Secure Storage using Maximally Recoverable Locally Repairable Codes

Tim Janz
Institute of Telecommunications
University of Stuttgart, Germany
tim.janz@inue.uni-stuttgart.de

Hedongliang Liu, Rawad Bitar
Institute of Communications Engineering
Technical University of Munich, Germany
{lia.liu, rawad.bitar}@tum.de

Frank R. Kschischang
Edward S. Rogers Sr. Department of
Electrical & Computer Engineering
University of Toronto, Canada
frank@ece.utoronto.ca

Abstract—**This paper considers data secrecy in distributed storage systems (DSSs) using maximally recoverable locally repairable codes (MR-LRCs). Conventional MR-LRCs are in general not secure against eavesdroppers who can observe the transmitted data during a global repair operation. This work enables nonzero secrecy dimension of DSSs encoded by MR-LRCs through a new repair framework. The key idea is to associate each local group with a central processing unit (CPU), which aggregates and transmits the contribution from the intact nodes of their group to the CPU of a group needing a global repair. The aggregation is enabled by so-called local polynomials that can be generated independently in each group. Two different schemes – direct repair and forwarded repair – are considered, and their secrecy dimension using MR-LRCs is derived. Positive secrecy dimension is enabled for several parameter regimes.**

I. INTRODUCTION

Locally repairable codes (LRCs) are used in distributed storage systems (DSSs) to protect data against loss due to node failures [1]–[4]. A class of LRCs introduced in [5], called *maximally recoverable locally repairable* codes (MR-LRCs [6]–[11], also known as *partial maximum-distance separable* (PMDS) codes [12]–[16]), can correct any erasure pattern that is information-theoretically correctable for a specified level of redundancy. In a DSS encoded by an MR-LRC, nodes are partitioned into local groups. Each group can repair a certain number of node failures locally, while additional failures can be also repaired by a so-called *global repair*. In this work, we consider DSSs encoded by MR-LRCs of block length N with g local groups, locality r, local distance δ and $h = N - gr$ global parities. This means that each group can repair up to $\delta - 1$ node failures by contacting r intact nodes within the same group. The h global parities permit repair of up to h additional failed nodes at arbitrary locations.

A desirable feature of DSSs is *secrecy*, i.e., the stored information should be kept secret even if some nodes are monitored by an eavesdropper. An LRC construction that maintains secrecy in the presence of a so-called (l_1, l_2)-*eavesdropper* was proposed in [17]. We assume a similar, though slightly altered, eavesdropper model where the eavesdropper has read-access to any l_1 nodes and, in addition, can observe all data traffic to—and the stored data within all nodes of—any l_2 local groups. In a DSS storing k independent symbols and having locality r, were an eavesdropper to observe any k independent symbols, it would be able to reconstruct the whole

DSS, resulting in no secrecy. Thus, to construct a DSS with a positive secrecy dimension, we assume throughout this paper that $l_1 + l_2 r < k$.

In this work, we study the secrecy of DSSs encoded with MR-LRCs. Fig. 1 illustrates a DSS that is encoded with an MR-LRC in the presence of an (l_1, l_2)-eavesdropper. We define the *secrecy dimension* as the minimum number of independent information symbols about which an (l_1, l_2)-eavesdropper cannot gain any information. The secrecy dimension of an LRC-equipped DSS has been characterized in [17], [18]. With the ability to globally repair failed nodes, DSSs encoded by MR-LRCs gain higher reliability than DSSs with LRCs. However, in the presence of an (l_1, l_2)-eavesdropper with $l_2 > 0$, the secrecy of the DSS comes under threat, possibly leading to zero secrecy dimension, since the global repair process enables the eavesdropper to obtain information from other groups.

The main contributions of this work are two new global repair schemes for DSSs encoded with MR-LRCs and the characterization of the secrecy dimensions achieved by these schemes. The main idea enabling a nonzero secrecy dimension is the introduction of a central processing unit (CPU) for each local group that serves as the interface for all communication with other local groups during a global repair. When a global repair is required, the CPU in each group having intact nodes uses so-called *local polynomials* to generate a symbol that is the contribution of that group to the global repair. The respective CPU then sends the group's contribution to the other CPUs according to the particular information-sharing scheme. The derived secrecy dimensions show that the DSSs, encoded with MR-LRCs and equipped with a global or forwarding global repair scheme, can achieve positive secrecy dimensions in presence of an (l_1, l_2)-eavesdropper.

II. PRELIMINARIES

For integers a, b with $a \le b$, let $[a, b]$ denote the set $\{a, a+1, \ldots, b-1, b\}$ and for $b \ge 1$, let $[b]$ denote the set $[1, b]$. The set of positive integers is denoted as $\mathbb{N}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Let $\mathbb{F}_{q^m}$ denote a finite extension field of degree m with base field $\mathbb{F}_q$, where q is a prime power. When the size is not relevant, we simply write $\mathbb{F}$ and we let $\mathbb{F}^* = \mathbb{F} \backslash \{0\}$. The rank of a matrix $\mathbf{M}$ with entries from $\mathbb{F}$ is denoted as $\mathrm{rk}(\mathbf{M})$. The Hadamard (element-wise) product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ is denoted
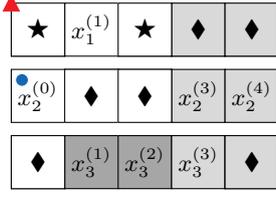
Fig. 1. Illustration of a DSS with $N = 15$ nodes and storing $k = 7$ independent symbols. The DSS is encoded by an MR-LRC with $g = 3$ groups, locality $r = 3$, local distance $\delta = 3$ (parities in light gray) and $h = 2$ global parities (in dark gray). The failed nodes marked with diamonds can be repaired locally while the failed nodes marked with stars need data from other groups to be repaired. The DSS is in presence of a $(1, 1)$-eavesdropper who can read the data stored on one node (marked by a blue circle) and the downloaded and stored data of any node in the top group (marked by a red triangle).

as $\mathbf{u} \odot \mathbf{v}$. The entropy of a discrete random variable $\mathsf{X} \in \mathcal{X}$ is defined as $\mathrm{H}(\mathsf{X}) = -\sum_{a \in \mathcal{X}: \mathrm{P_X}(a) > 0} \mathrm{P_X}(a) \log_{|\mathcal{X}|} \mathrm{P_X}(a)$. Throughout this paper, two different node indexings are used. The first is a DSS motivated indexing with group index $i \in [g]$ and node index $j \in [0, r + \delta - 2]$, e.g., in Fig. 1, the second node in the third group is denoted by $x_3^{(1)}$. The second indexing is only using one index $\mu \in [0, N - 1]$ and there is a bijective mapping $\varphi : \mathbb{N}_0 \times \mathbb{N} \to \mathbb{N}_0$ with $(i, j) \mapsto \mu = j + (r + \delta - 2)(i - 1)$. The inverse mapping φ^{-1} is written as $\varphi^{-1}(\mu) = (i(\mu), j(\mu))$, where $i(\mu) = \lceil \frac{\mu}{r+\delta-2} \rceil$ and $j(\mu) = \mu \mod (r + \delta - 2)$.

A. Linearized Reed–Solomon Codes

Let $\mathbb{F}_{q^m}[x; \sigma]$ denote the ring of skew polynomials with automorphism $\sigma : \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$ such that $\sigma(a) = a^q$ for $a \in \mathbb{F}_{q^m}$. This ring is endowed with the usual polynomial addition operation, but the multiplication, which is associative and distributes over addition, is generally non-commutative, being characterized by the property that $xa = \sigma(a)x$ for every $a \in \mathbb{F}_{q^m}$. For a vector $\mathbf{b} = (b_0, \ldots, b_{n-1}) \in \mathbb{F}_{q^m}^n$, let $\mathbf{V}_n^\sigma(\mathbf{b}) \in \mathbb{F}_{q^m}^{n \times n}$ be the σ-Vandermonde matrix as defined in [19, Def. 2]. The set $\Omega = \{b_0, \ldots, b_{n-1}\} \subseteq \mathbb{F}_{q^m}$ is a *P-independent set* if, and only if, $\mathrm{rk}(\mathbf{V}_n^\sigma(\mathbf{b})) = n$ [20, Lem. 12]. The following code family, which is based on σ-Vandermonde matrices and generalizes Reed–Solomon and Gabidulin codes [21], [22], was introduced in [23], [24].

*Definition 1 (**Linearized Reed–Solomon (LRS) Codes**):* Let $n = rg \geq k$ and $\mathbb{F}_{q^m}$ be such that $1 \leq g \leq q - 1$ and $r \leq m$ hold. Choose $\mathbf{a} = (a_1, a_1, \ldots, a_g) \in \mathbb{F}_{q^m}^g$ such that all elements a_i are from different conjugacy classes of $\mathbb{F}_{q^m}$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_g) \in \mathbb{F}_{q^m}^n$ where the entries in each $\boldsymbol{\beta}_i \in \mathbb{F}_{q^m}^r$, $i \in [g]$ are $\mathbb{F}_q$-linearly independent and let $\mathbf{b} \in \mathbb{F}_{q^m}^n$ with $b_\mu = a_{i(\mu)}(\beta_{i(\mu)}^{(j(\mu))})^{q-1}$ for $\mu \in [0, n-1]$. An (n, k) linearized Reed–Solomon code over $\mathbb{F}_{q^m}$ on $(\mathbf{b}, \boldsymbol{\beta})$ with respect to σ is given by

$$\mathcal{C}_{\mathrm{LRS}}^{\sigma, k}(\mathbf{b}, \boldsymbol{\beta}) := \{f(\mathbf{b}) \odot \boldsymbol{\beta} \mid f \in \mathbb{F}_{q^m}[x; \sigma], \deg(f) < k\}$$

with $f(\mathbf{b}) = (f(b_0), \ldots, f(b_{n-1})) \in \mathbb{F}_{q^m}^n$.

B. Skew Lagrange Polynomials and Lagrange Basis

Lagrange-type skew polynomials can be constructed by Newton interpolation for skew polynomials [25, Prop. 2.6].

*Definition 2 (**Skew Lagrange Polynomials**):* Let $\Omega = \{a_0, a_1, \ldots, a_{k-1}\} \subseteq \mathbb{F}_{q^m}$ be a P-independent set. A skew Lagrange polynomial $\ell_i^\Omega \in \mathbb{F}_{q^m}[x; \sigma]$ fulfills the constraints $\ell_i^\Omega(a_i) = 1$ and $\ell_i^\Omega(a_j) = 0$ for all $i, j \in [0, k-1], j \neq i$.

Given a skew polynomial $f \in \mathbb{F}_{q^m}[x; \sigma]$ of degree $k - 1$ evaluated on a P-independent set Ω, it can also be written in a form with the Lagrange basis $(\ell_0^\Omega, \ldots, \ell_{k-1}^\Omega) \in (\mathbb{F}_{q^m}[x; \sigma])^k$, instead of the form with the monomial basis $(1, x, \ldots, x^{k-1}) \in (\mathbb{F}_{q^m}[x; \sigma])^k$. More details on the transformation between the monomial and the Lagrange bases are given in Appendix A.

C. Maximally Recoverable Locally Repairable Codes

*Definition 3 (**MR-LRC** [5], [12]):* An LRC $\mathcal{C} \subseteq \mathbb{F}^n$ with g groups and local distance δ_i for $i \in [g]$ is said to be *maximally recoverable*, i.e., MR-LRC, if after puncturing at most $\delta_i - 1$ positions in each group, the punctured code is still MDS.

The MR-LRC construction used throughout this work is taken from [7] and defined next for completeness. It uses LRS codes and has been proven to be an MR-LRC in [7, Th. 2].

Construction 1 ([7, Constr. 1]): Let g be the number of local groups with equal locality r and local distance δ. Choose $\mathbb{F}_{q^m}$ such that $m \geq r$ and $q > \max(g, r + \delta - 2)$. The construction of the code has two steps:

1) *Outer code*: Choose an (n, k) LRS code $\mathcal{C}_{\mathrm{out}} \subseteq \mathbb{F}_{q^m}^n$ for $n = rg$.
2) *Local codes*: Choose any $(r + \delta - 1, r)$ MDS code $\mathcal{C}_{\mathrm{loc}, i} \subseteq \mathbb{F}_q^{r+\delta-1}$ which is linear over the local field $\mathbb{F}_q$ for $i \in [g]$.

The global code $\mathcal{C}_{\mathrm{glob}} \subseteq \mathbb{F}_{q^m}^N$ with $N = n + g(\delta - 1) = g(r + \delta - 1)$ is then defined by

$$\mathcal{C}_{\mathrm{glob}} = \{\mathbf{c}_{\mathrm{out}} \cdot \mathrm{diag}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_g) \mid \mathbf{c}_{\mathrm{out}} \in \mathcal{C}_{\mathrm{out}}\},$$

with $\mathbf{A}_i \in \mathbb{F}_q^{r \times (r+\delta-1)}$ being the generator matrix of $\mathcal{C}_{\mathrm{loc}, i}$ for $i \in [g]$. The number of global parities of this construction is $h = n - k = rg - k$.

To be able to securely store data in DSSs encoded by MR-LRCs in the presence of an (l_1, l_2)-eavesdropper, we use the following construction, which is similar to the construction from [17, Th. 33] based on Gabidulin codes.

Construction 2: Let k_e be the number of independent symbols that an (l_1, l_2)-eavesdropper observes in a DSS storing k independent symbols and $k_s = k - k_e$. Assume $k_s > 0$. The construction has the following steps:

1) Given a file $\mathbf{u}_s = (u_0, u_1, \ldots, u_{k_s-1})$ composed of k_s symbols in $\mathbb{F}_{q^m}$, generate $\mathbf{r} = (r_0, r_1, \ldots, r_{k_e-1})$, where the symbols r_i are independent and uniformly distributed over $\mathbb{F}_{q^m}$, for all $i \in [0, k_e - 1]$. Append $\mathbf{r}$ to $\mathbf{u}_s$ to obtain $\mathbf{u} = (\mathbf{r}, \mathbf{u}_s) \in \mathbb{F}_{q^m}^k$.
2) Encode $\mathbf{u}$ by the MR-LRC as in Construction 1.

As we will show in Section IV-A, with the two novel global repair schemes introduced in Section III-B, this construction has positive secrecy dimensions.

III. NOVEL GLOBAL REPAIR OF MR-LRCS

We first present a naive global repair scheme for MR-LRCs and point out that its secrecy dimension is zero. We then introduce a framework which allows for positive secrecy dimension in a DSS encoded by an MR-LRC in the presence of an (l_1, l_2)-eavesdropper.

A. The Naive Global Repair of MR-LRC

Consider a DSS encoded by Construction 2. In a naive global repair, each failed node downloads as many symbols from other nodes as needed for its own recovery. Note that a global repair is needed only if more than $\delta - 1$ nodes fail in some group. In this case, the nodes in this group need to download $k - \nu$ symbols from other groups to reconstruct their data, where ν is the number of intact nodes in their group. It follows from the definition of MR-LRCs that after puncturing the $\delta - 1$ failed nodes, the code is still an MDS code. Hence, as long as the failed nodes gather any k symbols from the intact nodes, they can reconstruct the whole data stored in the DSS (see also [7, Sec.III]).

The drawback of such a naive global repair is that if an (l_1, l_2)-eavesdropper with $l_2 > 0$ were to observe the nodes that require a global repair, it can also reconstruct the whole DSS after observing the global repair, leading to zero secrecy dimension.

B. Direct and Forwarded Global Repair

In our proposed global repair schemes, we assume that each group has a central processing unit (CPU) which coordinates the global repair process. If a global repair is needed in a group, the CPU of this group sends a request to the other CPUs. The CPU of each group, having sufficiently many intact nodes, collects the symbols from its group needed for the global repair and summarizes them into one symbol. This symbol is then sent to the CPU of the group that needs the global repair.

We introduce two schemes for the global repair process: direct repair and forwarded repair, which only differ in the manner in which each CPU sends its contribution to the CPU of the group needing a global repair. In the direct global repair scheme, the CPUs of the intact groups send their contribution directly to the CPU of the group that needs a repair. In the forwarded global repair scheme, each CPU of the intact groups forwards its contribution to the next CPU according to a forwarding list $\mathcal{F}$. At each CPU, the new contribution is the sum of its own contribution and, if applicable, the received contribution from the previous CPU in the forwarding list. Therefore, this scheme can be seen as an aggregate-and-forward scheme. In the forwarded repair scheme, each group receives at most one symbol. Hence, potentially increasing the secrecy dimension compared to the direct repair. The two schemes are illustrated in Fig. 2 with five groups and a global repair required in the first group.

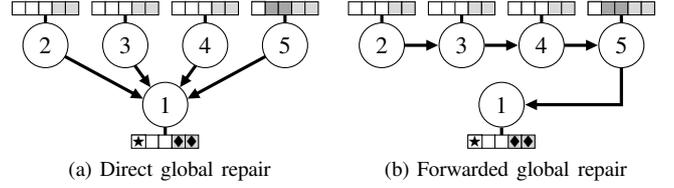(a) Direct global repair   (b) Forwarded global repair

Fig. 2. Illustration of two different global repair schemes where an erasure (star) in the first group is repaired with global repair. Each circle depicts a CPU of a group that coordinates a repair. The nodes in the groups are depicted as the little squares. The forwarding list for (b) is $\mathcal{F} = \{2, 3, 4, 5, 1\}$

C. Local Polynomials

We introduce in the following the key tool which enables each CPU to send at most one symbol to the group where the global repair is needed. This tool relies on the use of outer LRS codes in Construction 1.

We define a *minimal global repair set* as a set $\Delta_{\mathrm{gl}} \subseteq \{(i, j) \mid i \in [g], j \in [0, r + \delta - 2]\}$ of intact nodes storing independent symbols and $|\Delta_{\mathrm{gl}}| = k$. For a fixed group $s \in [g]$, it holds that $|\Delta_{\mathrm{gl}} \cap \{(s, j) \mid j \in [0, r+\delta-2]\}| \leq r$. In Fig. 1, we have $\Delta_{\mathrm{gl}} = \{(1,1), (2,0), (2,3), (2,4), (3,1), (3,2), (3,3)\}$.

Definition 4 (Local Polynomial): Let $\mathcal{C}_{\mathrm{glob}} \subseteq \mathbb{F}_{q^m}^N$ and $\mathcal{C}_{\mathrm{out}} \subseteq \mathbb{F}_{q^m}^n$ be the global and outer code from Construction 1. Fix a minimal global repair set Δ_{gl} of nodes. For a codeword $\mathbf{c} = (c_1^{(0)}, c_1^{(1)}, \ldots, c_g^{(r+\delta-2)}) \in \mathcal{C}_{\mathrm{glob}}$, the *local polynomial* $L_i \in \mathbb{F}_{q^m}[x; \sigma]$ of the i-th group has the following properties:

- $L_i(\tilde{b}_i^{(j)}) = c_i^{(j)} / \tilde{\beta}_i^{(j)}$ for all $(i, j) \in \Delta_{\mathrm{gl}}$,
- $L_i(\tilde{b}_s^{(t)}) = 0$ for all $s \neq i$ and $(s, t) \in \Delta_{\mathrm{gl}}$,

where $\tilde{\beta}_i^{(j)}$ are entries in $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} \operatorname{diag}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_g)$ and $\tilde{b}_i^{(j)} = a_i(\tilde{\beta}_i^{(j)})^{q-1}$.

Similar to the skew Lagrange polynomials, the local polynomials can be generated by the Newton interpolation [25, Prop. 2.6].

Theorem 1: We follow the notations in Definition 4. Let f be an encoding polynomial of the outer code $\mathcal{C}_{\mathrm{out}} = \mathcal{C}_{\mathrm{LRS}}^{\sigma,k}(\mathbf{b}, \boldsymbol{\beta})$. Let $\Delta_{\mathrm{gl},1} := \{i \mid (i, j) \in \Delta_{\mathrm{gl}}\}$. It holds that

$$f = \sum_{i \in \Delta_{\mathrm{gl},1}} L_i. \tag{1}$$

Proof: See Appendix B. ∎

IV. SECRECY DIMENSION OF MR-LRCS

This section investigates the secrecy dimension of a DSS encoded by Construction 2 with a direct or forwarded repair scheme. The following lemmas are needed to show the secrecy of Construction 2.

Lemma 1 (Secrecy Lemma [26]): Consider a DSS storing $\mathbf{u} = (\mathbf{u}_s, \mathbf{r})$ as in Construction 2. Let U_s, R and E be the random variables corresponding to $\mathbf{u}_s$, $\mathbf{r}$ and the symbols observed by an eavesdropper, respectively. If $H(\mathsf{E}) \leq H(\mathsf{R})$ and $H(\mathsf{R} \mid \mathsf{U}_s, \mathsf{E}) = 0$, then the eavesdropper cannot gain any information about $\mathbf{u}_s$, i.e., $I(\mathsf{U}_s; \mathsf{E}) = 0$.

Lemma 2: Let $\mathsf{K} = (\mathsf{K}_0, \mathsf{K}_1, \ldots, \mathsf{K}_{k-1})$ be a vector consisting of $k \in \mathbb{N}$ random variables. Consider a vector

$\mathsf{X} = (\mathsf{X}_0, \mathsf{X}_1, \ldots, \mathsf{X}_{m-1})$ consisting of $m \in \mathbb{N}$ random variables such that $\mathsf{X} = \mathbf{M}\mathsf{K}^{\mathsf{T}}$, where $\mathbf{M}$ is of size $m \times k$. It holds that $\mathsf{H}(\mathsf{X}) \leq k$. Furthermore, if the variables in K are i.i.d., then $\mathsf{H}(\mathsf{X}) = \mathrm{rk}(\mathbf{M})$.

Proof: See Appendix B. ∎

We now show that Construction 2 is information theoretically secure.

Theorem 2: Consider a DSS storing $\mathbf{u} = (\mathbf{u}_{\mathrm{s}}, \mathbf{r})$ as in Construction 2. We have

$$\mathsf{I}(\mathsf{U}_{\mathrm{s}}; \mathsf{E}) = 0,$$

where U_{s} and E are the random variables corresponding to the securely stored symbols and the observed symbols by the eavesdropper, respectively.

Proof: We prove the statement via Lemma 1. The first step is to show that $\mathsf{H}(\mathsf{E}) \leq \mathsf{H}(\mathsf{R})$ is fulfilled. By Construction 2, we choose $\mathbf{r}$ consisting of k_{e} random symbols. Thus, the first condition holds.

The second step is to show that $\mathsf{H}(\mathsf{R} \mid \mathsf{U}_{\mathrm{s}}, \mathsf{E}) = 0$. The eavesdropped symbols are $\mathbf{e} = \mathbf{M}\mathbf{c}_{\Delta_{\mathrm{gl}}}$, where the matrix $\mathbf{M} \in \mathbb{F}_{q^m}^{(k_{\mathrm{e}} \times k)}$ represents the k_{e} independent symbols that an eavesdropper has as constraints on the k encoded symbols as in Construction 2. The basis is the outer codeword at the global repair set without the column multipliers of the LRS code, i.e., $\mathbf{c}_{\Delta_{\mathrm{gl}}} := (\mathbf{c}_{\mathrm{out}} \odot \boldsymbol{\beta}^{-1})|_{\Delta_{\mathrm{gl}}}$. The matrix $\mathbf{M}$ can be transformed in the domain of the monomial coefficients of the encoding skew polynomial f by Definition 5 yielding

$$\mathbf{e}\mathbf{V}_k^{\sigma}(\mathbf{b})^{-1} = \underbrace{\mathbf{M}\mathbf{V}_k^{\sigma}(\mathbf{b})^{-1}}_{=:\mathbf{T}_{\mathrm{e}}}\mathbf{f},$$

where $\mathbf{f}$ denotes the coefficients of the encoding skew polynomial f and $\mathbf{b}$ are the code locators of the LRS code. The matrix $\mathbf{T}_{\mathrm{e}}$ gives k_{e} independent constraints on the polynomial coefficients $\mathbf{f}$, since $\mathbf{M}$ has rank k_{e} and $\mathbf{V}_k^{\sigma}(\mathbf{b})^{-1}$ has full rank. Together with the k_{s} coefficients of the information symbols, with random vector representation U_{s}, we have $k = k_{\mathrm{e}} + k_{\mathrm{s}}$ constraints on k coefficients of f. It remains to show that the k constraints are independent. The k_{s} coefficients can be written in a matrix $\mathbf{T}_{\mathrm{s}}$ such that $\mathbf{u}_{\mathrm{s}} = \mathbf{T}_{\mathrm{s}}\mathbf{f}$. The stacked matrix $\mathbf{T}$ consisting of $\mathbf{T}_{\mathrm{e}}$ and $\mathbf{T}_{\mathrm{s}}$ can be used to determine $\mathsf{H}(\mathsf{U}_{\mathrm{s}}, \mathsf{E})$ by Lemma 2. The matrix $\mathbf{T}_{\mathrm{s}}$ has k_{s} nonzero entries which are on the main diagonal. They contribute k_{s} to the rank of $\mathbf{T}$. If the Vandermonde matrix $\mathbf{V}_k^{\sigma}(\mathbf{b})^{-1}$ is punctured at the corresponding k_{s} columns, it still has rank $k - k_{\mathrm{s}}$ due to its structure. The overall rank of the punctured matrix $\mathbf{T}|_{k_{\mathrm{s}}}$ is therefore still $k - k_{\mathrm{s}} = k_{\mathrm{e}}$. Thus, the matrix $\mathbf{T}$ has rank k and all coefficients of f, including the random symbols $\mathbf{r}$, can be determined given $\mathbf{e}$ and $\mathbf{u}_{\mathrm{s}}$, i.e., $\mathsf{H}(\mathsf{R} \mid \mathsf{U}_{\mathrm{s}}, \mathsf{E}) = 0$ holds. ∎

Thus, for $k - k_{\mathrm{e}} > 0$, Construction 2 can guarantee a positive secrecy dimension against an (l_1, l_2)-eavesdropper that observes k_{e} independent symbols.

A. Secrecy Dimension for Direct and Forwarded Global Repair

The next step is to quantify the secrecy dimension of DSS using MR-LRCs with a direct or forwarded global repair. It

can be calculated by

$$k_{\mathrm{s}} = \mathsf{H}(\mathsf{K} \mid \mathsf{E}) = \mathsf{H}(\mathsf{K}) - \mathsf{H}(\mathsf{E}) = k - k_{\mathrm{e}},$$

which follows from $\mathsf{H}(\mathsf{K}, \mathsf{E}) = \mathsf{H}(\mathsf{K}) = k$ due to $\mathsf{E} = f(\mathsf{K})$. Thus, we quantify the secrecy dimension by calculating $\mathsf{H}(\mathsf{E})$.

1) Direct Global Repair: Before the secrecy dimension with the direct global repair scheme in the presence of an (l_1, l_2)-eavesdropper is calculated, some preliminary considerations should be made. First, note that the eavesdropper only gains knowledge during global repair if the global repair is performed in a group that is observed in an l_2-manner. Direct global repair does not reveal any information to the other groups which are only sending information. Second, if $l_2 = 0$, the secrecy dimension is $k_{\mathrm{s}} = k - l_1$ since the eavesdropper does not make any observations when global repairs are performed. Therefore, we only consider $l_2 \geq 1$. Third, the number of globally repairable erasures is bounded from above by the number of global parities $h = gr - k$. If more than h failed nodes need to be globally repaired, part of the data cannot be recovered.

Theorem 3 (Secrecy Dimension with Direct Global Repair): Consider a DSS encoded by Construction 2 with locality r, g groups and $h \leq r$ global parities in the presence of an (l_1, l_2)-eavesdropper with $l_2 \geq 1$ and $l_1 + l_2 r < k$. The secrecy dimension of the DSS with a direct global repair is

$$k_{\mathrm{s,dir}} = k - \underbrace{\left(l_2 r + l_1 - h + \sum_{i=1}^{g} \min(h, r - e_i) \right)}_{k_{\mathrm{e,dir}}}, \quad (2)$$

where e_i denotes the number of independent symbols that the eavesdropper is observing in the i-th group from the $l_2 r + l_1$ nodes, i.e., in the static case before the global repair process.

Proof:

The proof is given in Appendix C. ∎

Remark: Note that for $h \geq r$, the secrecy dimension will be zero since $k_{\mathrm{e,dir}} = k$. This can be verified by assuming that $h = r$, which means $k = n - h = gr - r$. Thus, it holds that $k_{\mathrm{e,dir}} = l_2 r + l_1 - r + \sum_{i=1}^{g} \min(r, r - e_i)$, where we have by definition of e_i that $k_{\mathrm{e,dir}} = (g - 1)r$.

2) Forwarded Global Repair: Before the secrecy dimension of a DSS with forwarded global repair is stated, we make some preliminary considerations. First, note that global repairs required by the groups observed in an l_2-manner does not add information to the eavesdropper, since the eavesdropper has observed the data stored in these groups before the nodes failed. Second, if two groups, that are observed in an l_2-manner, are next to each other in the forwarding list. The second group will only receive a symbol that is already known by the eavesdropper. Third, if the eavesdropper is at the beginning of the forwarding list, it does not receive a symbol and can thus not gain any knowledge during global repair. These special cases drive us to derive the secrecy dimension of a DSS with the forwarded global repair only for $g \geq 3$. For $g \leq 2$, the secrecy dimension is only determined by the static observations, i.e., $k_{\mathrm{s}} = k - (l_2 r + l_1)$.

Denote by $\mathcal{F}$ the forwarding list of the forwarded scheme and by $\mathcal{F}_{\text{up},i}$ the forwarding list containing the groups that are upstream with respect to the i-th group. Let $\mathcal{G}_{l_2}$ be the set of groups that are observed in an l_2-manner. If a group in $\mathcal{F}_{\text{up},i}$ is observed in an l_2-manner, let

$$\mathcal{F}'_{\text{up},i} := \mathcal{F}_{\text{up},i} \setminus \left\{ \bigcup_{\nu \in \mathcal{F}_{\text{up},i} \cap \mathcal{G}_{l_2}} \{ j \in \mathcal{F}_{\text{up},\nu} \} \cup \{ \nu \} \right\}.$$

In words, given a group that is observed in an l_2-manner, $\mathcal{F}'_{\text{up},i}$ is the set of groups that are upstream in the forwarding list $\mathcal{F}_{\text{up},i}$ between the l_2-observed group with index i and the next l_2-observed group (excluded) or until the end of the list (included). For example, in Fig. 2b, assume the group 3 and 5 were observed in an l_2-manner. Then the two lists are $\mathcal{F}_{\text{up},5} = \{2,3,4\}$ and $\mathcal{F}'_{\text{up},5} = \{4\}$.

*Theorem 4 (**Secrecy Dimension with Forwarded Global Repair**):* Consider a DSS encoded by Construction 2 with locality r, $g \geq 3$ groups and $h \leq r$ global parities in the presence of an (l_1, l_2)-eavesdropper with $l_2 \geq 1$ and $l_1 + l_2 r < k$.

The secrecy dimension of forwarded global repair is

$$k_{\text{s,fw}} = k - \underbrace{\left((l_2 r + l_1) + \sum_{i \in \mathcal{G}_{l_2}} \min(h, \sum_{j \in \mathcal{F}'_{\text{up},i}} (r - e_j)) \right)}_{k_{\text{e,fw}}},$$

(3)

where e_i denotes the number of independent symbols that the eavesdropper is observing in the i-th group from the $l_2 r + l_1$ nodes, i.e., in the static case before the global repair process.

Proof: The proof is given in Appendix C. ∎

B. Comparison of Direct and Forwarded Global Repair

We give two examples in which we compare the secrecy dimension of the introduced schemes.

Example 1: Consider the DSS depicted in Fig. 1 again. It has $g = 3$ groups, locality $r = 3$, local distance $\delta = 3$ and $h = 2$ global parities. The eavesdropper is a $(1,1)$-eavesdropper. The eavesdropped nodes and the failed nodes are depicted in Fig. 3. With the directed global repair illustrated in Fig. 3a, we can compute the secrecy dimension from (2) in Theorem 3: $k_{\text{s,dir}} = k - k_{\text{e,dir}} = 7 - 6 = 1$. Namely, one information symbol can be stored securely on the considered system with the direct scheme. With the forwarded global repair illustrated in Fig. 3b, from (3) in Theorem 4, we get that $k_{\text{s,fw}} = k - k_{\text{e,fw}} = 7 - 6 = 1$ information symbol can be stored securely on the considered system with the forwarded scheme.

Example 2: We now compare the secrecy dimension of the two global repair schemes for different number of groups g. Consider a DSS with locality $r = 7$ and $h = 3$ global parities in the presence of an $(0,1)$-eavesdropper. The DSS stores $k = r \cdot (g-1) + (r-h) = 7 \cdot (g-1) + 4$ symbols. The comparison is presented in Fig. 4, where the secrecy dimension of a DSS encoded by a conventional LRC with $h = 0$ global parity is also plotted. We can see that forwarded global repair has a higher secrecy dimension than direct global repair for $g > 3$.

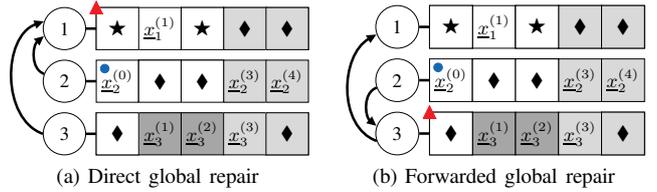(a) Direct global repair   (b) Forwarded global repair

Fig. 3. Illustration of the global repair schemes for the DSS from Fig. 1. The failed nodes marked with stars need to be repaired globally. In (a) they are repaired by direct global repair and in (b) forwarded global repair is used with the forwarding list $\mathcal{F} = \{2,3,1\}$. Both repairs are coordinated by the CPUs of the groups, depicted by the circles on the left. The secrecy rates of the DSS with respective repair schemes are given in Example 1.
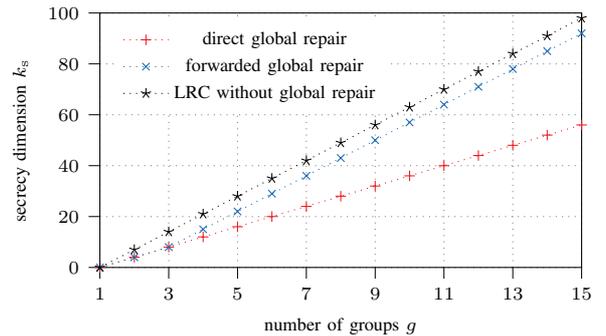

Fig. 4. Plot of the secrecy dimension of a DSS that uses an MR-LRC with forwarded global repair (blue) and direct global repair (red) for fixed parameters $l_2 = 1$, $l_1 = 0$ $r = 7$, $h = 3$. The secrecy dimensions are the same for $g \leq 3$. For $g > 3$ forwarded global repair has a higher secrecy dimension. In addition, the secrecy dimension of an LRC-coded DSS without global repair, i.e., $h = 0$ is plotted.

The secrecy dimension of the forwarding global repair scheme is only slightly below the secrecy dimension of a conventional LRC.

However, forwarded global repair has the drawback of an increasing latency in g, since each group, except the first group in the forwarding list, is waiting for the upstream contribution before sending its contribution.

V. CONCLUSION AND OUTLOOK

We have introduced a new repair framework for MR-LRCs with LRS codes. In the framework, we associate a central processing unit to each local group that uses local polynomials to summarize the global repair contribution from the local group. Two different global repair schemes are proposed and their secrecy dimensions in the presence of a passive eavesdropper are determined. For future research, it would be interesting to investigate the secrecy dimension of MR-LRCs for arbitrary repair graphs, i.e., global repair schemes that have an arbitrary global repair topology consisting of forwarding (line) and collecting (tree) structures [27]. Moreover, the general secrecy capacity of DSS that use MR-LRCs could be investigated.

REFERENCES

[1] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6925–6934, 2012.

[2] C. Huang, M. Chen, and J. Li, "Pyramid Codes: Flexible Schemes to Trade Space for Access Efficiency in Reliable Data Storage Systems," in *Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007)*, Jul. 2007, pp. 79–86.

[3] D. S. Papailiopoulos and A. G. Dimakis, "Locally Repairable Codes," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 5843–5855, Oct. 2014.

[4] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes With Local Regeneration and Erasure Correction," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4637–4660, Aug. 2014.

[5] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5245–5256, 2014.

[6] M. Chen, C. Huang, and J. Li, "On the maximally recoverable property for multi-protection group codes," in *2007 IEEE International Symposium on Information Theory*. IEEE, 2007, pp. 486–490.

[7] U. Martínez-Peñas and F. Kschischang, "Universal and dynamic locally repairable codes with maximal recoverability via sum-rank codes," *IEEE Transactions on Information Theory*, vol. PP, pp. 1–1, 06 2019.

[8] S. Gopi, V. Guruswami, and S. Yekhanin, "Maximally Recoverable LRCs: A Field Size Lower Bound and Constructions for Few Heavy Parities," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6066–6083, Oct. 2020.

[9] V. Guruswami, L. Jin, and C. Xing, "Constructions of Maximally Recoverable Local Reconstruction Codes via Function Fields," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6133–6143, Oct. 2020.

[10] H. Cai, Y. Miao, M. Schwartz, and X. Tang, "A Construction of Maximally Recoverable Codes With Order-Optimal Field Size," *IEEE Transactions on Information Theory*, vol. 68, no. 1, pp. 204–212, Jan. 2022.

[11] S. Gopi and V. Guruswami, "Improved Maximally Recoverable LRCs Using Skew Polynomials," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7198–7214, Nov. 2022.

[12] M. Blaum, J. L. Hafner, and S. Hetzler, "Partial-MDS codes and their application to raid type of architectures," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4510–4519, 2013.

[13] G. Calis and O. O. Koyluoglu, "A general construction for PMDS codes," *IEEE Communications Letters*, vol. 21, no. 3, pp. 452–455, 2016.

[14] R. Gabrys, E. Yaakobi, M. Blaum, and P. H. Siegel, "Constructions of partial MDS codes over small fields," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3692–3701, 2018.

[15] A. Neri and A.-L. Horlemann-Trautmann, "Random construction of partial MDS codes," *Designs, Codes and Cryptography*, vol. 88, no. 4, pp. 711–725, Apr. 2020.

[16] T. Bogart, A.-L. Horlemann-Trautmann, D. Karpuk, A. Neri, and M. Velasco, "Constructing Partial MDS Codes from Reducible Algebraic Curves," *SIAM Journal on Discrete Mathematics*, vol. 35, no. 4, pp. 2946–2970, Jan. 2021.

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

[18] A. Agarwal and A. Mazumdar, "Security in locally repairable storage," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6204–6217, 2016.

[19] H. Liu, H. Wei, A. Wachter-Zeh, and M. Schwartz, "Linearized reed-solomon codes with support-constrained generator matrix," in *2023 IEEE Information Theory Workshop (ITW)*. IEEE, 2023, pp. 7–12.

[20] T. Lam and A. Leroy, "Vandermonde and wronskian matrices over division rings," *Journal of Algebra*, vol. 119, no. 2, pp. 308–336, 1988. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0021869388900634

[21] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, 1960. [Online]. Available: https://doi.org/10.1137/0108018

[22] E. Gabidulin, "Theory of codes with maximum rank distance (translation)," *Problems of Information Transmission*, vol. 21, pp. 1–12, 01 1985.

[23] S. Liu, F. Manganiello, and F. R. Kschischang, "Construction and decoding of generalized skew-evaluation codes," in *2015 IEEE 14th Canadian Workshop on Information Theory (CWIT)*. IEEE, 2015, pp. 9–13.

[24] U. Martínez-Peñas, "Skew and linearized Reed-Solomon codes and maximum sum rank distance codes over any division ring," *Journal of Algebra*, vol. 504, 10 2017.

[25] U. Martínez-Peñas, M. Shehadeh, and F. R. Kschischang, "Codes in the sum-rank metric: Fundamentals and applications," *Foundations and Trends® in Communications and Information Theory*, vol. 19, no. 5, pp. 814–1031, 2022. [Online]. Available: http://dx.doi.org/10.1561/0100000120

[26] N. B. Shah, K. V. Rashmi, and P. V. Kumar, "Information-theoretically secure regenerating codes for distributed storage," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011, pp. 1–5.

[27] A. Patra and A. Barg, "Node repair on connected graphs," *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 3081–3095, 2022.

[28] W. Gander, "Change of basis in polynomial interpolation," *Numerical Linear Algebra with Applications*, vol. 12, no. 8, pp. 769–778, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.450

*Definition 5 (**Monomial and Lagrange Basis**):*
Let $f = f_0 + f_1 x + \ldots + f_{k-1} x^{k-1} \in \mathbb{F}_{q^m}[x; \sigma]$ be a skew polynomial in monomial basis with coefficient vector $\mathbf{f} = (f_0, f_1, \ldots, f_{k-1}) \in \mathbb{F}_{q^m}^k$. Let $\Omega = \{a_0, a_1, \ldots, a_{k-1}\} \subseteq \mathbb{F}_{q^m}$ be a P-independent set and $\Phi = \{p_0, p_1, \ldots, p_{k-1}\} \subseteq \mathbb{F}_{q^m}$ the set of evaluations of f such that $f(a_i) = p_i$ for all $i \in [0, k-1]$. Denote $\mathbf{p} = (p_0, p_1, \ldots, p_{k-1}) \in \mathbb{F}_{q^m}^k$. Let $\mathcal{L} = \{\ell_0, \ell_1, \ldots, \ell_{k-1}\}$ be a Lagrange basis on Ω as defined in Definition 2. A skew polynomial f can then be written as $f = p_0 \ell_0 + p_1 \ell_1 + \ldots + p_{k-1} \ell_{k-1}$, i.e., it has two representations $f = \mathbf{f} \cdot \mathbf{m}(x) = \mathbf{p} \cdot \boldsymbol{\ell}(x)$, where $\mathbf{m}(x) = (1, x, \ldots, x^{k-1})^\mathsf{T}$ and $\boldsymbol{\ell}(x) = (\ell_0, \ell_1, \ldots, \ell_{k-1})^\mathsf{T}$.

Lemma 3: The two representations from Definition 5 of a skew polynomial f have the $k \times k$ skew Vandermonde matrix $\mathbf{V}_k^\sigma(\mathbf{a}) \in \mathbb{F}_{q^m}^{k \times k}$ with $\mathbf{a} = (a_0, a_1, \ldots, a_{k-1}) \in \mathbb{F}_{q^m}^k$ as the transformation matrix. It holds that $\mathbf{f} \mathbf{V}_k^\sigma(\mathbf{a}) = \mathbf{p}$, and $\mathbf{m}(x) = \mathbf{V}_k^\sigma(\mathbf{a}) \boldsymbol{\ell}(x)$.

The proof is by definition of the two basis and the skew Vandermonde matrix and can be found in [28] for conventional polynomials.

Proof: From [25, Prop. 2.9], we know that multiplying the outer code $\mathcal{C}_{\text{out}}$ from the right with a block diagonal matrix $\mathbf{A} \in \mathbb{F}_q^{n \times t}$ can also be realized by adjusting the elements of $\boldsymbol{\beta} \in \mathbb{F}_{q^m}^n$ such that $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} \cdot \mathbf{A} \in \mathbb{F}_{q^m}^t$. We take at most r global codeword symbols $c_i^{(j)}$ from each group. Therefore, the corresponding $\tilde{\beta}_i^{(j)}$ are also $\mathbb{F}_q$-linearly independent since the generator matrices $\mathbf{A} \in \mathbb{F}_q^{r \times (r+\delta-1)}$ are MDS. The sum of local polynomials is equivalent to the encoding polynomial if their evaluations at k points are the same. We can consider

$$f(\tilde{b}_i^{(j)}) \tilde{\beta}_i^{(j)} = c_i^{(j)} \quad \text{for all } (i,j) \in \Delta_{\text{gl}}$$

at k positions. For a fixed $(i,j) \in \Delta_{\text{gl}}$, we have

$$
\begin{aligned}
f(\tilde{b}_i^{(j)}) \tilde{\beta}_i^{(j)} &= \sum_{m \in \Delta_{\text{gl},1}} L_m(\tilde{b}_i^{(j)}) \tilde{\beta}_i^{(j)} \\
&= L_i(\tilde{b}_i^{(j)}) \tilde{\beta}_i^{(j)} + \sum_{\substack{m \in \Delta_{\text{gl},1} \\ m \neq i}} L_m(\tilde{b}_i^{(j)}) \tilde{\beta}_i^{(j)} \\
&\overset{(a)}{=} c_i^{(j)} + 0 = c_i^{(j)},
\end{aligned}
$$

where (a) holds by Definition 4. Since Δ_{gl} has cardinality k, the sum of local polynomials is equal to f at k P-independent points and thus (1) holds. $\blacksquare$

Proof: We know that

$$\mathsf{H}(\mathsf{K}, \mathsf{X}) = \mathsf{H}(\mathsf{X}) + \mathsf{H}(\mathsf{K} \mid \mathsf{X})$$

by the chain rule of entropy and thus it holds that

$$\mathsf{H}(\mathsf{X}) \leq \mathsf{H}(\mathsf{K}, \mathsf{X})$$

with equality if, and only if, X essentially determines K, i.e., $\mathsf{H}(\mathsf{K} \mid \mathsf{X}) = 0$.
Furthermore, it holds that $\mathsf{X} = f(\mathsf{K}) = \mathbf{M} \mathsf{K}^\mathsf{T}$ which yields

$$\mathsf{H}(\mathsf{K}, \mathsf{X}) = \mathsf{H}(\mathsf{K}, \mathbf{M} \mathsf{K}^\mathsf{T}) = \mathsf{H}(\mathsf{K}) \leq k.$$

As a result,

$$\mathsf{H}(\mathsf{X}) = \mathsf{H}(\mathbf{M} \mathsf{K}^\mathsf{T}) \leq \mathsf{H}(\mathsf{K}) \leq k$$

holds. For the second part, we know that $\mathsf{H}(\mathsf{K}) = k$ since the random vector K consists of uniformly and independent distributed random variables. The entropy of X is

$$\mathsf{H}(\mathsf{X}) = \mathsf{H}(\mathbf{M} \mathsf{K}^\mathsf{T}) \leq k.$$

The rank of the matrix $\mathbf{M}$ determines how many symbols of the random vector X are independent and this can be expressed by $\mathsf{H}(\mathsf{X}) = \text{rk}(\mathbf{M})$. $\blacksquare$

The following lemma is used to determine the secrecy dimensions with a direct and forwarded global repair.

Lemma 4: Let $f \in \mathbb{F}_{q^m}[x; \sigma]$ be a skew polynomial of degree $k - 1$ in monomial basis with coefficient vector $\mathbf{f} = (f_0, f_1, \ldots, f_{k-1}) \in \mathbb{F}_{q^m}^k$. Let $\Omega = \{a_0, a_1, \ldots, a_{n-1}\} \subseteq \mathbb{F}_{q^m}$ be a P-independent set. Split the sets Ω into two subsets $\Omega_k = \{a_i \mid i \in [0, k-1]\}$ and $\Omega_{n-k} = \{a_i \mid i \in [k, n-1]\}$. Let $\mathcal{L} = \{\ell_0, \ell_1, \ldots, \ell_{k-1}\}$ be a Lagrange basis on Ω_k. The matrix

$$\mathbf{M} = \begin{pmatrix} \ell_0^{\Omega_k}(a_k) & \ell_1^{\Omega_k}(a_k) & \cdots & \ell_{k-1}^{\Omega_k}(a_k) \\ \ell_0^{\Omega_k}(a_{k+1}) & \ell_1^{\Omega_k}(a_{k+1}) & \cdots & \ell_{k-1}^{\Omega_k}(a_{k+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \ell_0^{\Omega_k}(a_{n-1}) & \ell_1^{\Omega_k}(a_{n-1}) & \cdots & \ell_{k-1}^{\Omega_k}(a_{n-1}) \end{pmatrix}$$

can be written as $\mathbf{M} = \left(\mathbf{V}_k^\sigma(\mathbf{a}_k)^{-1} \mathbf{V}_k^\sigma(\mathbf{a}_d) \right)^\mathsf{T}$ with $\mathbf{a}_k = (a_0, \ldots, a_{k-1})$ and $\mathbf{a}_d = (a_k, \ldots, a_{n-1})$. Moreover, $\mathbf{M}$ has full rank, i.e., $\text{rk}(\mathbf{M}) = \min(k, d)$.

Proof: It can be shown that the matrix $\mathbf{M}$ has full rank by decomposing $\mathbf{M}^\mathsf{T}$, which has the same rank as $\mathbf{M}$, into several matrices that are proven to have full rank. It holds that

$$\mathbf{M}^\mathsf{T} = \mathbf{L} \mathbf{V}_k^\sigma(\mathbf{a}_d)$$

with entries in $\mathbf{L}$ being the coefficient vectors of the Lagrange skew polynomials $\ell_i^{\Omega_k}, i \in [0, k-1]$, i.e., $L_{i,j} = \ell_{i,j}^{\Omega_k}$, where $\ell_{i,j}^{\Omega_k}$ is the j-th coefficient of the i-th polynomial $\ell_i^{\Omega_k}$. By Lemma 3, we know that $\mathbf{L} = \mathbf{V}_k^\sigma(\mathbf{a}_k)^{-1}$ since it holds that $\boldsymbol{\ell}(x) = \mathbf{m}(x) \mathbf{L}$. Overall, we have $\mathbf{M}^\mathsf{T} = \mathbf{V}_k^\sigma(\mathbf{a}_k)^{-1} \mathbf{V}_k^\sigma(\mathbf{a}_d)$. Both matrices have full rank and it holds that $\text{rk}(\mathbf{M}) = \text{rk}(\mathbf{M}^\mathsf{T}) = \min(k, d)$. $\blacksquare$

Remark: The above lemma also implies that submatrices of $\mathbf{M}$ have full rank since they are also a product of two Vandermonde matrices.

Before we give a general proof, the proof idea is illustrated with an example. Consider the DSS as depicted in Fig. 5. The global repair set for $x_1^{(0)}$ is $\Delta_{\mathrm{gl}} = \{(1,1),(2,0),(2,1),(3,0),(3,1)\}$. The static observations of the eavesdropper $\mathbf{e}_{\mathrm{st}}$ before the global repair can be summarized in a matrix $\mathbf{M}_{\mathrm{st}}$, where the basis is the outer codeword at the global repair set without the column multipliers of the LRS code, i.e., $\mathbf{c}_{\Delta_{\mathrm{gl}}} := (\mathbf{c}_{\mathrm{out}} \odot \boldsymbol{\beta}^{-1})|_{\Delta_{\mathrm{gl}}} = (c_{\mathrm{out},i}^{(j)}(\beta_i^{(j)})^{-1})_{(i,j) \in \Delta_{\mathrm{gl}}}$. For the DSS in Fig. 5, we have

$$\mathbf{M}_{\mathrm{st}} = \begin{pmatrix} \ell_{1,1}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

with $b' = b_1^{(0)}$ and $\mathbf{e}_{\mathrm{st}} = \mathbf{M}_{\mathrm{st}} \mathbf{c}_{\Delta_{\mathrm{gl}}}$. The eavesdropper can also observe the downloaded symbols to recover $x_1^{(0)}$. They can be summarized by $\mathbf{e}_{\mathrm{dl}} = \mathbf{M}_{\mathrm{dl}} \mathbf{c}_{\Delta_{\mathrm{gl}}}$, where $\mathbf{M}_{\mathrm{dl}}$ is

$$\mathbf{M}_{\mathrm{dl}} = \begin{pmatrix} \ell_{1,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 & 0 & 0 \\ 0 & \ell_{2,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 \\ 0 & 0 & 0 & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \end{pmatrix}.$$

By Lemma 2, the knowledge of the eavesdropper can be determined by calculating the rank of the stacked matrix $\mathbf{M}$, which consists of $\mathbf{M}_{\mathrm{st}}$ and $\mathbf{M}_{\mathrm{dl}}$, i.e.,

$$\mathbf{M} = \begin{pmatrix} \ell_{1,1}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \ell_{1,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 & 0 & 0 \\ 0 & \ell_{2,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 \\ 0 & 0 & 0 & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \end{pmatrix}.$$

The first row of $\mathbf{M}$ is linearly dependent on the last three rows. Thus, $\mathrm{rk}(\mathbf{M}) = \mathrm{rk}(\mathbf{M}')$ with

$$\mathbf{M}' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \ell_{1,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 & 0 & 0 \\ 0 & \ell_{2,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 \\ 0 & 0 & 0 & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \end{pmatrix}.$$

With proper row operations on $\mathbf{M}'$, we have $\mathrm{rk}(\mathbf{M}') = \mathrm{rk}(\mathbf{M}'')$, where

$$\mathbf{M}'' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 \\ 0 & 0 & 0 & 0 & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \end{pmatrix}.$$

It can be readily seen that $\mathrm{rk}(\mathbf{M}'') = 2 + \mathrm{rk}(\mathbf{M}''')$, where

$$\mathbf{M}''' = \begin{pmatrix} \ell_{2,1}^{\Delta_{\mathrm{gl}}}(b') & 0 & 0 \\ 0 & \ell_{3,0}^{\Delta_{\mathrm{gl}}}(b') & \ell_{3,1}^{\Delta_{\mathrm{gl}}}(b') \end{pmatrix}.$$
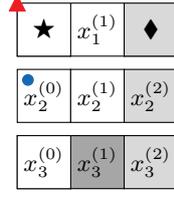

Fig. 5. Illustration of a DSS with $N = 9$ nodes and storing $k = 5$ independent symbols. The DSS is encoded by an MR-LRC with $g = 3$ groups, locality $r = 2$, local distance $\delta = 2$ (parities in light gray) and $h = 1$ global parities (in dark gray). The DSS is observed by a $(1,1)$-eavesdropper who can read the downloaded and stored data of any node in the top group (marked by a red triangle) and the data stored on one node (marked by a blue circle).

Overall, we have that $k_e = H(E) = \mathrm{rk}(\mathbf{M}) = 2 + \mathrm{rk}(\mathbf{M}''') = 2 + 2 = 4$. Thus, the secrecy dimension of the DSS shown in Fig. 5 is $k_s = k - k_e = 5 - 4 = 1$.

We now turn to the proof of Theorem 3.

Proof: Denote by $\mathcal{E}_1$ (w.r.t. $\mathcal{E}_2^{\mathrm{st}}$) the set of independent nodes that are observed by an eavesdropper in an l_1-(w.r.t. l_2-) manner in the static case without global repair, respectively. Without loss of generality, assume that the first l_2 groups are observed by the eavesdropper in the l_2-manner. Consider the worst case, there are h failed nodes that need to be globally repaired and they are, without loss of generality, in the first l_2 group at the first h positions with $j \in [0, h-1]$. We determine the entropy of the eavesdropped symbols $H(E)$ with Lemma 2. Assume that all possible local repairs are performed such that only erasures in the first group are left. By the definition of MR-LRCs (or PMDS codes), whenever a global repair is required, $|\Delta_{\mathrm{gl}}| = k$. In the following, we denote $\Delta_{\mathrm{gl}} = \{(i_1, j_1), \cdots, (i_k, j_k)\}$. Let the global repair set Δ_{gl} overlap as much as possible with the static eavesdropper observations $\mathcal{E}_1$ and $\mathcal{E}_2^{\mathrm{st}}$. The eavesdropped symbols $\mathbf{e}$ can be represented by $\mathbf{e} = \mathbf{M} \mathbf{c}_{\Delta_{\mathrm{gl}}}$, where $\mathbf{c}_{\Delta_{\mathrm{gl}}} := (\mathbf{c}_{\mathrm{out}} \odot \boldsymbol{\beta}^{-1})|_{\Delta_{\mathrm{gl}}}$ is the outer codeword at the global repair set without the column multipliers of the LRS code, and $\mathbf{M} = \begin{pmatrix} \mathbf{M}_{\mathrm{st}} \\ \mathbf{M}_{\mathrm{dl}} \end{pmatrix}$. The symbols in $\mathbf{e}_{\mathrm{st}} = \mathbf{M}_{\mathrm{st}} \mathbf{c}_{\Delta_{\mathrm{gl}}}$ are the observed stored symbols of which h are being repaired using the global repair set Δ_{gl}, and the matrix $\mathbf{M}_{\mathrm{st}} \in \mathbb{F}_{q^m}^{(l_2 r + l_1) \times k}$ can be expressed as

$$\mathbf{M}_{\mathrm{st}} = \begin{pmatrix} \ell_{i_1,j_1}^{\Delta_{\mathrm{gl}}}(b_1^{(0)}) & \cdots & \ell_{i_k,j_k}^{\Delta_{\mathrm{gl}}}(b_1^{(0)}) \\ \vdots & \ddots & \vdots \\ \ell_{i_1,j_1}^{\Delta_{\mathrm{gl}}}(b_1^{(h-1)}) & \cdots & \ell_{i_k,j_k}^{\Delta_{\mathrm{gl}}}(b_1^{(h-1)}) \\ & \widetilde{\mathbf{I}}_{\mathrm{st}} & \end{pmatrix}, \quad (4)$$

where the columns indexed by $(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}}) \cap \Delta_{\mathrm{gl}}$ of $\widetilde{\mathbf{I}}_{\mathrm{st}}$ form an identity matrix. The symbols observed by the eavesdropper during the global repair process are $\mathbf{e}_{\mathrm{dl}} = \mathbf{M}_{\mathrm{dl}} \mathbf{c}_{\Delta_{\mathrm{gl}}}$, where

$$\mathbf{M}_{\mathrm{dl}} = \begin{pmatrix} \ell_{i_1,j_1}^{\Delta_{\mathrm{gl}}}(b_1^{(0)}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & \ell_{i_k,j_k}^{\Delta_{\mathrm{gl}}}(b_1^{(0)}) \\ & \vdots & \\ \ell_{i_1,j_1}^{\Delta_{\mathrm{gl}}}(b_1^{(h-1)}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & \ell_{i_k,j_k}^{\Delta_{\mathrm{gl}}}(b_1^{(h-1)}) \end{pmatrix} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\}$$

It can be seen that the first h rows of $\mathbf{M}_{\mathrm{st}}$ are linearly dependent on $\mathbf{M}_{\mathrm{dl}}$. Hence, $\mathrm{rk}(\mathbf{M}) = \mathrm{rk}(\mathbf{M}')$, where $\mathbf{M}' = \begin{pmatrix} \widetilde{\mathbf{I}}_{\mathrm{st}} \\ \mathbf{M}_{\mathrm{dl}} \end{pmatrix}$ and $\mathrm{rk}(\mathbf{M}') = \mathrm{rk}(\widetilde{\mathbf{I}}_{\mathrm{st}}) + \mathrm{rk}(\mathbf{M}_{\mathrm{dl}}|_{\Delta_{\mathrm{gl}}\backslash(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}})})$.

Consider the matrix $\mathbf{M}_{\mathrm{dl}}|_{\Delta_{\mathrm{gl}}\backslash(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}})}$ groupwise for the i-th group. If the i-th group is fully punctured by the entries of $\widetilde{\mathbf{I}}_{\mathrm{st}}$, we have $e_i = r$ and there is no nonzero row in $\mathbf{M}_{\mathrm{dl}}|_{\Delta_{\mathrm{gl}}\backslash(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}})}$ corresponding to the i-th group. Otherwise, there are still $r - e_i$ columns corresponding to the i-th group. The rows corresponding to the i-th group are of the structure investigated in Lemma 4. They can be represented by the product of two Vandermonde matrices since they correspond to evaluations of the same polynomial at P-independent points. Therefore, the contribution of the i-th group is $\min(h, r - e_i)$. It holds that

$$\mathrm{rk}(\mathbf{M}_{\mathrm{dl}}|_{\Delta_{\mathrm{gl}}\backslash(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}})}) = \sum_{i=1}^{g} \min(h, r - e_i)$$

and we have

$$\mathrm{rk}(\mathbf{M}) = l_2 r + l_1 - h + \sum_{i=1}^{g} \min(h, r - e_i),$$

which gives us $\mathrm{H}(\mathsf{E})$ by Lemma 2. ∎

PROOF OF THEOREM 4

Proof: The proof is done in a similar manner as for Theorem 3. We follow the notations for $\mathcal{E}_1$, $\mathcal{E}_2^{\mathrm{st}}$ and Δ_{gl} as in the proof of Theorem 3. Let the l_1-observations of the eavesdropper be distributed in such a way that $|\mathcal{E}_1 \cap \Delta_{\mathrm{gl}}| = l_1$. Without loss of generality, assume that the global erasures occur in the first group. Consider the worst case that the first group is not observed by the eavesdropper in the l_2-manner. The eavesdropped symbols can be represented by $\mathbf{e} = \mathbf{M}\mathbf{c}_{\Delta_{\mathrm{gl}}}$ where $\mathbf{c}_{\Delta_{\mathrm{gl}}} := (\mathbf{c}_{\mathrm{out}} \odot \boldsymbol{\beta}^{-1})|_{\Delta_{\mathrm{gl}}}$ and $\mathbf{M} = \begin{pmatrix} \mathbf{M}_{\mathrm{st}} \\ \mathbf{M}_{\mathrm{dl}} \end{pmatrix} \in \mathbb{F}_{q^m}^{(l_2 r + l_1 + l_2 h) \times k}$. The matrix $\mathbf{M}_{\mathrm{st}} \in \mathbb{F}_{q^m}^{(l_2 r + l_1) \times k}$ represents the eavesdropper's observation in the static case, i.e., before the global repair, and $\mathbf{M}_{\mathrm{st}} = \widetilde{\mathbf{I}}_{\mathrm{st}}$, where the columns indexed by $(\mathcal{E}_1 \cup \mathcal{E}_2^{\mathrm{st}}) \cap \Delta_{\mathrm{gl}}$ of $\widetilde{\mathbf{I}}_{\mathrm{st}}$ form an identity matrix. Hence, it contributes $l_2 r + l_1$ to the rank of $\mathbf{M}$. The other part of $\mathbf{M}$,

namely $\mathbf{M}_{\mathrm{dl}} \in \mathbb{F}_{q^m}^{l_2 h \times k}$ summarizes the global repair symbols that are observed by the eavesdropper,

$$\mathbf{M}_{\mathrm{dl}} = \begin{pmatrix} \sum_{\nu \in \mathcal{F}_{\mathrm{up}, \mathcal{G}_{l_2}(1)}} \boldsymbol{\ell}_\nu^{\Delta_{\mathrm{gl}}}(b_1^{(0)}) \\ \vdots \\ \sum_{\nu \in \mathcal{F}_{\mathrm{up}, \mathcal{G}_{l_2}(l_2)}} \boldsymbol{\ell}_\nu^{\Delta_{\mathrm{gl}}}(b_1^{(h-1)}) \end{pmatrix},$$

where $\boldsymbol{\ell}_\nu^{\Delta_{\mathrm{gl}}}(b) = (\ell_{\nu;i_1,j_1}^{\Delta_{\mathrm{gl}}}(b), \ldots, \ell_{\nu;i_k,j_k}^{\Delta_{\mathrm{gl}}}(b))$ with $\ell_{\nu;i,j}^{\Delta_{\mathrm{gl}}}(b) = \ell_{i,j}^{\Delta_{\mathrm{gl}}}(b)$ for $\nu = i$ and $\ell_{\nu;i,j}^{\Delta_{\mathrm{gl}}}(b) = 0$ otherwise, for all $(i,j) \in \Delta_{\mathrm{gl}}$. For each repair and each group observed in an l_2-manner, one symbol, as the sum of all upstream symbols, is observed. Therefore, there are $l_2 h$ rows in $\mathbf{M}_{\mathrm{dl}}$. The matrix $\mathbf{M}_{\mathrm{dl}}$ can be reduced by Gaussian elimination to a matrix $\mathbf{M}'_{\mathrm{dl}}$ with $k - (l_2 r + l_1)$ nonzero columns by subtracting the rows of $\mathbf{M}_{\mathrm{st}}$. Moreover, the matrix $\mathbf{M}'_{\mathrm{dl}}$ can be further reduced to matrix $\mathbf{M}''_{\mathrm{dl}}$ such that only the columns corresponding to groups in $\mathcal{F}'_{\mathrm{up},i}$ are nonzero. The submatrices of $\mathbf{M}''$ for each l_2-manner observed group (only consider the nonzero columns) have a structure as described in Lemma 4 and they are of size $h \times (\sum_{j \in \mathcal{F}'_{\mathrm{up},i}} (r - e_j))$ with full rank. Thus, the rank of matrix $\mathbf{M}$ is

$$\mathrm{rk}(\mathbf{M}) = \mathrm{rk}\begin{pmatrix} \widetilde{\mathbf{I}}_{\mathrm{st}} \\ \mathbf{M}''_{\mathrm{dl}} \end{pmatrix}$$

$$= \left((l_2 r + l_1) + \sum_{i \in \mathcal{G}_{l_2}} \min(h, \sum_{j \in \mathcal{F}'_{\mathrm{up},i}} (r - e_j)) \right),$$

which gives us $\mathrm{H}(\mathsf{E})$ by Lemma 2. ∎