in the Appendix. The intuition behind the converse proofs is explained in Section IV. The coding schemes that achieve these optimal regions are presented in Section V. Finally conclusions are drawn in Section VI¹.

II. SYSTEM MODEL

A $(n, k, d, \alpha, \beta, \mathcal{B})$ -DSS consists of n storage nodes that store a file F of size \mathcal{B} across n nodes, with each node capable of storing up to α units of data. A data collector can connect to any $k < n$ nodes and must be able to reconstruct the entire file F . This is known as the MDS property of the DSS [3]. We focus on the scenario of single node failures in which at any given point any one node in the system could fail. For the repair of a failed node, any d out of the remaining $(n-1)$ *alive* nodes can be accessed by downloading up to $\beta \leq \alpha$ symbols to repair the failed node. The parameter $d\beta$ is referred to as the total repair bandwidth.

From an information theoretic perspective, the goal is to store a file F , whose entropy is \mathcal{B} , i.e.,

$$H(F) = \mathcal{B}. \quad (2)$$

We next introduce random variables corresponding to the data stored in the nodes and the random variables used in the file recovery and the repair process. Let W_i denote the content that is stored at node i , for $i = 1, 2, \dots, n$. Hence, the storage constraint implies

$$H(W_i) \leq \alpha, \quad i = 1, 2, \dots, n. \quad (3)$$

Due to the MDS property (i.e., the file B must be reconstructed from any $k \leq n$ nodes), we also have

$$H(F|W_{\{k\}}) = 0, \quad (4)$$

where $W_{\{k\}}$ is the data stored in any subset of k storage nodes. Next, we consider the repair of a failed node j from any d remaining nodes. We denote the data sent by node i to repair node j by S_{ij} . Due to the repair bandwidth constraint, we have

$$H(S_{ij}) \leq \beta, \quad (5)$$

and for *exact repair* of node j from the repair data of d nodes, we also have

$$H(W_j|S_{r_1j}, S_{r_2j}, \dots, S_{r_dj}) = 0, \quad \{r_i\}_{i=1}^d \in [1, n] \neq j. \quad (6)$$

Finally, we note that any repair data sent by node i , i.e., S_{ij} is a function of the data stored in node i , i.e., $H(S_{ij}|W_i) = 0$.

¹Parts of this paper will be presented at ITA-2014 [16] and ICC-2014 [17].

A. Adversarial Models

We next formalize the Type-I and Type-II security constraints, each corresponding to different capabilities of the wiretapping adversary.

- Type-I (node) security: in this setting, the wiretapper can read the storage content of any $l < k$ nodes. Hence, we require that the information leakage by revealing the data of any l storage nodes must be zero, i.e.,

$$I(F; W_{\{l\}}) = 0, \quad (7)$$

where $W_{\{l\}}$ is the data stored in any subset of l nodes.

- Type-II (repair) security: in this setting, the wiretapper can read the repair data of any $l < k$ nodes over time. Hence, for the repair of any l nodes to be secure, we require

$$I(F; S_{n_1}, S_{n_2}, \dots, S_{n_l}) = 0, \quad (8)$$

where S_{n_i} is the data (downloaded from d repair nodes) used in the repair of node n_i . Furthermore, it is worth noting that any DSS that is secure under Type-II secrecy constraint is also secure under Type-I secrecy constraint but the reverse statement is not true in general.

To illustrate these constraints, consider the $(n, k, d) = (3, 2, 2)$ -DSS. Then, the constraints regarding file-regeneration and exact repair are as follows:

File Regeneration:

$$H(F|W_1, W_2) = H(F|W_1, W_3) = H(F|W_2, W_3) = 0.$$

Exact Repair:

$$H(W_1|S_{21}, S_{31}) = H(W_2|S_{12}, S_{32}) = H(W_3|S_{13}, S_{23}) = 0.$$

Repair data functionality:

$$H(S_{12}, S_{13}|W_1) = H(S_{21}, S_{23}|W_2) = H(S_{31}, S_{32}|W_1) = 0.$$

For this example, the parameter l can be either 0 and 1. The non-trivial case is $l = 1$, for which the Type-I constraints can be written as:

$$I(F; W_1) = I(F; W_2) = I(F; W_3) = 0.$$

whereas the Type-II security constraints are:

$$I(F; S_{21}, S_{31}) = I(F; S_{12}, S_{32}) = I(F; S_{13}, S_{23}) = 0.$$

We next define the Type I (respectively Type-II) secrecy capacity of a DSS as the maximum file size that can be stored under the constraints placed on storage (3), file regeneration (4), repair bandwidth constraint (5), exact repair (6) and Type-I (respectively Type-II) secrecy constraint. Formally, we define the Type-I secrecy capacity as

$$\mathcal{B}_I^S = \max_{(3)-(6),(7)} H(F), \quad (9)$$

and the Type-II secrecy capacity as

$$\mathcal{B}_{II}^S = \max_{(3)-(6),(8)} H(F). \quad (10)$$

The study of distributed storage systems in the presence of a passive wiretapper was initiated in [9]. It was shown that for any (n, k, d) -DSS with either Type-I or Type-II secrecy constraint (characterized by the parameter l), the following is an upper bound on the maximum secure file size \mathcal{B}^S :

$$\mathcal{B}^S \leq \sum_{i=l}^{k-1} \min(\alpha, (d-i)\beta). \quad (11)$$

Intuitively, this result can be interpreted as follows. In the presence of a wiretapper, the maximum file size a DSS can store must necessarily reduce (compared to (1)) because l nodes are wiretapped. Since l nodes are compromised, at most $(k-l)$ nodes can help a data collector in recovering the entire file while keeping it secure from l nodes. Hence the summation is over $(k-l)$ nodes as opposed to k nodes [10]. To illustrate by an example, consider the $(3, 2, 2)$ -DSS with $l = 1$ for which the

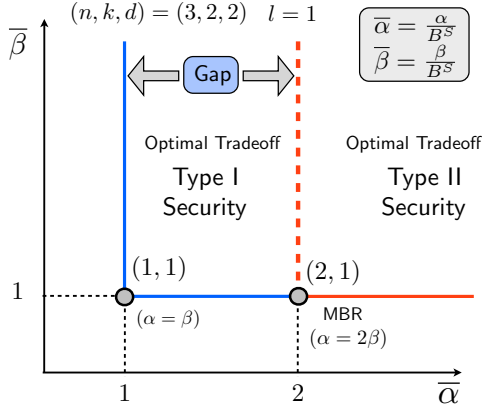


Fig. 3. Secure (α, β) tradeoff for $(3, 2, 2)$ -DSS and $l = 1$.

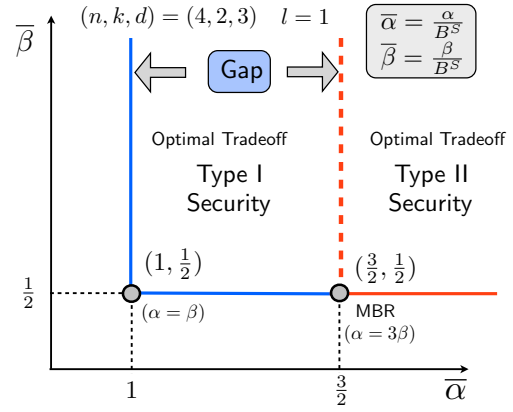


Fig. 4. Secure (α, β) tradeoff for $(4, 2, 3)$ -DSS and $l = 1$.

upper bound in (11) simplifies to

$$\mathcal{B}^S \leq \min(\alpha, \beta). \quad (12)$$

Normalizing (12) throughout by \mathcal{B}^S , and defining $\bar{\alpha} = \alpha/\mathcal{B}^S$, $\bar{\beta} = \beta/\mathcal{B}^S$, we can equivalently write $1 \leq \min(\bar{\alpha}, \bar{\beta})$. This bound gives rise to a region in the $(\bar{\alpha}, \bar{\beta})$ plane and is shown in Fig. 3 with the corner point $(\bar{\alpha}, \bar{\beta}) = (1, 1)$. Furthermore, it is not difficult to show that this tradeoff is achievable under Type-I security constraints by showing that it is possible to store a file of size $\mathcal{B}^S = 1$ using $\alpha = 1$ and $\beta = 1$. The scheme is shown in Figure 1. Hence, the upper bound in (12) along with the scheme of Figure 1 imply that the optimal (α, β) -tradeoff for $(3, 2, 2)$ -DSS, $l = 1$ in presence of a Type-I adversary is given by

$$\mathcal{B}_I^S \leq \min(\alpha, \beta). \quad (13)$$

As for the more stringent Type-II security is concerned, it has been shown by Shah *et al* in [10], using the Product-Matrix framework at the MBR point ($\alpha = d\beta$), that it is possible to store a file of size $\mathcal{B}^S = 1$ using $\alpha = 2$ and $\beta = 1$. This scheme (shown in Figure 2) shows that the following (α, β) -tradeoff region is achievable:

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{2}, \beta\right). \quad (14)$$

On the other hand, the bound in (12) implies that the optimal (α, β) -tradeoff region for the Type-II secrecy constraint is contained in:

$$\mathcal{B}_{II}^S \leq \min(\alpha, \beta). \quad (15)$$

Thus, there exists a significant gap between these two regions for the Type-II secrecy constraint. Furthermore, it is not clear whether the gap is due to the weakness of the upper bound or if there exists an improved achievable scheme that can close this gap. In this paper, we show that this gap is *fundamental* to the secure exact repair problem and the upper bound can be improved. By deriving a novel information theoretic converse, we show that for the $(3, 2, 2)$ -DSS, $l = 1$ and Type-II secrecy constraint, we show that the optimal secure storage-repair-bandwidth tradeoff is given by

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{2}, \beta\right). \quad (16)$$

III. MAIN RESULTS

In this section, we outline the main theorems that describe the optimal secure storage-vs-exact repair-bandwidth tradeoffs (in short referred to as (α, β) -tradeoff region) under the exact repair requirement.

Theorem 1: The optimal (α, β) -tradeoff regions for $(3, 2, 2)$ -DSS with $l = 1$ under exact repair are given by:

$$\mathcal{B}_I^S \leq \min(\alpha, \beta), \quad (17)$$

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{2}, \beta\right). \quad (18)$$

The converse proof for (17) follows directly from (11) and is therefore omitted. The main contribution is the converse proof for the bound (18) which is given in Section IV.

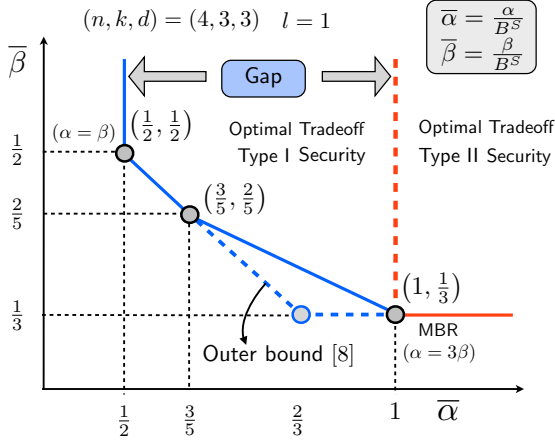


Fig. 5. Secure (α, β) tradeoff for $(4, 3, 3)$ -DSS and $l = 1$.

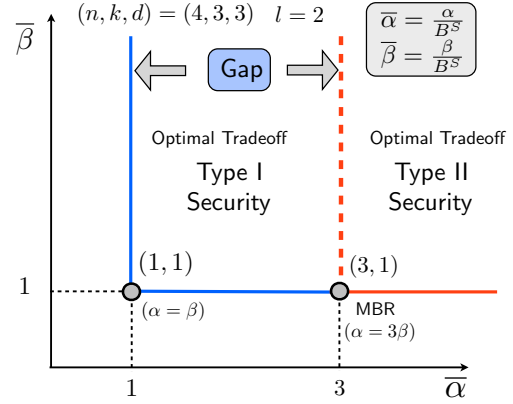


Fig. 6. Secure (α, β) tradeoff for $(4, 3, 3)$ -DSS and $l = 2$.

Theorem 2: The optimal (α, β) -tradeoff regions for $(4, 2, 3)$ -DSS with $l = 1$ under exact repair are given by:

$$\mathcal{B}_I^S \leq \min(\alpha, 2\beta) \quad (19)$$

$$\mathcal{B}_{II}^S \leq \min\left(\frac{2\alpha}{3}, 2\beta\right). \quad (20)$$

The bound in (19) follows directly from (11). The converse proof for (20) is given in Section VII-A in the Appendix.

Theorem 3: The optimal (α, β) -tradeoff regions for $(4, 3, 3)$ -DSS with $l = 1$ under exact repair are given by:

$$\mathcal{B}_I^S \leq \min\left(\min(\alpha, 2\beta) + \min(\alpha, \beta), \frac{\alpha + 6\beta}{3}\right) \quad (21)$$

$$\mathcal{B}_{II}^S \leq \min(\alpha, 3\beta). \quad (22)$$

The first term in (21) is directly from the upper bound given in (11). The converse proof for the second term in (21) is one novel contribution of this paper. This proof is leveraged off the proof of exact repair given for the $(4, 3, 3)$ -DSS without any secrecy constraints in [7]. This is presented in Section VII-B1 in the Appendix. A novel converse proof for the bound in (22) under the Type-II constraints is given in Section VII-B2 in the Appendix.

Theorem 4: The optimal (α, β) -tradeoff regions for $(4, 3, 3)$ -DSS with $l = 2$ under exact repair are given by:

$$\mathcal{B}_I^S \leq \min(\alpha, \beta) \quad (23)$$

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{3}, \beta\right). \quad (24)$$

The bound in (23) follows directly from (11). Hence the proof is skipped. The converse proof for (24) is given in Section VII-C in the Appendix.

Theorem 5: The optimal (α, β) -tradeoff regions for $(n, n-1, n-1)$ -DSS with $l = n-2$ under exact repair are given by:

$$\mathcal{B}_I^S \leq \min(\alpha, \beta) \quad (25)$$

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{n-1}, \beta\right). \quad (26)$$

This is an extension of the Theorems 1 and 4 to a general $(n, n-1, n-1)$ -DSS. This is the worst case scenario with respect to the $(n, n-1, n-1)$ -DSS since $l = k-1$ nodes are compromised and hence the file size that can be stored securely is the minimum among all possible $l < k$ scenarios. Similar to the Theorems 1-4 the proof of (25) is omitted as it is derived from (11). The proof for (26) is a novel extension of the converse proofs given for (18) (24) and is provided in the Appendix in Section VII-D.

Figs. 3-7 show the optimal (α, β) tradeoff regions described by Theorems 1-5. From these theorems, it is seen that the only efficient point in the optimal (α, β) -tradeoff region for the Type-II constraints is the MBR point where $\alpha = d\beta$. However, this is different from the optimal tradeoff-region achievable under Type-I constraints as seen in these results. Thus there is a gap between the optimal regions achievable under these two constraints i.e., the file size that can be securely stored under Type-II constraints is lower than the file size achieved under Type-I constraint.

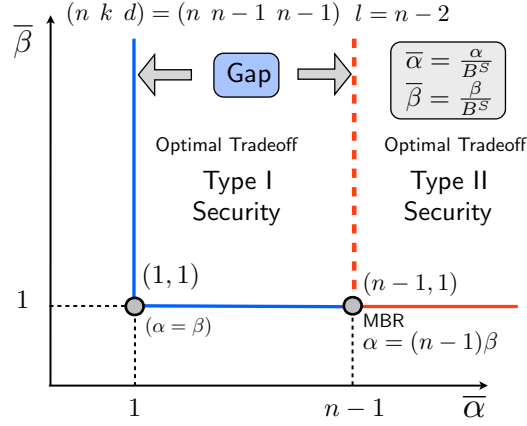


Fig. 7. Secure (α, β) tradeoff for $(n, n-1, n-1)$ -DSS and $l = n-2$.

Further notice that the gap between these two constraints with respect to the secure file size increases as n increases in the DSS. For the Type-I attacks, when $l = n-2$ and $k = n-1$, only one node can securely store the file and hence the maximum file size that can be securely stored is given by (25) which is independent of n , while it decreases due to the wiretapping process as seen in (26) (a more stringent constraint on the DSS).

IV. PROOF OF THEOREM 1 AND INTUITION BEHIND CONVERSE PROOFS

In this section, we present the proof of the Theorem 1 i.e., $(3, 2, 2)$ -DSS with $l = 1$ in order to explain the intuition behind the novel converse proofs that establish the secure storage-vs-bandwidth repair tradeoff region against Type-II attacks. The converse proofs for the other Theorems are presented in the Appendix.

Let F be the random variable that denotes the file stored on the DSS. For the $(3, 2, 2)$ -DSS with $l = 1$, the bound in (11) (corresponding to Type-I attack) is given by

$$H(F) \leq \min(\alpha, \beta). \quad (27)$$

In the Type-I attack, the stored content of any $l = 1$ node must not reveal any information about the file F . For instance, the content stored in node 1 must not reveal any information about F , i.e., the DSS must satisfy $I(F; W_1) = 0$. Using such Type-I security constraints, one can readily show that $H(F) \leq \min(\alpha, \beta)$ as follows:

$$H(F) = H(F|W_1) \quad (28)$$

$$= H(F, W_1) - H(W_1) \quad (29)$$

$$\leq H(F, W_1, W_2) - H(W_1) \quad (30)$$

$$= H(W_1, W_2) + \underbrace{H(F|W_1, W_2)}_{=0 \text{ (file regeneration)}} - H(W_1) \quad (31)$$

$$= H(W_1, W_2) - H(W_1) \quad (32)$$

$$= H(W_2|W_1) \quad (33)$$

Next, the conditional entropy $H(W_2|W_1)$ can be expanded in two distinct ways as follows:

(a) By bounding this term as

$$H(W_2|W_1) \leq H(W_2) \leq \alpha, \quad (34)$$

gives the bound $H(F) \leq \alpha$.

(b) On the other hand, we can also bound this term as

$$H(W_2|W_1) \leq H(W_2, S_{32}|W_1) \quad (35)$$

$$= H(S_{32}|W_1) + \underbrace{H(W_2|S_{32}, W_1)}_{=0 \text{ (repair of node 2)}} \quad (36)$$

$$= H(S_{32}|W_1) \quad (37)$$

$$\leq H(S_{32}) \leq \beta, \quad (38)$$

which gives the bound $H(F) \leq \beta$.

While this bound $H(F) \leq \min(\alpha, \beta)$ is tight for the case of Type-I attack, it turns out to be strictly sub-optimal for the case of more severe Type-II attack. In particular, for the Type-II attack, in order to secure the DSS from an eavesdropper that wiretaps the repair process of any single node ($l = 1$), the DSS must satisfy the following three security constraints:

$$I(F; S_{21}, S_{31}) = I(F; S_{12}, S_{32}) = I(F; S_{13}, S_{23}) = 0, \quad (39)$$

to indicate the secure repair of nodes 1, 2 and 3 respectively. If the eavesdropper can read (S_{21}, S_{31}) , i.e., the repair data for W_1 , then two cases can arise:

- if there is no redundancy in the repair process ($\alpha = d\beta$), then it is clear that if the stored data is secure, then the repair data will also be secure.
- However, for $\alpha < d\beta$, more information about the file could be leaked in general. This aspect is not captured by the proof for the Type-I attack and one needs to carefully deal with the security of the repair process.

We have the following sequence of bounds for secure repair of node 1,

$$H(F) = H(F|S_{21}, S_{31}) \quad (40)$$

$$= H(F, S_{21}, S_{31}) - H(S_{21}, S_{31}) \quad (41)$$

$$\leq H(F, W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}) \quad (42)$$

$$= H(W_2, S_{21}, S_{31}) + \underbrace{H(F|W_2, S_{21}, S_{31})}_{=0 \text{ (file regeneration)}} - H(S_{21}, S_{31}) \quad (43)$$

$$= H(W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}) \quad (44)$$

$$= H(W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}, W_1) + \underbrace{H(W_1|S_{21}, S_{31})}_{=0 \text{ (repair of node 1)}} \quad (45)$$

$$= H(W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}, W_1) \quad (46)$$

$$= H(W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}, W_1, S_{12}, S_{13}) + \underbrace{H(S_{12}, S_{13}|W_1, S_{21}, S_{31})}_{=0} \quad (47)$$

$$\leq H(W_2, S_{21}, S_{31}) - H(S_{21}, S_{31}, S_{12}, S_{13}), \quad (48)$$

where (43) follows from the file regeneration requirement from nodes (1, 2), (47) follows from the fact that (S_{12}, S_{13}) are functions of W_1 . Similarly, for secure repair of node 2, we have

$$H(F) \leq H(W_1, S_{12}, S_{32}) - H(S_{12}, S_{32}, S_{21}, S_{23}). \quad (49)$$

Adding (48) and (49), we have

$$2H(F) \leq H(W_2, S_{21}, S_{31}) + H(W_1, S_{12}, S_{32}) - H(S_{21}, S_{31}, S_{12}, S_{13}) - H(S_{12}, S_{32}, S_{21}, S_{23}) \quad (50)$$

$$= H(W_2, S_{21}, S_{31}) + H(W_1, S_{12}, S_{32}) - 2H(S_{21}, S_{12}) - H(S_{31}, S_{13}|S_{21}, S_{12}) - H(S_{23}, S_{32}|S_{21}, S_{12}) \quad (51)$$

$$\leq H(W_2, S_{21}, S_{31}) + H(W_1, S_{12}, S_{32}) - H(S_{21}, S_{12}) - H(S_{12}, S_{21}, S_{13}, S_{31}, S_{23}, S_{32}) \quad (52)$$

$$= H(W_2, S_{21}, S_{31}) + H(W_1, S_{12}, S_{32}) - H(S_{21}, S_{12}) - H(S_{12}, S_{21}, S_{13}, S_{31}, S_{23}, S_{32}, W_2) \quad (53)$$

$$\leq H(W_2, S_{21}, S_{31}) + H(W_1, S_{12}, S_{32}) - H(S_{21}, S_{12}) - H(W_2, S_{21}, S_{31}) \quad (54)$$

$$= H(W_1, S_{12}, S_{32}) - H(S_{21}, S_{12}) \quad (55)$$

$$\leq H(W_1, W_3, S_{12}, S_{32}) - H(S_{21}, S_{12}) \quad (56)$$

$$= H(W_3, S_{12}) + \underbrace{H(W_1, S_{32}|W_3, S_{12})}_{=0} - H(S_{21}, S_{12}) \quad (57)$$

$$= H(W_3, S_{12}) - H(S_{21}, S_{12}) \quad (58)$$

$$\leq H(W_3, S_{12}) - H(S_{12}) \quad (59)$$

$$\leq H(W_3) + H(S_{12}) - H(S_{12}) \quad (60)$$

$$= H(W_3) \quad (61)$$

$$\leq \alpha. \quad (62)$$

where

- (53) is obtained by noting that W_2 is a function of the repair data S_{12}, S_{32} ,
- (54) follows from the fact that conditioning reduces entropy,
- (58) is obtained by noting the fact that (W_1, W_2) can be obtained from (W_3, S_{12}) . To note this, first observe that (S_{32}, S_{31})

are functions of W_3 , hence using (S_{32}, S_{12}) , we can repair W_2 (and subsequently obtain S_{21}). Thus, using (S_{21}, S_{31}) , we can repair W_1 .

Hence from (62), we have $H(F) \leq \frac{\alpha}{2}$. Thus along with the bound $H(F) \leq \beta$, we have the proof for

$$H(F) \leq \min\left(\frac{\alpha}{2}, \beta\right) \quad (63)$$

Interestingly, this tradeoff between (α, β) only has one efficient point (see Fig. 3), corresponding to the case when $\alpha = 2\beta$ (or the MBR point for the $(3, 2, 2)$ DSS). Extending this approach to more general (n, k, d) -DSS is far from trivial. The proof presented above essentially couples two security constraints corresponding to the secure repair of nodes 1 and 2. In general, we believe that it suffices to consider the coupling between $\binom{k}{l}$ security constraints, (each constraint corresponding to the secure repair of l nodes). As we show in the Appendix, this statement turns out to be true for the $(4, 2, 3)$, $(4, 3, 3)$ and the $(n, n-1, n-1)$ DSS and helps us in establishing the corresponding optimal tradeoff regions.

V. ACHIEVABILITY PROOFS

In this section, we present coding schemes that achieve the bounds mentioned in Theorems 1-5. We first present the achievability of the MBR point under Type-II secrecy constraint. Later, the coding schemes that achieve a various spectrum of points beyond the MBR point in the presence of a Type-I adversary are presented.

A. Achievability under Type-II security constraint

As mentioned earlier, the MBR point is the only valid point in the (α, β) tradeoff region in the presence of Type-II adversary. The MBR point is defined by the (α, β) relationship $\alpha = d\beta$ [3]. Substituting this in (11), we get that the optimal secure file size \mathcal{B}^S must satisfy

$$\mathcal{B}^S \leq \left(kd - \binom{k}{2}\right)\beta - \left(ld - \binom{l}{2}\right)\beta. \quad (64)$$

Notice that (64) is identical to the tradeoff regions specified by Theorems 1-5 with the corresponding values of l . This equivalence is further explained below.

Secure codes that achieve this MBR point for a general (n, k, d) -DSS with any $l < k$ compromised nodes have been described in [10] for both Type-I and Type-II attacks (since $\alpha = d\beta$ at the MBR point, the eavesdropper does not get any additional information by wiretapping the repair process i.e., Type-I and Type-II attacks are equivalent at the MBR point). For example, in the $(3, 2, 2)$ -DSS with $l = 1$, the MBR point in (64) is simplified as $\mathcal{B}_{II}^S = \beta; \alpha = 2\beta$, which is the $\alpha = 2, \beta = 1, \mathcal{B}_{II}^S = 1$ point shown in Fig. 3, whose achievability is given in Fig. 2. In the $(4, 3, 3)$ -DSS with $l = 1$, (64) is given by $\mathcal{B}_{II}^S = 3\beta; \alpha = 3\beta$, which is also the $\alpha = 3, \beta = 1, \mathcal{B}_{II}^S = 3$ point shown in Fig. 5. Along similar lines, with $l = 2$, the MBR point in (64) is given by $\mathcal{B}_{II}^S = \beta; \alpha = 3\beta$, i.e., the $\alpha = 3, \beta = 1, \mathcal{B}_{II}^S = 1$ point shown in Fig. 6. For the $(4, 2, 3)$ -DSS, with $l = 1$, we have $\alpha = 3, \beta = 1$ and $\mathcal{B}_{II}^S = 2$, i.e., MBR point in Fig. 4. For the $(n, n-1, n-1)$ -DSS, with $l = n-2$, we have $\alpha = n-1, \beta = 1$ and $\mathcal{B}_{II}^S = 1$, i.e., the MBR point in Fig. 7. For more information on the codes that achieve the MBR point, please see [10].

B. Achievability under Type-I security constraint

At the MBR point, the Type-I and Type-II security constraints are equivalent and hence the MBR point is achievable under both these constraints [10]. Beyond this MBR point, a spectrum of points in the secure storage-vs-exact-repair-bandwidth tradeoff region are achievable in the presence of a Type-I adversary. We next present the various coding schemes that achieve these points.

1) $(\alpha, \beta) = (1, 1), \mathcal{B}_I^S = 1$ in a $(3, 2, 2)$ -DSS, $l = 1$: The coding scheme that achieves this point is shown in Fig. 1. In this scheme, $X \in \mathbb{F}_q$ and K the secure key, is a random variable that is uniformly distributed over \mathbb{F}_q such that $q > 3$. From Fig. 1, it is easy to see that the repair and the reconstruction processes are straightforward.

2) $(\alpha, \beta) = (2, 1), \mathcal{B}_I^S = 2$ in a $(4, 2, 3)$ -DSS, $l = 1$: This point corresponds to the normalized (α, β) pair i.e., $(\bar{\alpha}, \bar{\beta}) = (1, \frac{1}{2})$ in the Fig. 4. Let a_1, a_2 denote the message symbols to be stored in the DSS. Choose keys k_1 and k_2 to securely store the data on the DSS. Here $a_1, a_2, k_1, k_2 \in \mathbb{F}_5$. Let x_1, x_2, x_3, x_4 denote 4 linearly independent combinations of the message and key symbols. They are stored in the DSS as shown in Table I. Notice that the eavesdropper cannot infer any information by observing the data contents on any single storage node. Further, the repair and reconstruction processes are straightforward.

3) $(\alpha, \beta) = (1, 1), \mathcal{B}_I^S = 2$ in a $(4, 3, 3)$ -DSS, $l = 1$: This point corresponds to the normalized (α, β) pair i.e., $(\bar{\alpha}, \bar{\beta}) = (\frac{1}{2}, \frac{1}{2})$ in the Fig. 5. Below, we present a coding scheme that can store a file of size $\mathcal{B}^S = 2$ securely in a $(4, 3, 3)$ -DSS with $l = 1$ when each node stores $\alpha = 1$ unit of data and sends $\beta = 1$ unit of data for the repair of failed nodes.

Let $a_1, a_2 \in \mathbb{F}_2$ denote the message symbols to be stored in the DSS. Choose a key k that is uniformly distributed over \mathbb{F}_2 in order to securely store the data on the DSS. The resulting DSS is shown in Table. II. Since the eavesdropper is unaware

TABLE I
A (4, 2, 3) SECURE EXACT REPAIR CODE FOR $(\alpha, \beta) = (2, 1)$, $\mathcal{B}^S = 2$.

	first symbol	second symbol
node 1	x_1	x_2
node 2	x_3	x_4
node 3	$x_1 + x_3$	$x_2 + x_4$
node 4	$x_1 + x_4$	$x_2 + x_3$

TABLE II
A (4, 3, 3) SECURE EXACT REPAIR CODE FOR $(\alpha, \beta) = (1, 1)$, $\mathcal{B}^S = 2$.

node 1	node 2	node 3	node 4
$a_1 + k$	$a_1 + k$	$a_1 + a_2 + k$	k

of the key k , it cannot decode either a_1 or a_2 from any of the nodes in the DSS, thereby securing the DSS against Type-I attacks.

Further, it is easy to see that the file (a_1, a_2) can be recovered from any $k = 3$ nodes (by using 3 linearly independent combinations of 3 symbols a_1, a_2 and k). Consider the repair process of the 1st node. The nodes 2, 3 and 4 send their data contents in order to repair this failed node. See that a_1 and k can be recovered from these three data symbols and that $a_1 + k$, the contents of the 1st node can be restored exactly.

4) $(\alpha, \beta) = (3, 2)$, $\mathcal{B}_I^S = 5$ in a (4, 3, 3)-DSS, $l = 1$: This point corresponds to the normalized (α, β) pair i.e., $(\bar{\alpha}, \bar{\beta}) = (\frac{3}{5}, \frac{2}{5})$ in the Fig. 5. This can be shown through a coding scheme that can securely store a file of size $\mathcal{B}^S = 5$ over $n = 4$ nodes where each node stores 3 units of data and sends 2 units of data for repair of other nodes. Let $(a_1, a_2, a_3, a_4, a_5) \in \mathbb{F}_q$ denote the message symbols to be stored in the DSS. Let (k_1, k_2, k_3) be the secure keys that are used by the DSS to keep the data secure from the eavesdropper. It is assumed that these keys are uniformly distributed over \mathbb{F}_q for some $q > 8$.

Let $A = [a_1, a_2, a_3, a_4, a_5]^T$ and $K = [k_1, k_2, k_3]^T$ represent the message and the secure key vectors. Let

$$X = [A \ K] \begin{bmatrix} G_A \\ G_K \end{bmatrix} = G_A A + G_K K, \quad (65)$$

be the 1×12 vector of symbols to be stored on a DSS. Let $\begin{bmatrix} G_A \\ G_K \end{bmatrix}$ be a 8×12 Vandermonde matrix such that the i th row is defined as $[1, i, i^2, i^3, \dots, i^{11}]$ where $i = 1, 2, \dots, 8$. Here, G_A is a 5×12 sub matrix obtained from the above Vandermonde matrix given by

$$\begin{bmatrix} 1, 1, 1^2, \dots, 1^{11} \\ 1, 2, 2^2, \dots, 2^{11} \\ \vdots \\ 1, 5, 5^2, \dots, 5^{11} \end{bmatrix} \quad (66)$$

and G_K is a 3×12 sub matrix given by

$$\begin{bmatrix} 1, 6, 6^2, \dots, 6^{11} \\ 1, 7, 7^2, \dots, 7^{11} \\ 1, 8, 8^2, \dots, 8^{11} \end{bmatrix} \quad (67)$$

Since we use a 8×12 Vandermonde matrix, we define all the elements over \mathbb{F}_9 . The resulting elements of X are stored in the DSS as shown in Table III.

In order to achieve secrecy, i.e., the eavesdropper should not get any information about $\{a_j\}_{j=1}^5$ by observing the contents of any single node, the following condition must hold true.,

$$I(A; [X]_3) = 0 \quad (68)$$

where $[X]_3$ is any 3 elements of the vector X (since any 3 elements of X are linearly independent combinations of A and K). We can represent $[X]_3$ as

$$[X]_3 = [A \ K] \begin{bmatrix} \hat{G}_A \\ \hat{G}_K \end{bmatrix}_{8 \times 3}, \quad (69)$$

TABLE III
A (4, 3, 3) SECURE EXACT REPAIR CODE FOR $(\alpha, \beta) = (3, 2)$, $\mathcal{B}^S = 5$.

	first symbol	second symbol	third symbol
node 1	x_1	x_2	$x_9 = x_6 + x_7$
node 2	x_3	x_4	$x_{10} = x_1 + x_8$
node 3	x_5	x_6	$x_{11} = x_2 + x_3$
node 4	x_7	x_8	$x_{12} = x_4 + x_5$

TABLE IV
REPAIR PROCEDURE OF A (4, 3, 3) SECURE EXACT REPAIR CODE WHEN NODE 1 FAILS.

	first symbol	second symbol
node 2	x_3	$x_{10} = x_1 + x_8$
node 3	x_6	$x_{11} = x_2 + x_3$
node 4	x_7	x_8

where \hat{G}_A, \hat{G}_K are sub matrices of $\begin{bmatrix} G_A \\ G_K \end{bmatrix}$ that correspond to $[X]_3$. Notice that these sub matrices are also Vandermonde matrices. Thus we have the following set of inequalities,

$$\begin{aligned}
 I(A; [X]_3) &= I(A; \hat{G}_A A + \hat{G}_K K) \\
 &= H(\hat{G}_A A + \hat{G}_K K) - H(\hat{G}_A A + \hat{G}_K K | A) \\
 &= H(\hat{G}_A A + \hat{G}_K K) - H(\hat{G}_K K) \\
 &\leq 3 - \text{rank}(\hat{G}_K) \\
 &= 3 - 3 = 0,
 \end{aligned}$$

The last two inequalities follow from the fact that the sub-Vandermonde matrices have a maximum rank of 3.

Repair Process: Since this is a cyclic construction, every node is equivalent. Without loss of generality, it suffices to consider the repair of any single node. Consider the case when node 1 fails and the other nodes participate in the repair process. Since $\beta = 2$, each node sends 2 symbols to aid the repair of the 1st node. The repair process of node 1 is shown in Table IV. The first symbol of node 1 is recovered by combining the second symbols sent by nodes 2 and 4. The second symbol of node 1 is obtained by a combination of the first symbol sent by node 2 and the second symbol sent by node 3. Finally, the third symbol of node 1 is recovered by a combination of the first symbols sent by nodes 3 and 4. Hence, upon receiving the 6 symbols, it can be seen that the 1st node can exactly recover its original symbols, thereby satisfying the requirements of an exact repair code.

Remark 1: This code construction can be easily extended to the exact repair case in the absence of an eavesdropper. In the absence of an eavesdropper (no requirement for keys), this code can store a file size of 8 ($\{x_j\}_{j=1}^8$, no linear combinations required) in $n = 4$ nodes with each node storing $\alpha = 3$ units of data and sending $\beta = 2$ units for the repair process. Thus this code achieves a normalized storage-bandwidth tradeoff point $(\frac{3}{8}, \frac{1}{4})$ which is shown to be optimal in [7]. Further, this code is also efficient in terms of the disk reads required for a repair process compared to the code proposed for the exact repair of a (4, 3, 3)-DSS in [7]. It is seen that the number of disk reads required in this proposed code is 6 (6 symbols obtained directly from the nodes) while it is 9 in the case of the code proposed in [7] (see [7] for more details). This behavior is akin to the codes proposed in [13] that are designed to be data-read and download efficient during a repair process.

5) $(\alpha, \beta) = (1, 1)$, $\mathcal{B}^S = 1$ in a (4, 3, 3)-DSS, $l = 2$: Let $a \in \mathbb{F}_q$ denote the message symbol to be stored on the DSS. Since $l = 2$, any two nodes need to be secure from the eavesdropping attack. Hence we choose two keys k_1, k_2 that are uniformly distributed in \mathbb{F}_q in order to securely store a . Let $X = [x_1, x_2, x_3, x_4]^T$ denote the storage contents of the 4 nodes in the DSS such that

$$X = [G_A \ G_K] \begin{bmatrix} a \\ k_1 \\ k_2 \end{bmatrix} = G_A a + G_K \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}, \quad (70)$$

where G_A and G_K are chosen such that $\text{rank}([G_A \ G_K]) = 3$ and $\text{rank}(G_K) = 2$ (for instance Vandermonde matrices similar to Section V-B4).

$\{x_i\}_{i=1}^4$ are the linear combinations of the message symbol a and the secure keys k_1, k_2 . The rank of the concatenated

matrix $[G_A \ G_K]$ ensures that the file can be recovered from any 3 out of the 4 nodes in the DSS (3 linearly independent combinations of 3 symbols), thereby satisfying the reconstruction property of the DSS. By following the similar arguments in Section V-B4, it can be shown that the rank of the matrix G_K which is 2, ensures that the storage contents on any two nodes are secure from a Type-I attack by an eavesdropper.

Remark 2: The above coding scheme can be extended to a general $(n, n-1, n-1)$ -DSS with $l = n-2$ to achieve the $(\alpha, \beta) = (1, 1)$, $\mathcal{B}^S = 1$ point under exact repair and Type-I secrecy as shown in Fig. 7.

Remark 3: Although the same $(\alpha, \beta) = (3, 1)$ pair is achievable under Type-II secrecy when $l = 1$ or $l = 2$ for the $(4, 3, 3)$ -DSS, it is seen that the maximum file size that can be securely stored when $l = 1$ is $\mathcal{B}^S = 3$ while it is $\mathcal{B}^S = 1$ when $l = 2$. Intuitively, when 2 nodes in the DSS are compromised, the secrecy constraints are stringent compared to the case when only one node is compromised and hence the file size that can be securely stored decreases when the number of compromised nodes increases.

VI. CONCLUSION

Securing distributed storage systems against passive eavesdropping attacks is addressed in this paper. A complete characterization of the storage-bandwidth tradeoff region is provided for a (n, k, d) -DSS for $n \leq 4$ and $l < k$ under exact repair and Type-I, Type-II secrecy constraints. These results have been extended to the $(n, n-1, n-1)$ -DSS when $n-2$ nodes are compromised. Novel regenerating codes that are read and download repair efficient are presented in order to achieve these optimal secure tradeoff regions. The converse proof for the optimal tradeoff region under Type-I secrecy constraint was leveraged off the proof presented for the exact repair tradeoff region in the absence of an eavesdropper. Novel converse proofs were presented under the Type-II secrecy constraints, thereby characterizing the maximum file size that can be securely stored in a DSS. As expected, the file size that can be securely stored decreases when the number of compromised nodes increases in the DSS. The gap in the file size that can be securely stored under Type-I and Type-II attacks increases as n increases, thereby indicating the severe limitations of the DSS under Type-II attacks. Further, the secure storage-vs-exact repair bandwidth tradeoff region obtained for the Type-II attacks indicates that the MBR point is the only achievable point in the corresponding tradeoff region. Extending these results to a general (n, k, d) -DSS is part of our ongoing work.

ACKNOWLEDGEMENT

This work was supported in part by L3 Communications. The authors would also like to thank Dr. Soheil Mohajer for helpful discussions regarding the converse proof of $(4, 2, 3)$ -DSS.

REFERENCES

- [1] K. V. Rashmi, N. B. Shah, D. Gu, H. Kuang, D. Borthakur and K. Ramchandran, "A Solution to the Network Challenges of Data Recovery in Erasure-coded Distributed Storage Systems: A Study on the Facebook Warehouse Cluster", in arXiv:1309.0186, Sept. 2013.
- [2] K. V. Rashmi, N. B. Shah and P. V. Kumar, "Enabling node repair in any erasure code for distributed storage", in arXiv:1101.0133, Jun. 2011.
- [3] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. Wainwright and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sept. 2010.
- [4] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1204–1216, Jul. 2000.
- [5] N. B. Shah, K. V. Rashmi, P. V. Kumar and K. Ramchandran, "Distributed storage codes with repair-by-transfer and non-achievability of interior points on the storage-bandwidth tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.
- [6] V. Cadambe, S. Jafar, H. Maleki, K. Ramchandran and C. Suh, "Asymptotic interference alignment for optimal repair of MDS codes in distributed storage," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2974–2987, May. 2013.
- [7] C. Tian, "Rate region of the $(4,3,3)$ exact-repair regenerating codes," in *Proc. Intern. Symp. Inf Theory, ISIT*, Istanbul, Turkey, Jun. 2013.
- [8] O. O. Koyluoglu, A. S. Rawat, and S. Vishwanath, "Secure Cooperative Regenerating Codes for Distributed Storage Systems," in arXiv:1210.3664, Oct. 2012.
- [9] S. Pawar, S. El Rouayheb, and K. Ramchandran, "Securing dynamic distributed storage systems against eavesdropping and adversarial attacks," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6734–6753, Mar. 2012.
- [10] N. B. Shah, K. V. Rashmi, and P. V. Kumar, "Information-theoretically secure regenerating codes for distributed storage," in *Proc. IEEE Global Commun. Conf., GLOBECOM*, Houston, TX, Dec. 2011.
- [11] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," in arXiv:1210.6954, Aug. 2013.
- [12] S. Goparaju, S. El Rouayheb, R. Calderbank and H. Vincent Poor, "Data Secrecy in Distributed Storage Systems under Exact Repair," in *Proc. IEEE International Symposium on Network Coding, NETCOD*, Calgary, Canada, Jun. 2013.
- [13] K. V. Rashmi, N. B. Shah and K. Ramchandran, "A Piggybacking Design Framework for Read-and Download-efficient Distributed Storage Codes", in *Proc. IEEE International Symposium on Information Theory, ISIT*, Istanbul, Turkey, Jun. 2013.
- [14] F. Cheng, R. W. Yeung, and K. W. Shum, "Imperfect Secrecy in Wiretap Channel II", in arXiv:1202.0859, Dec. 2012.
- [15] K. Rashmi, N. B. Shah, P. V. Kumar, and K. Ramchandran, "Explicit Construction of Optimal Exact Regenerating Codes for Distributed Storage", in *Proc. Allerton Conference on Control, Computing, and Communication*, Urbana-Champaign, IL, 2009, pp. 1243–1249.
- [16] R. Tandon, S. Amuru, T. C. Clancy and R. M. Buehrer, "Distributed Storage Systems with Secure and Exact Repair-New Results", in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2014.
- [17] R. Tandon, S. Amuru, T. C. Clancy and R. M. Buehrer, "On Secure Distributed Storage Systems with Exact Repair", in *Proc. IEEE International Conference on Communications (ICC)*, Sydney, Australia, May 2014.

VII. APPENDIX

A. Proof of Theorem 2: (4, 2, 3)-DSS, $l = 1$

In this section, we present the proof for the Type-II setting for the the (4, 2, 3)-DSS and $l = 1$. In particular we will show that

$$\mathcal{B}_{II}^S \leq \min\left(\frac{2\alpha}{3}, 2\beta\right). \quad (71)$$

To this end, we focus on proving that $\mathcal{B}_{II}^S \leq \frac{2\alpha}{3}$.

- File regeneration from any $k = 2$ nodes:

$$H(F|W_1, W_2) = 0 \quad (72)$$

$$H(F|W_1, W_3) = 0 \quad (73)$$

$$H(F|W_1, W_4) = 0 \quad (74)$$

$$H(F|W_2, W_3) = 0 \quad (75)$$

$$H(F|W_2, W_4) = 0 \quad (76)$$

$$H(F|W_3, W_4) = 0. \quad (77)$$

- Exact repair requirements:

$$H(W_1|S_{21}, S_{31}, S_{41}) = 0 \quad (78)$$

$$H(W_2|S_{12}, S_{32}, S_{42}) = 0 \quad (79)$$

$$H(W_3|S_{13}, S_{23}, S_{43}) = 0 \quad (80)$$

$$H(W_4|S_{14}, S_{24}, S_{34}) = 0. \quad (81)$$

- Secure repair of any $l = 1$ node:

$$I(F; S_{21}, S_{31}, S_{41}) = 0 \quad (82)$$

$$I(F; S_{12}, S_{32}, S_{42}) = 0 \quad (83)$$

$$I(F; S_{13}, S_{23}, S_{43}) = 0 \quad (84)$$

$$I(F; S_{14}, S_{24}, S_{34}) = 0. \quad (85)$$

- Repair data from a node is a function of stored data:

$$H(S_{12}, S_{13}, S_{14}|W_1) = 0 \quad (86)$$

$$H(S_{21}, S_{23}, S_{24}|W_2) = 0 \quad (87)$$

$$H(S_{31}, S_{32}, S_{34}|W_3) = 0 \quad (88)$$

$$H(S_{41}, S_{42}, S_{43}|W_4) = 0. \quad (89)$$

For secure repair of node 1, we have

$$H(F) = H(F|S_{21}, S_{31}, S_{41}) \quad (90)$$

$$= H(F, S_{21}, S_{31}, S_{41}) - H(S_{21}, S_{31}, S_{41}) \quad (91)$$

$$\stackrel{(149)}{=} H(F, S_{21}, S_{31}, S_{41}) - H(S_{21}, S_{31}, S_{41}, S_{12}) \quad (92)$$

$$= H(F, S_{21}, S_{31}, S_{41}) - H(S_{12}, S_{21}) - H(S_{31}, S_{41}|S_{12}, S_{21}). \quad (93)$$

Similarly, for secure repair of node 2, we have

$$H(F) = H(F|S_{12}, S_{32}, S_{42}) \quad (94)$$

$$= H(F, S_{12}, S_{32}, S_{42}) - H(S_{12}, S_{32}, S_{42}) \quad (95)$$

$$\stackrel{(150)}{=} H(F, S_{12}, S_{32}, S_{42}) - H(S_{12}, S_{31}, S_{41}, S_{21}) \quad (96)$$

$$= H(F, S_{12}, S_{32}, S_{42}) - H(S_{12}, S_{21}) - H(S_{32}, S_{42}|S_{12}, S_{21}). \quad (97)$$

Adding (90) and (94), we obtain

$$2H(F) = H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{12}, S_{32}, S_{42}) - 2H(S_{12}, S_{21}) - H(S_{31}, S_{41}|S_{12}, S_{21}) - H(S_{32}, S_{42}|S_{12}, S_{21}) \quad (98)$$

$$\leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{12}, S_{32}, S_{42}) - H(S_{12}, S_{21}) - H(S_{31}, S_{41}, S_{32}, S_{42}, S_{12}, S_{21}) \quad (99)$$

$$\leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{12}, S_{32}, S_{42}) - H(S_{21}) - H(S_{31}, S_{41}, S_{32}, S_{42}, S_{12}, S_{21}). \quad (100)$$

Notice that,

$$H(S_{31}, S_{41}, S_{32}, S_{42}, S_{12}, S_{21}) = H(S_{31}, S_{41}, S_{32}, S_{42}, S_{12}, S_{21}, W_1, W_2, F) \quad (101)$$

$$\geq H(F, S_{21}, S_{31}, S_{41}) \quad (102)$$

$$= H(S_{21}, S_{31}, S_{41}) + H(F|S_{21}, S_{31}, S_{41}) \quad (103)$$

$$\stackrel{(153)}{=} H(S_{21}, S_{31}, S_{41}) + H(F). \quad (104)$$

Using (101) in (98), we get

$$2H(F) \leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{12}, S_{32}, S_{42}) - H(S_{21}) - (H(S_{21}, S_{31}, S_{41}) + H(F)). \quad (105)$$

We bound the first term in the above equation as follows

$$H(F, S_{21}, S_{31}, S_{41}) \leq H(F, S_{21}, S_{31}, S_{41}, W_4) \quad (106)$$

$$\stackrel{(74)}{=} H(S_{21}, S_{31}, S_{41}, W_4) \quad (107)$$

$$= H(W_4) + H(S_{41}|W_4) + H(S_{21}, S_{31}|S_{41}, W_4) \quad (108)$$

$$\leq H(W_4) + H(S_{21}, S_{31}|S_{41}) \quad (109)$$

$$= H(W_4) + H(S_{21}, S_{31}, S_{41}) - H(S_{41}) \quad (110)$$

$$\leq \alpha + H(S_{21}, S_{31}, S_{41}) - H(S_{41}). \quad (111)$$

By symmetry, we can show that

$$H(F, S_{12}, S_{32}, S_{42}) = H(F, S_{21}, S_{31}, S_{41}). \quad (112)$$

Hence, we can bound

$$H(F, S_{12}, S_{32}, S_{42}) \leq \alpha + H(S_{21}, S_{31}, S_{41}) - H(S_{41}). \quad (113)$$

Substituting (111) and (113) in (105), we have

$$2H(F) \leq 2\alpha + 2H(S_{21}, S_{31}, S_{41}) - 2H(S_{41}) - H(S_{21}) - H(F) - H(S_{21}, S_{31}, S_{41}) \quad (114)$$

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{41}) - H(S_{21}). \quad (115)$$

Similarly, we have

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{21}) - H(S_{31}) \quad (116)$$

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{21}) - H(S_{41}) \quad (117)$$

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{31}) - H(S_{21}) \quad (118)$$

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{31}) - H(S_{41}) \quad (119)$$

$$3H(F) \leq 2\alpha + H(S_{21}, S_{31}, S_{41}) - 2H(S_{41}) - H(S_{31}) \quad (120)$$

$$(121)$$

By using the fact that

$$H(S_{21}, S_{31}, S_{41}) \leq H(S_{21}) + H(S_{31}) + H(S_{41}), \quad (122)$$

and summing all the above 6 inequalities, we have

$$3H(F) \leq 2\alpha \implies \mathcal{B}_{II}^S \leq \frac{2\alpha}{3}. \quad (123)$$

B. Proof of Theorem 3: (4, 3, 3)-DSS, $l = 1$

1) *Converse for Type-I secrecy constraint:* We note that for (4, 3, 3)-DSS and $l = 1$, under Type-I secrecy constraint, the optimal (α, β) -tradeoff is given by

$$\mathcal{B}_I^S \leq \min \left(\min(\alpha, 2\beta) + \min(\alpha, \beta), \frac{\alpha + 6\beta}{3} \right). \quad (124)$$

We note that the following bound

$$\mathcal{B}_I^S \leq \min(\alpha, 2\beta) + \min(\alpha, \beta). \quad (125)$$

follows directly from (11). Hence, in the remainder of this section, we will show that

$$\mathcal{B}_I^S \leq \frac{\alpha + 6\beta}{3}. \quad (126)$$

This proof of this bound closely follows a recent result in [7] which presents the optimal (α, β) -tradeoff for the (4, 3, 3)-DSS in the absence of secrecy constraint. The key difference is carefully incorporating the Type-I secrecy constraint as we will highlight during the proof. We proceed (similar to [7]) as follows:

$$\begin{aligned} & 8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) \\ & \geq 4H(S_{31}, S_{21}, W_4) + 4H(S_{32}, W_4) \\ & \geq 4H(S_{31}, S_{21}, W_4) + 4H(S_{32}, W_4) - 2I(S_{21}; W_3|W_4) - 2I(W_3; W_4|S_{32}, S_{34}, S_{24}) \end{aligned} \quad (127)$$

$$\stackrel{(a)}{=} 4H(S_{31}, S_{21}, W_4) + 2H(W_4) + 2H(W_3, W_4, S_{21}) - 2H(W_3, W_4) + 2H(S_{32}, S_{34}, S_{24}) + 2H(W_3, W_4, S_{24}, S_{32}, S_{34}) - 2H(W_4, S_{32}, S_{34}, S_{24})$$

$$\stackrel{(b)}{\geq} 4H(S_{31}, S_{21}, W_4) + 2H(W_4) + 2 \left[H(F) + H(W_4) \right] - 2H(W_3, W_4) + 2H(S_{32}, S_{34}, S_{24}) + 2H(W_3, W_4, S_{24}, S_{32}, S_{34}) - 2H(W_4, S_{32}, S_{34}, S_{24}), \quad (128)$$

where

- (a) follows by expanding all the mutual information expressions; then using the fact that (S_{32}, S_{34}) is a function of W_3 ; and finally using the symmetry (henceforth referred in short by (*sym*)), which implies that $H(S_{32}, W_4) = H(S_{24}, W_3) = H(S_{21}, W_4)$. In general we have $H(\mathcal{S}, \mathcal{W}) = H(\pi(\mathcal{S}), \pi(\mathcal{W}))$ where $\mathcal{S} \subseteq \{S_{ij}\}_{i,j=1}^4, i \neq j, \mathcal{W} \subseteq \{W_j\}_{j=1}^4$ and π is a permutation of the indices i, j [7].
- (b) is the key inequality that invokes the secrecy constraint; and can be proved as follows

$$\begin{aligned} H(W_3, W_4, S_{21}) &= H(F, W_3, W_4, S_{21}) \\ &\geq H(F, W_4) \\ &= H(F) + H(W_4) - I(F; W_4) \\ &= H(F) + H(W_4). \end{aligned}$$

in which the first equality follows from the fact that the file F can be recovered from (W_3, W_4, S_{21}) and the last equality follows from the Type-I secrecy constraint, i.e., $I(F; W_4) = 0$.

Next, we lower bound the following term that appears in (128):

$$\begin{aligned} & H(S_{31}, S_{21}, W_4) + H(S_{32}, S_{34}, S_{24}) \stackrel{(sym)}{=} H(S_{34}, S_{24}, W_1) + H(S_{32}, S_{34}, S_{24}) \\ & = H(W_1|S_{34}, S_{24}) + H(S_{32}|S_{34}, S_{24}) + 2H(S_{34}, S_{24}) \\ & \geq H(W_1, S_{32}|S_{34}, S_{24}) + 2H(S_{34}, S_{24}) \\ & = H(W_1, S_{32}, S_{34}, S_{24}) + H(S_{34}, S_{24}) \\ & \stackrel{(rep)}{=} H(W_1, W_2, W_4) + H(S_{34}, S_{24}) \\ & = H(W_1, W_2, W_4, F) + H(S_{34}, S_{24}) \\ & \geq H(W_4, F) + H(S_{34}, S_{24}) \\ & = H(W_4) + H(F) + H(S_{34}, S_{24}), \end{aligned} \quad (129)$$

where the fourth equality is obtained by noticing that the file F can be recovered using $(W_1, S_{32}, S_{34}, S_{24})$ and the fact that S_{ij} is a function of W_i . This is shown below.

$$\begin{aligned}
H(W_1, S_{32}, S_{34}, S_{24}) &= H(W_1, S_{14}, S_{32}, S_{34}, S_{24}) \\
&= H(W_1, W_4, S_{32}) \\
&= H(W_1, S_{12}, W_4, S_{42}, S_{32}) \\
&= H(W_1, W_2, W_4),
\end{aligned}$$

This property is henceforth denoted by (*rep*) or repair in the rest of the converse proof. The equation (129) follows from Type-I secrecy constraint. Substituting (129) in (128), we have the following

$$\begin{aligned}
&8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) \\
&\geq 2H(S_{31}, S_{21}, W_4) + 2H(W_4) + 4\left[H(F) + H(W_4)\right] - 2H(W_3, W_4) + 2H(W_3, W_4, S_{24}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) + 2H(S_{34}, S_{24}) \\
&\stackrel{(sym)}{=} 2H(S_{31}, S_{21}, W_4) + 2H(W_4) + 4\left[H(F) + H(W_4)\right] - 2H(W_3, W_4) + 2H(W_3, W_4, S_{24}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) + 2H(S_{31}, S_{21}) \\
&\geq 2H(S_{31}, S_{21}, W_4) + 2H(S_{31}, S_{21}, W_4) - 2H(W_3, W_4) + 2H(W_3, W_4, S_{24}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) + 4\left[H(F) + H(W_4)\right] \\
&= 4H(S_{31}, S_{21}, W_4) + 4\left[H(F) + H(W_4)\right] - 2H(W_3, W_4) + 2H(W_3, W_4, S_{24}) - 2H(W_4, S_{32}, S_{34}, S_{24}). \tag{130}
\end{aligned}$$

Next, we have the following inequality:

$$\begin{aligned}
H(W_3, W_4, S_{24}) - H(W_3, W_4) &= H(S_{24}|W_3, W_4) \\
&\geq H(S_{24}|S_{13}, W_3, W_4) \\
&= H(S_{24}, S_{13}, W_3, W_4) - H(S_{13}, W_3, W_4). \tag{131}
\end{aligned}$$

Using (131) to further lower bound (130), we get

$$\begin{aligned}
8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) &\geq 4H(S_{31}, S_{21}, W_4) + 4\left[H(F) + H(W_4)\right] \\
&\quad - 2H(S_{13}, W_3, W_4) + 2H(W_3, W_4, S_{24}, S_{13}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) \\
&\stackrel{(rep)}{=} 2H(S_{31}, S_{21}, W_4) + 2H(S_{31}, S_{21}, W_4, W_1) \\
&\quad - 2H(S_{13}, W_3, W_4) + 2H(W_3, W_4, S_{24}, S_{13}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) + 4\left[H(F) + H(W_4)\right] \\
&\stackrel{(sym)}{=} 2H(S_{31}, S_{21}, W_4) + 2H(S_{13}, S_{23}, W_4, W_3) \\
&\quad - 2H(S_{13}, W_3, W_4) + 2H(W_3, W_4, S_{24}, S_{13}) \\
&\quad - 2H(W_4, S_{32}, S_{34}, S_{24}) + 4\left[H(F) + H(W_4)\right]. \tag{132}
\end{aligned}$$

Next, we have the following inequality

$$\begin{aligned}
H(S_{13}, S_{23}, W_4, W_3) + H(W_3, W_4, S_{24}, S_{13}) - H(S_{13}, W_3, W_4) &= H(S_{23}|S_{13}, W_3, W_4) + H(S_{24}|S_{13}, W_3, W_4) \\
&\quad + H(S_{13}, W_3, W_4) \\
&\geq H(S_{23}, S_{24}|S_{13}, W_3, W_4) + H(S_{13}, W_3, W_4) \\
&= H(S_{23}, S_{24}, S_{13}, W_3, W_4). \tag{133}
\end{aligned}$$

Using (133) in (132), we obtain

$$\begin{aligned}
8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) &\geq 2H(S_{31}, S_{21}, W_4) + 4\left[H(F) + H(W_4)\right] \\
&\quad + 2H(S_{23}, S_{24}, S_{13}, W_3, W_4) - 2H(W_4, S_{32}, S_{34}, S_{24}) \\
&\geq 2H(S_{31}, S_{21}, W_4) + 4\left[H(F) + H(W_4)\right] \\
&\quad + 2H(S_{23}, S_{24}, S_{13}, W_3, W_4) - 2H(W_4, S_{32}, S_{34}, S_{24}, S_{14}) \\
&\stackrel{(c)}{=} 2H(S_{31}, S_{21}, W_4) + 4\left[H(F) + H(W_4)\right] \\
&\quad + 2H(S_{23}, S_{24}, S_{13}, W_3, W_4) - 2H(S_{32}, S_{34}, S_{24}, S_{14}), \tag{134}
\end{aligned}$$

where (c) follows as W_4 is a function of (S_{14}, S_{24}, S_{34}) . Next, we have

$$\begin{aligned}
2H(S_{23}, S_{24}, S_{13}, W_3, W_4) &\stackrel{(sym)}{=} H(S_{23}, S_{24}, S_{14}, W_3, W_4) + H(S_{34}, S_{32}, S_{14}, W_2, W_4) \\
&\stackrel{(rep)}{=} H(S_{23}, S_{24}, S_{14}, S_{32}, S_{34}, W_3, W_4) \\
&\quad + H(S_{34}, S_{32}, S_{14}, S_{24}, S_{23}, W_2, W_4) \\
&= H(W_3, W_4 | S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\quad + H(W_2, W_4 | S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\quad + 2H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\geq H(W_2, W_3, W_4, S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\quad + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\geq H(W_2, W_3, W_4) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&= H(W_2, W_3, W_4, F) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\geq H(W_4, F) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&= \left[H(W_4) + H(F)\right] + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}), \tag{135}
\end{aligned}$$

where (135) follows from the Type-I secrecy constraint. Using (135) in (134) gives

$$\begin{aligned}
8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) &\geq 2H(S_{31}, S_{21}, W_4) + 5\left[H(F) + H(W_4)\right] \\
&\quad - 2H(S_{32}, S_{34}, S_{24}, S_{14}) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\stackrel{(sym)}{=} 5\left[H(F) + H(W_4)\right] \\
&\quad + H(S_{31}, S_{21}, W_4) + H(S_{14}, S_{24}, W_3) \\
&\quad - 2H(S_{32}, S_{34}, S_{24}, S_{14}) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&\stackrel{(rep)}{=} 5\left[H(F) + H(W_4)\right] \\
&\quad + H(S_{31}, S_{21}, W_4) + H(S_{14}, S_{24}, S_{32}, S_{34}, W_3) \\
&\quad - 2H(S_{32}, S_{34}, S_{24}, S_{14}) + H(S_{14}, S_{24}, S_{34}, S_{23}, S_{32}) \\
&= 5\left[H(F) + H(W_4)\right] \\
&\quad + H(S_{31}, S_{21}, W_4) + H(W_3 | S_{14}, S_{24}, S_{32}, S_{34}) \\
&\quad + H(S_{23} | S_{14}, S_{24}, S_{32}, S_{34}) \\
&\geq 5\left[H(F) + H(W_4)\right] \\
&\quad + H(S_{31}, S_{21}, W_4) + H(W_3, S_{23} | S_{14}, S_{24}, S_{32}, S_{34}). \tag{136}
\end{aligned}$$

Finally, using symmetry, we write $H(S_{31}, S_{21}, W_4)$ as:

$$\begin{aligned}
H(S_{31}, S_{21}, W_4) &\stackrel{(sym)}{=} H(S_{14}, S_{34}, W_2) \\
&\stackrel{(rep)}{=} H(W_2, S_{14}, S_{34}, S_{23}, S_{24}). \tag{137}
\end{aligned}$$

On the other hand, we also have

$$\begin{aligned}
H(W_3, S_{23}|S_{14}, S_{24}, S_{32}, S_{34}) &= H(W_3, S_{23}, S_{14}, S_{24}, S_{32}, S_{34}) - H(S_{14}, S_{24}, S_{32}, S_{34}) \\
&\stackrel{(sym)}{=} H(W_3, S_{23}, S_{14}, S_{24}, S_{32}, S_{34}) - H(S_{14}, S_{24}, S_{32}, S_{23}) \\
&= H(W_3, S_{32}|S_{14}, S_{34}, S_{23}, S_{24}) \\
&\geq H(W_3, S_{32}|W_2, S_{14}, S_{34}, S_{23}, S_{24}).
\end{aligned} \tag{138}$$

Using (137) and (138), we have from (136)

$$\begin{aligned}
8H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) &\geq 5 \left[H(F) + H(W_4) \right] + H(W_2, W_3, S_{32}, S_{14}, S_{24}, S_{34}, S_{23}) \\
&\stackrel{(rep)}{=} 5 \left[H(F) + H(W_4) \right] + H(W_2, W_3, W_4, S_{32}, S_{14}, S_{24}, S_{34}, S_{23}) \\
&\geq 5 \left[H(F) + H(W_4) \right] + H(W_2, W_3, W_4) \\
&= 5 \left[H(F) + H(W_4) \right] + H(W_2, W_3, W_4, F) \\
&\geq 5 \left[H(F) + H(W_4) \right] + H(W_4, F) \\
&= 5 \left[H(F) + H(W_4) \right] + H(W_4) + H(F) \\
&= 6 \left[H(F) + H(W_4) \right],
\end{aligned} \tag{139}$$

$$= 6 \left[H(F) + H(W_4) \right], \tag{140}$$

where (139) follows from Type-I secrecy constraint. Hence, from (140), we have

$$6H(F) \leq 2H(W_4) + 4H(S_{31}, S_{21}) + 4H(S_{32}) \tag{141}$$

$$\leq 2\alpha + 12\beta, \tag{142}$$

which implies that $3H(F) \leq \alpha + 6\beta$ and hence we have the proof of the bound:

$$\mathcal{B}_I^S \leq \frac{\alpha + 6\beta}{3}. \tag{143}$$

2) *Converse for Type-II Constraint:* For this case, we will show that the existing bound (11) $\mathcal{B}_{II}^S \leq \min(\alpha, 2\beta) + \min(\alpha, \beta)$ is strictly loose and the secure storage-repair bandwidth tradeoff is given by

$$\mathcal{B}_{II}^S \leq \min(\alpha, 3\beta). \tag{144}$$

To this end, we focus on proving that $\mathcal{B}_{II}^S \leq \alpha$.

- File regeneration from any $k = 3$ nodes:

$$H(F|W_1, W_2, W_3) = 0 \tag{145}$$

$$H(F|W_1, W_2, W_4) = 0 \tag{146}$$

$$H(F|W_1, W_3, W_4) = 0 \tag{147}$$

$$H(F|W_2, W_3, W_4) = 0. \tag{148}$$

- Exact repair requirements:

$$H(W_1|S_{21}, S_{31}, S_{41}) = 0 \tag{149}$$

$$H(W_2|S_{12}, S_{32}, S_{42}) = 0 \tag{150}$$

$$H(W_3|S_{13}, S_{23}, S_{43}) = 0 \tag{151}$$

$$H(W_4|S_{14}, S_{24}, S_{34}) = 0. \tag{152}$$

- Secure repair of any $l = 1$ node:

$$I(F; S_{21}, S_{31}, S_{41}) = 0 \tag{153}$$

$$I(F; S_{12}, S_{32}, S_{42}) = 0 \tag{154}$$

$$I(F; S_{13}, S_{23}, S_{43}) = 0 \tag{155}$$

$$I(F; S_{14}, S_{24}, S_{34}) = 0. \tag{156}$$

- Repair data from a node is a function of stored data:

$$H(S_{12}, S_{13}, S_{14}|W_1) = 0 \quad (157)$$

$$H(S_{21}, S_{23}, S_{24}|W_2) = 0 \quad (158)$$

$$H(S_{31}, S_{32}, S_{34}|W_3) = 0 \quad (159)$$

$$H(S_{41}, S_{42}, S_{43}|W_4) = 0. \quad (160)$$

For secure repair of node 1, we have

$$\begin{aligned} H(F) &\stackrel{(153)}{=} H(F|S_{21}, S_{31}, S_{41}) \\ &= H(F, S_{21}, S_{31}, S_{41}) - H(S_{21}, S_{31}, S_{41}) \\ &= H(F, S_{21}, S_{31}, S_{41}) - H(S_{21}) - H(S_{31}|S_{21}) - H(S_{41}|S_{31}, S_{21}). \end{aligned} \quad (161)$$

Similarly, for secure repair of node 2, we have

$$\begin{aligned} H(F) &\stackrel{(154)}{=} H(F|S_{12}, S_{32}, S_{42}) \\ &\stackrel{(150)}{=} H(F|S_{12}, S_{32}, S_{42}, W_2) \\ &\stackrel{(158)}{=} H(F|S_{12}, S_{32}, S_{42}, W_2, S_{21}, S_{23}) \\ &\leq H(F|S_{21}, S_{23}, S_{42}) \\ &= H(F, S_{21}, S_{23}, S_{42}) - H(S_{21}, S_{23}, S_{42}) \\ &= H(F, S_{21}, S_{23}, S_{42}) - H(S_{21}) - H(S_{23}|S_{21}) - H(S_{42}|S_{23}, S_{21}). \end{aligned} \quad (162)$$

Adding (161) and (162), we obtain

$$\begin{aligned} 2H(F) &\leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) - 2H(S_{21}) - H(S_{31}|S_{21}) - H(S_{23}|S_{21}) \\ &\quad - H(S_{41}|S_{31}, S_{21}) - H(S_{42}|S_{23}, S_{21}) \\ &\leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) - 2H(S_{21}) - H(S_{31}, S_{23}|S_{21}) \\ &\quad - H(S_{41}|S_{31}, S_{21}) - H(S_{42}|S_{23}, S_{21}) \\ &= H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) - H(S_{21}) - H(S_{21}, S_{31}, S_{23}) \\ &\quad - H(S_{41}|S_{31}, S_{21}) - H(S_{42}|S_{23}, S_{21}) \\ &= H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) - H(S_{21}, S_{31}, S_{23}) \\ &\quad - H(S_{21}) - H(S_{41}|S_{21}) - H(S_{42}|S_{21}) + I(S_{41}; S_{31}|S_{21}) + I(S_{42}; S_{23}|S_{21}) \\ &\leq H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) - H(S_{21}, S_{31}, S_{23}) \\ &\quad - H(S_{21}) - H(S_{41}, S_{42}|S_{21}) + I(S_{41}; S_{31}|S_{21}) + I(S_{42}; S_{23}|S_{21}) \\ &= H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) \\ &\quad - H(S_{21}, S_{23}, S_{31}) - H(S_{41}, S_{42}, S_{21}) + I(S_{41}; S_{31}|S_{21}) + I(S_{42}; S_{23}|S_{21}) \\ &= \left[H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) \right] \\ &\quad - 2H(S_{21}, S_{23}, S_{31}) + I(S_{41}; S_{31}|S_{21}) + I(S_{42}; S_{23}|S_{21}), \end{aligned} \quad (163)$$

where in (163), we have replaced $H(S_{41}, S_{42}, S_{21})$ by $H(S_{31}, S_{23}, S_{21})$ by invoking the following symmetry property:

$$H(S_{41}, S_{42}, S_{21}) = H(S_{31}, S_{32}, S_{21}) = H(S_{31}, S_{23}, S_{21}).$$

Next, consider the term $H(F, S_{21}, S_{31}, S_{41})$ appearing in (163):

$$\begin{aligned} H(F, S_{21}, S_{31}, S_{41}) &\leq H(F, W_4, S_{23}, S_{21}, S_{31}, S_{41}) \\ &= H(W_4, S_{23}, S_{21}, S_{31}) + H(F, S_{41}|W_4, S_{23}, S_{21}, S_{31}) \\ &= H(W_4, S_{23}, S_{21}, S_{31}) \end{aligned} \quad (164)$$

$$\begin{aligned} &= H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(W_4; S_{21}, S_{31}, S_{23}) \\ &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{21}, S_{31}, S_{23}) \end{aligned} \quad (165)$$

$$\begin{aligned} &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{21}, S_{31}) \\ &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{31}|S_{21}) \\ &\leq \alpha + H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{31}|S_{21}). \end{aligned} \quad (166)$$

where (164) follows from the fact that S_{41} is a function of W_4 ; and the fact that the file F can be recovered from $W_4, S_{23}, S_{21}, S_{31}$. Equation (165) follows from the fact that S_{41} is a function of W_4 and by using data processing inequality. Similarly, for the term $H(F, S_{21}, S_{23}, S_{42})$ in (163):

$$\begin{aligned} H(F, S_{21}, S_{23}, S_{42}) &\leq H(F, W_4, S_{21}, S_{23}, S_{42}, S_{31}) \\ &= H(W_4, S_{23}, S_{21}, S_{31}) + H(F, S_{42}|W_4, S_{23}, S_{21}, S_{31}) \\ &= H(W_4, S_{23}, S_{21}, S_{31}) \\ &= H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(W_4; S_{21}, S_{31}, S_{23}) \\ &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{42}; S_{21}, S_{31}, S_{23}) \\ &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{42}; S_{21}, S_{23}) \\ &\leq H(W_4) + H(S_{23}, S_{21}, S_{31}) - I(S_{42}; S_{23}|S_{21}) \\ &\leq \alpha + H(S_{23}, S_{21}, S_{31}) - I(S_{42}; S_{23}|S_{21}). \end{aligned} \quad (167)$$

Adding (166) and (167), we arrive at

$$\left[H(F, S_{21}, S_{31}, S_{41}) + H(F, S_{21}, S_{23}, S_{42}) \right] \leq 2\alpha + 2H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{31}|S_{21}) - I(S_{42}; S_{23}|S_{21}). \quad (168)$$

Substituting (168) in (163), we obtain

$$\begin{aligned} 2H(F) &\leq 2\alpha + 2H(S_{23}, S_{21}, S_{31}) - I(S_{41}; S_{31}|S_{21}) - I(S_{42}; S_{23}|S_{21}) \\ &\quad - 2H(S_{21}, S_{23}, S_{31}) + I(S_{41}; S_{31}|S_{21}) + I(S_{42}; S_{23}|S_{21}) \end{aligned} \quad (169)$$

$$= 2\alpha, \quad (170)$$

which yields $H(F) \leq \alpha$ and hence we have the proof for

$$\mathcal{B}_{II}^S \leq \alpha. \quad (171)$$

C. Proof of Theorem 4: (4, 3, 3)-DSS, $l = 2$

Here, we focus on the case of (4, 3, 3)-DSS and $l = 2$. For this case, we will show that the secure storage-repair bandwidth tradeoff is given by

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{3}, \beta\right). \quad (172)$$

For secure repair of any $l = 2$ nodes, we have the following $\binom{4}{2} = 6$ constraints:

$$I(F; S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) = 0 \quad (173)$$

$$I(F; S_{21}, S_{31}, S_{41}, S_{13}, S_{23}, S_{43}) = 0 \quad (174)$$

$$I(F; S_{12}, S_{32}, S_{42}, S_{13}, S_{23}, S_{43}) = 0 \quad (175)$$

$$I(F; S_{21}, S_{31}, S_{41}, S_{14}, S_{24}, S_{34}) = 0 \quad (176)$$

$$I(F; S_{12}, S_{32}, S_{42}, S_{14}, S_{24}, S_{34}) = 0 \quad (177)$$

$$I(F; S_{13}, S_{23}, S_{43}, S_{14}, S_{24}, S_{34}) = 0. \quad (178)$$

With these secure repair constraints, and exact repair requirements, we will now show that

$$\mathcal{B}_{II}^S \leq \frac{\alpha}{3}. \quad (179)$$

For the secure repair of nodes 1 and 2, we have

$$\begin{aligned} H(F) &\stackrel{(173)}{=} H(F|S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \\ &= H(F, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) - H(S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}). \end{aligned} \quad (180)$$

We first focus on the term $H(F, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42})$ and bound it as follows:

$$\begin{aligned} H(F, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) &\leq H(F, W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \\ &= H(W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) + H(F|W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \\ &= H(W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) + H(F|W_4, W_1, W_2, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \end{aligned} \quad (181)$$

$$= H(W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \quad (182)$$

$$= H(W_4, S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}, S_{23}, S_{13}) \quad (183)$$

$$= H(W_4, S_{21}, S_{31}, S_{12}, S_{32}, S_{23}, S_{13}) \quad (184)$$

$$= H(W_4, U_{123}), \quad (185)$$

where (181) follows from the fact that W_1 can be obtained from (S_{21}, S_{31}, S_{41}) and W_2 can be obtained from (S_{12}, S_{32}, S_{42}) . Next, (182) follows from (146). (183) follows from the fact that W_2 (and hence S_{23}) is a function of (S_{12}, S_{32}, S_{42}) . Similarly, W_1 (and hence S_{13}) is a function of (S_{21}, S_{31}, S_{41}) . Finally, (184) follows from the fact that (S_{41}, S_{42}) are functions of W_4 . In (185), we have defined

$$U_{123} \triangleq (S_{21}, S_{12}, S_{31}, S_{13}, S_{32}, S_{23}). \quad (186)$$

Substituting (185) in (180), we have

$$\begin{aligned} H(F) &\leq H(W_4, U_{123}) - H(S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}) \\ &\stackrel{(a)}{=} H(W_4, U_{123}) - H(S_{21}, S_{31}, S_{41}, S_{12}, S_{32}, S_{42}, S_{23}, S_{13}) \\ &= H(W_4, U_{123}) - H(S_{21}, S_{12}, S_{31}, S_{13}, S_{32}, S_{23}) - H(S_{41}, S_{42}|S_{21}, S_{12}, S_{31}, S_{13}, S_{32}, S_{23}) \\ &\stackrel{(b)}{=} H(W_4, U_{123}) - H(U_{123}) - H(S_{41}, S_{42}|U_{123}) \end{aligned}$$

where (a) follows from the fact that W_2 (and hence S_{23}) is a function of (S_{12}, S_{32}, S_{42}) . Similarly, W_1 (and hence S_{13}) is a function of (S_{21}, S_{31}, S_{41}) . In (b), we have used the variable U_{123} as defined in (186).

In summary, for secure repair of nodes 1 and 2, we have

$$H(F) \leq H(W_4, U_{123}) - H(U_{123}) - H(S_{41}, S_{42}|U_{123}). \quad (187)$$

Similarly, for secure repair of nodes 1 and 3, we have

$$H(F) \leq H(W_4, U_{123}) - H(U_{123}) - H(S_{41}, S_{43}|U_{123}), \quad (188)$$

and for secure repair of nodes 2 and 3, we can obtain

$$H(F) \leq H(W_4, U_{123}) - H(U_{123}) - H(S_{42}, S_{43}|U_{123}). \quad (189)$$

Summing up (187), (188) and (189), we obtain

$$3H(F) \leq 3H(W_4, U_{123}) - 3H(U_{123}) - H(S_{41}, S_{42}|U_{123}) - H(S_{41}, S_{43}|U_{123}) - H(S_{42}, S_{43}|U_{123}). \quad (190)$$

We next note that for any four random variables (X, Y, Z, U) , the following inequality holds:

$$H(X, Y|U) + H(X, Z|U) + H(Y, Z|U) \geq 2H(X, Y, Z|U). \quad (191)$$

This inequality can be proved as follows:

$$\begin{aligned} H(X, Y|U) + H(X, Z|U) + H(Y, Z|U) &= H(X, Y|U) + H(Z|U) + H(X|Z, U) + H(Y, Z|U) \\ &\geq H(X, Y|U) + H(Z|X, Y, U) + H(X|Y, Z, U) + H(Y, Z|U) \\ &= 2H(X, Y, Z|U). \end{aligned}$$

Using the inequality (191) by letting $X = S_{41}$, $Y = S_{42}$, $Z = S_{43}$ and $U = U_{123}$ in (190), we obtain

$$\begin{aligned} 3H(F) &\leq 3H(W_4, U_{123}) - 3H(U_{123}) - H(S_{41}, S_{42}|U_{123}) - H(S_{41}, S_{43}|U_{123}) - H(S_{42}, S_{43}|U_{123}) \\ &\leq 3H(W_4, U_{123}) - 3H(U_{123}) - 2H(S_{41}, S_{42}, S_{43}|U_{123}) \\ &= 3H(W_4, U_{123}) - H(U_{123}) - 2H(S_{41}, S_{42}, S_{43}, U_{123}) \\ &= 3H(W_4, U_{123}) - H(U_{123}) - 2H(W_4, S_{41}, S_{42}, S_{43}, U_{123}) + 2H(W_4|S_{41}, S_{42}, S_{43}, U_{123}) \\ &= 3H(W_4, U_{123}) - H(U_{123}) - 2H(W_4, S_{41}, S_{42}, S_{43}, U_{123}) \tag{192} \\ &\leq 3H(W_4, U_{123}) - H(U_{123}) - 2H(W_4, U_{123}) \\ &= H(W_4, U_{123}) - H(U_{123}) \\ &\leq H(W_4) + H(U_{123}) - H(U_{123}) \\ &= H(W_4) \\ &\leq \alpha, \tag{193} \end{aligned}$$

where (192) follows from the fact that $H(W_4|S_{41}, S_{42}, S_{43}, U_{123}) = 0$. Hence, from (193), we have shown that $3H(F) \leq \alpha$, and thus we have the proof for

$$\mathcal{B}_{II}^S \leq \frac{\alpha}{3}. \tag{194}$$

D. Proof of Theorem 5: $(n, n-1, n-1)$ -DSS, $l = (n-2)$.

In this section, we present the proof for the Type-II setting for the more general $(n, n-1, n-1)$ -DSS and $l = n-2$. In particular, we will show that

$$\mathcal{B}_{II}^S \leq \min\left(\frac{\alpha}{n-1}, \beta\right). \tag{195}$$

To this end, we focus on proving that $\mathcal{B}_{II}^S \leq \frac{\alpha}{n-1}$. From Type-II security requirement we require:

$$I(F; S_{\pi_1}, S_{\pi_2}, \dots, S_{\pi_{n-2}}) = 0, \tag{196}$$

where S_{π_r} is the repair data (from the remaining $d = (n-1)$ alive nodes) that is used to repair the π_r th node. Note that there are $\binom{n}{n-2} = n(n-1)/2$ such constraints; each corresponding to the secure repair of a set of $l = (n-2)$ nodes.

Let us consider the first $k = (n-1)$ nodes, i.e., nodes $1, 2, \dots, n-1$. For secure repair of any $l = (n-2)$ out of these $(n-1)$ nodes, we have $\binom{k}{l} = \binom{n-1}{n-2} = (n-1)$ constraints. Before describing these constraints, we note that the repair data (coming from the remaining $(n-1)$ alive nodes) for node i is given by:

$$S_i = (S_{1i}, \dots, S_{(i-1)i}, S_{(i+1)i}, \dots, S_{ni}). \tag{197}$$

Using this, we define

$$S_{[1:n-1]} \triangleq (S_1, S_2, \dots, S_{n-1}), \tag{198}$$

where $S_{[1:n-1]}$ is the collective repair data that is used to repair the first $k = (n-1)$ nodes. Next, we define

$$U_{ij} \triangleq (S_{ij}, S_{ji}), \tag{199}$$

where U_{ij} consists of the repair data S_{ij} that node i sends in repair of node j , and the repair data S_{ji} that node j sends in the repair of node i . Using this, we define for any set $A \subset \{1, \dots, n\}$:

$$S_A^{(j)} \triangleq \{U_{ij} : i \in A\} \tag{200}$$

$$U_A \triangleq \{U_{ij} : (i, j) \in A, i \neq j\}. \tag{201}$$

With these definitions in place, we can write the $(n - 1)$ Type-II secrecy constraints for the first $(n - 1)$ nodes as follows:

$$I(F; S_{[1:n-1]} \setminus \{S_1\}) = 0 \quad (202)$$

$$I(F; S_{[1:n-1]} \setminus \{S_2\}) = 0 \quad (203)$$

\vdots

$$I(F; S_{[1:n-1]} \setminus \{S_{n-1}\}) = 0. \quad (204)$$

where we have defined

$$S_{[1:n-1]} \setminus \{S_i\} \triangleq (S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_{n-1}), \quad (205)$$

for $i = 1, 2, \dots, (n - 1)$.

Using the constraint (202) (i.e., secure repair of nodes $(2, 3, \dots, n - 1)$), we have the following:

$$\begin{aligned} H(F) &= H(F|S_{[1:n-1]} \setminus \{S_1\}) \\ &= H(F, S_{[1:n-1]} \setminus \{S_1\}) - H(S_{[1:n-1]} \setminus \{S_1\}). \end{aligned} \quad (206)$$

Let us focus on the first term appearing in (206):

$$\begin{aligned} H(F, S_{[1:n-1]} \setminus \{S_1\}) &\leq H(F, W_n, S_{[1:n-1]} \setminus \{S_1\}) \\ &= H(W_n, S_{[1:n-1]} \setminus \{S_1\}) + H(F|W_n, S_{[1:n-1]} \setminus \{S_1\}) \\ &= H(W_n, S_2, S_3, \dots, S_{n-1}) + H(F|W_n, S_2, S_3, \dots, S_{n-1}) \\ &\leq H(W_n, S_2, S_3, \dots, S_{n-1}) + H(F|W_n, W_2, W_3, \dots, W_{n-1}) \\ &= H(W_n, S_2, S_3, \dots, S_{n-1}) \\ &\leq H(W_n, S_1, S_2, S_3, \dots, S_{n-1}) \\ &= H(W_n, U_{[1:n-1]}, S_1, S_2, S_3, \dots, S_{n-1}) \\ &= H(W_n, U_{[1:n-1]}) + H(S_1, S_2, S_3, \dots, S_{n-1}|W_n, U_{[1:n-1]}) \\ &= H(W_n, U_{[1:n-1]}) + H(S_{n1}, S_{n2}, S_{n3}, \dots, S_{n(n-1)}|W_n, U_{[1:n-1]}) \\ &= H(W_n, U_{[1:n-1]}), \end{aligned} \quad (207)$$

where (207) follows from the fact that $(S_{n1}, S_{n2}, \dots, S_{n(n-1)})$ are all functions of W_n .

Next, we focus on the second term appearing in (206):

$$\begin{aligned} H(S_{[1:n-1]} \setminus \{S_1\}) &= H(S_2, \dots, S_{n-1}) \\ &= H(S_2, \dots, S_{n-1}, S_{21}, S_{2n}, \dots, S_{(n-1)1}, S_{(n-1)n}) \end{aligned} \quad (208)$$

$$= H(U_{[1:n-1]}, S_{[2,3,\dots,n-1]}^{(n)}) \quad (209)$$

$$= H(U_{[1:n-1]}) + H(S_{[2,3,\dots,n-1]}^{(n)}|U_{[1:n-1]}), \quad (210)$$

where (208) follows from the fact that W_i (and hence (S_{i1}, S_{in})) is a function of S_i . Thus, as we have S_2 , we can add S_{21}, S_{2n} ; similarly, as we have S_i , we can add (S_{i1}, S_{in}) without increasing the entropy, for $i = 2, 3, \dots, n$. Finally, (209) follows by directly expanding all the terms S_2, S_3, \dots, S_{n-1} and compactly expressing all the variables by using the definitions of $U_{[1:n-1]}$ and $S_{[2,3,\dots,n-1]}^{(n)}$ which were defined in (201) and (200).

Using (207) and (210) in (206), we obtain

$$H(F) \leq H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - H\left(S_{[2,3,\dots,n-1]}^{(n)}|U_{[1:n-1]}\right). \quad (211)$$

In summary, from the secure repair constraint of nodes $\{1, \dots, n - 1\} \setminus \{1\}$, we have

$$H(F) \leq H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - H\left(S_{[1:n-1] \setminus \{1\}}^{(n)}|U_{[1:n-1]}\right). \quad (212)$$

Similarly, for the secure repair of nodes $\{1, \dots, n - 1\} \setminus \{i\}$, we can obtain

$$H(F) \leq H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - H\left(S_{[1:n-1] \setminus \{i\}}^{(n)}|U_{[1:n-1]}\right). \quad (213)$$

There are total of $(n - 1)$ such bounds for $i = 1, 2, \dots, (n - 1)$. Summing up these $(n - 1)$ bounds, we obtain

$$(n - 1)H(F) \leq (n - 1)H(W_n, U_{[1:n-1]}) - (n - 1)H(U_{[1:n-1]}) - \sum_{i=1}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right). \quad (214)$$

We next focus on the summand appearing in (214) for which we have the following inequality:

$$\sum_{i=1}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) \geq (n - 2)H\left(S_{[1:n-1]}^{(n)}|U_{[1:n-1]}\right), \quad (215)$$

which can be shown readily as follows:

$$\sum_{i=1}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) = \sum_{i=2}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) + H\left(S_{[1:n-1]\setminus\{1\}}^{(n)}|U_{[1:n-1]}\right) \quad (216)$$

$$= \sum_{i=2}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) + H\left(S_{[2,3,\dots,n-1]}^{(n)}|U_{[1:n-1]}\right) \quad (217)$$

$$= \sum_{i=2}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) + H(U_{2n}, U_{3n}, \dots, U_{(n-1)n}|U_{[1:n-1]}) \quad (218)$$

$$\begin{aligned} &= \sum_{i=2}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) + H(U_{2n}|U_{[1:n-1]}) \\ &\quad + H(U_{3n}|U_{2n}, U_{[1:n-1]}) + \dots + H(U_{(n-1)n}|U_{2n}, U_{3n}, \dots, U_{(n-2)n}, U_{[1:n-1]}) \\ &\geq \sum_{i=2}^{n-1} H\left(S_{[1:n-1]\setminus\{i\}}^{(n)}|U_{[1:n-1]}\right) + H\left(U_{2n}|S_{[1:n-1]\setminus\{2\}}^{(n)}, U_{[1:n-1]}\right) \\ &\quad + H\left(U_{3n}|S_{[1:n-1]\setminus\{3\}}^{(n)}, U_{[1:n-1]}\right) + \dots + H\left(U_{(n-1)n}|S_{[1:n-1]\setminus\{(n-1)\}}^{(n)}, U_{[1:n-1]}\right) \\ &= (n - 2)H\left(S_{[1:n-1]}^{(n)}|U_{[1:n-1]}\right). \end{aligned} \quad (219)$$

This completes the proof for the bound (215).

Using (215) to further bound (214), we obtain

$$\begin{aligned} (n - 1)H(F) &\leq (n - 1)H(W_n, U_{[1:n-1]}) - (n - 1)H(U_{[1:n-1]}) - (n - 2)H\left(S_{[1:n-1]}^{(n)}|U_{[1:n-1]}\right) \\ &= (n - 1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - (n - 2)H\left(S_{[1:n-1]}^{(n)}, U_{[1:n-1]}\right) \\ &= (n - 1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - (n - 2)H\left(S_{[1:n-1]}^{(n)}, W_n, U_{[1:n-1]}\right) \quad (220) \\ &\leq (n - 1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) - (n - 2)H(W_n, U_{[1:n-1]}) \\ &= H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &\leq H(W_n) + H(U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &= H(W_n) \\ &\leq \alpha, \end{aligned} \quad (221)$$

where (220) follows from the fact that W_n is a function of $S_{[1:n-1]}^{(n)}$. To note this, we observe that $S_{[1:n-1]}^{(n)}$ among other variables, consists of $(S_{1n}, S_{2n}, \dots, S_{(n-1)n})$, which is precisely the repair data for regenerating the information stored in node n (i.e., W_n).

Hence, (221) implies that $(n - 1)H(F) \leq \alpha$, and hence we have the proof for the bound:

$$\mathcal{B}_{II}^S \leq \frac{\alpha}{n - 1}. \quad (222)$$