Secure Distributed Storage Systems: Local Repair with Minimum Bandwidth Regeneration

(Invited Paper)

Ankit Singh Rawat*, Natalia Silberstein†, O. Ozan Koyluoglu‡ and Sriram Vishwanath*
*Dept. of ECE, The University of Texas at Austin, Austin, TX 78712, USA, E-mail: {ankitsr, sriram}@utexas.edu
†Dept. of CS, Technion – Israel Institute of Technology, Haifa 32000, Israel, E-mail: natalys@cs.technion.ac.il
‡Dept. of ECE, The University of Arizona, Tucson, AZ 85721, USA, E-mail: ozan@email.arizona.edu

***Abstract*—This paper addresses the issue of securing information stored on a distributed storage system from a passive eavesdropping attack. The security notion is perfect secrecy, *i.e.*, the system is said to be secure only if the mutual information between the stored information and the observations at the adversary is zero. The paper summarizes state of the art on securing repair-efficient distributed storage systems. Then, storage systems that employ locally repairable codes with minimum bandwidth regenerating codes as local codes (MBR-LRCs) are investigated. A secure file size upper bound and a construction of secure MBR-LRCs are provided. These two are shown to match under special cases, establishing the secrecy capacity of these systems.**

I. Introduction

Designing efficient mechanisms to store large volumes of data is an important problem as more and more information is being generated and stored. Distributed storage system (DSS), where data is stored on a network of storage nodes, has emerged as the prevalent approach to address this ever growing need for cloud storage. DSS must store data with redundancy in order to tolerate inevitable node (disk) failures. Erasure codes as opposed to simple replication allow for resilience against node failures at the cost of small storage overhead. Classical erasure codes exhibit high communication bandwidths when the content of a failed node has to be reconstructed. Regenerating codes [1] are proposed to have such bandwidth-efficient node repairs. Multiple alternate classes of codes that enable efficient node repair have recently been proposed in [1]–[10] and references therein. These codes mainly consider *repair-bandwidth* [1], *i.e.*, the amount of data downloaded during a node repair from remaining nodes in the system, and/or *locality* [6], *i.e.*, the number of remaining nodes contacted in the event of a node failure to perform node repair.

Another important issue that is crucial for successful implementation of a DSS is its resilience to adversarial attacks. In these attacks, an adversary may attempt to gain access to the valuable information stored on the system or may want to modify the stored information in order to disrupt the functions that utilize this stored information. In this paper, we focus on the first kind of adversarial attacks where a passive eavesdropper tries to obtain the stored information by observing a certain number of nodes in the system. For this eavesdropper model, we study the perfect secrecy capacity of a DSS, *i.e.*, the maximum amount of information that can be stored on the system without leaking any information to the eavesdropper. While addressing the secrecy capacity of a DSS, we restrict ourselves to the coding schemes that ensure reliability against node failures and enable efficient node repairs as these properties are instrumental for practical storage systems. In particular, we explore the secrecy capacity of a DSS with minimum bandwidth regenerating locally repariable codes (MBR-LRCs) [9], [11], where the system exhibits properties of minimum distance optimal locally repairable codes (LRCs) [7] together with a node regeneration with minimum repair bandwidth [1].

The rest of the paper is organized as follows. Sec. II introduces our system model, preliminaries, and necessary background material, where we summarize the existing literature on secrecy capacity of regenerating codes, LRCs, and MSR-LRCs, a family of codes that combine minimum storage regenerating (MSR) codes with LRCs to allow for both small locality and regeneration efficiencies. In Sec. III, we focus on secrecy capacity of MBR-LRCs [9], [11], where the underlying codes are LRCs with minimum bandwidth regenerating (MBR) codes as their local codes, and detail the secrecy capacity achieving code constructions.

II. System Model, Preliminaries, and Background

A. System model

We consider a DSS that stores a file $\mathbf{f}$ that is $\mathcal{M}$ symbols long (over a finite field $\mathbb{F}$) on n live nodes. The file $\mathbf{f} = (f_1, \ldots, f_{\mathcal{M}})$ is first encoded into n data blocks $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, each of length α over $\mathbb{F}$. Each data block is stored on a different node. During a node repair process, a newcomer node contacts d out of the remaining $n-1$ nodes and downloads β symbols from each contacted node.

B. Regenerating codes

In [1], Dimakis *et al.* establish a trade-off between repair-bandwidth $d\beta$ and per node storage α for (n,k)-DSS, where (i) content of any k nodes are sufficient to reconstruct the original file $\mathbf{f}$; and (ii) *any* d out of remaining $n-1$ nodes allows node repair. The codes that achieve this repair-bandwidth vs. storage trade-off are termed *regenerating codes*; in particular, the codes that attain two extreme points of the trade-off are

referred to as *minimum storage regenerating (MSR)* codes and *minimum bandwidth regenerating (MBR)* codes, respectively. Explicit constructions for regenerating codes that enable exact repair are presented in [2], [5], [12] and references therein.

C. Locally repairable codes

An $(n, \mathcal{M}, \alpha)_{\mathbb{F}_q}$ vector code $\mathcal{C}$ such that $|\mathcal{C}| = q^{\mathcal{M}}$, is said to be a *locally repairable code* and denoted by (r, δ, α)-LRC, if every encoded block in an n-blocks long codeword $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_n) \in \mathcal{C}$ is α symbols long, and for each $i \in [n]$, there exists a set $\Gamma(i) \subseteq [n]$ with the following three properties: (1) $i \in \Gamma(i)$, (2) $|\Gamma(i)| \leq r+\delta-1$, and (3) minimum Hamming distance of $\mathcal{C}|_{\Gamma(i)}$, the code obtained by puncturing $\mathcal{C}$ over blocks in $[n]\backslash\Gamma(i)$, is at least δ.

Distinct sets in $\{\Gamma(i)\}_{i\in[n]}$ are called *local groups*. For an $(n, \mathcal{M}, r, \delta, \alpha)$-LRC $\mathcal{C}$, its minimum distance satisfies [8], [9]

$$d_{\min}(\mathcal{C}) \leq n - \left\lceil \frac{\mathcal{M}}{\alpha} \right\rceil + 1 - \left(\left\lceil \frac{\mathcal{M}}{r\alpha} \right\rceil - 1 \right) (\delta - 1).$$

Explicit constructions for minimum distance optimal *scalar* LRCs, *i.e.*, $\alpha = 1$, are presented in [6], [10], [13]–[15]. In [7], Papailiopoulos *et al.* design $d_{\min}$-optimal *vector* LRCs, *i.e.*, $\alpha > 1$, with a single local parity, *i.e.*, δ = 2. [8], [9] present codes that achieve this bound for general (α, δ) and have g disjoint local groups with $\{\mathcal{G}_i\}_{i\in[g]}$ denoting the set of indices of nodes in g local groups. It follows from property (3) of an (r, δ, α)-LRC that a node can be repaired by contacting at most $d = r$ nodes from a local group it belongs to, with repair-bandwidth $d\alpha$ (here, $\beta = \alpha$). An LRC where sub-codes $\mathcal{C}|_{\Gamma(i)}$ for each $i \in [n]$ are MSR codes (MBR codes) is referred to as an MSR-LRC (MBR-LRC). For an MSR-LRC or MBR-LRC, it possible to lower repair-bandwidth by allowing for $d > r$ and $\beta < \alpha$ [8], [9]. We denote such a code by an $(r, \delta, \alpha, \beta, d)$-LRC.

D. Gabidulin codes

Gabidulin codes [16] is a family of maximum rank distance (MRD) codes which is shown to be an important component of construction of secure and locally repairable DSS in [8], [15]. For an $[N, K, D]_{q^m}$ Gabidulin code with $m > N$, encoding process of a message vector $(f_1, \ldots, f_K) \in \mathbb{F}_{q^m}^K$ comprises following two steps: 1) Construct a data linearized polynomial of the form $f(y) = \sum_{i=1}^{K} f_i y^{q^{i-1}}$. 2) Evaluate $f(y)$ at N linearly independent (over $\mathbb{F}_q$) points in $\mathbb{F}_{q^m}$, $\{y_1, \ldots, y_N\} \subseteq \mathbb{F}_{q^m}$, to obtain a Gabidulin codeword $\mathbf{c} = (c_1, \ldots, c_N) = (f(y_1), \ldots, f(y_N)) \in \mathbb{F}_{q^m}^N$.

Next, we state properties of linearized polynomials which prove instrumental in their application to DSS setting.

Property 1. A linearized polynomial $f(y) = \sum_{i=1}^{K} f_i y^{q^{i-1}}$ satisfies $f(ay_1+by_2) = af(y_1)+bf(y_2)$, for any $y_1, y_2 \in \mathbb{F}_{q^m}$ and $a, b \in \mathbb{F}_q$.

Property 2. Given evaluations of $f(\cdot)$ at any K linearly independent (over $\mathbb{F}_q$) points in $\mathbb{F}_{q^m}$, say $\{z_1, \ldots, z_K\}$, one can recover $f(\cdot)$ and therefore reconstruct data vector $(f_1, \ldots, f_K)$ by performing polynomial interpolation.

E. Eavesdropper model and proof of secrecy

Consider an (ℓ_1, ℓ_2) eavesdropper, which can access the stored data of nodes in the set $\mathcal{E}_1$, and additionally can access both the stored and downloaded data at the nodes in the set $\mathcal{E}_2$ with $\ell_1 = |\mathcal{E}_1|$ and $\ell_2 = |\mathcal{E}_2|$. This eavesdropper model, defined in [17], generalizes the eavesdropper model considered in [18]. A DSS is said to achieve a secure file size of $\mathcal{M}^s$ against an (ℓ_1, ℓ_2) eavesdropper, if for any sets $\mathcal{E}_1$ and $\mathcal{E}_2$ of size ℓ_1 and ℓ_2, respectively, we have $I(\mathbf{f}^s; \mathbf{e}) = 0$. Here, $\mathbf{f}^s$ denotes the secure information of size $\mathcal{M}^s$, and $\mathbf{e}$ represents the eavesdropper's observation vector.

The following lemma [17], [19] allows one to establish the perfect secrecy for a given coding scheme for DSS.

Lemma 1. *Consider a system with information symbols $\mathbf{f}^s$, random symbols $\mathbf{r}$ (independent of $\mathbf{f}^s$), and an eavesdropper with its observations given by $\mathbf{e}$. If $H(\mathbf{e}) \leq H(\mathbf{r})$ and $H(\mathbf{r}|\mathbf{f}^s, \mathbf{e}) = 0$, then $I(\mathbf{f}^s; \mathbf{e}) = 0$.*

F. Secure regenerating codes

In [18], Pawar *et al.* establish the following upper bound on the secure file size when the eavesdropper observes the content of ℓ nodes.

$$\mathcal{M}^s \leq \sum_{i=\ell+1}^{k} \min\{(d-i+1)\beta, \alpha\}. \tag{1}$$

At MBR point when $d = n-1$, [18] shows the tightness of this bound. [17] proposes product matrix based secure coding scheme achieving this bound for any $\ell < k$ at the MBR point with general d. However, at the MSR point the coding scheme proposed in [17] can only store a secure file of size $(k-\ell_1-\ell_2)(\alpha-\ell_2\beta)$. Note that the bound in (1) reduces to $\mathcal{M}^s \leq (k-\ell_1-\ell_2)\alpha$ at the MSR point, which concludes that the coding scheme from [17] characterizes secrecy capacity only when $\ell_2 = 0$. In [8], the following improved bound on secrecy capacity at MSR point is obtained.

Theorem 1. *For an (n, k)-DSS employing an MSR code, we have*

$$\mathcal{M}^s \leq \sum_{i=\ell_1+1}^{k-\ell_2} \left(\alpha - I(\mathbf{x}_i; \mathbf{d}_{i,k-\ell_2+1}, \ldots, \mathbf{d}_{i,k})\right), \tag{2}$$

where $\mathbf{x}_i$ and $\mathbf{d}_{i,j}$ denote the data stored on node i and the data downloaded from node i to perform node repair at node j, respectively.

For linear exact-MSR codes with $d = n-1$, a lower bound on the mutual information terms in (2) is obtained for $\ell_2 \leq 2$ in [8]. In [20], Goparaju *et al.* generalize this lower bound to $\ell_2 < k$ case, resulting in the following bound.

Corollary 1. *For a linear (n, k) exact-MSR code with $d = n-1$, secrecy capacity against an (ℓ_1, ℓ_2)-eavesdropper satisfies*

$$\mathcal{M}^s \leq (k-\ell_1-\ell_2)\left(1 - \frac{1}{n-k}\right)^{\ell_2} \alpha. \tag{3}$$

In [8], we present a secure coding scheme at the MSR point.

This construction first precodes the secure file and random symbols using a Gabidulin code (similar to the classical secret sharing scheme [21]) and then encodes the resulting symbols with zigzag codes [5] (an MSR code). This scheme achieves the secure file size in the right hand side of (3) for any ℓ_1 and ℓ_2 when $\mathcal{E}_2 \cap [k] \geq (\ell_2 - 1)$. The characterization of secrecy capacity in general remains open.

G. Secure locally repairable codes

For LRCs with a *single* local parity per local group, *i.e.*, $\delta = 2$, to perform node repair a newcomer node downloads all the data stored on r surviving nodes from its own local group. Therefore, all the information in the group is revealed to an eavesdropper that observes the data downloaded during a single node repair. In [8], we characterize the secrecy capacity for $d_{\min}$-optimal LRCs with single parity per local group:

Theorem 2. *The secrecy capacity of an $(r, \delta = 2, \alpha, \beta = \alpha, d = r)$-LRC against an (ℓ_1, ℓ_2)-eavesdropper is*

$$\mathcal{M}^s = [\mu r + h - (\ell_2 r + \ell_1)]^+ \alpha, \tag{4}$$

where $[a]^+$ denotes $\max\{a, 0\}$, and $\mu \geq 0$ and $0 \leq h \leq r$ are positive integers such that $n - d_{\min} + 1 = \mu(r+1) + h$.

For LRCs with *multiple* parities per local group, *i.e.*, $\delta > 2$, the secrecy capacity depends on the node repair model employed by the system. For naïve repair model, *i.e.*, when a newcomer contacts r out of $r+\delta-2$ surviving nodes in its local group and downloads *all* the data stored on these r nodes, we get the same characterization of secrecy capacity as presented in Theorem 2 with $\mu \geq 0$ and $0 \leq h < r+\delta-1$ denoting two positive integers such that $n - d_{\min} + 1 = \mu(r + \delta - 1) + h$.

On the other hand, if regenerating codes are employed as local codes per group (when $\delta > 2$), and repair-bandwidth efficient node repairs are performed, then one can improve the secrecy capacity of DSS against eavesdropping attacks. In [8], we consider MSR-LRCs that have MSR codes as their local codes. In particular, for a special set of parameters, secrecy capacity is characterized as follows [8].

Theorem 3. *For an $(r, \delta > 2, \alpha, \beta, d)$-MSR-LRC, the secrecy capacity against an (ℓ_1, ℓ_2)-eavesdropper with $\ell_2 r + \ell_1 \leq \mu r + \min\{h, r\}$ is given by*

$$\mathcal{M}^s = (\mu r + \min\{h, r\} - \ell_2 - \ell_1)\alpha - \ell_2(r-1)\beta. \tag{5}$$

An achievability scheme for MSR-LRC was presented in [8]. This scheme, based on Gabidulin codes, generalizes the construction for secure MSR codes. Moreover, [8] presents a general bound on $\mathcal{M}^s$ for all (ℓ_1, ℓ_2) with $\ell_1 + \ell_2 < k$. However, there is a gap between the bound and secure file size of the scheme in [8] for ℓ_1 and ℓ_2 not covered by Theorem 3.

III. Secrecy in DSS Employing MBR-LRCs

In this section, we consider secure MBR-LRCs. Recall that MBR-LRCs is a family of $(r, \delta, \alpha, \beta, d)$-LRCs where each local code, *i.e.*, sub-code obtained by puncturing the code outside a local group, is an MBR code. The concept of MBR-LRCs is first introduced in [9]. For an MBR-LRC, we have the following upper bound on its minimum distance [9].

$$d_{\min} \leq n - P^{(\text{inv})}(\mathcal{M}) + 1, \tag{6}$$

where the function $P^{(\text{inv})}(\cdot)$ is defined as

$$P^{(\text{inv})}(v_1 \mathcal{M}^{\text{loc}} + v_0) = v_1(r + \delta - 1) + \nu \tag{7}$$

for some $v_1 \geq 0$, $1 \leq v_0 \leq \mathcal{M}^{\text{loc}}$, where ν is uniquely determined from $\alpha(\nu-1) - \binom{\nu-1}{2}\beta < v_0 \leq \alpha\nu - \binom{\nu}{2}\beta$ and $\mathcal{M}^{\text{loc}} = r\alpha - \binom{r}{2}\beta$. We refer the reader to [9] for a detailed introduction to MBR-LRCs. In [11], Kamath *et al.* present the first explicit construction of distance optimal MBR-LRCs (w.r.t. (6)) for a wide range of system parameters.

In what follows, we assume for simplicity that $(r+\delta-1)|n$. Note that the distance optimal MBR-LRCs necessarily have disjoint local groups [9] and that, for MBR-LRCs, the amount of data downloaded during a node repair is exactly equal to what is eventually stored on the newcomer; as a result, without of loss of generality, we can assume that $\ell_1 > 0$ and $\ell_2 = 0$. Assuming that an MBR-LRC has g disjoint local groups, we denote an $(\ell_1, 0)$-eavesdropper with $\mathcal{E}_1 = \cup_{i=1}^{g} \mathcal{E}_1^i$, where $\mathcal{E}_1^i$ represents the sets of indices of eavesdropped nodes in i-th local group. Let $|\mathcal{E}_1^i| = l_1^i$, which implies that $\ell_1 = \sum_{i=1}^{g} l_1^i$. A data collector contacts to nodes in the set $\mathcal{K} = \cup_{i=1}^{g} \mathcal{K}_i$ with $|\mathcal{K}| \leq n - d_{\min} + 1$ to reconstruct the file. Here, $\mathcal{K}_i$ denotes the set of indices of nodes that the data collector contacts in i-th local group. The following lemma provides an upper bound on the secrecy capacity of an $(r, \delta, \alpha, d, \beta)$-LRC with g local groups [8].

Lemma 2. *For a DSS employing an $(r, \delta, \alpha, \beta, d)$-LRC that is secure against an (ℓ_1, ℓ_2)-eavesdropper, we have*

$$\mathcal{M}^s \leq \sum_{i=1}^{g} H(\mathbf{x}_{\mathcal{K}_i} | \mathbf{x}_{\mathcal{E}_1^i}) \;\; \textit{for all } \left(\{\mathcal{E}_1^i, \mathcal{K}_i\}_{i=1}^{g}\right) \in \mathcal{X}, \tag{8}$$

where $\mathcal{X}$ denotes the set of tuples $(\{\mathcal{E}_1^i, \mathcal{K}_i\}_{i=1}^{g})$ that are allowed under our model.

Next, we utilize Lemma 2 to obtain an explicit upper bound on the secrecy capacity for MBR-LRCs. Let vector $\mathbf{l}_1 = (l_1^1, \ldots, l_1^g)$ denote a pattern of eavesdropped nodes. We assume that μ and h are integers such that $n - d_{\min} + 1 = \mu(r+\delta-1)+h$, where $\mu \geq 0$ and $0 \leq h < r+\delta-1$. We define $\tilde{h} = \min\{r, h\}$ and assume $l_1 < \mu r + \tilde{h}$ (otherwise the secrecy capacity is necessarily zero as implied by the following).

Theorem 4. *For an $(r, \delta > 2, \alpha, \beta, d)$-MBR-LRC, the secrecy capacity against an $(\ell_1, 0)$-eavesdropper satisfies*

$$\mathcal{M}^s \leq \rho \sum_{i=\xi+1}^{r-1} (d-i)\beta + (\mu - \rho) \sum_{i=\xi}^{r-1} (d-i)\beta + \sum_{i=\nu}^{\tilde{h}-1} (d-i)\beta,$$

where (ξ, ρ, ν) is a tuple of positive integers such that $\xi < r$, $\rho \leq \mu$, $\nu \leq \min\{h, \xi\}$, and $\ell_1 = \xi\mu + \rho + \nu$.

Proof. For an MBR-LRC, we have

$$H(\mathbf{x}_{\mathcal{K}_i} | \mathbf{x}_{\mathcal{E}_1^i}) \leq \sum_{j=l_1^i}^{\min\{|\mathcal{K}_i|, r\}-1} (d-j)\beta. \tag{9}$$

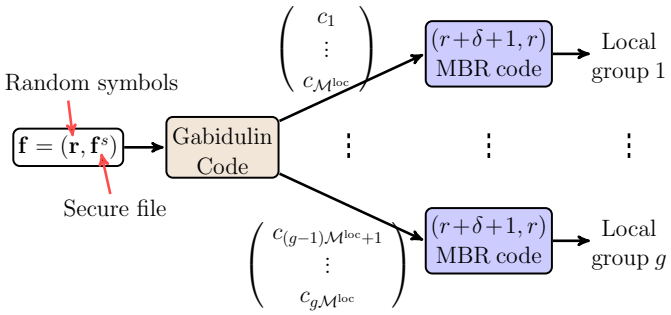

Fig. 1: Description of construction for secure MBR-LRC.

Now, we consider the data collector defined by $\mathcal{K}_1 = \mathcal{G}_1, \mathcal{K}_2 = \mathcal{G}_2, \ldots, \mathcal{K}_\mu = \mathcal{G}_\mu$; $\mathcal{K}_{\mu+2} = \cdots = \mathcal{K}_g = \emptyset$; and $\mathcal{K}_{\mu+1} \subset \mathcal{G}_{\mu+1}$ such that $|\mathcal{K}_{\mu+1}| = h$; and an eavesdropper with eavesdropping pattern $\mathbf{l}_1$ so that $l_1^1 = \cdots = l_1^\rho = \xi + 1$; $l_1^{\rho+1} = \cdots = l_1^\mu = \xi$; $l_1^{\mu+2} = \cdots = l_1^g = 0$; and $l_1^{\mu+1} = \nu$. By using Lemma 2 and (9), we obtain the upper bound. □

Next, we present an achievability scheme for secure MBR-LRCs. (See Fig. 1.)

Construction I: Assume that $n = g(r+\delta-1)$ and $\ell_1 = ag + b$, for $a \geq 0$ and $g > b \geq 0$. Let $\mathbf{r}$ be a random vector which contains

$$\kappa(\ell_1) = g\sum_{i=0}^{a-1}(d-i)\beta + (\ell_1 - ga)(d-a)\beta$$

i.i.d. random symbols distributed uniformly in $\mathbb{F}_{q^m}$. Assume further $\kappa(\ell_1) \leq \mathcal{M} = \mu\mathcal{M}^{\rm loc} + \tilde{h}\alpha - \binom{\tilde{h}}{2}\beta$. Given $\mathbf{f}^s = (f_1^s, \ldots, f_{\mathcal{M}^s}^s)$, a file containing $\mathcal{M}^s = \mathcal{M} - \kappa(\ell_1)$ symbols from $\mathbb{F}_{q^m}$, define $\mathbf{f} = (\mathbf{r}, \mathbf{f}^s) = (f_1 = r_1, f_2 = r_2, \ldots, f_{\mathcal{M}-\mathcal{M}^s} = r_{\mathcal{M}-\mathcal{M}^s}, f_{\mathcal{M}-\mathcal{M}^s+1} = f_1^s, \ldots, f_{\mathcal{M}} = f_{\mathcal{M}^s}^s)$. Then, encode $\mathcal{M}$ symbols long $\mathbf{f}$ using the two stage encoding process for MBR-LRCs presented in [11]:

Step 1: Encode $\mathbf{f}$ to a codeword of an $[N = g\mathcal{M}^{\rm loc}, K = \mathcal{M}, D = g\mathcal{M}^{\rm loc} - \mathcal{M} + 1]$ Gabidulin code over $\mathbb{F}_{q^m}$. (See Section II-D for encoding process of a Gabidulin code.)

Step 2: Partition $g\mathcal{M}^{\rm loc}$ symbols of the Gabidulin codeword into g disjoint groups of size $\mathcal{M}^{\rm loc}$ each. Then, apply an $(r+\delta-1, r)$ MBR code (over $\mathbb{F}_q$) with file size $\mathcal{M}^{\rm loc}$ inside each local group.

In the following, we show the security of this scheme and a special case result.

Proposition 1. *An MBR-LRC $\mathcal{C}$ obtained from Construction I achieves security against an $(\ell_1, 0)$-eavesdropper. Moreover, $\mathcal{C}$ attains the upper bound of Theorem 4 if $\mu + 1 = g$ or $\ell_1 \leq \mu + \min\{h, 1\}$.*

Proof. In order to establish the claim in the proposition, we need to show that $I(\mathbf{e}; \mathbf{f}) = 0$, where $\mathbf{e}$ denotes the observations of any $(\ell_1, 0)$-eavesdropper. We apply Lemma 1 for this. It follows from the rank accumulation profile [9] of local MBR codes that $H(\mathbf{e}) \leq \kappa(\ell_1) = H(\mathbf{r})$. Then, it remains to show $H(\mathbf{r}|\mathbf{f}^s, \mathbf{e}) = 0$ holds as well, i.e., the second requirement in Lemma 1. For this, we outline a decoding mechanism for random symbols $\mathbf{r}$ given $\mathbf{f}^s$ and $\mathbf{e}$. Consider that the eavesdropper observes maximum possible number of independent symbols, *i.e.*, $|\mathbf{e}| = \kappa(\ell_1)$. Since $(r+\delta-1, r)$ MBR codes utilized in the second step of encoding process of MBR-LRC from [11] have their encoding coefficients over $\mathbb{F}_q$, it follows from Property 1 of linearized polynomials that all encoded symbols and therefore, symbols in $\mathbf{e}$ are evaluations of data polynomial $f(\cdot)$. From these, we can remove the contribution of $\mathbf{f}^s$ to obtain $\widetilde{\mathbf{e}}$, which are evaluations of $\widetilde{f}(y) = \sum_{i=1}^{\kappa(\ell_1)} r_i y^{q^{i-1}}$ at $\kappa(\ell_1)$ linearly independent (over $\mathbb{F}_q$) points. Now it follows from Property 2 of linearized polynomials that these evaluations, *i.e.*, $\widetilde{\mathbf{e}}$, are sufficient to recover $\widetilde{f}(\cdot)$ by performing polynomial interpolation. Therefore, one can obtain $\mathbf{r}$ given $\mathbf{f}^s$ and $\mathbf{e}$. □

References

[1] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran. Network coding for distributed storage systems. *IEEE Trans. Inf. Theory*, 56(9):4539–4551, 2010.

[2] K. V. Rashmi, N. B. Shah, and P. V. Kumar. Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction. *IEEE Trans. Inf. Theory*, 57(8):5227–5239, 2011.

[3] A. Datta and F. Oggier. An overview of codes tailor-made for networked distributed data storage. *CoRR*, abs/1109.2317, 2011.

[4] D. S. Papailiopoulos, A. G. Dimakis, and V. R. Cadambe. Repair optimal erasure codes through hadamard designs. *IEEE Trans. Inf. Theory*, 59(5):3021–3037, 2013.

[5] I. Tamo, Z. Wang, and J. Bruck. Zigzag codes: MDS array codes with optimal rebuilding. *IEEE Trans. Inf. Theory*, 59(3):1597–1616, 2013.

[6] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Trans. Inf. Theory*, 58(11):6925–6934, 2012.

[7] D. S. Papailiopoulos and A. G. Dimakis. Locally repairable codes. In *Proc. IEEE ISIT*, 2012.

[8] 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*, 60(1):212–236, 2014.

[9] G. M. Kamath, N. Prakash, V. Lalitha, and P.V. Kumar. Codes with local regeneration. *CoRR*, abs/1211.1932, 2012.

[10] I. Tamo and A. Berg. A family of optimal locally recoverable codes. *CoRR*, abs/1311.3284, 2013.

[11] G M. Kamath, N. Silberstein, N. Prakash, A. S. Rawat, V. Lalitha, O. O. Koyluoglu, P. V. Kumar, and S. Vishwanath. Explicit MBR all-symbol locality codes. In *Proc. IEEE ISIT*, 2013.

[12] Z. Wang, I. Tamo, and J. Bruck. Long MDS codes for optimal repair bandwidth. In *Proc. of IEEE ISIT*, Jul. 2012.

[13] C. Huang, M. Chen, and J. Li. Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems. In *Proc. NCA*, pages 79–86, 2007.

[14] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar. Optimal linear codes with a local-error-correction property. In *Proc. IEEE ISIT*, 2012.

[15] N. Silberstein, A. S. Rawat, and S. Vishwanath. Error resilience in distributed storage via rank-metric codes. In *Proc. Allerton*, 2012.

[16] E. M. Gabidulin. Theory of codes with maximum rank distance. *Problems of Information Transmission*, 21:1 – 12, July 1985.

[17] N. Shah, K. Rashmi, and P. V. Kumar. Information-theoretically secure regenerating codes for distributed storage. In *Proc. GLOBECOM*, 2011.

[18] S. Pawar, S. El Rouayheb, and K. Ramchandran. Securing dynamic distributed storage systems against eavesdropping and adversarial attacks. *IEEE Trans. Inf. Theory*, 57(10):6734–6753, 2011.

[19] A. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54(8):1355 – 1387, October 1975.

[20] N. Shah, K. Rashmi, P. V. Kumar, and K. Ramchandran. Interference alignment in regenerating codes for distributed storage: Necessity and code constructions. *IEEE Trans. Inf. Theory*, 58(4):2134–2158, 2012.

[21] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11):612 – 613, November 1979.