Toward Optimal Secure Distributed Storage Systems With Exact Repair

Ravi Tandon, *Member, IEEE*, SaiDhiraj Amuru, *Member, IEEE*, Thomas Charles Clancy, *Senior Member, IEEE*, and Richard Michael Buehrer, *Senior Member, IEEE*

***Abstract*— Distributed storage systems (DSSs) in the presence of an external wiretapper are considered. A DSS is parameterized by (n, k, d), in which the data are stored across n nodes (each with storage capacity α), and must be recoverable by accessing the contents stored on any k out of n nodes. If a node fails, any $d \geq k$ out of $(n-1)$ nodes help in the repair (regeneration) of the failed node (by sending $d\beta$ units of repair data, where $\beta \leq \alpha$), so that the data can still be recovered from the DSS. For such a (n, k, d)-DSS, security from the two types of wiretappers is investigated: 1) Type-I (node data) wiretapper, which can read the data stored on any $\ell < k$ nodes and 2) Type-II (repair data) wiretapper, which can read the data that is used to repair a set of ℓ failed nodes. The focus of this paper is on the optimal tradeoff between the storage (α) and the repair bandwidth ($d\beta$) in presence of a Type-I/Type-II wiretapper and the practically relevant constraint of exact repair in which a failed node must be replaced by its exact replica. In this paper, several new results and outer bounds for the storage-versus-exact-repair-bandwidth tradeoff(s) are obtained for the Type-I and Type-II security problems. Furthermore, new outer bounds are presented for the Type-II problem, which hold for general (n, k, d, ℓ) parameters. It is shown that these outer bounds strictly improve upon the existing cutset-based outer bounds. The key technical contribution of this paper is in developing novel information theoretic converse proofs for these problems. From our optimal characterization results, we show that in a Type-II setting, the only efficient point in the storage-versus-exact-repair-bandwidth tradeoff is the minimum bandwidth regenerating (MBR) point corresponding to $\alpha = d\beta$. This is in sharp contrast to the Type-I setting in which the optimal tradeoff allows a spectrum of operating points beyond the MBR point.**

***Index Terms*— Distributed storage, secure storage, network coding, information theory.**

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 distributed 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 and its variations) have been used by Facebook, OceanStore, RAID-6 [2] and others to introduce redundancy into the storage system. However, when using classical erasure codes 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 $k \leq d \leq n-1$) 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], it is shown that in order to store a file of size $\mathcal{B}$, the parameters of a DSS must necessarily satisfy

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

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). Furthermore, it was also 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

Manuscript received February 17, 2014; revised November 10, 2015; accepted February 19, 2016. Date of publication March 21, 2016; date of current version May 18, 2016. This work was supported by the L3 Communications. This paper was presented at the Information Theory and Applications 2014 and the IEEE International Conference on Communications 2014.

R. Tandon is with the Department of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ 24061 USA (e-mail: tandonr@vt.edu).

S. Amuru was with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA. He is now with Samsung R&D Institute, Bangalore 560037, India (e-mail: adhiraj@vt.edu).

T. C. Clancy is with the Hume Center for National Security and Technology, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: tcc@vt.edu).

R. M. Buehrer is with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: rbuehrer@vt.edu).

Communicated by M. Langberg, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2544340

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 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). Novel code constructions that achieve points in the (α, β) tradeoff beyond the MSR and MBR points are developed in [7] and [8]. Beyond these results, Tian [9], [10] has recently characterized the optimal exact-repair tradeoff for the $(4, 3, 3)$-DSS through a novel computer aided proof where it has been shown that there is a gap between the optimal tradeoffs for functional repair and exact repair. Novel bounds for the $(5, 4, 4)$-DSS are presented in [11] and [15] and for general (n, k, d)-DSS in [12]–[14], and [16].

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. In this paper, we focus on information theoretic security against two types of wiretappers, (a) Type-I wiretapper, which can read the storage contents of any ℓ nodes and (b) Type-II wiretapper, which can read the contents of the repair data (and thereby the storage content as well) of any ℓ nodes. Throughout this paper, we assume that $\ell < k$ since k is the minimum number of nodes required to reconstruct the entire file of size $\mathcal{B}$. Else, if $\ell \geq k$, 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 [21]. 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 information theoretic security in such systems was introduced in [18] where it was shown that any (n, k, d)-DSS in presence of either Type-I or Type-II wiretapper must satisfy

$$\mathcal{B}^S_{\text{Type-II}} \leq \mathcal{B}^S_{\text{Type-I}} \leq \sum_{i=\ell}^{k-1} \min(\alpha, (d-i)\beta), \tag{2}$$

where $\mathcal{B}^S_{\text{Type-I}}$ (respectively $\mathcal{B}^S_{\text{Type-II}}$) indicates the maximum amount of data that can be securely stored in the DSS in the presence of a Type-I (respectively Type-II) wiretapper. It was also shown that the above bound is achievable for the special case of $\alpha = d\beta$ corresponding to the bandwidth limited regime (henceforth referred to by the MBR point). 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 [17]. Optimal exact repair codes that are secure against eavesdropping have been explored for the storage and bandwidth limited regimes in [19]–[22]. The codes developed in [19] achieve the bound in (2) for the bandwidth limited regime ($\alpha = d\beta$) for all (n, k, d) configurations with any $\ell < k$. In [20], the optimality of the codes proposed for the MSR point in [19] was proved under some special conditions.

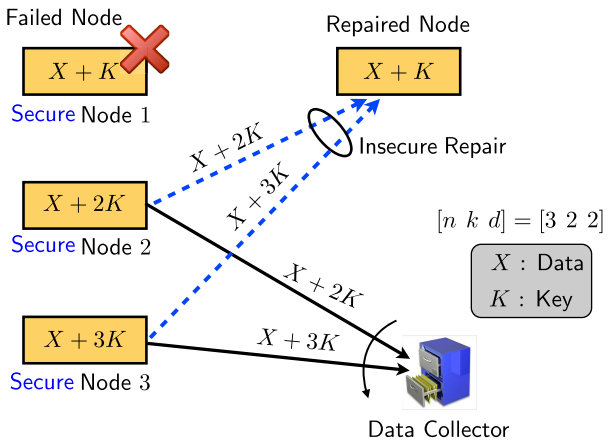

Fig. 1. $(3, 2, 2)$-DSS with Type-I (node) security.

While the outer bound in (2) takes into account the security constraints and is strictly smaller than the one in (1), however, it is far from clear if the bound is achievable for all (α, β). Moreover, the bound in (2) does not explicitly differentiate between the constraints for a Type-I or a Type-II wiretapper. Note that if the data is secure from a Type-II wiretapper, then it is also secure from a weaker Type-I wiretapper but the reverse statement may not be true in general. As an example, consider the $(3, 2, 2)$-DSS, for which the outer bound in (2) boils down to $\mathcal{B}^S_{\text{Type-I}} \leq \min(\alpha, \beta)$. Consider the $(3, 2, 2)$-DSS in presence of a Type-I wiretapper which can read the contents of any $\ell = 1$ node. For this setting, a secure exact repair code shown in Fig. 1 for $(\alpha, \beta) = (1, 1)$. 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. Since this secure code satisfies $(\mathcal{B}^S_{\text{Type-I}}, \alpha, \beta) = (1, 1, 1)$, the bound $\mathcal{B}^S_{\text{Type-I}} \leq \min(\alpha, \beta)$ is satisfied with equality.

Now, consider the same $(3, 2, 2)$-DSS in the presence of a more powerful Type-II wiretapper that can observe the repair data of any $\ell = 1$ node. It is clear that the code in Fig. 1 will

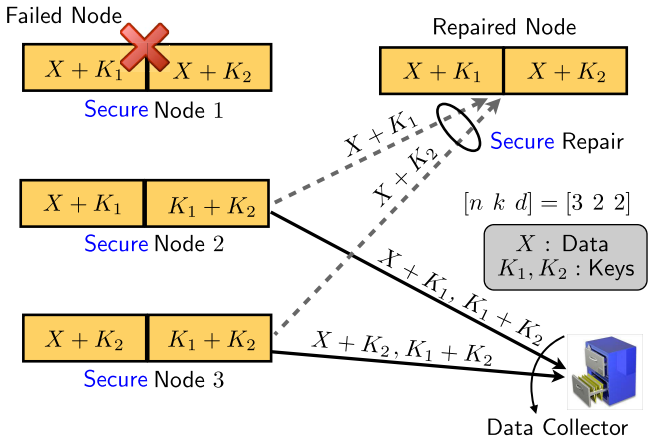

Fig. 2. $(3, 2, 2)$-DSS with Type-II (repair) security.

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 wiretapper is shown in Fig. 2. It is seen that the storage capacity is increased to $\alpha = 2$ compared to the Type-I security problem (where $\alpha = 1$). Since the wiretapper 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 Fig. 2 requires higher storage per-node ($\alpha = 2$) compared to the secure node scheme in Fig. 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 $\mathcal{B}^S_{\text{Type-II}} = 1$, 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 $\mathcal{B}^S_{\text{Type-II}} = 1$ while ensuring secure exact repair? From the results of this paper, this question is answered 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 Fig. 2 is optimal for the Type-II security problem. In particular, we show that in order for a DSS to be secure against a Type-II wiretapper, the maximum secure file size $\mathcal{B}^S_{\text{Type-II}}$ must satisfy $\mathcal{B}^S_{\text{Type-II}} \leq \min(\alpha/2, \beta)$ which reveals that the existing bound in (2) is strictly loose.

We next summarize the main contributions of this paper.

1) We characterize the optimal storage-vs-exact-repair-bandwidth trade-off(s) for Type-I and Type-II security problems for the $(n, 2, d)$-DSS with $\ell = 1$, $(n, n-1, n-1)$-DSS with $\ell = n-2$ and all (n, k, d)-DSS with $n \leq 4$ and $\ell < k$. The Type-II tradeoff region is shown to be strictly smaller than the Type-I tradeoff region which indicates the severity of the constraints imposed by the security of the repair process. Furthermore, our results show that the optimal tradeoff in presence of a Type-II wiretapper consists of only a single most efficient point corresponding to the bandwidth limited regime of $\alpha = d\beta$.
2) We also present new outer bounds for the secure storage-vs-exact-repair-bandwidth trade-off region for any (n, k, d) DSS with $1 \leq \ell < k$ in presence of a Type-II wiretapper. We show that these new bounds strictly improve upon the cutset-based bounds presented in [18].
3) The key technical contributions of this paper are the converse proofs that lead to new outer bounds on the storage-vs-bandwidth tradeoff region against Type-I and Type-II wiretappers. In contrast to prior works, our proofs capture the dependency in the repair data and the stored data between various nodes in the DSS and also explicitly differentiate between the security constraints that must be satisfied in presence of a Type-I or a Type-II wiretapper.

The rest of this paper is organized as follows. In Section II we describe the DSS system parameters and the attack models corresponding to the Type-I and Type-II wiretappers. Our main results on the secure storage-vs-exact-repair-bandwidth trade-offs are presented in Section III. The intuition behind the converse proofs is explained in Section IV through a representative example and the remaining proofs are presented in the Appendix. The coding schemes that achieve these optimal regions are presented in Section V. Finally conclusions are drawn in Section VI.

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 file reconstruction 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.

The goal is to characterize the maximum amount of data that can be stored on a DSS subject to file reconstruction and node regeneration constraints. In other words, from an information theoretic perspective, this is tantamount to finding the maximum entropy of the file F i.e., $H(F) = \mathcal{B}$ subject to the above constraints. We next introduce random variables corresponding to the data stored across 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 $i = 1, 2 \ldots, n$. Hence, the storage constraint implies

$$H(W_i) \leq \alpha, \quad i = 1, 2, \ldots, n. \tag{3}$$

Due to the file reconstruction property, i.e. the file F must be reconstructed from any $k \leq n$ nodes, we also have

$$H(F|W_{\mathsf{B}}) = 0, \tag{4}$$

where W_{B} is the data stored in any subset $\mathsf{B} \subseteq \{1, 2, \ldots, n\}$ with $|\mathsf{B}| = k$ nodes ($|\mathsf{B}|$ indicates size of the set B). 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

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

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

$$H(W_j|S_j) = 0, \tag{6}$$

where S_j indicates the *total repair data* from any $d \in [1, n] \neq j$ nodes that aids in the repair process of the jth node. 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., $H(S_{ij}|W_i) = 0$.

A. Wiretapper Models

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

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

$$I(F; W_{\mathcal{A}}) = 0, \tag{7}$$

  where $W_{\mathcal{A}}$ is the data stored in any subset $\mathcal{A} \subseteq \{1, 2, \ldots, n\}$ with $|\mathcal{A}| \leq \ell$ nodes.
- Type-II (repair) security: in this setting, the wiretapper can read the repair data of any $\ell < k$ nodes. Hence, for the repair of any ℓ nodes to be secure, we require

$$I(F; S_{\mathcal{A}}) = 0, \tag{8}$$

  where $S_{\mathcal{A}}$ is the total repair data for the nodes in any subset $\mathcal{A} \subseteq \{1, 2, \ldots, n\}$ with $|\mathcal{A}| \leq \ell$.

Remark 1: It is worth noting that any DSS that is secure under Type-II secrecy constraint is also secure under Type-I secrecy constraint. This is due to the fact that $W_{\mathcal{A}}$ can be repaired from $S_{\mathcal{A}}$, i.e., $H(W_{\mathcal{A}}|S_{\mathcal{A}}) = 0$, which together with data processing inequality implies $I(F; W_{\mathcal{A}}) \leq I(F; S_{\mathcal{A}})$, but the reverse statement is not true in general.

Remark 2: Note that the security problem is non-trivial only when $\ell < k$. Otherwise if $\ell \geq k$, then the maximum secure file size under both Type-I and Type-II security constraints will be 0, since the wiretapper must be able to reconstruct the file from any k nodes due to the file reconstruction property of the DSS.

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

$$H(F|W_i, W_j) = 0, \quad \forall i, j \in [1, n], \ i \neq j.$$

Exact repair:

$$H(W_i|S_i) = 0, \quad \forall i \in [1, n].$$

Repair data functionality:

$$H(S_{i1}, S_{i2}, \ldots, S_{i(i-1)}, S_{i(i+1)}, \ldots, S_{in}|W_i) = 0 \quad \forall i \in [1, n].$$

For this example, the parameter ℓ can be either 0 and 1. For $\ell = 1$, the Type-I constraints can be written as:

$$I(F; W_1) = I(F; W_2) = \ldots = I(F; W_n) = 0.$$

whereas the Type-II security constraints are:

$$I(F; S_1) = I(F; S_2) = \ldots = I(F; S_n) = 0.$$

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 reconstruction (4), repair bandwidth (5), exact repair (6) and Type-I (respectively Type-II) security.

Formally, we define the Type-I secrecy capacity as

$$\mathcal{B}^S_{\text{Type-I}} \triangleq \max_{(3)-(6),(7)} H(F), \tag{9}$$

and the Type-II secrecy capacity as

$$\mathcal{B}^S_{\text{Type-II}} \triangleq \max_{(3)-(6),(8)} H(F). \tag{10}$$

The study of distributed storage systems in the presence of a passive wiretapper was initiated in [18]. 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 sizes:

$$\mathcal{B}^S_{\text{Type-II}} \leq \mathcal{B}^S_{\text{Type-I}} \leq \sum_{i=\ell}^{k-1} \min(\alpha, (d-i)\beta). \tag{11}$$

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 $(k - \ell)$ nodes can help a data collector in recovering the entire file while keeping it secure from ℓ nodes. Hence the summation is over $(k - \ell)$ nodes as opposed to k nodes [19]. To illustrate by an example, consider the $(n, 2, d)$-DSS with $\ell = 1$ for which the upper bound in (11) simplifies to

$$\mathcal{B}^S_{\text{Type-I}} \leq \min(\alpha, (d-1)\beta). \tag{12}$$

Normalizing (12) throughout by $\mathcal{B}^S_{\text{Type-I}}$, and defining $\overline{\alpha} = \alpha/\mathcal{B}^S_{\text{Type-I}}$, $\overline{\beta} = \beta/\mathcal{B}^S_{\text{Type-I}}$, we can equivalently write $1 \leq \min(\overline{\alpha}, (d-1)\overline{\beta})$. This bound gives rise to a region in the $(\overline{\alpha}, \overline{\beta})$ plane and is shown in Fig. 3 with the corner point $(\overline{\alpha}, \overline{\beta}) = (1, \frac{1}{d-1})$. Furthermore, it is not difficult to show that this tradeoff is indeed achievable under Type-I security constraints by showing that it is possible to store a file of size $\mathcal{B}^S_{\text{Type-I}} = d - 1$ using $\alpha = d - 1$ and $\beta = 1$. An example achievability scheme for a $(3, 2, 2)$-DSS is shown in Fig. 1. The achievability schemes for a general $(n, 2, d)$-DSS will be discussed in Section V. Hence, the upper bound in (12) along with the achievability schemes (that will be shown in in Section V) imply that the optimal (α, β)-tradeoff for $(n, 2, d)$-DSS, $\ell = 1$ in presence of a Type-I adversary is given by

$$\mathcal{B}^S_{\text{Type-I}} \leq \min(\alpha, (d-1)\beta). \tag{13}$$

As for the more stringent Type-II security is concerned, it has been shown by Shah *et al.* in [19], using the

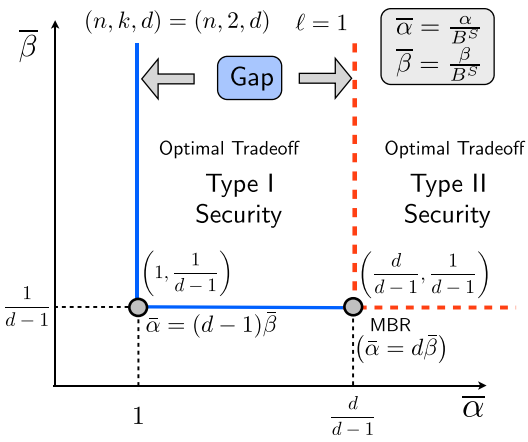

Fig. 3. Optimal secure (α,β) tradeoff for $(n,2,d)$-DSS and $\ell=1$.

Product-Matrix framework at the MBR point ($\alpha=d\beta$), that it is possible to store a file of size $\mathcal{B}^S_{\text{Type-II}}=d-1$ using $\alpha=d$ and $\beta=1$. This scheme (shown in Fig. 2) shows that the following (α,β)-tradeoff region is achievable:

$$\mathcal{B}^S_{\text{Type-II}}\leq \min\left(\frac{(d-1)\alpha}{d},(d-1)\beta\right). \tag{14}$$

On the other hand, the outer bound in (11) is the same for both Type-I and Type-II security problems, i.e.

$$\mathcal{B}^S_{\text{Type-II}}\leq \min(\alpha,(d-1)\beta). \tag{15}$$

Thus, there exists a gap between what is achievable i.e., (14) and the outer bound (15) for the Type-II secrecy problem. Specifically, it indicates that $d/(d-1)$ times more storage is necessary under the stringent Type-II security constraints. 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 the optimal secure storage-repair-bandwidth tradeoff for $(n,2,d)$-DSS, $\ell=1$ and Type-II secrecy constraint, is given by (14).

Remark 3: Note that for the non-secure version of the problem, the MSR and MBR points were defined in [3] and have been extensively studied. For the exact repair problem with security constraints however, as shown in this work, the storage-vs-repair bandwidth tradeoff region is no longer the same for several (n,k,d) parameters, and the definitions of MSR and MBR points defined for the non-secure setting no longer hold true. Instead, we focus on the fundamental tradeoff between (α,β) and the maximum file size that can be stored under both Type-I and Type-II security constraints. In what follows, we use the term MBR point to only signify the fact that α and β satisfy the relationship $\alpha=d\beta$.

III. Main Results

In this section, we present our main results that describe the secure storage-vs-exact repair-bandwidth tradeoffs (in short referred to as (α,β)-tradeoff region).

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

$$\mathcal{B}^S_{\text{Type-I}}\leq \min(\alpha,(d-1)\beta), \tag{16}$$

$$\mathcal{B}^S_{\text{Type-II}}\leq \min\left(\alpha(1-\frac{1}{d}),(d-1)\beta\right). \tag{17}$$

The converse proof for (16) follows directly from (11) and is therefore omitted. The main contribution is the converse proof for the bound (17) which is given in Section IV highlighting the difference in the derivation of the bounds under Type-I and Type-II security. We also discuss how the stringent Type-II security constraints reduce the file size that can be securely stored.

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

$$\mathcal{B}^S_{\text{Type-I}}\leq \min(\alpha,\beta) \tag{18}$$

$$\mathcal{B}^S_{\text{Type-II}}\leq \min\left(\frac{\alpha}{n-1},\beta\right). \tag{19}$$

This is the worst case scenario with respect to the $(n,n-1,n-1)$-DSS since $\ell=k-1$ nodes are compromised and hence the file size that can be stored securely is the minimum among all possible $\ell<k$ scenarios. The proof of (18) is omitted as it follows directly from (11). The proof for (19) is provided in the Appendix in Section VI-A.

Theorem 3: In addition to the cut-set bounds in [18], any general (n,k,d)-DSS with $1\leq\ell<k$ must satisfy the following under Type-II security constraints:

$$\mathcal{B}^S_{\text{Type-II}}\leq (k-\ell)\alpha-\frac{(k-\ell)\alpha}{d}. \tag{20}$$

For $1\leq\ell\leq\min(n-d,k/2)$, the above outer bound can be further strengthened to

$$\mathcal{B}^S_{\text{Type-II}}\leq (k-\ell)\alpha-\frac{\ell(k-\ell)\alpha}{d}. \tag{21}$$

These bounds are proved in the Appendix in Section VI-B and they strictly improve upon the cutset-bounds for general (n,k,d,ℓ) parameter values.

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

$$\mathcal{B}^S_{\text{Type-I}}\leq \min\left(\min(\alpha,2\beta)+\min(\alpha,\beta),\frac{\alpha+6\beta}{3}\right) \tag{22}$$

$$\mathcal{B}^S_{\text{Type-II}}\leq \min(\alpha,3\beta). \tag{23}$$

The above theorem characterizes the optimal secure tradeoff for both the Type-I and Type-II problems for the $(4,3,3)$-DSS when $\ell=1$. Note that the optimal tradeoffs when $\ell=2$ follow from Theorem 2. We note here that when $\ell=0$, i.e. without any security constraints the exact repair tradeoff region was obtained by Tian in [9] through a novel computer aided proof. However, recently a simpler proof for the same exact repair problem was obtained in [11]. Our proof for Theorem 4, given in Section VI-C, utilizes the methodology of [11] while incorporating Type-I and Type-II security constraints.

A. Illustration and Comparisons

Table I summarizes the results presented in Theorems 1-4. Figs. 3-7 show the (α,β) tradeoff regions described by these

TABLE I
SUMMARY OF RESULTS

DSS (n,k,d,ℓ)	Type-I Security $(\mathcal{B}^S_{\text{Type-I}})$	Type-II Security $(\mathcal{B}^S_{\text{Type-II}})$
$(n,2,d,1)$	$\min(\alpha,(d-1)\beta)$ [18] (optimal)	$\min(\alpha(1-\frac{1}{d}),(d-1)\beta)$ Theorem 1 (optimal)
$(n,n-1,n-1,n-2)$	$\min(\alpha,\beta)$ [18] (optimal)	$\min(\frac{\alpha}{n-1},\beta)$ Theorem 2 (optimal)
(n,k,d,ℓ)	$\sum_{i=\ell}^{k-1}\min(\alpha,(d-i)\beta)$ [18] (outer bound)	$\begin{cases}(k-\ell)\alpha-\frac{(k-\ell)\alpha}{d} & \forall 1\le\ell<k\\ (k-\ell)\alpha-\frac{\ell(k-\ell)\alpha}{d} & \text{if } \ell\le\min(n-d,k/2)\end{cases}$ Theorem 3 (outer bound)
$(4,3,3,1)$	$\min\left(\min(\alpha,2\beta)+\min(\alpha,\beta),\frac{\alpha+6\beta}{3}\right)$ Theorem 4 (optimal)	$\min(\alpha,3\beta)$ Theorem 4 (optimal)

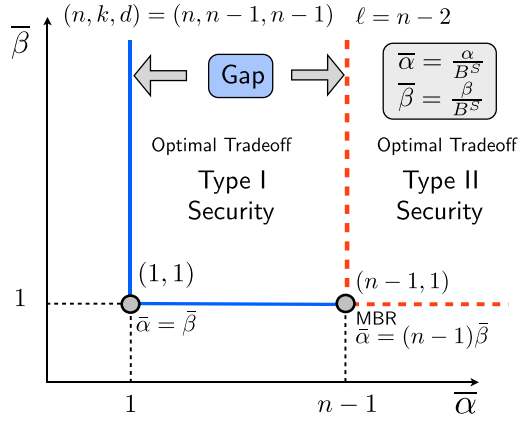

Fig. 4. Optimal secure (α,β) tradeoff for $(n,n-1,n-1)$-DSS and $\ell=n-2$.

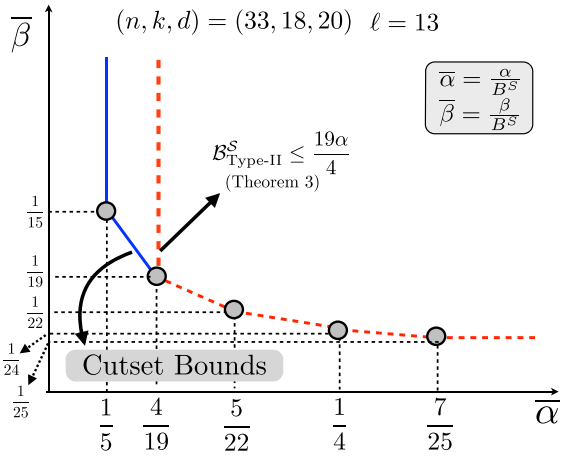

Fig. 5. Bounds on secure (α,β) tradeoff for $(33,18,20)$-DSS, $\ell=13$.

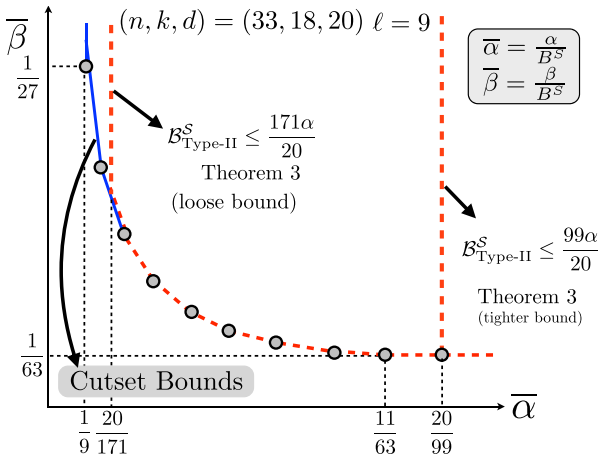

Fig. 6. Bounds on secure (α,β) tradeoff for $(33,18,20)$-DSS, $\ell=9$.

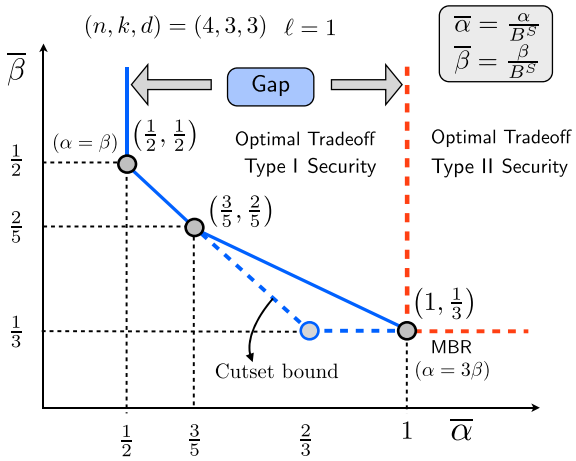

Fig. 7. Optimal secure (α,β) tradeoffs for $(4,3,3)$-DSS and $\ell=1$.

Theorems. We note that in the figures, $\mathcal{B}^S$ corresponds to a shorthand notation for $\mathcal{B}^S_{\text{Type-I}}$ (respectively $\mathcal{B}^S_{\text{Type-II}}$) for the Type-I (respectively Type-II) secrecy constraints.

From these figures, it is seen that the only efficient point in the optimal (α,β)-tradeoff region for the Type-II constraints in the case of $(n,2,d)$, $(n,n-1,n-1)$ and $(4,3,3)$ DSS systems is the MBR point corresponding to $\alpha=d\beta$. However, this is different from the optimal tradeoff-region under Type-I constraints as seen in these figures. Thus there is a gap between the optimal regions under these two constraints i.e., the file size that can be securely stored under Type-II constraints can

be strictly smaller than the file size that can be stored under Type-I security constraints. Further notice that in the case of the $(n, n-1, n-1)$-DSS, the gap between these two constraints with respect to the secure file size increases as n increases. For the Type-I constraint, when $\ell = n-2$, only one node can securely store the file and hence the maximum file size that can be securely stored is given by $\min(\alpha, \beta)$ which is independent of n. On the other hand, in the case of a Type-II wiretapper, the maximum file size that can be stored in the DSS is given by $\min\left(\frac{\alpha}{n-1}, \beta\right)$ which decreases as n increases.

Interestingly, the (α, β) tradeoff region for the cases considered in this paper for the Type-I security constraints satisfy the following relationship

$$\mathcal{B}^{S}_{\text{Type-I}} = \mathcal{B}^{\text{Non Secure}} - \ell\alpha, \tag{24}$$

where $\mathcal{B}^{\text{Non Secure}}$ denotes the maximum file size that can be stored in the absence of a wiretapper (or in other words corresponding to $\ell = 0$). However, at this point, it is unclear if this intriguing relationship can be generalized to any (n, k, d) DSS with Type-I security and exact repair constraints. This is especially because, the non-secure and secure tradeoff regions for general (n, k, d) parameters are not known.

We next illustrate the usefulness of our bounds for general (n, k, d)-DSS which were presented in Theorem 3. To this end we consider the $(33, 18, 20)$-DSS as a representative example and two values of ℓ i.e., $\ell = 9$ and $\ell = 13$. Figs. 5 and 6 show the new bounds together with the cutset bounds for the Type-II security problem. As mentioned in Theorem 3, any general (n, k, d) DSS must satisfy $\mathcal{B}^{S}_{\text{Type-II}} \leq (k-\ell)(1-\frac{1}{d})\alpha$ which for the case when $\ell = 13$ corresponds to $\mathcal{B}^{S}_{\text{Type-II}} \leq \frac{19\alpha}{4}$. The same bound for the case when $\ell = 9$ corresponds to $\mathcal{B}^{S}_{\text{Type-II}} \leq \frac{171\alpha}{20}$. However, this bound can be further tightened as $\ell = 9$ satisfies the condition $\ell \leq \min(n-d, k/2)$, therefore the tightened outer bound given in Theorem 3 applies to this case. Specifically, this bound corresponds to $\mathcal{B}^{S}_{\text{Type-II}} \leq \frac{99\alpha}{20}$ which is shown in Fig. 6.

IV. Proof of Theorem 1 and Intuition Behind Converse Proofs

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

A. Revisiting the Cut-Set Bound With Security Constraints

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

$$H(F) \leq \mathsf{min}(\alpha, (d-1)\beta). \tag{25}$$

In the presence of a Type-I wiretapper, the stored content of any $\ell = 1$ 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 $H(F) \leq \min(\alpha, (d-1)\beta)$ as follows:

$$\begin{aligned} H(F) &= H(F|W_1) = H(F, W_1) - H(W_1) && (26)\\ &\leq H(F, W_1, W_2) - H(W_1) && (27)\\ &= H(W_1, W_2) + \underbrace{H(F|W_1, W_2)}_{=0 \text{ (file reconstruction)}} - H(W_1) && (28)\\ &= H(W_1, W_2) - H(W_1) = H(W_2|W_1). && (29) \end{aligned}$$

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

(a) By bounding this term as

$$H(W_2|W_1) \leq H(W_2) \leq \alpha, \tag{30}$$

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

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

$$\begin{aligned} H(W_2|W_1) &\leq H(W_2, S_2|W_1) && (31)\\ &= H(S_2|W_1) + \underbrace{H(W_2|S_2, W_1)}_{=0 \text{ (repair of node 2 from } S_2)} && (32)\\ &= H(S_2|W_1) && (33)\\ &\leq H(S_2 \backslash S_{12}) \leq (d-1)\beta, && (34) \end{aligned}$$

which gives the bound $\mathcal{B}^{S}_{\text{Type-I}} \leq (d-1)\beta$. In the above, $S_2 \backslash S_{12}$ indicates the set of repair data from any $d-1$ nodes excluding the repair data from node 1 to node 2.

B. Tightening the Cut-Set Bound for Secure Repair (Type-II)

While the bound $H(F) \leq \min(\alpha, (d-1)\beta)$ is tight for the case of Type-I security constraint, it turns out to be strictly sub-optimal for the case of more severe Type-II problem. In particular, in presence of Type-II wiretapper, with $\ell = 1$, the DSS must satisfy the following security constraints:

$$I(F; S_1) = I(F; S_2) = \ldots = I(F; S_n) = 0. \tag{35}$$

If the wiretapper can read S_1, i.e., the repair data for node 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 scenario 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,

$$\begin{aligned} \mathcal{B}^{S}_{\text{Type-II}} &= H(F|S_1) = H(F, S_1) - H(S_1)\\ &\leq H(F, W_2, S_1) - H(S_1)\\ &= \underbrace{H(F|W_2, S_1)}_{=0 \text{ (file reconstruction)}} + H(W_2, S_1) - H(S_1)\\ &= H(W_2, S_1) - H(S_1). && (36) \end{aligned}$$

Recall that S_1 is the repair data of node 1 i.e., data from any d nodes that aid in the repair of node 1. We also note

that any outer bound (converse) that holds for a subsystem of $d+1$ nodes must also be a valid outer bound for the original DSS with n nodes. Therefore in what follows, to prove the improved bound, we focus on the first $d+1$ nodes, and the resulting $(d+1, 2, d)$-DSS. Note that in this subsystem of $d+1$ nodes, every node must be repairable by the other unique d nodes. Hence, the total repair data for node 1 i.e., S_1 can be written as $S_1 = (S_{21}, S_{31}, \ldots, S_{(d+1)1})$. More importantly, observe that there is some dependency between S_1 and W_2 through S_{21}, since S_{21} is a function of W_2. It is precisely this dependency in the incoming repair data of some nodes and the stored data at other nodes which is not captured by the cut-set bounds presented in the previous section. We next show how to exploit this dependency in order to tighten the outer bound by presenting an improved upper bound on the term $H(W_2, S_1)$ in (36).

To this end, we focus on the term $H(S_1) + H(W_2)$ and bound it as follows;

$$\begin{aligned} H(S_1) + H(W_2) &\overset{(i)}{=} H(S_1, S_{21}) + H(W_2, S_{21}) \\ &= 2H(S_{21}) + H(S_1|S_{21}) + H(W_2|S_{21}) \\ &\geq 2H(S_{21}) + H(S_1, W_2|S_{21}) \\ &= H(S_{21}) + H(S_1, W_2), \end{aligned} \tag{37}$$

where (i) follows from the fact that S_1 denotes the total incoming repair data for node 1, and that S_{21} is a function of W_2. Therefore, we have

$$H(W_2, S_1) - H(S_1) \leq H(W_2) - H(S_{21}), \tag{38}$$

which when substituted in (36) we have

$$\mathcal{B}^S_{\text{Type-II}} \leq H(W_2) - H(S_{21}). \tag{39}$$

Writing the above relationship for all the d nodes that aid in the repair of node 1, and summing all the resulting inequalities we have

$$d\mathcal{B}^S_{\text{Type-II}} \leq \sum_{i=2}^{d+1} H(W_i) - \sum_{i=2}^{d+1} H(S_{i1}). \tag{40}$$

Based on the repair properties of a DSS we have

$$H(W_1) \leq H(S_1) = H(S_{21}, \ldots, S_{(d+1)1}) \leq \sum_{i=2}^{d+1} H(S_{i1}). \tag{41}$$

By using the above inequality in (40), we have

$$\begin{aligned} d\mathcal{B}^S_{\text{Type-II}} &\leq \sum_{i=2}^{d+1} H(W_i) - H(W_1) \\ &= \sum_{i=1}^{d+1} H(W_i) - 2H(W_1) \end{aligned}$$

Repeating this argument for all $d+1$ nodes, we arrive at the following inequalities:

$$d\mathcal{B}^S_{\text{Type-II}} \leq \sum_{i=1}^{d+1} H(W_i) - 2H(W_j), \quad j = 1, 2, \ldots, (d+1). \tag{42}$$

Summing all the above $(d+1)$ inequalities, we obtain

$$\begin{aligned} (d+1)d\mathcal{B}^S_{\text{Type-II}} &\leq (d-1)\sum_{i=1}^{d+1} H(W_i) \\ &\leq (d-1)(d+1)\alpha \end{aligned} \tag{43}$$

which implies $\mathcal{B}^S_{\text{Type-II}} \leq \frac{(d-1)}{d}\alpha$ thereby completing the proof for Theorem 1. Interestingly, this tradeoff between (α, β) only has one efficient point (see Fig. 3), corresponding to the case when $\alpha = d\beta$. Exploiting the dependency in the repair data of a set of nodes and the stored data across other nodes is the key technique behind the above converse proof as well as proofs which are presented for more general (n, k, d) parameters as detailed in the Appendix.

C. Converse Proof Mechanism

The converse proofs for the various Theorems presented in the Appendix rely on a novel usage of Han's inequality [28] on subsets of random variables which enable to obtain tighter bounds for the (α, β) tradeoff region in the secure setting. We first present the Han's inequality and then briefly illustrate the proof mechanism and how the Han's inequality assists in obtaining tighter storage-versus-repair bandwidth tradeoff's.

Han's inequality: Let $(X_1, X_2, \ldots, X_L)$ be a set of L random variables. For any subset $S \subseteq \{1, 2, \ldots, L\}$ let $X(S)$ denote the subset $\{X_i : i \in S\}$ of random variables. Let

$$H_s^{(L)} = \frac{1}{\binom{L}{s}} \sum_{S:|S|=s} \frac{H(X(S))}{s}.$$

We have the following result by Han:

$$H_1^{(L)} \geq H_2^{(L)} \geq \ldots \geq H_L^{(L)}. \tag{44}$$

1) Outer Bound Proof Mechanism: We briefly mention the general steps involved in the converse proofs for the various Theorems mentioned above. For more details, we refer the reader to the Appendix.

1) We first obtain bounds on the file size that can be securely stored in the DSS under the security, reconstruction and regeneration constraints. These constraints result in upper bounds on the maximum file size which are typically of the form: $H(W_i) - H(S^i_{\mathcal{A}})$, where $\mathcal{A}$ is a subset of the nodes that aid in the repair process of node i. In other words, $|\mathcal{A}| \leq d$.
2) In order to obtain tighter bounds, we use Han's inequality and lower bound $H(S^i_{\mathcal{A}})$ in terms of $H(S_i)$, i.e., the entire repair data for the ith node.
3) In addition, we also invoke the symmetric property induced by the exact nature of the repair process to finally obtain the outer bounds on the maximum file size that can be securely stored.

V. Achievability Proofs

In this section, we present coding schemes that achieve the bounds mentioned in Theorems 1, 2 and 4.

A. Achievability Under Type-II Security Constraint

As mentioned earlier, the MBR point is the only efficient 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] and corresponds to the bandwidth-limited regime. Substituting this in (11), we get that the optimal secure file size $\mathcal{B}^S_{\text{Type-II}}$ must satisfy

$$\mathcal{B}^S_{\text{Type-II}} \leq \left(kd - \binom{k}{2}\right)\beta - \left(\ell d - \binom{\ell}{2}\right)\beta. \qquad (45)$$

Notice that (45) is identical to the tradeoff regions specified by Theorems 1, 2 and 4 with the corresponding values of ℓ. Codes that achieve this point for a general (n, k, d)-DSS with any $\ell < k$ compromised nodes have been described in [19].

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 [19]. 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) Achievability for Theorem 1: From Theorem 1, it suffices to show that the $(\alpha, \beta) = (d-1, 1)$, $\mathcal{B}^S_{\text{Type-I}} = d - 1$ is achievable in a $(n, 2, d)$-DSS when $\ell = 1$. Secure codes that achieve this (α, β) pair have been described in [19].

2) Achievability for Theorem 2: From Theorem 2, it suffices to show the achievability of the following (α, β) pair $(1, 1)$, $\mathcal{B}^S_{\text{Type-I}} = 1$. Let $a \in \mathbb{F}_q$ be the message symbol that is stored in the DSS and $k_1, k_2, \ldots, k_{n-2} \in \mathbb{F}_q$ be the keys that ensure the secure storage of a on the DSS. It is assumed that these keys and the message symbol are uniformly distributed over $\mathbb{F}_q$ for some $q > n - 1$.

Let $X = [x_1, \ldots, x_{n-1}]$ denote the random linear combinations of a and the keys $[k_1, \ldots, k_{n-2}]$. Let $x_n = \sum_{i=1}^{n-1} x_i$ denote the parity symbol of the symbols x_i. Each of these n symbols is stored in a single node on the DSS because $\alpha = 1$. By noticing the fact that $\beta = 1$, i.e., node i sends x_i to repair any other node, the repair procedure is straightforward. Along similar lines, the reconstruction procedure is also clear. Further, it is clear that the file $F = a$ is secure from any $(n-2)$ nodes, because any $(n-1)$ message symbols x_i are linearly independent combinations of the message symbol a and the keys $k_1, k_2, \ldots, k_{n-2}$.

3) Achievability for Theorem 4: From Theorem 4, it suffices to show the achievability of the following (α, β) pairs:

- P1: $(\alpha, \beta) = (1, 1)$, $\mathcal{B}^S_{\text{Type-I}} = 2$
- P2: $(\alpha, \beta) = (3, 2)$, $\mathcal{B}^S_{\text{Type-I}} = 5$
- P3: $(\alpha, \beta) = (3, 1)$, $\mathcal{B}^S_{\text{Type-I}} = 3$

We note that achievability of P3 follows from [19] directly. Hence, we focus on P1 and P2.

a) Achievability of P1: This point corresponds to the normalized (α, β) pair i.e., $(\bar{\alpha}, \bar{\beta}) = (\frac{1}{2}, \frac{1}{2})$ in the Fig. 7. Below, we present a coding scheme that can store a file of size $\mathcal{B}^S_{\text{Type-I}} = 2$ securely in a $(4, 3, 3)$-DSS with $\ell = 1$ when each node stores $\alpha = 1$ unit of data and sends $\beta = 1$ unit of data for the repair of failed nodes.

TABLE II

A $(4, 3, 3)$ SECURE EXACT REPAIR CODE FOR $(\alpha, \beta) = (1, 1)$, $\mathcal{B}^S_{\text{TYPE-I}} = 2$

node 1	node 2	node 3	node 4
$W_1 = a_1 + k$	$W_2 = a_2 + k$	$W_3 = a_1 + a_2 + k$	$W_4 = k$

TABLE III

A $(4, 3, 3)$ SECURE EXACT REPAIR CODE FOR $(\alpha, \beta) = (3, 2)$, $\mathcal{B}^S_{\text{TYPE-I}} = 5$

	first symbol	second symbol	third symbol
node 1 (W_1)	x_1	x_3	x_5
node 2 (W_2)	x_2	x_4	x_7
node 3 (W_3)	$x_1 + x_2$	x_6	x_8
node 4 (W_4)	$x_3 + x_4$	$x_5 + x_6$	$x_7 + x_8$

TABLE IV

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

	first symbol	second symbol
node 2	x_2	x_4
node 3	$x_1 + x_2$	x_6
node 4	$x_3 + x_4$	$x_5 + x_6$

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 $\mathbb{F}_2$ and also independent of a_1, a_2. The resulting DSS is shown in Table. II. Since the wiretapper is unaware 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. In particular, it is straightforward to check that $I(F; W_i) = I(a_1, a_2; W_i) = 0$, $i = 1, 2, 3, 4$. The reconstruction and the repair procedures are straightforward.

b) Achievability of P2: This point corresponds to the normalized (α, β) pair i.e., $(\bar{\alpha}, \bar{\beta}) = (\frac{3}{5}, \frac{2}{5})$ in the Fig. 7. This can be shown through a coding scheme that can securely store a file of size $\mathcal{B}^S_{\text{Type-I}} = 5$ 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 $X = [x_1, x_2, \ldots, x_8]$ denote the 8 random linear combinations of the message symbols A and the keys K. These new symbols X are stored on the DSS as shown in Table III. When any node fails, the other three nodes send the two symbols that match the color of the symbols present on the failed node. An example when node 1 fails is shown in Table IV. The reconstruction process i.e., the recovery of $[x_1, \ldots, x_8]$ from any three nodes is straightforward.

In order to achieve secrecy, i.e., the wiretapper should not get any information about $\{a_j\}_{j=1}^5$ by observing the contents

of any single node, i.e., the following condition must hold true.,

$$I(F; W_i) = I(A; W_i) = 0, \quad i = 1, 2, 3, 4, \tag{46}$$

which can be readily verified as follows,

$$\begin{aligned} I(A; W_i) &= H(W_i) - H(W_i|A) \\ &\overset{(a)}{=} H(W_i) - H(K) \\ &\leq 3\log(q) - H(k_1, k_2, k_3) \\ &\overset{(b)}{=} 3\log(q) - 3\log(q) = 0, \end{aligned} \tag{47}$$

where (a) follows from the fact that given the message symbols A and the data contents W_i, we can recover the keys K, (b) follows from the fact that keys are independent and uniformly distributed in $\mathbb{F}_q$.

VI. Conclusion

Securing distributed storage systems against two types of wiretapping attacks is addressed in this paper. In particular, the storage-versus-repair bandwidth tradeoff is investigated in presence of a Type-I wiretapper which can only read the stored content of a set of nodes. Secondly, we also focus on a more capable Type-II wiretapper that can read the repair data of a set of nodes. A complete characterization of the storage-bandwidth tradeoff region is provided for a (n, k, d)-DSS for $n \leq 4$ and $\ell < k$ under exact repair and Type-I, Type-II secrecy constraints. Furthermore, the optimal tradeoff regions were also obtained for the $(n, 2, d)$ DSS and the $(n, n-1, n-1)$-DSS when any $\ell = 1$ node and $\ell = n-2$ nodes are compromised respectively. Also, new outer bounds were presented for the generic (n, k, d) DSS in presence of a Type-II wiretapper and were shown to be strictly better than the existing cutset bounds. Our results show that there exists a gap between the optimal tradeoff regions in the presence of Type-I and Type-II wiretappers, thereby highlighting the fundamental difference between these problems. The proof techniques utilized in this work try to exploit the dependency in the repair data of some nodes and stored data on some other nodes which in turn leads to improved outer bounds and also provides optimal characterizations in some scenarios. Improving upon the outer bounds for a general (n, k, d)-DSS for both Type-I and Type-II problems is part of our ongoing work.

Appendix

A. Proof of Theorem 2: (n, n–1, n–1)-DSS, $\ell = (n–2)$

In this section, we present the proof of the outer bound for the $(n, n-1, n-1)$-DSS in presence of a Type-II wiretapper when $\ell = n-2$. In particular, we will show that

$$\mathcal{B}^S_{\text{Type-II}} \leq \min\left(\frac{\alpha}{n-1}, \beta\right). \tag{48}$$

To this end, we focus on proving that $\mathcal{B}^S_{\text{Type-II}} \leq \frac{\alpha}{n-1}$. Let $S_{[1:n]} = (S_1, S_2, \ldots, S_n)$ denote the total repair data of all the n nodes and $S_{[1:n]} \backslash (S_i, S_j)$ denote the repair data of exactly $(n-2)$ nodes excluding nodes i and j where $i \neq j$. From Type-II security requirement for $\ell = (n-2)$ we require:

$$I(F; S_{[1:n]} \backslash (S_i, S_j)) = 0, \ \forall i \neq j. \tag{49}$$

Note that there are $\binom{n}{n-2} = n(n-1)/2$ such constraints; each corresponding to the secure repair of a set of $\ell = (n-2)$ nodes. Let us consider the first $k = (n-1)$ nodes, i.e., nodes $1, 2, \ldots, n-1$. For secure repair of any $\ell = (n-2)$ out of these $(n-1)$ nodes, we have $\binom{k}{\ell} = \binom{n-1}{n-2} = (n-1)$ constraints, which are given by:

$$I(F; S_{[1:n-1]} \setminus S_1) = 0 \tag{50}$$

$$I(F; S_{[1:n-1]} \setminus S_2) = 0 \tag{51}$$

$$\vdots$$

$$I(F; S_{[1:n-1]} \setminus S_{n-1}) = 0. \tag{52}$$

Next, in order to explicitly capture the dependency between the repair data for node i and node j, we define the following variable:

$$U_{ij} \triangleq (S_{ij}, S_{ji}), \tag{53}$$

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 also define for any set $A \subseteq \{1, \ldots, n\}$:

$$S^{(j)}_A \triangleq \{U_{ij} : i \in A\} \tag{54}$$

$$U_A \triangleq \{U_{ij} : (i, j) \in A, i \neq j\}. \tag{55}$$

It can be seen that the variable $S^{(j)}_A$ consists of two terms: a) the total outgoing data from the set of nodes in A to repair node j and b) the total outgoing data from node j in the repair of set of nodes in the set A. The variable U_A consists of all pair-wise repair data between all distinct pairs of nodes present in the set A. With these definitions, we now proceed to the proof of the outer bound.

Using the constraint (50) (i.e., secure repair of nodes $(2, 3, \ldots, n-1)$), we have the following:

$$\begin{aligned} H(F) &= H(F|S_{[1:n-1]} \setminus S_1) \\ &= H(F, S_{[1:n-1]} \setminus S_1) - H(S_{[1:n-1]} \setminus S_1). \end{aligned} \tag{56}$$

Let us focus on the first term appearing in (56):

$$\begin{aligned} &H(F, S_{[1:n-1]} \setminus S_1) \\ &\leq H(F, W_n, S_{[1:n-1]} \setminus S_1) \\ &= H(W_n, S_{[1:n-1]} \setminus S_1) + H(F|W_n, S_{[1:n-1]} \setminus S_1) \\ &= H(W_n, S_2, S_3, \ldots, S_{n-1}) + H(F|W_n, S_2, S_3, \ldots, S_{n-1}) \\ &\leq H(W_n, S_2, S_3, \ldots, S_{n-1}) + \underbrace{H(F|W_n, W_2, W_3, \ldots, W_{n-1})}_{=0 \text{ (file reconstruction)}} \\ &= H(W_n, S_2, S_3, \ldots, S_{n-1}) \\ &\leq H(W_n, S_1, S_2, S_3, \ldots, S_{n-1}) \\ &\overset{(a)}{=} H(W_n, U_{[1:n-1]}, S_1, S_2, S_3, \ldots, S_{n-1}) \\ &= H(W_n, U_{[1:n-1]}) + \underbrace{H(S_1, S_2, S_3, \ldots, S_{n-1}|W_n, U_{[1:n-1]})}_{=0} \\ &= H(W_n, U_{[1:n-1]}), \end{aligned} \tag{57}$$

where (a) follows from the fact that $U_{[1:n-1]}$ is a function of $(S_{[1:n-1]})$ and (57) follows from the fact that $(S_1, S_2, \ldots, S_{(n-1)})$ are all functions of $(W_n, U_{[1:n-1]})$.

Next, we focus on the second term appearing in (56):

$$\begin{aligned} &H(S_{[1:n-1]} \setminus S_1) \\ &= H(S_2, \ldots, S_{n-1}) \\ &= H(S_2, \ldots, S_{n-1}, S_{21}, S_{2n}, S_{31}, S_{3n}, \ldots, S_{(n-1)1}, S_{(n-1)n}) \qquad (58) \\ &= H(U_{[1:n-1]}, S^{(n)}_{[2,3,\ldots,n-1]}) \qquad (59) \\ &= H(U_{[1:n-1]}) + H(S^{(n)}_{[2,3,\ldots,n-1]} | U_{[1:n-1]}), \qquad (60) \end{aligned}$$

where (58) 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, (59) 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 (55) and (54).

Using (57) and (60) in (56), we obtain

$$\begin{aligned} H(F) \leq\ & H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ & - H\left(S^{(n)}_{[2,3,\ldots,n-1]} | U_{[1:n-1]}\right). \end{aligned} \qquad (61)$$

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

$$\begin{aligned} H(F) \leq\ & H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ & - H\left(S^{(n)}_{[1:n-1]\setminus\{1\}} | U_{[1:n-1]}\right). \end{aligned} \qquad (62)$$

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

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

Summing up these $(n-1)$ bounds, we obtain

$$\begin{aligned} (n-1)H(F) \leq\ & (n-1)H(W_n, U_{[1:n-1]}) \\ & - (n-1)H(U_{[1:n-1]}) \\ & - \sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}} | U_{[1:n-1]}\right). \end{aligned} \qquad (64)$$

We next focus on the summand appearing in (64) for which we have the following lower bound;

$$\sum_{i=1}^{n-1} H\left(S^{(n)}_{[1:n-1]\setminus\{i\}} | U_{[1:n-1]}\right) \geq (n-2) H\left(S^{(n)}_{[1:n-1]} | U_{[1:n-1]}\right). \qquad (65)$$

This bound follows by a direct application of the Han's inequality [28]. More specifically, using the conditional version of this inequality (conditioned on $U_{[1:n-1]}$) for subsets of size $n-1$ and $n-2$ in the set $S^{(n)}_{[1:n-1]}$, i.e. $H^{(n-1)}_{n-2} \geq H^{(n-1)}_{n-1}$ leads directly to (65).

Substituting (65) in (64), we obtain

$$\begin{aligned} &(n-1)H(F) \\ &\leq (n-1)H(W_n, U_{[1:n-1]}) - (n-1)H(U_{[1:n-1]}) \\ &\quad - (n-2)H\left(S^{(n)}_{[1:n-1]} | U_{[1:n-1]}\right) \\ &= (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &\quad - (n-2)H\left(S^{(n)}_{[1:n-1]}, U_{[1:n-1]}\right) \\ &= (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &\quad - (n-2)H\left(S^{(n)}_{[1:n-1]}, W_n, U_{[1:n-1]}\right) \qquad (66) \\ &\leq (n-1)H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &\quad - (n-2)H\left(W_n, U_{[1:n-1]}\right) \\ &= H(W_n, U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &\leq H(W_n) + H(U_{[1:n-1]}) - H(U_{[1:n-1]}) \\ &= H(W_n) \\ &\leq \alpha, \qquad (67) \end{aligned}$$

where (66) follows from the fact that W_n is a function of $S^{(n)}_{[1:n-1]}$. To note this, we observe that $S^{(n)}_{[1:n-1]}$ 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).

Hence, (67) implies that $(n-1)H(F) \leq \alpha$, and hence we have the proof for the bound:

$$\mathcal{B}^S_{\text{Type-II}} \leq \frac{\alpha}{n-1}. \qquad (68)$$

B. Proof of Theorem 3: (n, k, d)-DSS, $1 \leq \ell < k$

In this section, we present new outer bounds for the Type-II setting for the general (n, k, d)-DSS and $1 \leq \ell < k$. We first present a bound that any (n, k, d) DSS must satisfy under the Type-II security constraints and later strengthen the bound under special conditions.

We next define some notations and a symmetry assumption that will be used in subsequent proofs.

1) $S_{[\mathcal{A}]}$: total repair data for the nodes in set $\mathcal{A}$.
2) $W_{[\mathcal{A}]}$: total data stored across the nodes in set $\mathcal{A}$.
3) $S^{[\mathcal{B}]}_{[\mathcal{A}]}$: outgoing repair data from nodes in set $\mathcal{A}$ to the nodes in set $\mathcal{B}$.
4) An (n, k, d)-DSS is said to be symmetric [9], [11] if for any subset of storage contents $\mathcal{A} \subseteq (W_1, W_2, \ldots, W_n)$ and subset of repair data $\mathcal{B} \subseteq (S_{ij})^n_{i,j=1}$, and a permutation Π of the integers $1, 2, \ldots, n$,

$$H(\mathcal{A}, \mathcal{B}) = H(\Pi(\mathcal{A}), \Pi(\mathcal{B})). \qquad (69)$$

We note that the symmetry assumption is made without any loss in generality, as any asymmetric secure code which satisfies the constraints of file reconstruction, repair and security (either Type-I/Type-II) can be symmetrized by augmenting its $n!$ independent copies, each copy corresponding to a permutation of the node labels. The resulting symmetric code and the original asymmetric code would achieve the same (α, β) tradeoff upto a scaling factor of $n!$. Also note that the resulting symmetric code will also satisfy the desired security

constraints due to the fact that all augmented copies of the original asymmetric code satisfy the security constraints and are independent of each other.

Due to the Type-II security constraints, the repair data of any set of ℓ nodes must not reveal any information about the file F. We start with one such constraint which gives the following:

$$\begin{aligned}\mathcal{B}^{S}_{\text{Type-II}} &= H(F|S_{[1:\ell]})\\ &= H(F, S_{[1:\ell]}) - H(S_{[1:\ell]})\\ &\le H(F, W_{[\ell+1:k]}, S_{[1:\ell]}) - H(S_{[1:\ell]})\\ &= H(W_{[\ell+1:k]}, S_{[1:\ell]})\\ &\quad + \underbrace{H(F|W_{[\ell+1:k]}, S_{[1:\ell]})}_{=0 \text{ (file reconstruction)}} - H(S_{[1:\ell]})\\ &= H(W_{[\ell+1:k]}, S_{[1:\ell]}) - H(S_{[1:\ell]}). \end{aligned} \tag{70}$$

We now try to capture the dependency between $W_{[\ell+1:k]}$ and $S_{[1:\ell]}$ by first observing that $S^{[1:\ell]}_{[\ell+1:k]}$ can be obtained from either one of these random variables. Here, $S^{[1:\ell]}_{[\ell+1:k]}$ corresponds to the total repair data sent from the nodes $\ell+1, \ldots, k$ to repair the set of nodes $1, 2, \ldots, \ell$. We use this observation to first obtain the following bound

$$\begin{aligned} &H(W_{[\ell+1:k]}) + H(S_{[1:\ell]})\\ &\quad= H\left(W_{[\ell+1:k]}, S^{[1:\ell]}_{[\ell+1:k]}\right) + H\left(S^{[1:\ell]}_{[\ell+1:k]}, S_{[1:\ell]}\right)\\ &\quad= 2H\left(S^{[1:\ell]}_{[\ell+1:k]}\right) + H\left(W_{[\ell+1:k]}\middle|S^{[1:\ell]}_{[\ell+1:k]}\right)\\ &\qquad+ H\left(S_{[1:\ell]}\middle|S^{[1:\ell]}_{[\ell+1:k]}\right)\\ &\quad\ge 2H\left(S^{[1:\ell]}_{[\ell+1:k]}\right) + H\left(W_{[\ell+1:k]}, S_{[1:\ell]}\middle|S^{[1:\ell]}_{[\ell+1:k]}\right)\\ &\quad= H\left(S^{[1:\ell]}_{[\ell+1:k]}\right) + H\left(W_{[\ell+1:k]}, S_{[1:\ell]}, S^{[1:\ell]}_{[\ell+1:k]}\right)\\ &\quad= H\left(S^{[1:\ell]}_{[\ell+1:k]}\right) + H(W_{[\ell+1:k]}, S_{[1:\ell]}). \end{aligned} \tag{71}$$

Therefore, from (71), we have

$$\begin{aligned} &H(W_{[\ell+1:k]}, S_{[1:\ell]}) - H(S_{[1:\ell]})\\ &\qquad\le H(W_{[\ell+1:k]}) - H\left(S^{[1:\ell]}_{[\ell+1:k]}\right). \end{aligned} \tag{72}$$

Substituting the above inequality in (70), we have

$$\begin{aligned}\mathcal{B}^{S}_{\text{Type-II}} &\le H(W_{[\ell+1:k]}) - H(S^{[1:\ell]}_{[\ell+1:k]})\\ &\le H(W_{[\ell+1:k]}) - H(S^{[1]}_{[\ell+1:k]})\\ &\overset{(*)}{=} H(W_{[1:k-\ell]}) - H(S^{[1]}_{[\ell+1:k]}). \end{aligned} \tag{73}$$

where $(*)$ follows from symmetry i.e., $H(W_{[\ell+1:k]}) = H(W_{[1:k-\ell]})$. Now consider the second term in the above equation, $H(S^{[1]}_{[\ell+1:k]})$. It consists of $k-\ell$ terms $(S_{(\ell+1)1}, \ldots, S_{k1})$. By symmetry, we have

$$\begin{aligned} H\left(S^{[1]}_{[\ell+1:k]}\right) &= \frac{\sum_{\mathcal{A}:|\mathcal{A}|=(k-\ell)\ \mathcal{A}\subset\{2,\ldots,d+1\}} H\left(S^{[1]}_{[\mathcal{A}]}\right)}{\binom{d}{k-\ell}}\\ &= (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=(k-\ell)\ \mathcal{A}\subset\{2,\ldots,d+1\}} \frac{H\left(S^{[1]}_{[\mathcal{A}]}\right)}{k-\ell}}{\binom{d}{k-\ell}}\right)\\ &\overset{(a)}{\ge} (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=d\ \mathcal{A}\subseteq[2,\ldots,d+1]} \frac{H\left(S^{[1]}_{[\mathcal{A}]}\right)}{d}}{\binom{d}{d}}\right)\\ &= (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=d\ \mathcal{A}\subseteq[2,\ldots,d+1]} \frac{H(S_1)}{d}}{\binom{d}{d}}\right)\\ &= \frac{(k-\ell)}{d}H(S_1)\\ &\ge \frac{(k-\ell)}{d}H(W_1). \end{aligned} \tag{74}$$

where (a) follows from the Han's inequality [28]. By using this inequality in (73), we have

$$\begin{aligned}\mathcal{B}^{S}_{\text{Type-II}} &\le H(W_{[1:k-l]}) - \frac{(k-\ell)}{d}H(W_1)\\ &\le H(W_1) + \sum_{i=2}^{k-\ell} H(W_i) - \frac{(k-\ell)}{d}H(W_1)\\ &= \sum_{i=2}^{k-\ell} H(W_i) + \left(1 - \frac{(k-\ell)}{d}\right)H(W_1)\\ &\le (k-\ell-1)\alpha + \left(1 - \frac{(k-\ell)}{d}\right)\alpha\\ &= (k-\ell)\alpha - \frac{(k-\ell)}{d}\alpha. \end{aligned} \tag{75}$$

1) Tighter Outer Bound for $\ell \le min(n-d, k/2)$: We next strengthen this bound for the special case in which $\ell \le n-d$ and $\ell \le k/2$. For this setting, we start with the bound in (73):

$$\mathcal{B}^{S}_{\text{Type-II}} \le H(W_{[\ell+1:k]}) - H(S^{[1:\ell]}_{[\ell+1:k]}). \tag{76}$$

The condition $n \ge d+\ell$ can be interpreted as follows: we have d distinct nodes in addition to the ℓ nodes that are considered in the above equation and these d nodes can all participate in the repair of ℓ nodes. In other words, the d repair nodes are distinct from the ℓ nodes. Hence, due to symmetry we

can write

$$
\begin{aligned}
H\left(S_{[\ell+1:k]}^{[1:\ell]}\right) &= \frac{\sum_{\mathcal{A}:|\mathcal{A}|=(k-\ell)} H\left(S_{[\mathcal{A}]}^{[1:\ell]}\right)}{\binom{d}{k-\ell}} \\
&= (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=(k-\ell)} \frac{H\left(S_{[\mathcal{A}]}^{[1:\ell]}\right)}{k-\ell}}{\binom{d}{k-\ell}}\right) \\
&\overset{(b)}{\geq} (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=d} \frac{H\left(S_{[\mathcal{A}]}^{[1:\ell]}\right)}{d}}{\binom{d}{d}}\right) \\
&= (k-\ell)\left(\frac{\sum_{\mathcal{A}:|\mathcal{A}|=d} \frac{H(S_1,S_2,\ldots,S_\ell)}{d}}{\binom{d}{d}}\right) \\
&= \frac{(k-\ell)}{d} H(S_1,\ldots,S_\ell) \\
&\geq \frac{(k-\ell)}{d} H(W_1,\ldots,W_\ell) \\
&= \frac{(k-\ell)}{d} H(W_{[1:\ell]}) \\
&\overset{(c)}{\geq} \frac{(k-\ell)}{d}\frac{\ell}{(k-\ell)} H(W_{[\ell+1:k]}) \\
&= \frac{\ell}{d} H(W_{[\ell+1:k]}), \qquad (77)
\end{aligned}
$$

where (b) and (c) follow from Han's inequality. Note that in order for us to apply Han's inequality in (c), we require that $\ell \leq (k-\ell)$, which leads to the condition $\ell \leq k/2$.

Therefore, by using (76), and (77), we have

$$
\begin{aligned}
\mathcal{B}^S_{\text{Type-II}} &\leq H(W_{[\ell+1:k]}) - \frac{\ell}{d} H(W_{[\ell+1:k]}) \\
&= \left(1-\frac{\ell}{d}\right) H(W_{[\ell+1:k]}) \\
&\leq \left(1-\frac{\ell}{d}\right)(k-\ell)\alpha \\
&= (k-\ell)\alpha - \frac{\ell(k-\ell)}{d}\alpha, \qquad (78)
\end{aligned}
$$

which gives the strengthened bound described in Theorem 3 when $\ell \leq \min(n-d, k/2)$.

C. Proof of Theorem 4: (4, 3, 3)-DSS, $\ell = 1$

In this section, we present converse proofs for the optimal tradeoff regions of a $(4,3,3)$ DSS with $\ell = 1$ under both Type-I and Type-II security constraints. Since $\ell = 1$, we have the following sequence of bounds which hold for both Type-I and Type-II security constraints;

$$
\begin{aligned}
H(F) &= H(F|W_1) \\
&\leq H(F, W_2, W_3|W_1) \\
&= H(W_2, W_3|W_1) + \underbrace{H(F|W_1, W_2, W_3)}_{=0 \text{ (file reconstruction)}} \\
&= H(W_2, W_3|W_1) \\
&= H(W_2|W_1) + H(W_3|W_1, W_2) \\
&= H(W_2) + H(W_3) - I(W_1; W_2) - I(W_3; W_1, W_2) \\
&\overset{(*)}{=} 2H(W_1) - I(W_1; W_2) - I(W_3; W_1, W_2)
\end{aligned}
$$

where $(*)$ follows from symmetry in the DSS, i.e., $H(W_1) = H(W_2) = H(W_3)$.

Next, we note that the outgoing repair data from each node is a function of its stored content. Hence, using the data processing inequality, we have the following;

$$
\begin{aligned}
H(F) &\leq 2H(W_1) - I(W_1; W_2) - I(W_3; W_1, W_2) \\
&\leq 2H(W_1) - I(S_1^2; S_2^1) - I(S_3^{[1,2]}; S_{[1,2]}^3), \\
&= 2H(W_1) - \{H(S_1^2) + H(S_2^1) - H(S_1^2, S_2^1)\} \\
&\quad - \{H(S_3^{[1,2]}) + H(S_{[1,2]}^3) - H(S_3^{[1,2]}, S_{[1,2]}^3)\} \\
&= 2H(W_1) - \underbrace{\{H(S_2^1) + H(S_3^{[1,2]})\}}_{\text{T1}} \\
&\quad - \underbrace{\{H(S_1^2) + H(S_{[1,2]}^3)\}}_{\text{T2}} \\
&\quad + \underbrace{\{H(S_1^2, S_2^1) + H(S_3^{[1,2]}, S_{[1,2]}^3)\}}_{\text{T3}}, \qquad (79)
\end{aligned}
$$

where $S_{[\mathcal{A}]}^{[\mathcal{B}]}$ indicates the repair data from nodes in set $[\mathcal{A}]$ to the nodes in $[\mathcal{B}]$. We bound the above equation in three parts by bounding the terms T1, T2 and T3 as follows.

For the first part T1, we have

$$
\begin{aligned}
\text{T1} &= H(S_2^1) + H(S_3^{[1,2]}) \\
&= H(S_2^1) + H(S_3^{[1,2]}) + H(S_4^{[1,2,3]}) - H(S_4^{[1,2,3]}) \\
&\geq H(S_2^1, S_3^{[1,2]}, S_4^{[1,2,3]}) - H(S_4^{[1,2,3]}) \\
&\overset{(i)}{=} H(S_{[2,3,4]}^1, S_{[3,4]}^2, S_4^3) - H(S_4^{[1,2,3]}) \\
&= H(S_1, W_1, S_{[3,4]}^2, S_4^3) - H(S_4^{[1,2,3]}) \\
&\geq H(W_1, S_1, S_1^2, S_{[3,4]}^2, S_4^3) - H(S_4^{[1,2,3]}) \\
&\geq H(S_1, S_{[1,3,4]}^2, W_2, S_4^3) - H(S_4^{[1,2,3]}) \\
&\geq H(S_1, W_2, S_1^3, S_2^3, S_4^3) - H(S_4^{[1,2,3]}) \\
&\overset{(ii)}{\geq} H(F, S_1, W_2, W_3) - H(S_4^{[1,2,3]}) \\
&\geq H(F, S_1) - H(S_4^{[1,2,3]}) \\
&\geq H(F, S_1) - H(W_4) \\
&\overset{(*)}{=} H(F, S_1) - H(W_1), \qquad (80)
\end{aligned}
$$

where (i) follows by regrouping the terms, (ii) follows from the reconstruction property of the DSS i.e., F can be recovered from W_1, W_2, W_3 and $(*)$ follows from symmetry.

For the term T2 $= H(S_1^2) + H(S_{[1,2]}^3)$ in (79), we get the following bounds;

$$\begin{aligned} H(S_1^2) &\overset{(iii)}{=} \frac{1}{3}\left(H(S_1^2) + H(S_3^2) + H(S_4^2)\right) \\ &\geq \frac{1}{3}H(S_1^2, S_3^2, S_4^2) \\ &= \frac{1}{3}H(S_2) \\ &\overset{(iv)}{=} \frac{1}{3}H(S_1), \end{aligned}$$

where (iii), (iv) follow from symmetry. We also have

$$\begin{aligned} H(S_{[1,2]}^3) &\overset{(v)}{=} \frac{1}{3}\left(H(S_{[1,2]}^3) + H(S_{[1,4]}^3) + H(S_{[2,4]}^3)\right) \\ &\overset{(vi)}{\geq} \frac{1}{3}(2H(S_{[1,2,4]}^3)) \\ &= \frac{2}{3}H(S_3) \\ &\overset{(vii)}{=} \frac{2}{3}H(S_1), \end{aligned}$$

where (v) follows from symmetry and (vi) follows from application of Han's inequality, and (vii) also follows from symmetry. Hence, from the above two inequalities, we can bound T2 as

$$\begin{aligned} \text{T2} &= H(S_1^2) + H(S_{[1,2]}^3) \\ &\geq \frac{1}{3}H(S_1) + \frac{2}{3}H(S_1) \\ &= H(S_1). \end{aligned} \tag{81}$$

Finally, to bound the term T3 $= H(S_1^2, S_2^1) + H(S_3^{[1,2]}, S_{[1,2]}^3)$ in (79), we first note that for any 3 random variables X, Y and Z, the following inequality holds true and can be proved easily

$$H(Z) \leq H(X, Z) + H(Y, Z) - H(X, Y, Z). \tag{82}$$

We use this inequality to upper bound the term $H(S_1^2, S_2^1)$ as

$$\begin{aligned} H(S_1^2, S_2^1) &\leq H(S_1) + H(S_2) - H(S_1, S_2) \\ &\overset{(*)}{=} 2H(S_1) - H(S_1, S_2), \end{aligned} \tag{83}$$

where we applied (82) with $Z = (S_1^2, S_2^1)$, $X = S_1$ and $Y = S_2$ and $(*)$ follows from symmetry.

Next, applying (82) with $Z = (S_3^{[1,2]}, S_{[1,2]}^3)$, $X = (S_1, S_2)$, and $Y = S_3$, we get

$$\begin{aligned} H(S_3^{[1,2]}, S_{[1,2]}^3) &\leq H(S_1, S_2) + H(S_3) - H(S_1, S_2, S_3) \\ &= H(S_1, S_2) + H(S_1) - \underbrace{H(F, S_1, S_2, S_3)}_{\text{file reconstruction}} \\ &\leq H(S_1, S_2) + H(S_1) - H(F, S_1). \end{aligned} \tag{84}$$

Combining, (83) and (84) we have

$$\begin{aligned} \text{T3} &= H(S_1^2, S_2^1) + H(S_3^{[1,2]}, S_{[1,2]}^3) \\ &\leq 2H(S_1) - H(S_1, S_2) + H(S_1, S_2) + H(S_1) - H(F, S_1) \\ &= 3H(S_1) - H(F, S_1). \end{aligned} \tag{85}$$

Finally, substituting (80), (81) and (85) in (79), we have

$$\begin{aligned} H(F) &\leq 2H(W_1) - \text{T1} - \text{T2} + \text{T3} \\ &\leq 2H(W_1) - H(F, S_1) + H(W_1) - H(S_1) \\ &\quad + 3H(S_1) - H(F, S_1) \\ &= 3H(W_1) - 2H(F, S_1) + 2H(S_1). \end{aligned} \tag{86}$$

1) Type-II Security: For Type-II security, since $\ell = 1$, we have $H(F, S_1) = H(S_1) + H(F|S_1) = H(S_1) + H(F)$. Hence, we have using (86),

$$\begin{aligned} H(F) &\leq 3H(W_1) - 2H(F, S_1) + 2H(S_1) \\ &= 3H(W_1) - 2H(F) - 2H(S_1) + 2H(S_1) \\ &\leq 3\alpha - 2H(F) \end{aligned} \tag{87}$$

which implies that

$$3H(F) \leq 3\alpha, \tag{88}$$

and thereby gives the bound for the Type-II security

$$\mathcal{B}_{\text{Type-II}}^S \leq \alpha. \tag{89}$$

2) Type-I Security: We note that for $(4, 3, 3)$-DSS and $\ell = 1$, under Type-I secrecy constraint, Theorem 4 states that the optimal (α, β)-tradeoff is given by

$$\mathcal{B}_{\text{Type-I}}^S \leq \min\left(\min(\alpha, 2\beta) + \min(\alpha, \beta), \frac{\alpha + 6\beta}{3}\right). \tag{90}$$

We note that the bound $\mathcal{B}_{\text{Type-I}}^S \leq \min(\alpha, 2\beta) + \min(\alpha, \beta)$ follows directly from the existing cutset bound (11). Hence, it remains to show that $\mathcal{B}_{\text{Type-I}}^S \leq \frac{\alpha+6\beta}{3}$.

Note that $H(F, S_1) \geq H(F, W_1)$ since W_1 is a function of the S_1. Further, for the Type-I security constraints, we have $H(F, W_1) = H(W_1) + H(F|W_1) = H(W_1) + H(F)$. By using (86) and the fact that $H(F, S_1) \geq H(F, W_1)$, we have

$$\begin{aligned} H(F) &\leq 3H(W_1) - 2H(F, S_1) + 2H(S_1) \\ &\leq 3H(W_1) - 2H(F, W_1) + 2H(S_1) \\ &= 3H(W_1) - 2H(F) - 2H(W_1) + 2H(S_1) \\ &= H(W_1) - 2H(F) + 2H(S_1) \\ &= H(W_1) - 2H(F) + 2H(S_{21}, S_{31}, S_{41}) \\ &\leq \alpha + 2(3\beta) - 2H(F), \end{aligned} \tag{91}$$

which implies that $3H(F) \leq \alpha + 6\beta$ and hence we have the proof of the bound for the Type-I security

$$\mathcal{B}_{\text{Type-I}}^S \leq \frac{\alpha + 6\beta}{3}. \tag{92}$$

ACKNOWLEDGEMENT

The authors would like to thank Dr. Soheil Mohajer for several helpful discussions.

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 *Proc. USENIX HotStorage*, Sep. 2013, pp. 1–5.

[2] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Enabling node repair in any erasure code for distributed storage," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Jul./Aug.. 2011, pp. 1235–1239.

[3] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[4] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, 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 nonachievability of interior points on the storage-bandwidth tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.

[6] V. R. Cadambe, S. A. 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, B. Sasidharan, V. Aggarwal, V. A. Vaishampayan, and P. V. Kumar, "Layered exact-repair regenerating codes via embedded error correction and block designs," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1933–1947, Apr. 2015.

[8] S. Goparaju, S. El Rouayheb, and R. Calderbank. (Feb. 2014). "New codes and inner bounds for exact repair in distributed storage systems." [Online]. Available: http://arxiv.org/abs/1402.2343

[9] C. Tian, "Rate region of the (4, 3, 3) exact-repair regenerating codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1426–1430.

[10] C. Tian, "Characterizing the rate region of the (4,3,3) exact-repair regenerating codes," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 967–975, May 2014.

[11] S. Mohajer and R. Tandon, "Exact repair for distributed storage systems: Partial characterization via new outer bounds," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2015, pp. 130–135.

[12] S. Mohajer and R. Tandon, "New bounds on the (n, k, d) storage systems with exact repair," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 2056–2060.

[13] M. Elyasi, S. Mohajer, and R. Tandon, "Linear exact repair rate region of $(k + 1, k, k)$ distributed storage systems: A new approach," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 2061–2065.

[14] S. Mohajer, M. Elyasi, and R. Tandon, "Linear exact repair rate region of $(k + 1, k, k)$ distributed storage systems: A new approach," *IEEE Trans. Inf. Theory*, to be published.

[15] B. Sasidharan, K. Senthoor, and P. V. Kumar. (Dec. 2013). "An improved outer bound on the storage-repair-bandwidth tradeoff of exact-repair regenerating codes." [Online]. Available: http://arxiv.org/abs/1312.6079

[16] I. M. Duursma. (Jun. 2014). "Outer bounds for exact repair codes." [Online]. Available: http://arxiv.org/abs/1406.4852

[17] O. O. Koyluoglu, A. S. Rawat, and S. Vishwanath, "Secure cooperative regenerating codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5228–5244, Sep. 2014.

[18] S. Pawar, S. El Rouayheb, and K. Ramchandran, "Securing dynamic distributed storage systems against eavesdropping and adversarial attacks," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6734–6753, Oct. 2012.

[19] 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, USA, Dec. 2011, pp. 1–5.

[20] B. Sasidharan, P. V. Kumar, N. B. Shah, K. V. Rashmi, and K. Ramachandran, "Optimality of the product-matrix construction for secure MSR regenerating codes," in *Proc. Int. Symp. Commun. Control Signal Process. (ISCCSP)*, Athens, Greece, May 2014, pp. 10–14.

[21] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 212–236, Jan. 2014.

[22] S. Goparaju, S. El Rouayheb, R. Calderbank, and H. V. Poor, "Data secrecy in distributed storage systems under exact repair," in *Proc. IEEE Int. Symp. Netw. Coding (NETCOD)*, Calgary, AB, Canada, Jun. 2013, pp. 1–6.

[23] K. V. Rashmi, N. B. Shah, and K. Ramchandran, "A piggybacking design framework for read-and download-efficient distributed storage codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jun. 2013, pp. 331–335.

[24] F. Cheng, R. W. Yeung, and K. W. Shum, "Imperfect secrecy in wiretap channel II," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 628–636, Jan. 2015.

[25] K. V. Rashmi, N. B. Shah, P. V. Kumar, and K. Ramchandran, "Explicit construction of optimal exact regenerating codes for distributed storage," in *Proc. Allerton Conf. Control, Comput., Commun.*, Champaign, IL, USA, 2009, pp. 1243–1249.

[26] R. Tandon, S. D. Amuru, T. C. Clancy, and R. M. Buehrer, "Distributed storage systems with secure and exact repair—New results," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2014, pp. 1–6.

[27] R. Tandon, S. Amuru, T. C. Clancy, and R. M. Buehrer, "On secure distributed storage systems with exact repair," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 3908–3912.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

Ravi Tandon (S'03–M'09) is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Arizona. Prior to joining the University of Arizona in Fall 2015, he was a Research Assistant Professor at Virginia Tech with positions in the Bradley Department of ECE, Hume Center for National Security and Technology and at the Discovery Analytics Center in the Department of Computer Science.

He received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Kanpur in 2004 and the Ph.D. degree in Electrical and Computer Engineering from the University of Maryland, College Park (UMCP) in 2010. From 2010 to 2012, he was a post-doctoral research associate in the Department of Electrical Engineering at Princeton University.

He is a recipient of the Best Paper Award at IEEE GLOBECOM 2011. He was nominated for the Graduate School Best Dissertation Award, and also for the ECE Distinguished Dissertation Fellowship Award at the University of Maryland, College Park. His current research interests include information theory and its applications to wireless networks, communications, security and privacy, distributed storage systems, machine learning and data mining.

SaiDhiraj Amuru (S'12–M'15) is a Chief Engineer at the Samsung R&D Institute, Bangalore. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2009. He received the Ph.D. degree in Electrical and Computer Engineering from Virginia Tech in Sept. 2015. His Ph.D. advisor at Virginia Tech was Dr. R. Michael Buehrer. From 2009 to 2011, he was with Qualcomm, India, as a Modem Engineer. He visited the Networks, Economics, Communication Systems, Informatics and Multimedia Research Lab, University of California, Los Angeles, during the summer of 2014. His research interests include cognitive radio, statistical signal processing, and online learning.

Thomas Charles Clancy (S'02–M'06–SM'10) is an Associate Professor of Electrical and Computer Engineering at Virginia Tech, Director of the Hume Center for National Security and Technology, the L-3 Communications Faculty Fellow in Cybersecurity of the College of Engineering, and Co-Director of the NSF Security and Software Engineering Research Center. Dr. Clancy received his M.S. in Electrical Engineering from the University of Illinois and his Ph.D. in Computer Science from the University of Maryland where his studies focused on information-theoretic foundations of communications and security. He is author to over 100 peer-reviewed publications and is a Senior Member of the IEEE. His research interests are focused in wireless security and electronic warfare.

Richard Michael Buehrer (S'89–M'91–SM'04) joined Virginia Tech from Bell Labs as an Assistant Professor with the Bradley Department of Electrical and Computer Engineering in 2001. He is currently a Professor of Electrical Engineering and is the director of Wireless Virginia Tech, a comprehensive research group focusing on wireless communications. During 2009 Dr. Buehrer was a visiting researcher at the Laboratory for Telecommunication Sciences (LTS) a federal research lab which focuses on telecommunication challenges for national defense. While at LTS, his research focus was in the area of cognitive radio with a particular emphasis on statistical learning techniques.

His current research interests include position location networks, iterative receiver design, dynamic spectrum sharing, cognitive radio, communication theory, Multiple Input Multiple Output (MIMO) communications, intelligent antenna techniques, Ultra Wideband, spread spectrum, interference avoidance, and propagation modeling. His work has been funded by the National Science Foundation, the Defense Advanced Research Projects Agency, Office of Naval Research, and several industrial sponsors.

Dr. Buehrer has authored or co-authored over 50 journal and approximately 150 conference papers and holds 11 patents in the area of wireless communications. In 2010 he was co-recipient of the Fred W. Ellersick MILCOM Award for the best paper in the unclassified technical program. He is currently a Senior Member of IEEE, and an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE WIRELESS COMMUNICATIONS LETTERS. He was formerly an associate editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGIES, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON EDUCATION. In 2003 he was named Outstanding New Assistant Professor by the Virginia Tech College of Engineering and received the Dean's Award for Teaching Excellence in 2014.