Distributed Storage Systems with Secure and Exact Repair - New Results

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 (DSS) in the presence of a passive eavesdropper are considered in this paper. A typical DSS is characterized by 3 parameters (n, k, d) where, a file is stored in a distributed manner across n nodes such that it can be recovered entirely from any k out of n nodes. Whenever a node fails, $d \in [k, n)$ nodes participate in the repair process. In this paper, we study the exact repair capabilities of a DSS, where a failed node is replaced with its exact replica. Securing this DSS from a passive eavesdropper capable of wiretapping the repair process of any $l < k$ nodes, is the main focus of this paper. Specifically, we characterize the optimal secure storage-vs-exact-repair-bandwidth tradeoff region for the $(4, 2, 3)$ DSS when $l = 1$ and the $(n, n-1, n-1)$ DSS when $l = n-2$.**

I. Introduction

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 commodity nodes/disks that when collectively used are capable of recovering the entire file. Replication-based schemes to ensure data reliability incur huge storage overhead and therefore are not scalable [1]. Erasure codes can be used to introduce data redundancy at a low storage overhead but only at the expense of bandwidth intensive repair procedures [2]. To overcome these issues, the concept of regenerating codes for distributed storage systems was introduced by Dimakis *et al.* [3]. A typical distributed storage system (DSS) consists of n storage nodes each with a storage capacity of α units of data such that the entire file of size $\mathcal{B}$ can be recovered by accessing any $k < n$ nodes. This is called as the reconstruction property of the DSS. Whenever a node fails, $d \in [k, n)$ nodes 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 referred to as the per-node repair bandwidth.

Two types of repair procedures have predominantly been studied, namely a) functional repair and b) exact repair. 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 functional repair, the contents of the repaired node may not necessarily be identical to the failed node. 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 useful in many practical applications where the data has to be stored intact. The file recovery process is also easier in this case as the reconstruction procedure need not change whenever a failed node is replaced.

In [3], by using the concepts of network coding [4], the authors show that the parameters of a DSS must satisfy

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

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) which is in general achievable only for the functional-repair case [3]. 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) namely, the minimum storage regenerating case (MSR) and the minimum bandwidth regenerating (MBR) case have been studied extensively [5], [6]. Beyond these points, the optimal exact-repair tradeoff for the $(4, 3, 3)$-DSS was characterized in [7] where it has been shown that there is a gap between the optimal tradeoffs for functional and exact repair.

Besides providing fault tolerance and efficient repair mechanisms, the design of DSSs must also take data security into consideration. Due to the distributed and dynamic nature of storage nodes, several security threats may possibly arise. To this end, two types of eavesdropping attacks have been studied in the literature, namely (a) Type-I attack, in which the eavesdropper can read the storage contents of any l nodes in the DSS and (b) Type-II attack, in which the eavesdropper can read the contents of the repair data of any l nodes in the DSS. Note that the Type-II attack is a stronger attack in comparison to Type-I as the eavesdropper can reconstruct the stored content of any compromised node from its repair data. Throughout this paper, we assume that $l < k$ since k is the minimum number of nodes required to reconstruct the file of size $\mathcal{B}$. Else, if $l \geq k$, the eavesdropper can recover the file by using the reconstruction property of the DSS.

The focus of this paper is on incorporating two practically relevant aspects into the design of DSS, namely exact repair *and* secure repair. Clearly, such a system would be resilient to both Type-I and Type-II attacks from an eavesdropper. Optimal exact repair codes that are secure against eavesdropping of repair data have been explored for the MSR and MBR points in [9]-[12]. The codes developed in [11] achieve the MBR point for all (n, k, d) configurations with any $l < k$. The MSR code in [11] was shown to be optimal for Type-II attacks with $l = 1$ in [9]. The maximum file size $\mathcal{B}$ that can be securely stored using a restricted class of linear MSR codes with exact repair was studied in [12].

In this paper, we present novel converse proofs that characterize the optimal secure exact repair region for the $(4, 2, 3)$-DSS when $l = 1$ and the $(n, n-1-n-1)$-DSS when $l = n-2$ under Type-II attacks. Our results show that there exists a significant gap between the optimal storage-vs-repair bandwidth tradeoff for Type-I and Type-II attacks. It is also noteworthy that for the $(n, n-1-n-1)$-DSS without any security constraints, characterizing the optimal (α, β) tradeoff with exact repair is an open problem. However, as we show that under an additional constraint on the security of the repair data of any $l = n-2$ nodes, the optimal (α, β) tradeoff can be characterized. Additional results for the $(3, 2, 2)$-DSS and $(4, 3, 3)$-DSS can be found in [13].

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 storing up to α units of data. A data collector connects to any $k < n$ nodes in order to reconstruct the file F. This is known as the regeneration property of the DSS [3]. We focus on single node failures in which at any given point only one node in the system could fail. For the repair of a failed node, any d out of the remaining $(n-1)$ *alive* nodes send $\beta \leq \alpha$ units of data in order to aid the repair process. 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., $H(F) = \mathcal{B}$. Let W_i denote the storage content at node i, for $i = 1, 2 \ldots, n$. Hence, due to the storage constraint, we have

$$H(W_i) \leq \alpha, \quad \forall\ i = 1, 2, \ldots, n. \tag{2}$$

Due to the regeneration property of the DSS, we also have

$$H\left(F | W_{\{k\}}\right) = 0, \tag{3}$$

where $W_{\{k\}}$ is the data stored in any subset of k storage nodes. Let S_{ij} denote the data sent by node i to repair node j. Due to the repair bandwidth constraint, we have

$$H(S_{ij}) \leq \beta. \tag{4}$$

Furthermore, for *exact repair* of node j from d nodes, we also have

$$H(W_j | S_{r_1 j}, S_{r_2 j}, \ldots, S_{r_d j}) = 0,\ \{r_i\}_{i=1}^{d} \in [1, n] \neq j. \tag{5}$$

Since S_{ij} is a function of the data stored in node i, we have $H(S_{ij}|W_i) = 0$. For the repair of any $l < k$ nodes to be secure (i.e., security against a Type-II attack), we require

$$I\left(F; S_{n_1}, S_{n_2}, \ldots, S_{n_l}\right) = 0, \tag{6}$$

where S_{n_i} is the total repair data for node n_i.

The secrecy capacity of a DSS under Type-II attacks is defined as the maximum file size that can be stored under the storage (2), file regeneration (3), repair bandwidth (4), exact repair (5) and Type-II secrecy constraint (6). Formally,

$$\mathcal{B}_{II}^{S} = \max_{(2)-(6)} H(F). \tag{7}$$

The study of distributed storage systems in the presence of a passive eavesdropper was initiated in [10]. It was shown that for any (n, k, d)-DSS with Type-II secrecy constraint (characterized by the parameter l), the following is an upper bound on the maximum secure file size $\mathcal{B}_{II}^{S}$:

$$\mathcal{B}_{II}^{S} \leq \sum_{i=l}^{k-1} \mathsf{min}(\alpha, (d-i)\beta). \tag{8}$$

Intuitively, this upper bound on the secrecy capacity can be interpreted as follows: in presence of a eavesdropper, as l nodes are compromised, at most $(k-l)$ nodes can help in recovering the entire file while keeping it secure from the eavesdropper. Hence the summation (compared to (1)) is over $(k-l)$ nodes as opposed to k nodes [11]. We next show that there exists a significant gap between the optimal secure storage-vs-exact repair-bandwidth tradeoff region and the upper bound (8).

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 and Type-II security constraints.

Theorem 1: The optimal (α, β)-tradeoff region for $(4, 2, 3)$-DSS with $l = 1$ under exact repair and Type-II security constraints is given by:

$$\mathcal{B}_{II}^{S} \leq \min\left(\frac{2\alpha}{3}, 2\beta\right). \tag{9}$$

Theorem 2: The optimal (α, β)-tradeoff region for the $(n, n-1, n-1)$-DSS with $l = n-2$ under exact repair and Type-II security constraints is given by:

$$\mathcal{B}_{II}^{S} \leq \min\left(\frac{\alpha}{n-1}, \beta\right). \tag{10}$$

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.

Figs. 1, 2 show the optimal (α, β) tradeoff regions described by Theorems 1-2. 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 minimum bandwidth regenerating (MBR) point which corresponds to $\alpha = d\beta$ [3]. However, this is different from the optimal tradeoff-region achievable under Type-I constraints (which is given by (8), see [13]). 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 that

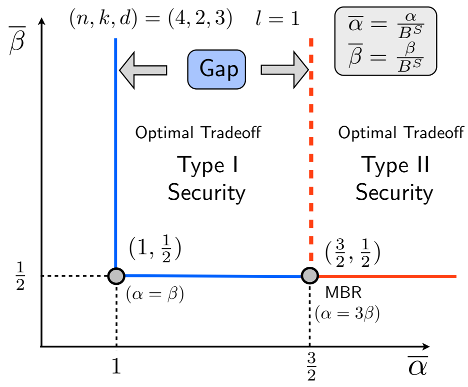

Fig. 1: Secure (α,β) tradeoff for $(4,2,3)$-DSS and $l=1$.

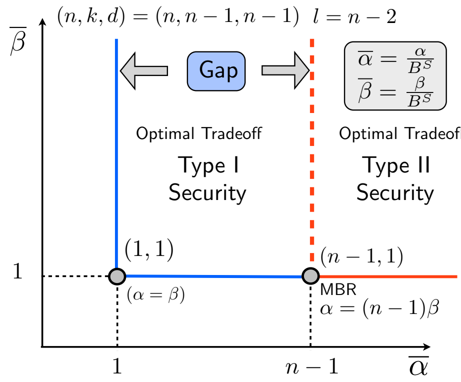

Fig. 2: Secure (α,β) tradeoff for $(n,n-1,n-1)$-DSS and $l=n-2$.

can be securely stored under Type-I constraint. Further notice that the gap between these two constraints with respect to the secure file size increases as n increases in the DSS (as the maximum file size that can be securely stored decreases due to the wiretapping process as seen in (10)).

In order to show the achievability of the (α,β)-tradeoffs described in Theorems 1, and 2, it suffices to show that

- for the $(4,2,3)$-DSS, a file of size $\mathcal{B}^S_{II}=2$ can be stored using $(\alpha,\beta)=(3,1)$,
- for the $(n,n-1,n-1)$-DSS, a file of size $\mathcal{B}^S_{II}=1$ can be stored using $(\alpha,\beta)=(n-1,1)$.

Secure codes that achieve these MBR points have already been described in [11] for Type-II attacks. Therefore, we focus on the converse proofs, i.e., the proofs of the upper bounds on the secrecy capacity. These are the main contributions of the paper and are the topics of the next section.

IV. Converse Proofs

For a (n,k,d)-DSS, we need to consider $\binom{k}{l}$ joint security and exact repair constraints in order to characterize the (α,β) tradeoff region because the repair data of any l nodes can be read by an eavesdropper.

A. *Proof of Theorem 1:* $(4,2,3)$*-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

$$\mathcal{B}^S_{II}\leq\min\left(\frac{2\alpha}{3},2\beta\right). \tag{11}$$

To this end, we focus on proving that $\mathcal{B}^S_{II}\leq\frac{2\alpha}{3}$. We first recollect the problem constraints:

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

$$H(F|W_1,W_2)=0 \tag{12}$$
$$H(F|W_1,W_3)=0 \tag{13}$$
$$H(F|W_1,W_4)=0 \tag{14}$$
$$H(F|W_2,W_3)=0 \tag{15}$$
$$H(F|W_2,W_4)=0 \tag{16}$$
$$H(F|W_3,W_4)=0. \tag{17}$$

- Exact repair requirements:

$$H(W_1|S_{21},S_{31},S_{41})=0 \tag{18}$$
$$H(W_2|S_{12},S_{32},S_{42})=0 \tag{19}$$
$$H(W_3|S_{13},S_{23},S_{43})=0 \tag{20}$$
$$H(W_4|S_{14},S_{24},S_{34})=0. \tag{21}$$

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

$$I(F;S_{21},S_{31},S_{41})=0 \tag{22}$$
$$I(F;S_{12},S_{32},S_{42})=0 \tag{23}$$
$$I(F;S_{13},S_{23},S_{43})=0 \tag{24}$$
$$I(F;S_{14},S_{24},S_{34})=0. \tag{25}$$

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

$$H(S_{12},S_{13},S_{14}|W_1)=0 \tag{26}$$
$$H(S_{21},S_{23},S_{24}|W_2)=0 \tag{27}$$
$$H(S_{31},S_{32},S_{34}|W_3)=0 \tag{28}$$
$$H(S_{41},S_{42},S_{43}|W_4)=0. \tag{29}$$

For secure repair of node 1, we have

$$\begin{aligned}H(F)&=H(F|S_{21},S_{31},S_{41})\\&=H(F,S_{21},S_{31},S_{41})-H(S_{21},S_{31},S_{41})\\&\overset{(18)}{=}H(F,S_{21},S_{31},S_{41})-H(S_{21},S_{31},S_{41},S_{12})\\&=H(F,S_{21},S_{31},S_{41})-H(S_{12},S_{21})\\&\qquad-H(S_{31},S_{41}|S_{12},S_{21}).\end{aligned} \tag{30}$$

Similarly, for secure repair of node 2, we have

$$\begin{aligned}H(F)&=H(F|S_{12},S_{32},S_{42})\\&=H(F,S_{12},S_{32},S_{42})-H(S_{12},S_{32},S_{42})\\&\overset{(19)}{=}H(F,S_{12},S_{32},S_{42})-H(S_{12},S_{31},S_{41},S_{21})\\&=H(F,S_{12},S_{32},S_{42})-H(S_{12},S_{21})\\&\qquad-H(S_{32},S_{42}|S_{12},S_{21}).\end{aligned} \tag{31}$$

Adding (30) and (31), we obtain

$$\begin{aligned}2H(F) &= H(F,S_{21},S_{31},S_{41}) + H(F,S_{12},S_{32},S_{42}) \\ &\quad - 2H(S_{12},S_{21}) - H(S_{32},S_{42}|S_{12},S_{21}) \\ &\qquad - H(S_{31},S_{41}|S_{12},S_{21}) \\ &\le H(F,S_{21},S_{31},S_{41}) + H(F,S_{12},S_{32},S_{42}) \\ &\quad - H(S_{12},S_{21}) - H(S_{31},S_{41},S_{32},S_{42},S_{12},S_{21}) \\ &\le H(F,S_{21},S_{31},S_{41}) + H(F,S_{12},S_{32},S_{42}) \\ &\quad - H(S_{21}) - H(S_{31},S_{41},S_{32},S_{42},S_{12},S_{21}). \end{aligned} \tag{32}$$

Notice that,

$$\begin{aligned}&H(S_{31},S_{41},S_{32},S_{42},S_{12},S_{21}) \\ &= H(S_{31},S_{41},S_{32},S_{42},S_{12},S_{21},W_1,W_2,F) \\ &\ge H(F,S_{21},S_{31},S_{41}) \\ &= H(S_{21},S_{31},S_{41}) + H(F|S_{21},S_{31},S_{41}) \\ &\overset{(22)}{=} H(S_{21},S_{31},S_{41}) + H(F). \end{aligned} \tag{33}$$

Substituting (33) in (32), we get

$$\begin{aligned}2H(F) &\le H(F,S_{21},S_{31},S_{41}) + H(F,S_{12},S_{32},S_{42}) \\ &\quad - H(S_{21}) - (H(S_{21},S_{31},S_{41}) + H(F)). \end{aligned} \tag{34}$$

We bound the first term in the above equation as follows

$$\begin{aligned}&H(F,S_{21},S_{31},S_{41}) \\ &\le H(F,S_{21},S_{31},S_{41},W_4) \overset{(14)}{=} H(S_{21},S_{31},S_{41},W_4) \\ &= H(W_4) + H(S_{41}|W_4) \\ &\quad + H(S_{21},S_{31}|S_{41},W_4) \\ &\le H(W_4) + H(S_{21},S_{31}|S_{41}) \\ &\le \alpha + H(S_{21},S_{31},S_{41}) - H(S_{41}). \end{aligned} \tag{35}$$

By symmetry, we can show that

$$H(F,S_{12},S_{32},S_{42}) = H(F,S_{21},S_{31},S_{41}). \tag{36}$$

Hence, we can bound

$$H(F,S_{12},S_{32},S_{42}) \le \alpha + H(S_{21},S_{31},S_{41}) - H(S_{41}). \tag{37}$$

Substituting (35) and (37) in (34), we have

$$\begin{aligned}2H(F) &\le 2\alpha + 2H(S_{21},S_{31},S_{41}) - 2H(S_{41}) \\ &\quad - H(S_{21}) - H(F) - H(S_{21},S_{31},S_{41}) \\ 3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{41}) - H(S_{21}). \end{aligned} \tag{38}$$

Similarly, we have

$$\begin{aligned}3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{21}) - H(S_{31}) \\ 3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{21}) - H(S_{41}) \\ 3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{31}) - H(S_{21}) \\ 3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{31}) - H(S_{41}) \\ 3H(F) &\le 2\alpha + H(S_{21},S_{31},S_{41}) - 2H(S_{41}) - H(S_{31}) \end{aligned}$$

By using the fact that

$$H(S_{21},S_{31},S_{41}) \le H(S_{21}) + H(S_{31}) + H(S_{41}), \tag{39}$$

and summing all the above 6 inequalities, we have

$$3H(F) \le 2\alpha \implies \mathcal{B}_{II}^S \le \frac{2\alpha}{3}. \tag{40}$$

The bound $\mathcal{B}_{II}^S \le 2\beta$ is directly obtained from (8). Using these two inequalities, we have the bound given in (9).

B. Proof of Theorem 2: $(n, n-1, n-1)$-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

$$\mathcal{B}_{II}^S \le \min\left(\frac{\alpha}{n-1}, \beta\right). \tag{41}$$

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

$$I(F; S_{\pi_1}, S_{\pi_2}, \ldots, S_{\pi_{n-2}}) = 0, \tag{42}$$

where S_{π_r} is the repair data (from the remaining $d = (n-1)$ alive nodes) that is used to repair the π_rth 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 $1, 2, \ldots, n-1$. 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:

$$S_i = (S_{1i}, \ldots, S_{(i-1)i}, S_{(i+1)i}, \ldots, S_{ni}). \tag{43}$$

Using this, we define

$$S_{[1:n-1]} \triangleq (S_1, S_2, \ldots, S_{n-1}), \tag{44}$$

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

$$U_{ij} \triangleq (S_{ij}, S_{ji}), \tag{45}$$

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 $A \subset \{1, \ldots, n\}$:

$$S_A^{(j)} \triangleq \{U_{ij} : i \in A\} \tag{46}$$

$$U_A \triangleq \{U_{ij} : (i,j) \in A, i \neq j\}. \tag{47}$$

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

$$I(F; S_{[1:n-1]} \setminus \{S_1\}) = 0 \tag{48}$$

$$I(F; S_{[1:n-1]} \setminus \{S_2\}) = 0 \tag{49}$$

$$\vdots$$

$$I(F; S_{[1:n-1]} \setminus \{S_{n-1}\}) = 0. \tag{50}$$

where we have defined

$$S_{[1:n-1]} \setminus \{S_i\} \triangleq (S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_{n-1}), \tag{51}$$

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

Using the constraint (48) (i.e., secure repair of nodes $(2, 3, \ldots, n-1)$), we have the following:

$$\begin{aligned} H(F) &= H(F|S_{[1:n-1]} \setminus \{S_1\}) \\ &= H(F, S_{[1:n-1]} \setminus \{S_1\}) - H(S_{[1:n-1]} \setminus \{S_1\}). \end{aligned} \tag{52}$$

Let us focus on the first term appearing in (52):

$$\begin{aligned} &H(F, S_{[1:n-1]} \setminus \{S_1\}) \\ &\quad \leq H(F, W_n, S_{[1:n-1]} \setminus \{S_1\}) \\ &\quad = H(W_n, S_{[1:n-1]} \setminus \{S_1\}) + H(F|W_n, S_{[1:n-1]} \setminus \{S_1\}) \\ &\quad = H(W_n, S_2, \ldots, S_{n-1}) + H(F|W_n, S_2, \ldots, S_{n-1}) \\ &\quad \leq H(W_n, S_2, \ldots, S_{n-1}) + H(F|W_n, S_2, \ldots, S_{n-1}) \\ &\quad = H(W_n, S_2, \ldots, S_{n-1}) \\ &\quad \leq H(W_n, S_1, S_2, \ldots, S_{n-1}) \\ &\quad = H(W_n, U_{[1:n-1]}, S_1, S_2, \ldots, S_{n-1}) \end{aligned}$$

$$\begin{aligned} &= H(W_n, U_{[1:n-1]}) + H(S_1, S_2, \ldots, S_{n-1}|W_n, U_{[1:n-1]}) \\ &= H(W_n, U_{[1:n-1]}) \\ &\qquad + H(S_{n1}, S_{n2}, \ldots, S_{n(n-1)}|W_n, U_{[1:n-1]}) \\ &= H(W_n, U_{[1:n-1]}), \end{aligned} \tag{53}$$

where (53) follows from the fact that $(S_{n1}, S_{n2}, \ldots, S_{n(n-1)})$ are all functions of W_n.

Next, we focus on the second term appearing in (52):

$$\begin{aligned} &H(S_{[1:n-1]} \setminus \{S_1\}) \\ &\quad = H(S_2, \ldots, S_{n-1}) \\ &\quad = H(S_2, \ldots, S_{n-1}, S_{21}, S_{2n}, \ldots, S_{(n-1)1}, S_{(n-1)n}) && (54) \\ &\quad = H(U_{[1:n-1]}, S^{(n)}_{[2,3,\ldots,n-1]}) && (55) \\ &\quad = H(U_{[1:n-1]}) + H(S^{(n)}_{[2,3,\ldots,n-1]}|U_{[1:n-1]}), && (56) \end{aligned}$$

where (54) 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 $i = 2, 3, \ldots, n$. Finally, (55) follows by directly expanding all the terms $S_2, S_3, \ldots, S_{n-1}$ and compactly expressing all the variables by using the definitions of $U_{[1:n-1]}$ and $S^{(n)}_{[2,3,\ldots,n-1]}$ which were defined in (47) and (46).

Using (53) and (56) in (52), we obtain

$$\begin{aligned} H(F) \leq\ & H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ & - H\left(S^{(n)}_{[2,3,\ldots,n-1]}|U_{[1:n-1]}\right). \end{aligned} \tag{57}$$

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

$$\begin{aligned} H(F) \leq\ & H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ & - H\left(S^{(n)}_{[1:n-1]\setminus\{1\}}|U_{[1:n-1]}\right). \end{aligned} \tag{58}$$

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

$$\begin{aligned} H(F) \leq\ & H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ & - H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right). \end{aligned} \tag{59}$$

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

$$\begin{aligned} (n-1)H(F) \leq\ & (n-1)H(W_n, U_{[1:n-1]}) \\ & - (n-1)H(U_{[1:n-1]}) \\ & - \sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right). \end{aligned} \tag{60}$$

We next focus on the summand appearing in (60), i.e., $\sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right)$. We next prove the following inequality which presents a lower bound on this summand:

$$\begin{aligned} &\sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &\qquad \geq (n-2)H\left(S^{(n)}_{[1:n-1]}|U_{[1:n-1]}\right). \end{aligned} \tag{61}$$

This inequality can be proved readily as follows:

$$\begin{aligned} &\sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &\qquad + H\left(S^{(n)}_{[1:n-1]\setminus\{1\}}|U_{[1:n-1]}\right) \\ &= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &\qquad + H\left(S^{(n)}_{[2,3,\ldots,n-1]}|U_{[1:n-1]}\right) \\ &= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &\qquad + H\left(U_{2n}, \ldots, U_{(n-1)n}|U_{[1:n-1]}\right) \\ &= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right) \\ &\quad + H\left(U_{2n}|U_{[1:n-1]}\right) \\ &\quad + H\left(U_{3n}|U_{2n}, U_{[1:n-1]}\right) \\ &\quad \vdots \\ &\quad + H\left(U_{(n-1)n}|U_{2n}, U_{3n}, \ldots, U_{(n-2)n}, U_{[1:n-1]}\right) \end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right)\\
&\quad + H\left(U_{2n}|S^{(n)}_{[1:n-1]\setminus\{2\}}, U_{[1:n-1]}\right)\\
&\quad + H\left(U_{3n}|S^{(n)}_{[1:n-1]\setminus\{3\}}U_{[1:n-1]}\right)\\
&\quad\ \vdots\\
&\quad + H\left(U_{(n-1)n}|S^{(n)}_{[1:n-1]\setminus\{(n-1)\}}, U_{[1:n-1]}\right)\\
&= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}|U_{[1:n-1]}\right)\\
&\quad + \sum_{i=2}^{n-1} H\left(U_{in}|S^{(n)}_{[1:n-1]\setminus\{i\}}, U_{[1:n-1]}\right)\\
&= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}}, U_{in}|U_{[1:n-1]}\right)\\
&= \sum_{i=2}^{n-1} H\left(S^{(n)}_{[1:n-1]}|U_{[1:n-1]}\right)\\
&= (n-2)H\left(S^{(n)}_{[1:n-1]}|U_{[1:n-1]}\right). \qquad (62)
\end{aligned}
$$

This completes the proof for the bound (61).

Using (61) to further bound (60), we obtain

$$
\begin{aligned}
&(n-1)H(F)\\
&\leq (n-1)H(W_n, U_{[1:n-1]}) - (n-1)H(U_{[1:n-1]})\\
&\quad - (n-2)H\left(S^{(n)}_{[1:n-1]}|U_{[1:n-1]}\right)\\
&= (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]})\\
&\quad - (n-2)H\left(S^{(n)}_{[1:n-1]}, U_{[1:n-1]}\right)\\
&= (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]})\\
&\quad - (n-2)H\left(S^{(n)}_{[1:n-1]}, W_n, U_{[1:n-1]}\right) \qquad (63)\\
&\leq (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]})\\
&\quad - (n-2)H\left(W_n, U_{[1:n-1]}\right)\\
&= H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]})\\
&\leq H(W_n) + H(U_{[1:n-1]}) - H(U_{[1:n-1]})\\
&= H(W_n)\\
&\leq \alpha, \qquad (64)
\end{aligned}
$$

where (63) follows from the fact that W_n is a function of $S^{(n)}_{[1:n-1]}$. To note this, we observe that $S^{(n)}_{[1:n-1]}$ among other variables, consists of $(S_{1n}, S_{2n}, \ldots, S_{(n-1)n})$, which is precisely the repair data for regenerating the information stored in node n (i.e., W_n). Finally, (64) follows from the storage constraint, i.e., $H(W_n) \leq \alpha$.

Hence, (64) implies that $(n-1)H(F) \leq \alpha$, and hence we have the proof for the bound:

$$
\mathcal{B}^S_{II} \leq \frac{\alpha}{n-1}. \qquad (65)
$$

V. 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 the $(4,2,3)$, $(n, n-1, n-1)$ distributed storage systems under exact repair and Type-II secrecy constraints when $l = 1, (n-2)$ respectively. Novel converse proofs that characterize these optimal tradeoff regions are presented. Our results show that the file size that can be securely stored decreases when the number of compromised nodes increases in the DSS. For the $(n, n-1, n-1)$ system, 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. A complete characterization of the optimal tradeoff region under Type-I and II adversaries is available in the longer version of this paper [13]. Extending these results to a general (n, k, d) DSS is part of our ongoing work.

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] 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.

[10] 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.

[11] 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.

[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] R. Tandon, S. Amuru, T. C. Clancy and R. M. Buehrer, “Towards Optimal Secure Distributed Storage Systems with Exact Repair”, in arXiv:1310.0054, Oct. 2013.