PAPER

Secure Regenerating Codes Based on Rashmi-Shah-Kumar MBR Codes*

Masazumi KURIHARA[†a)] ***and*** **Hidenori KUWAKADO**[††b)]**,** ***Members***

SUMMARY In this paper, we present a construction of (n, k, d, m) secure regenerating codes for distributed storage systems against eavesdroppers that can observe either data stored in at most m storage nodes or downloaded data for repairing at most m failed nodes in a network where $m < k \le d \le n-1$. The (n, k, d, m) secure regenerating code is based on an (n, k, d) minimum bandwidth regenerating (MBR) code, which was proposed by Rashmi, Shah and Kumar as optimal exact-regenerating codes, for all values of the parameters (n, k, d). The (n, k, d, m) secure regenerating codes have the security as a secret sharing scheme such that even if an eavesdropper knows either data stored in at most m storage nodes or downloaded data for repairing at most m failed nodes, no information about data leaks to the eavesdropper.

key words: *secret sharing, secure regenerating code, distributed storage*

1. Introduction

Distributed storage systems provide reliable access to data by storing the data redundantly in a collection of individually unreliable storage nodes in a network such that the data can be reconstructed from active nodes even if a small set of nodes fails. For instance, such a reliable distributed storage system can be constructed by using maximum distance separable (MDS) codes such as Reed-Solomon (RS) codes. This is optimal in the tradeoff between redundancy and reliability. With an (n, k) RS code, a data file is encoded to n shares in such a way that the data file can be reconstructed from any k shares, and the n shares are stored across n storage nodes in the network. On the other hand, it is desirable to regenerate (i.e., repair) a failed node in order to maintain such a reliable distributed storage system.

Dimakis et al. [1] proposed the concept of a *regenerating code* that has not only the property of reconstructing the data file but also that of repairing a failed node (i.e., regenerating the data as was stored in the failed node). Under the concept of regenerating codes, a data-collector is permitted to connect to any k active nodes to reconstruct the data file, and a failed node is permitted to connect to any d active nodes to repair itself. The authors showed that regenerating codes reduce the repair-bandwidth when $d \ge k$. Further, they showed that there exists a fundamental tradeoff between the storage capacity of each node and the repair-bandwidth. In Sect. 2, the definition of regenerating codes will be described in detail. In the research results for regenerating codes by many researchers, Rashmi et al. [3] presented the general construction of exact-regenerating codes such as (a) minimum-bandwidth-regenerating (MBR) codes for all values of (n, k, d) and (b) minimum-storage-regenerating (MSR) codes for all values of $(n, k, d \ge 2k-1)$.

Regenerating codes may be similar to secret sharing schemes. The secret sharing scheme produces shares in such a way that a share does not give any information about the data. However, in general, the secret sharing scheme does not have the property of regenerating the share as was stored in a failed node. On the other hand, in the concept of a regenerating code proposed by Dimakis et al. [1], the regenerating code does not have the property of the security.

Combining a regenerating code with a secret sharing scheme has been proposed in [2], [4]–[7], [9]. Pawar et al. [2] proposed the secrecy capacity for secure distributed storage systems against passive eavesdroppers and secure regenerating codes based on MBR codes, where the codes are confined to the case $n = d + 1$. In [4], [5], [7], [9], secure regenerating codes based on MSR codes were proposed. In this paper, we propose a construction of (n, k, d, m) secure regenerating codes based on (n, k, d) MBR codes for any values of (n, k, d). The proposed code is an extended version of secure regenerating codes given in [6]. Previous secure regenerating codes [6] are confined to the case $n = d + 1$. We show that by using the proposed secure regenerating codes an eavesdropper can not obtain information about the data not only from shares stored in storage nodes but also from downloaded data for repair. The proposed secure regenerating code includes all values of parameters of secure regenerating codes given in [2, Sect. V] and [6].

Shah et al. [9] have recently presented the secure regenerating code based on an MBR code [3]. In a construction of a secure regenerating code, their code and our code are based on the same MBR code. However, their code is different from our code. The difference between their code and our code will be described in Sect. 4 in detail.

This paper is organized as follows: In Sect. 2, we describe the storage system and the regenerating codes. And then, we define the secrecy capacity for the system and show its upper bound. In Sect. 3, we describe (n, k, d) MBR codes proposed by Rashmi et al. [3]. In Sect. 4, we propose a con-

Manuscript received May 15, 2012.
Manuscript revised October 19, 2012.

[†]The author is with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu-shi, 182-8585 Japan.
[††]The author is with the Graduate School of Engineering, Kobe University, Kobe-shi, 657-8501 Japan.

*The extended abstract of this paper was presented in ISITA2012[11].

a) E-mail: kurihara@uec.ac.jp
b) E-mail: kuwakado@kobe-u.ac.jp

DOI: 10.1587/transfun.E96.A.635

struction of (n,k,d,m) secure regenerating code based on the (n,k,d) MBR code. In Sect. 5, we evaluate the security of the proposed (n,k,d,m) secure regenerating code and present our results as main theorems. In the result, we show that the proposed secure regenerating codes have the security of a ramp secret sharing scheme, and achieve the upper bound of the secrecy capacity. In Sect. 6, we prove the main theorem. Finally, a conclusion is given in Sect. 7.

2. Regenerating Codes and Secrecy

In this section we describe regenerating codes associated with parameters (n,k,d,α,β,B) and secret capacity for secure regenerating codes.

Assume that there are n storage nodes such as node 1, node 2, ..., node n, in a network.

2.1 (n,k,d,α,β,B) Regenerating Codes

In a distributed storage system, a data file (a *message*) is encoded to *n shares* so that the data file is reconstructed from any subset of k shares, and n shares are stored across n storage nodes in the network. Let B be the size of the data file and let α be that of a share per node where the data size is measured in terms of symbols over a finite field $\mathbb{F}_q$ with q elements. The share size α is equal to the storage capacity of a node. In this system, an end-user, which is called a *data-collector*, is permitted to connect to any k active storage nodes and download a share from each node to reconstruct the data file. This process is termed as the reconstruction property.

When a storage node fails, the failed node loses the share of itself. Under the concept of regenerating codes [1], the failed node is permitted to connect to any d active nodes, where $d \geq k$, and download data for repair purposes from each node in order to regenerate a share. In this paper, we consider *exact-regeneration* described in detail in [3, Sect. I-D] as a regeneration. The exact-regeneration means that the fail node regenerates exactly the same share as was stored in itself prior to failure. The d active nodes aiding in the repair are termed as *helper-nodes*. The downloaded data from each helper-node is called a *piece* for the failed node, and let β be the size of a piece. A vector consisting of d pieces downloaded from d helper-nodes is called a *piece-vector* for the failed node, and the size $d\beta$ of a piece-vector is called a *repair-bandwidth*. As a result, the failed node can regenerate the same share of itself from the piece-vector. This process is termed as the regeneration property.

From the above description about regenerating codes, a regenerating code is characterized by the parameters (n,k,d,α,β,B) where $k \leq d \leq n-1$, and such a code is written as an (n,k,d,α,β,B) regenerating code. It is desirable to minimize both of α and β for fixed k and d. It is not, however, possible to minimize both of α and β simultaneously since there is a tradeoff between choices of the parameters α and β [1]. The regenerating code with parameters (α,β) obtained by first minimizing α and then minimizing β is called a minimum storage regenerating (MSR) code. The parameters of an MSR code satisfy the following equations:

$$\alpha = \frac{B}{k} \text{ and } \beta = \frac{\alpha}{(d-k+1)}. \tag{1}$$

Reversing the order, the regenerating code with parameters (α,β) obtained by first minimizing β and then minimizing α is called a minimum bandwidth regenerating (MBR) code. The parameters of an MBR code satisfy the following equations:

$$\beta = \frac{2B}{k(2d-k+1)} \text{ and } \alpha = d\beta \tag{2}$$

For fixed d, an MBR code minimizes the repair-bandwidth $d\beta$ to repair a failed node.

2.2 Secrecy

Let S be a random variable with the uniform distribution over $\mathbb{F}_q^{L_S}$, representing a secret $\underline{S} \in \mathbb{F}_q^{L_S}$ where $L_S \leq B$. Let $H(S)$ denote the entropy of the random variable S. The base of the logarithm of the entropy is q, and then the entropy $H(S)$ of S is equal to L_S. A secret $\underline{S}$ is encoded to *n shares* $\underline{c}_1,\ldots,\underline{c}_n \in \mathbb{F}_q^{\alpha}$ so that the secret $\underline{S}$ can be reconstructed from any subset of k shares $\underline{c}_{i_1},\ldots,\underline{c}_{i_k}$. For each $i \in \{1,\ldots,n\}$, a share $\underline{c}_i$ is stored in node i in the network. Let C_i be a random variable representing a share $\underline{c}_i$. The reconstruction property then can be written as follows: for k random variables $C_{i_1}\cdots C_{i_k}$ representing any k shares $\underline{c}_{i_1},\ldots,\underline{c}_{i_k}$,

$$H(S|C_{i_1},\ldots,C_{i_k}) = 0. \tag{3}$$

Assume that node f fails where $f \in \{1,\ldots,n\}$. The failed node f connects any d helper-nodes $h_1,\ldots,h_d$ of the remaining $n-1$ nodes in the network to regenerate the share $\underline{c}_f$ of itself. Each helper-node h_p, $p \in \{1,\ldots,d\}$, computes a *piece* $\underline{d}_{f,h_p} \in \mathbb{F}_q^{\beta}$ for the failed node from the share $\underline{c}_{h_p}$ of itself, and send it to the failed node. As a result, the failed node obtains a *piece-vector* $\underline{d}_f = [\underline{d}_{f,h_1}^t,\ldots,\underline{d}_{f,h_d}^t]^t \in \mathbb{F}_q^{d\beta}$, which consists of d pieces, and can compute the same share $\underline{c}_f$ from it. Let D_{f,h_p} and D_f be random variables representing a piece $\underline{d}_{f,h_p}$ and a piece-vector $\underline{d}_f$, respectively. The regeneration property can be written as

$$H(C_f|D_f) = 0. \tag{4}$$

Thus, we have $H(S|C_f) \geq H(S|D_f)$ from Eq. (4) because C_f which is dependent on S is uniquely determined from D_f, that is, the Markov chain $S \to D_f \to C_f$ holds. Note that it is not always true that $H(D_f|C_f) = 0$.

We consider two secrecy conditions for shares and piece-vectors, respectively. Assume that there are at least $\max\{l,m\}$ repaired nodes in the network where l and m are nonnegative integers that are strictly less than k. First, the secrecy condition for m random variables $C_{i_1},\ldots,C_{i_m}$ representing any m shares $\underline{c}_{i_1},\ldots,\underline{c}_{i_m}$ is defined as

$$H(S|C_{i_1},\ldots,C_{i_m}) = H(S),\ {}^{\forall}i_1,\ldots,i_m \in \{1,\ldots,n\}. \tag{5}$$

The parameter m will be used in a construction of an (n,k,d,m) secure regenerating code in Sect. 4. Next, the secrecy condition for l random variables $D_{i_1},\ldots,D_{i_l}$ representing any l piece-vectors $\underline{d}_{i_1},\ldots,\underline{d}_{i_l}$ is defined as

$$H(S|D_{i_1},\ldots,D_{i_l}) = H(S),\ {}^{\forall}i_1,\ldots,i_l \in \{1,\ldots,n\}. \quad (6)$$

We notice that if $l \geq m$, then from the regenerating property by Eq. (4),

$$\begin{aligned} H(S) &\geq H(S|C_{i_1},\ldots,C_{i_l}) && (7)\\ &\geq H(S|D_{i_1},\ldots,D_{i_l}) = H(S). && (8)\end{aligned}$$

Thus we have

$$H(S|C_{i_1},\ldots,C_{i_l}) = H(S) \quad (9)$$

if $H(S|D_{i_1},\ldots,D_{i_l}) = H(S)$ and $l \geq m$. In this paper, from the above definitions of the two secrecy conditions, we assume that an eavesdropper can observe either at most l piece-vectors or at most m shares in the network.

We propose the following definition, which refines that of the secrecy capacity given by Pawar et al. [2, Sect. V-A], of the secrecy capacity. Given a repairable and secure distributed storage system with a collection of the parameters (n,k,d,α,β) concerned with the reconstruction and the regeneration and the parameters (l,m) concerned with the secrecy, its *secrecy capacity*, denoted by $C_S = C_S(n,k,d,\alpha,\beta;l,m)$, is defined to be the maximum amount of data that can be stored in this system such that the reconstruction property and the two secrecy conditions are simultaneously satisfied for all possible data-collectors and eavesdroppers, that is,

$$\begin{aligned} C_S &= C_S(n,k,d,\alpha,\beta;l,m)\\ &:= \sup H(S). \quad (10)\\ &\quad H(S|C_{i_1},\ldots,C_{i_k}) = 0,\ {}^{\forall}i_1,\ldots,i_k \in \{1,\ldots,n\}\\ &\quad H(S|D_{f_1},\ldots,D_{f_l}) = H(S),\ {}^{\forall}f_1,\ldots,f_l \in \{1,\ldots,n\}\\ &\quad H(S|C_{j_1},\ldots,C_{j_m}) = H(S),\ {}^{\forall}j_1,\ldots,j_m \in \{1,\ldots,n\}\end{aligned}$$

The above definition of the secrecy capacity is similar to that of the secrecy capacity given by Pawar et al. [2, Sect. V-A, Eq. (6)]. The secrecy capacity of Eq. (10) must satisfy the two secrecy conditions for piece-vectors and shares. On the other hand, the secrecy capacity given by Pawar et al. satisfies the only secrecy condition for piece-vectors. In the definition of the secrecy capacity by Eq. (10), when $l \geq m$, the definition of the secrecy capacity is identical to that given by Pawar et al. from the result of Eq. (9).

Under the passive eavesdropper model given by Pawar et al. [2, Sect. V], an eavesdropper can observe only at most l piece-vectors in the network. On the other hand, we assume that an eavesdropper can observe either at most l piece-vectors or at most m shares in the network. That is, the secrecy capacity given by Pawar et al. is subject only to the two conditions of Eqs. (3) and (6). On the other hand, the secrecy capacity of Eq. (10) is subject to the three conditions of Eqs. (3), (5) and (6).

Theorem 1 (Upper Bound): For a repairable and secure distributed storage system with a collection of parameters (n,k,d,α,β) and (l,m), the secrecy capacity is upper bounded as

$$C_S(n,k,d,\alpha,\beta;l,m) \leq \sum_{j=\max\{l,m\}+1}^{k} \min\{(d-j+1)\beta,\alpha\}. \quad (11)$$

Proof: See Appendix A. □

For an MBR code, the repair-bandwidth $d\beta$ is equal to the storage size α, that is, $d\beta = \alpha$. Thus, if a function which determines C_f from D_f for repair is a bijection, then the piece-vector D_f is also determined from the share C_f, that is, $H(D_f|C_f) = 0$. Therefore, it is true that $H(S|C_{i_1},\ldots,C_{i_m}) = H(S)$ implies $H(S|D_{i_1},\ldots,D_{i_m}) = H(S)$, and the converse is also true because of the regeneration property. Hence, for an MBR code, we can assume that $l = m$ without loss of generality for the parameters m and l in Eqs. (5) and (6), respectively. Because the case of $l = m$ is included in the case of $l \geq m$, for an MBR code, the secrecy capacity of Eq. (10) is identical to that given by Pawar et al. [2, Eq. (6)]. Moreover, the upper bound of Eq. (11) is also identical to that given by Pawar et al. [2, Eq. (7)].

For the MBR code [3] used in this paper as an underlying code, the function which uniquely determines C_f from D_f for repair is bijective. In this paper, hence, we can assume that $l = m$ without loss of generality for the parameters m and l in Eqs. (5) and (6), respectively. When $l = m$, the upper bound of Eq. (11) is simplified to

$$C_S(n,k,d,\alpha,\beta;m,m) \leq \sum_{j=m+1}^{k} (d-j+1)\beta, \quad (12)$$

because of $d\beta = \alpha$.

Remark 2: For an MSR code, in general, the repair-bandwidth $d\beta$ is greater than the storage size α when $k \geq 2$, that is, $d\beta > \alpha$. In such a situation, the secrecy capacity of Eq. (10) is useful to estimate the secrecy ability of secure MSR codes. The pair of the secrecy capacity of Eq. (10) and its upper bound of Eq. (11) are useful to estimate the secrecy ability of secure MSR codes [10]. However, secure MSR codes are outside of this paper, that is, this paper only discusses secure MBR codes.

3. Rashmi-Shah-Kumar (n,k,d) MBR Codes

Rashmi, Shah and Kumar [3] proposed the first constructions of general and optimal exact-regenerating codes such as (a) (n,k,d) MBR codes for all values of (n,k,d), and (b) (n,k,d) MSR codes for all values of (n,k,d) where $d \geq 2k-2$. In general, from the definition of regenerating codes in Sect. 2.1, all n,k,d satisfy $k \leq d \leq n-1$.

In this section, an (n,k,d) MBR code over $\mathbb{F}_q$ is introduced because an (n,k,d,m) secure regenerating code based on the (n,k,d) MBR code will be proposed in the next section.

3.1 Message Matrix M

The parameters $(n, k, d, B, \alpha, \beta)$ of an (n, k, d)MBR code over $\mathbb{F}_q$ satisfy the following relations [3, Sect. IV]:

$$\alpha = d,\ \beta = 1,\ B = \frac{1}{2}k(2d - k + 1), \tag{13}$$

moreover, we assume that $n < q$ and the data file is composed of B message symbols in $\mathbb{F}_q$. Note that the parameters α and B are uniquely determined from the parameters d and k, and the parameter β is a fixed value as $\beta = 1$.

Let M_1 be a $k \times k$ symmetric matrix constructed so that the $k(k+1)/2$ components in the upper-triangular half of the matrix are filled up by $k(k+1)/2$ distinct message symbols drawn from the set of the B message symbols of the data file. The remaining $k(d-k)$ message symbols are used to fill up a second $k \times (d-k)$ matrix M_2. Let $\mathbf{O}$ denote the $(d-k)\times(d-k)$ zero matrix with all zero components. A *message matrix* M is then defined as the $d \times d$ symmetric matrix given by

$$M = \begin{bmatrix} M_1 & M_2 \\ M_2^t & \mathbf{O} \end{bmatrix} \tag{14}$$

$$= \begin{bmatrix} u_{1,1} & \cdots & u_{1,k} & u_{1,k+1} & \cdots & u_{1,d} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u_{k,1} & \cdots & u_{k,k} & u_{k,k+1} & \cdots & u_{k,d} \\ u_{k+1,1} & \cdots & u_{k+1,k} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u_{d,1} & \cdots & u_{d,k} & 0 & \cdots & 0 \end{bmatrix}, \tag{15}$$

where M_2^t is the transpose of the matrix M_2. From the definition of the message matrix M with components $u_{i,j}$, note that $u_{i,j} = u_{j,i}$ for all $i, j \in \{1, \ldots, d\}$, and $u_{i,j} = 0$ for all $i, j \in \{k+1, \ldots, d\}$.

3.2 Encoding, Share and Reconstruction

For each $i \in \{1, \ldots, n\}$, assign a unique and public symbol x_i in $\mathbb{F}_q$ to node i in such a way that the following two conditions are satisfied.

Condition for x_i:

1. For any $i \in \{1, \ldots, n\}$, $x_i \neq 0$.
2. For any $i, j \in \{1, \ldots, n\}$, $x_i \neq x_j$ if $i \neq j$.

For the message matrix M, the *share* $\underline{c}_i$ stored in node i is then defined as

$$\underline{c}_i = [c_{i,1}, c_{i,2}, \ldots, c_{i,d}]^t := M\underline{\phi}_i \in \mathbb{F}_q^d \tag{16}$$

where $\underline{\phi}_i = \left[1, x_i, x_i^2, \ldots, x_i^{d-1}\right]^t \in \mathbb{F}_q^d$ is a *coding vector* associated with node i and all operations are done over $\mathbb{F}_q$. For each $i \in \{1, \ldots, n\}$, the coding vector $\underline{\phi}_i$ associated with node i is also assigned uniquely and publicly. Note that the size of a share is d from Eq. (13) and Eq. (16). Thus, the message matrix M with B distinct message symbols is encoded to n shares $\underline{c}_1, \ldots, \underline{c}_n$.

Note that all the B message symbols can be reconstructed from any k shares by using the reconstruction method proposed by Rashmi et al. [3, Theorem 3].

3.3 Regeneration, Piece and Piece-Vector

In this section, we describe the regeneration method proposed by Rashmi et al. [3, Theorem 2]. Suppose that a node f fails and helper-nodes $h_1, \ldots, h_d$ are active. For each index $p \in \{1, \ldots, d\}$, the helper-node h_p computes a *piece* $\underline{d}_{f,h_p}$ for the failed node from the share $\underline{c}_{h_p}$ of itself and the coding vector $\underline{\phi}_f$ of the failed node as follows:

$$\underline{d}_{f,h_p} := \underline{c}_{h_p}^t \underline{\phi}_f \in \mathbb{F}_q,\ p = 1, \ldots, d, \tag{17}$$

and sends it to the failed node. Note that $\underline{d}_{f,h_p} = \underline{d}_{f,h_p}^t$ since a piece $\underline{d}_{f,h_p}$ is a scalar in $\mathbb{F}_q$. As a result, the failed node obtains the *piece-vector* $\underline{d}_f$ consisting of d pieces as follows:

$$\underline{d}_f := \left[\underline{d}_{f,h_1}, \underline{d}_{f,h_2}, \ldots, \underline{d}_{f,h_d}\right]^t \in \mathbb{F}_q^d. \tag{18}$$

Note that the sizes of a piece and a piece-vector are one and d, respectively, from Eqs. (13), (17) and (18). Using the piece-vector $\underline{d}_f$ and the d coding vectors $\underline{\phi}_{h_1}, \ldots, \underline{\phi}_{h_d}$ associated with the helper-nodes $h_1, \ldots, h_d$, the failed node can compute the share $\underline{c}_f$ as follows:

$$\underline{c}_f = \begin{bmatrix} \underline{\phi}_{h_1}^t \\ \vdots \\ \underline{\phi}_{h_d}^t \end{bmatrix}^{-1} \underline{d}_f. \tag{19}$$

Note that the $d \times d$ matrix $[\underline{\phi}_{h_1}, \ldots, \underline{\phi}_{h_d}]^t$ is nonsingular since the determinant of the matrix is the Vandermonde determinant from the conditions for x_i. On the other hand, from the relation between $\underline{c}_f$ and $\underline{d}_f$ in Eq. (19), $\underline{d}_f$ is also uniquely determined from $\underline{c}_f$ by using the matrix $[\underline{\phi}_{h_1}, \ldots, \underline{\phi}_{h_d}]^t$, that is, $H(D_f|C_f) = 0$. Thus, we have $H(S|C_f) \leq H(S|D_f)$, and then the following lemma holds.

Lemma 3: For random variables D_f and C_f representing a piece-vector $\underline{d}_f$ and a share $\underline{c}_f$ for a failed node f respectively, $H(S|D_f) = H(S|C_f)$.

Proof: It holds that $H(S|C_f) \geq H(S|D_f)$ from the regeneration property, i.e., $H(C_f|D_f) = 0$. On the other hand, since $\underline{d}_f$ is also uniquely determined from $\underline{c}_f$ by Eq. (19), i.e., $H(D_f|C_f) = 0$, it holds that $H(S|C_f) \leq H(S|D_f)$. This completes the proof of the lemma. □

From Lemma 3, we have the following corollary.

Corollary 4: For any m such that $m \leq n$ and any $i_1, \ldots, i_m \in \{1, \ldots, n\}$, $H(S|C_{i_1}, \ldots, C_{i_m}) = H(S|D_{i_1}, \ldots, D_{i_m})$.

Proof: From the relation between a piece-vector and a share in Eq. (19), for $i_1, \ldots, i_m \in \{1, \ldots, n\}$, it holds that $H(D_{i_1}, \ldots, D_{i_m}|C_{i_1}, \ldots, C_{i_m}) = 0$ and $H(C_{i_1}, \ldots, C_{i_m}|D_{i_1}, \ldots,$

D_{i_m})= 0. Therefore, this completes the proof of the corollary. □

Finally, we show the following lemma about a relation between shares and coding vectors.

Lemma 5: For any $i, j \in \{1, \ldots, n\}$, $\underline{c}_i^t \underline{\phi}_j = \underline{c}_j^t \underline{\phi}_i$, where $\underline{c}_i$ and $\underline{c}_j$ are the shares of nodes i and j, respectively. Similarly, $\underline{\phi}_i$ and $\underline{\phi}_j$ are the coding vectors associated with nodes i and j, respectively.

Proof: Since the size of a piece $\underline{c}_i^t \underline{\phi}_j$ is one symbol in $\mathbb{F}_q$, that is, $\underline{c}_i^t \underline{\phi}_j$ is a scalar, $(\underline{c}_i^t \underline{\phi}_j)^t = \underline{c}_i^t \underline{\phi}_j$. Moreover, since $(\underline{c}_i^t \underline{\phi}_j)^t = \underline{c}_i^t \underline{\phi}_j$ and $M = M^t$, we have $\underline{c}_i^t \underline{\phi}_j = (\underline{c}_i^t \underline{\phi}_j)^t = \underline{\phi}_j^t \underline{c}_i = \underline{\phi}_j^t M \underline{\phi}_i = \underline{\phi}_j^t M^t \underline{\phi}_i = (M \underline{\phi}_j)^t \underline{\phi}_i = \underline{c}_j^t \underline{\phi}_i$. This completes the proof of the lemma. □

4. (n, k, d, m) Secure Regenerating Codes

In this section, we propose an (n, k, d, m) secure regenerating code based on the (n, k, d) MBR code. The new parameter m ($0 \le m \le k$) is a secrecy parameter. The (n, k, d, m) secure regenerating code satisfies Eq. (2) because it can be considered as the (n, k, d) MBR code. Hence, we can assume that $l = m$ in Eqs. (5) and (6) without loss of generality. Furthermore, the (n, k, d, m) secure regenerating code satisfies Eqs. (5) and (6) where $l = m$, that is, the parameter m means the perfect secrecy condition such that no information about secret is obtained from either at most m shares or at most m piece-vectors.

4.1 Construction

We define two kinds of vectors $\underline{r}_j$, $1 \le j \le d$, and $\underline{s}_j$, $m+1 \le j \le d$, which are subvectors of the j-th column vector of the upper-triangular matrix of the message matrix M with components $u_{i,j}$ in Eq. (15), characterized by the parameter m.

First, let

$$\underline{r}_j := [u_{1,j}, u_{2,j}, \ldots, u_{j,j}]^t \in \mathbb{F}_q^j, \quad 1 \le j \le m,$$

and let

$$\left.\begin{array}{l} \underline{s}_j := [u_{1,j}, u_{2,j}, \ldots, u_{j-m,j}]^t \in \mathbb{F}_q^{j-m}, \\ \underline{r}_j := [u_{j-m+1,j}, u_{j-m+2,j}, \ldots, u_{j,j}]^t \in \mathbb{F}_q^m, \end{array}\right\}, \; m+1 \le j \le k,$$

where the above vectors are subvectors of the j-th column vector of the upper-triangular matrix of M_1. Let

$$\left.\begin{array}{l} \underline{s}_j := [u_{1,j}, u_{2,j}, \ldots, u_{k-m,j}]^t \in \mathbb{F}_q^{k-m}, \\ \underline{r}_j := [u_{k-m+1,j}, u_{k-m+2,j}, \ldots, u_{k,j}]^t \in \mathbb{F}_q^m, \end{array}\right\}, k+1 \le j \le d,$$

where the above vectors are subvectors of the $(j-k)$-th column vector of M_2.

Next, let $\underline{R}_1$ be the vector consisting of the k vectors $\underline{r}_j$, $1 \le j \le k$, as follows:

$$\underline{R}_1 := [\underline{r}_1^t, \underline{r}_2^t, \ldots, \underline{r}_k^t]^t \in \mathbb{F}_q^{L_{R_1}}, \tag{20}$$

and let $\underline{R}_2$ be the vector consisting of the $d-k$ vectors $\underline{r}_j$, $k+1 \le j \le d$, as follows:

$$\underline{R}_2 := [\underline{r}_{k+1}^t, \underline{r}_{k+2}^t, \ldots, \underline{r}_d^t]^t \in \mathbb{F}_q^{L_{R_2}}, \tag{21}$$

The length L_{R_1} of the vector $\underline{R}_1$ and that L_{R_2} of the vector $\underline{R}_2$ are respectively given as

$$L_{R_1} = \sum_{j=1}^{m} (k - j + 1) = \frac{1}{2} m(2k - m + 1), \tag{22}$$

$$L_{R_2} = m(d - k). \tag{23}$$

Let $\underline{R}$ be the vector consisting of $\underline{R}_1$ and $\underline{R}_2$ as follows:

$$\underline{R} := [\underline{R}_1^t, \underline{R}_2^t]^t \in \mathbb{F}_q^{L_R}, \tag{24}$$

and let $\underline{S}$ be the vector consisting the $d-m$ vectors $\underline{s}_j$, $m+1 \le j \le d$, as follows:

$$\underline{S} := [\underline{s}_{m+1}^t, \underline{s}_{m+2}^t, , \ldots, \underline{s}_d^t]^t \in \mathbb{F}_q^{L_S}. \tag{25}$$

The length L_R of the vector $\underline{R}$ and that L_S of the vector $\underline{S}$ are respectively given as

$$L_R = L_{R_1} + L_{R_2} = \frac{1}{2} m(2d - m + 1), \tag{26}$$

$$\begin{aligned} L_S &= \sum_{j=m+1}^{k} (k - j + 1) + (k - m)(d - k) \\ &= \frac{1}{2}(m - k)(m - (2d - k + 1)). \end{aligned} \tag{27}$$

Note that L_R and L_S satisfy $B = L_S + L_R$.

Finally, we propose a construction of an (n, k, d, m) secure regenerating code based on an (n, k, d) MBR code by substituting secret and random symbols for the components of the message matrix M as follows: the vector $\underline{S}$ consists of L_S secret symbols that are chosen from a finite field $\mathbb{F}_q$ uniformly and independently, and the vector $\underline{R}$ consists of L_R random symbols that are independent random elements uniformly distributed over $\mathbb{F}_q$. Then let $\underline{S}$ be called a *secret*, and let $\underline{R}$ be called a *random vector*. We assume that L_R random symbols are independent of L_S secret symbols. The message matrix M with B distinct entries is then filled up with all B symbols of $\underline{S}$ and $\underline{R}$. The (n, k, d, m) secure regenerating code is then defined by the above setting the secret $\underline{S}$ and the random vector $\underline{R}$ for the message matrix M. Thus, for each $i \in \{1, \ldots, n\}$, the share $\underline{c}_i$ stored in node i for the secret $\underline{S}$ is derived from Eq. (16).

Example 6: Let $(k, d, m) = (4, 6, 2)$. The size B of the message is then given as $B = 18$. The lengths L_{R_1} and L_{R_2} are given as $L_{R_1} = 7$ and $L_{R_2} = 4$, respectively. Thus $L_R = L_{R_1} + L_{R_2} = 11$. And then, the length L_S is given as $L_S = 7$. The message matrix M is filled up with eighteen symbols of the random vector $\underline{R} = [r_1, \ldots, r_{11}]^t \in \mathbb{F}_q^{11}$ and seven symbols of the secret $\underline{S} = [s_1, \ldots, s_7]^t \in \mathbb{F}_q^7$ as follows:

$$M = \left[\begin{array}{c|c} M_1 & M_2 \\ \hline M_2^t & \mathbf{O} \end{array}\right] = \left[\begin{array}{cccc|cc} r_1 & r_2 & s_1 & s_2 & s_4 & s_6 \\ r_2 & r_3 & r_4 & s_3 & s_5 & s_7 \\ s_1 & r_4 & r_5 & r_6 & r_8 & r_{10} \\ s_2 & s_3 & r_6 & r_7 & r_9 & r_{11} \\ \hline s_4 & s_5 & r_8 & r_9 & 0 & 0 \\ s_6 & s_7 & r_{10} & r_{11} & 0 & 0 \end{array}\right],$$

where the vectors $\underline{r}_j$ and $\underline{s}_j$ are given as follows:

$$\begin{aligned} &\underline{r}_1 = [u_{1,1}]^t = [r_1]^t,\ \underline{r}_2 = [u_{1,2}, u_{2,2}]^t = [r_2, r_3]^t, \\ &\underline{r}_3 = [u_{2,3}, u_{3,3}]^t = [r_4, r_5]^t,\ \underline{r}_4 = [u_{3,4}, u_{4,4}]^t = [r_6, r_7]^t, \\ &\underline{r}_5 = [u_{3,5}, u_{4,5}]^t = [r_8, r_9]^t,\ \underline{r}_6 = [u_{3,6}, u_{4,6}]^t = [r_{10}, r_{11}]^t, \\ &\underline{s}_3 = [u_{1,3}]^t = [s_1]^t,\ \underline{s}_4 = [u_{1,4}, u_{2,4}]^t = [s_2, r_3]^t, \\ &\underline{s}_5 = [u_{1,5}, u_{2,5}]^t = [r_4, r_5]^t,\ \underline{s}_6 = [u_{1,6}, u_{2,6}]^t = [r_6, r_7]^t, \end{aligned}$$

and

$$\begin{aligned} &\underline{R}_1 = [\underline{r}_1^t, \underline{r}_2^t, \underline{r}_3^t, \underline{r}_4^t]^t = [r_1, \ldots r_7]^t, \\ &\underline{R}_2 = [\underline{r}_5^t, \underline{r}_6^t]^t = [r_8, \ldots r_{11}]^t, \\ &\underline{S} = [\underline{s}_3^t, \underline{s}_4^t, \underline{s}_5^t, \underline{s}_6^t]^t = [s_1, \ldots s_7]^t. \end{aligned}$$

Shah et al. [9] have recently presented the $\{\ell, \ell'\}$ secure Product-Matrix MBR code ($\{\ell, \ell'\}$ secure PM-MBR code)[†] based on the MBR code [3]. For an MBR code, the secrecy model of their code is identical to that of our code. In a construction a of secure regenerating code, their code and our code use the same MBR code [3] as an underlying code. When $\ell = m$ and $\ell' = 0$, their code with parameters $[n, k, d]$ and our (n, k, d, m) secure regenerating code have the same secrecy ability. However, their code is different from our code since the massage matrix of their code is different from that of our code. We denote the same points and the different points between their code and our code in the case of the same secrecy ability below.

1. Their code and our code have the same features in the following:
   a. The size of submatrices M_1, M_2.
   b. The arrangement of submatrices in the message matrix M.
   c. The number of secret symbols contained in each of submatrices.
   d. The number of random symbols contained in each of submatrices.
2. Their code and our code differ in the position of random symbols and that of secret symbols. In their construction method [9, Sect.III-B] for $\{\ell = m, \ell' = 0\}$ secure PM-MBR code, the $\left(md - \binom{m}{2}\right)$ random symbols are substituted for the $\left(md - \binom{m}{2}\right)$ components of the first m rows of the message matrix M, where $\left(md - \binom{m}{2}\right) = L_R$.

4.2 Reconstruction and Regeneration

We consider a reconstruction and a regeneration for an (n, k, d, m) secure regenerating code. When we treat B symbols of L_S secret symbols and L_R random symbols in the message matrix M for the (n, k, d, m) secure regenerating code as B message symbols of the underlying (n, k, d) MBR code, the (n, k, d, m) secure regenerating code becomes identical to the (n, k, d) MBR code. Thus, a data-collector can reconstruct the message matrix M with the secret $\underline{S}$ and the random vector $\underline{R}$ from any k shares by using the reconstruction method [3, Theorem 3] for the (n, k, d) MBR code. After that, the data-collector can obtain the secret $\underline{S}$ from the reconstructed message matrix M. Consequently, the (n, k, d, m) secure regenerating code satisfies Eq. (3). Similarly, a failed node can repair the share from a piece-vector consisting of d pieces from any d helper-nodes by using the regenerating method described in Sect. 3.3. Consequently, the (n, k, d, m) secure regenerating code satisfies Eq. (4).

5. Evaluation

In this section, we evaluate the security of an (n, k, d, m) secure regenerating code proposed in Sect. 4. We introduce a parameter t, $0 \le t \le n$, to evaluate the proposed codes.

Considering the (n, k, d, m) secure regenerating code with a secret $\underline{S}$, the secret $\underline{S}$ is encoded to n shares $\underline{c}_1, \ldots, \underline{c}_n$ by using Eq. (16).

First, we consider the security of the (n, k, d, m) secure regenerating code for any t shares, that is, we evaluate the conditional entropy $H(S|C_{i_1}, \ldots, C_{i_t})$ of a random variable S representing a secret $\underline{S}$ given t random variables $C_{i_1}, \ldots, C_{i_t}$ representing any t shares $\underline{c}_{i_1}, \ldots, \underline{c}_{i_t}$. For simplifying the notation of the index, let $(i_1, i_2, \ldots, i_t) = (1, 2, \ldots, t)$, without loss of generality. We show the following first main theorem.

Theorem 7: For t random variables $C_1, \ldots, C_t$ representing any t shares $\underline{c}_1, \ldots, \underline{c}_t$,

$$H(S|C_1, \ldots, C_t) = \frac{g(t)}{L_S} H(S), \tag{28}$$

where the function $g(t)$ of an integer variable t is defined by

$$g(t) := \begin{cases} L_S, & (0 \le t \le m), \\ \frac{1}{2}(t-k)(t-(2d-k+1)), & (m+1 \le t \le k-1), \\ 0, & (k \le t \le n). \end{cases} \tag{29}$$

Notice that $H(S) = L_S$ from the definition of the secret $\underline{S}$. In particular, when $t = m$, Eq. (28) is identical to Eq. (5) because of $g(m) = L_S$. Furthermore, when $t = k$, Eq. (28) is identical to Eq. (3) because of $g(k) = 0$.

Proof: See Sect. 6 and Proof 18. □

The function $g(t)$ of a variable t is a convex and monotonically decreasing function in the range $m \le t \le k$. In particular, $g(t)$ equals L_S when $t = m$, and equals 0 when $t = k$.

Next, we consider the security of the (n, k, d, m) secure

[†]See [9] about the details of $\{\ell, \ell'\}$ secure PM-MBR code. The parameters ℓ and ℓ' are different from the parameter l used in Eq. (6).

regenerating code for any t piece-vectors, that is, we evaluate the conditional entropy $H(S|D_{i_1},\ldots,D_{i_t})$ of a random variable S representing a secret $\underline{S}$ given t random variables $D_{i_1},\ldots,D_{i_t}$ representing t piece-vectors $\underline{d}_{i_1},\ldots,\underline{d}_{i_t}$ of any t failed nodes $i_1,\ldots,i_t$. For simplifying the notation of the index, let $(i_1,i_2,\ldots,i_t)=(1,2,\ldots,t)$, without loss of generality. We show the following second main theorem.

Theorem 8: For t random variables $D_1,\ldots,D_t$ representing t piece-vectors $\underline{d}_1,\ldots,\underline{d}_t$ of any t failed nodes $1,\ldots,t$,

$$H(S|D_1,\ldots,D_t)=\frac{g(t)}{L_S}H(S), \tag{30}$$

where the function $g(t)$ is defined by Eq. (29). In particular, when $t=m$, Eq. (30) is identical to Eq. (6) because of $g(m)=L_S$.

Proof: From Corollary 4 and Theorem 7, $H(S|D_1,\ldots,D_t)=H(S|C_1,\ldots,C_t)=g(t)H(S)/L_S$. □

From Theorems 7 and 8, the (n,k,d,m) secure regenerating codes have the secrecy property of a ramp secret sharing scheme, and we have the following remark.

Remark 9: We mentioned the secrecy capacity and its upper bound for an MBR code at last part in Sect. 2. For an MBR code, we can assume that $l=m$ without loss of generality, and have the inequation (12). For the parameters $(\alpha=d,\beta=1)$ of the (n,k,d,m) secure regenerating code, the inequation (12) is represented as

$$C_S(n,k,d,d,1;m,m)\le\sum_{j=m+1}^{k}(d-j+1)=L_S. \tag{31}$$

On the other hand, the (n,k,d,m) secure regenerating code satisfies the properties of reconstruction and regeneration in Eqs. (3) and (4), respectively, as mentioned in Sect. 4.2. Moreover, from Theorems 7 and 8, when $t=m$, the (n,k,d,m) secure regenerating code satisfies the secrecy properties of Eqs. (5) and (6), respectively. Hence, the (n,k,d,m) secure regenerating code achieves the upper bound of the secrecy capacity $C_S(n,k,d,d,1;m,m)$ because of $H(S)=L_S$.

Finally, apart from the argument about the security of the ramp type of Eqs. (28) and (30), we assume that an eavesdropper can obtain n pieces $\underline{d}_{f,1},\ldots,\underline{d}_{f,n}$ for a failed node f in a network. Note that each piece $\underline{d}_{f,i}$, $i\in\{1,\ldots,n\}$, for the failed node f is computed in node i. And then, we show the security of the (n,k,d,m) secure regenerating code for pieces such that no information of a secret leaks to an eavesdropper even if the eavesdropper can obtain n pieces $\underline{d}_{f,1},\ldots,\underline{d}_{f,n}$ for the failed node f.

Theorem 10: If $m\ge 1$ for an (n,k,d,m) secure regenerating code, then for n random variables $D_{f,1},\ldots,D_{f,n}$ representing n pieces $\underline{d}_{f,1},\ldots,\underline{d}_{f,n}$,

$$H(S|D_{f,1},\ldots,D_{f,n})=H(S). \tag{32}$$

Proof: For any d nodes $h_1,\ldots,h_d$, let $\underline{d}_f=[\underline{d}_{f,h_1},\ldots,\underline{d}_{f,h_d}]^t$. Let D_{f,h_p} and D_f be random variables representing a piece $\underline{d}_{f,h_p}$ and a piece-vector $\underline{d}_f$, respectively, and then let $D_f=[D_{f,h_1},\ldots,D_{f,h_d}]$. From Theorem 8, $H(S|D_f)=H(S)$ when $m\ge 1$. For any node $j\in\{1,\ldots,n\}\setminus\{h_1,h_2,\ldots,h_d\}$, $H(D_{f,j}|D_f)=0$ because $\underline{d}_{f,j}$ is uniquely computed from the piece-vector $\underline{d}_f$ and the coding vector $\underline{\phi}_j$ associated with node j as follows. First, from Eq. (19), $\underline{c}_f$ can be uniquely computed from $\underline{d}_f$. Next, from Lemma 5, we have that $\underline{c}_f^t\underline{\phi}_j=\underline{c}_j^t\underline{\phi}_f=\underline{d}_{f,j}$. Thus, each piece $\underline{d}_{f,j}$ can be uniquely computed from $\underline{d}_f$ and $\underline{\phi}_j$, that is, $H(D_{f,j}|D_f)=0$. As a result, we have $H(S)=H(S|D_f)=H(S|D_f,D_{f,j})$. Therefore, $H(S)=H(S|D_f)=H(S|D_f,D_{f,j_1})=H(S|D_f,D_{f,j_1},D_{f,j_2})=\cdots=H(S|D_f,D_{f,j_1},\ldots,D_{f,j_{n-d}})=H(S|D_{f,1}\cdots D_{f,n})$, where $j_1,\ldots,j_{n-d}\in\{1,\ldots,n\}\setminus\{h_1,\ldots,h_d\}$. This completes the proof of the lemma. □

6. Proof of Theorem 7

In this section, we define vectors $\underline{u}^{(1)},\underline{u}^{(2)},\underline{u}^{(3)},\underline{v}^{(1)},\underline{v}^{(2)},\underline{w},\underline{v}$, which are dependent to the parameter t introduced in Sect. 5 where $0\le t\le k$, in order to prove Theorem 7. These vectors are mainly used in the proof of Theorem 15.

6.1 Vectors Consisting of Components of the Message Matrix M

We define three kinds of vectors $\underline{u}_j^{(1)}$, $1\le j\le k$, $\underline{u}_j^{(2)}$, $k+1\le j\le d$, and $\underline{u}_j^{(3)}$, $t+1\le j\le d$, which are related to the j-th column vector of upper-triangular matrix of the message matrix M with components $u_{i,j}$, characterized by the parameter t. The way of definition of the vectors is the same as that in Sect. 4. The vectors $\underline{u}_j^{(1)}$, $\underline{u}_j^{(2)}$ and $\underline{u}_j^{(3)}$ are corresponding to the vectors $\underline{r}_j$, $1\le j\le k$, $\underline{r}_j$, $k+1\le j\le d$, and $\underline{s}_j$, $m+1\le j\le d$, in Sect. 4, respectively.

First, let

$$\underline{u}_j^{(1)}:=[u_{1,j},u_{2,j},\ldots,u_{j,j}]^t\in\mathbb{F}_q^j,\quad 1\le j\le t,$$

and let

$$\left.\begin{aligned}\underline{u}_j^{(3)}&:=[u_{1,j},u_{2,j},\ldots,u_{j-t,j}]^t\in\mathbb{F}_q^{j-t},\\ \underline{u}_j^{(1)}&:=[u_{j-t+1,j},u_{j-t+2,j},\ldots,u_{j,j}]^t\in\mathbb{F}_q^{t}\end{aligned}\right\},\ t+1\le j\le k,$$

where the above vectors are subvectors of the j-th column vector of the upper-triangular matrix M_1. Let

$$\left.\begin{aligned}\underline{u}_j^{(3)}&:=[u_{1,j},u_{2,j},\ldots,u_{k-t,j}]^t\in\mathbb{F}_q^{k-t},\\ \underline{u}_j^{(2)}&:=[u_{k-t+1,j},u_{k-t+2,j},\ldots,u_{k,j}]^t\in\mathbb{F}_q^{t}\end{aligned}\right\},\ k+1\le j\le d,$$

where the above vectors are subvectors of the $(j-k)$-th column vector of M_2.

Next, from the above defined vectors, the vector $\underline{u}^{(1)}$ consisting of the k vectors $\underline{u}_j^{(1)}$, $1\le j\le k$, is defined by

$$\underline{u}^{(1)}:=[(\underline{u}_1^{(1)})^t,(\underline{u}_2^{(1)})^t,\ldots,(\underline{u}_k^{(1)})^t]^t\in\mathbb{F}_q^{L_1}, \tag{33}$$

where $(\underline{u}_j^{(1)})^t$ is the transpose of the vector $\underline{u}_j^{(1)}$, and the vector $\underline{u}^{(2)}$ consisting of the $d-k$ vectors $\underline{u}_j^{(2)}$, $k+1 \le j \le d$, is defined by

$$\underline{u}^{(2)} := [(\underline{u}_{k+1}^{(2)})^t, (\underline{u}_{k+2}^{(2)})^t, \ldots, (\underline{u}_d^{(2)})^t]^t \in \mathbb{F}_q^{L_2}. \tag{34}$$

By substituting t for m in Eq. (22) and Eq. (23), the lengths L_1 and L_2 of the vectors $\underline{u}^{(1)}$ and $\underline{u}^{(2)}$ are respectively given by

$$L_1 = \frac{1}{2}t(2k-t+1), \tag{35}$$

$$L_2 = t(d-k). \tag{36}$$

Finally, the vector $\underline{u}^{(3)}$ consisting of the $d-t$ vectors $\underline{u}_j^{(3)}$, $t+1 \le j \le d$, is defined by

$$\underline{u}^{(3)} := [(\underline{u}_{t+1}^{(3)})^t, (\underline{u}_{t+2}^{(3)})^t, \ldots, (\underline{u}_d^{(3)})^t]^t \in \mathbb{F}_q^{B-L_1-L_2}. \tag{37}$$

Example 11: (Example 6 continued) Let $(k,d,t) = (4,6,3)$. The size B of the message is equal to that of Example 6, i.e., $B = 18$ because of $(k,d) = (4,6)$ in both examples. On the other hand, the lengths L_1 and L_2 are given as $L_1 = 9$ and $L_2 = 6$, respectively. And then, the length of the vector $\underline{u}^{(3)}$ is equal to $B - L_1 - L_2 = 3$. The message matrix M is given as follows:

$$M = \left[\begin{array}{c|c} M_1 & M_2 \\ \hline M_2^t & \mathbf{O} \end{array}\right] = \left[\begin{array}{cccc|cc} u_{1,1} & u_{1,2} & u_{1,3} & u_{1,4} & u_{1,5} & u_{1,6} \\ u_{1,2} & u_{2,2} & u_{2,3} & u_{2,4} & u_{2,5} & u_{2,6} \\ u_{1,3} & u_{2,3} & u_{3,3} & u_{3,4} & u_{3,5} & u_{3,6} \\ u_{1,4} & u_{2,4} & u_{3,4} & u_{4,4} & u_{4,5} & u_{4,6} \\ \hline u_{1,5} & u_{2,5} & u_{3,5} & u_{4,5} & 0 & 0 \\ u_{1,6} & u_{2,6} & u_{3,6} & u_{4,6} & 0 & 0 \end{array}\right],$$

where the vectors $\underline{u}_j^{(1)}$, $\underline{u}_j^{(2)}$ and $\underline{u}_j^{(3)}$ are given as follows:

$$\underline{u}_1^{(1)} = [u_{1,1}]^t,\ \underline{u}_2^{(1)} = [u_{1,2}, u_{2,2}]^t,\ \underline{u}_3^{(1)} = [u_{1,3}, u_{2,3}, u_{3,3}]^t,$$
$$\underline{u}_4^{(3)} = [u_{1,4}]^t,\ \underline{u}_4^{(1)} = [u_{2,4}, u_{3,4}, u_{4,4}]^t,$$
$$\underline{u}_5^{(3)} = [u_{1,5}]^t,\ \underline{u}_5^{(2)} = [u_{2,5}, u_{3,5}, u_{4,5}]^t,$$
$$\underline{u}_6^{(3)} = [u_{1,6}]^t,\ \underline{u}_6^{(2)} = [u_{2,6}, u_{3,6}, u_{4,6}]^t,$$

and

$$\underline{u}^{(1)} = [(\underline{u}_1^{(1)})^t, (\underline{u}_2^{(1)})^t, (\underline{u}_3^{(1)})^t, (\underline{u}_4^{(1)})^t)]^t \in \mathbb{F}_q^9,$$
$$\underline{u}^{(2)} = [(\underline{u}_5^{(2)})^t, (\underline{u}_6^{(2)})^t]^t \in \mathbb{F}_q^6,$$
$$\underline{u}^{(3)} = [(\underline{u}_4^{(3)})^t, (\underline{u}_5^{(3)})^t, (\underline{u}_6^{(3)})^t]^t \in \mathbb{F}_q^3.$$

6.2 Vectors Consisting of Components of t Shares

In this section, by using the same parameter t used in the previous Sect. 6.1, we define four vectors $\underline{v}^{(1)}, \underline{v}^{(2)}, \underline{w}$, and $\underline{v}$.

For any t shares $\underline{c}_{i_1}, \underline{c}_{i_2}, \ldots, \underline{c}_{i_t}$, we consider making replacement all components of them to new three vectors $\underline{v}^{(1)}, \underline{v}^{(2)}$, and $\underline{w}$. For simplifying the notation of the index, let $(i_1, i_2, \ldots, i_t) = (1, 2, \ldots, t)$.

Let A be a set of all components of the t shares $\underline{c}_1, \ldots, \underline{c}_t$. The cardinality of the set A is equal to dt. We consider a direct sum decomposition of A such that A is a direct sum of A_1 and A_2 so that for any element in A_2, the element can be represented by a linear combination of elements in A_1 over $\mathbb{F}_q$. To construct such a pair of subsets A_1 and A_2 of A, we define three kinds of vectors $\underline{v}_j^{(1)}$, $\underline{v}_j^{(2)}$ and $\underline{w}_j$ below.

For any t shares $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_t$, let $C(t)$ be a $t \times d$ matrix with components of the t shares as follows:

$$C(t) = \left[\begin{array}{ccccccccc} c_{1,1} & \cdots & c_{1,t-1} & c_{1,t} & \cdots & c_{1,k} & c_{1,k+1} & \cdots & c_{1,d} \\ c_{2,1} & \cdots & c_{2,t-1} & c_{2,t} & \cdots & c_{2,k} & c_{2,k+2} & \cdots & c_{2,d} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c_{t,1} & \cdots & c_{t,t-1} & c_{t,t} & \cdots & c_{t,k} & c_{t,k+1} & \cdots & c_{t,d} \end{array}\right], \tag{38}$$

where the i-th row of the matrix $C(t)$ is the vector of a share $\underline{c}_i^t$, $1 \le i \le t$.

First, for the j-th column of the matrix $C(t)$, we define the following three kinds of vectors $\underline{v}_j^{(1)}$, $1 \le j \le k$, $\underline{v}_j^{(2)}$, $k+1 \le j \le d$, and $\underline{w}_j$, $1 \le j \le t-1$. Let

$$\left.\begin{array}{l} \underline{v}_j^{(1)} := [c_{1,j}, c_{2,j}, \ldots, c_{j,j}]^t \in \mathbb{F}_q^j, \\ \underline{w}_j := [c_{j+1,j}, c_{j+2,j}, \ldots, c_{t,j}]^t \in \mathbb{F}_q^{t-j} \end{array}\right\},\ 1 \le j \le t-1,$$

and let

$$\underline{v}_j^{(1)} := [c_{1,j}, c_{2,j}, \ldots, c_{t,j}]^t \in \mathbb{F}_q^t,\ \ t \le j \le k,$$
$$\underline{v}_j^{(2)} := [c_{1,j}, c_{2,j}, \ldots, c_{t,j}]^t \in \mathbb{F}_q^t,\ \ k+1 \le j \le d.$$

Next, from the above defined vectors, let $\underline{v}^{(1)}$ be a vector consisting of the k vectors $\underline{v}_j^{(1)}$, $1 \le j \le k$, as follows:

$$\underline{v}^{(1)} := [(\underline{v}_1^{(1)})^t, (\underline{v}_2^{(1)})^t, \ldots, (\underline{v}_k^{(1)})^t]^t \in \mathbb{F}_q^{L_1}, \tag{39}$$

and let $\underline{v}^{(2)}$ be a vector consisting of the $d-k$ vectors $\underline{v}_j^{(2)}$, $k+1 \le j \le d$, as follows:

$$\underline{v}^{(2)} := [(\underline{v}_{k+1}^{(2)})^t, (\underline{v}_{k+2}^{(2)})^t, \ldots, (\underline{v}_d^{(2)})^t]^t \in \mathbb{F}_q^{L_2}. \tag{40}$$

Let $\underline{v}$ be a vector consisting of $\underline{v}^{(1)}$ and $\underline{v}^{(2)}$ as follows:

$$\underline{v} := [(\underline{v}^{(1)})^t, (\underline{v}^{(2)})^t]^t \in \mathbb{F}_q^{L_1+L_2}. \tag{41}$$

The lengths L_1 and L_2 correspond to Eq. (35) and Eq. (36) respectively. Let $\underline{w}$ be a vector consisting of the $t-1$ vectors $\underline{w}_j$, $1 \le j \le t-1$, as follows:

$$\underline{w} := [\underline{w}_1^t, \underline{w}_2^t, \ldots, \underline{w}_{t-1}^t]^t \in \mathbb{F}_q^{dt-L_1-L_2}. \tag{42}$$

Example 12: (Example 11 continued) Let $(k,d,t) = (4,6,3)$. The parameters (B, L_1, L_2) and the matrix $C(t)$ are given as follows: $(B, L_1, L_2) = (18, 9, 6)$ and

$$C(t) = \left[\begin{array}{cccccc} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} & c_{1,5} & c_{1,6} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} & c_{2,5} & c_{2,6} \\ c_{3,1} & c_{3,2} & c_{3,3} & c_{3,4} & c_{3,5} & c_{3,6} \end{array}\right], \tag{43}$$

where $t = 3$, and the vectors $\underline{v}_j^{(1)}$, $\underline{v}_j^{(2)}$, and $\underline{w}_j$ are given as follows:

$$\underline{v}_1^{(1)} = [c_{1,1}]^t,\ \ \underline{w}_1 = [c_{2,1}, c_{3,1}]^t,$$

$$\underline{v}_2^{(1)} = [c_{1,2}, c_{2,2}]^t, \quad \underline{vw}_2 = [c_{3,2}]^t,$$
$$\underline{v}_3^{(1)} = [c_{1,3}, c_{2,3}, c_{3,3}]^t, \quad \underline{v}_4^{(1)} = [c_{1,4}, c_{2,4}, c_{3,4}]^t,$$
$$\underline{v}_5^{(2)} = [c_{1,5}, c_{2,5}, c_{3,5}]^t, \quad \underline{v}_6^{(2)} = [c_{1,6}, c_{2,6}, c_{3,6}]^t,$$

and

$$\underline{v}^{(1)} = [(\underline{v}_1^{(1)})^t, (\underline{v}_2^{(1)})^t, (\underline{v}_3^{(1)})^t, (\underline{v}_4^{(1)})^t]^t \in \mathbb{F}_q^9,$$
$$\underline{v}^{(2)} = [(\underline{v}_5^{(2)})^t, (\underline{v}_6^{(2)})^t]^t \in \mathbb{F}_q^6,$$
$$\underline{w} = [\underline{w}_1^t, \underline{w}_2^t]^t \in \mathbb{F}_q^3.$$

Finally, let A_1 be a set consisting of all components of the vector $\underline{v}$, and let A_2 be a set consisting of all components of the vector $\underline{w}$. From the definition of $\underline{v}$ and $\underline{w}$, it is true that the set A is the direct sum of A_1 and A_2. Note that the cardinality of A_1 is equal to $L_1 + L_2$, and that of A_2 is equal to $dt - L_1 - L_2$.

We show the relation between the sets A_1 and A_2 in the following lemma.

Lemma 13: For any element in A_2, the element can be represented by a linear combination of elements in A_1 over $\mathbb{F}_q$. In other words, each component of the vector $\underline{w}$ can be represented by a linear combination of components of the vector $\underline{v}$ over $\mathbb{F}_q$.
Proof: See Appendix B. □

Let V and W be random variables representing a vector $\underline{v}$ of Eq. (41) and a vector $\underline{w}$ of Eq. (42), respectively. From the definitions of vectors of $\underline{v}$ and $\underline{w}$, the random variable V and W are defined from only t random variables $C_1, \ldots, C_t$. We have the following lemma from Lemma 13.

Lemma 14: For t random variables $C_1, \ldots, C_t$ representing any t shares $\underline{c}_1, \ldots, \underline{c}_t$, the conditional entropy $H(W|V)$ of W given V is given by $H(W|V) = 0$.
Proof: From Lemma 13, each component of $\underline{w}$ can be uniquely determined from components of $\underline{v}$. This completes the proof of the lemma. □

We show the conditional entropy $H(S|V)$ of S given V in the following theorem.

Theorem 15: For t random variables $C_1, \ldots, C_t$ representing any t shares $\underline{c}_1, \ldots, \underline{c}_t$, the conditional entropy $H(S|V)$ of S given V is given by

$$H(S|V) = \frac{g(t)}{L_S} H(S), \tag{44}$$

where the function $g(t)$ is defined by Eq. (29).
Proof: See Appendix C. □

Example 16: (Example 6 continued) Let $(k, d, m) = (4, 6, 2)$. The parameters (B, L_S, L_R) are then given as follows: $(B, L_S, L_R) = (18, 7, 11)$. For each t such that $0 \le t \le n$, the conditional entropy $H(S|V)$ of S given V is as follows:

$$H(S|V) = \begin{cases} 7, & (0 \le t \le 2), \\ 3, & (t = 3), \\ 0, & (4 \le t \le n). \end{cases}$$

□

From Lemma 14, we have the following theorem.

Theorem 17: For t random variables $C_1, \ldots, C_t$ representing any t shares $\underline{c}_1, \ldots, \underline{c}_t$,

$$H(S|C_1, \ldots, C_t) = H(S|V). \tag{45}$$

Proof:

$$H(S|C_1, \ldots, C_t) = H(S|V, W) = H(S|V). \tag{46}$$

The first equality follows from the relation between $[\underline{v}^t, \underline{w}^t]$ and $[\underline{c}_1^t, \ldots, \underline{c}_t^t]$ such that the vector $[\underline{v}^t, \underline{w}^t]$ is given by rearranging components of the vector $[\underline{c}_1^t, \ldots, \underline{c}_t^t]$ which consists of t shares. The second equality follows from Lemma 14. □

Finally, a proof of Theorem 7 is derived from the above argument as follows:

Proof 18 (Proof of Theorem 7): From Theorems 15 and 17, Theorem 7 is valid. □

7. Conclusions

In this paper, we have presented a construction of the (n, k, d, m) secure regenerating codes based on the (n, k, d) MBR codes given by Rashmi et al. [3] for all values of the parameters (n, k, d). The (n, k, d, m) secure regenerating codes have the security of a ramp secret sharing scheme and achieve the upper bound of the secrecy capacity. The complexities of encoding, reconstruction and regeneration are the same as that of the underlying MBR code.

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, vol.56, no.9, pp.4539–4551, Sept. 2010.
[2] S. Pawar, S.E. 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. 2011.
[3] 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, vol.57, no.8, pp.5227–5239, Aug. 2011.
[4] M. Kurihara and H. Kuwakado, "On regenerating codes and secret sharing for distributed storage," IEICE Technical Report, IT2010-56, Jan. 2011. (in Japanese)
[5] M. Kurihara and H. Kuwakado, "On an extended version of Rashmi-Shah-Kumar regenerating codes and secret sharing for distributed storage," IEICE Technical Report, IT2010-114, March 2011. (in Japanese)
[6] M. Kurihara and H. Kuwakado, "On ramp secret sharing schemes for distributed storage systems under repair dynamics," IEICE Technical Report, IT2011-17, July 2011. (in Japanese)
[7] H. Kuwakado and M. Kurihara, "Computationally-secure regenerating code," Proc. Computer Security Symposium 2011, pp.131–136, 19–21, Oct. 2011.
[8] M. Kurihara and H. Kuwakado, "Ramp secret sharing schemes based on MBR codes," IEICE Technical Report, ISEC2011-43, Nov. 2011. (in Japanese)

[9] N.B. Shah, K.V. Rashmi, and P.V. Kumar, "Information-theoretically secure regenerating codes for distributed storage," IEEE Global Communications Conference (GLOBECOM) 2011, Dec. 2011.
[10] M. Kurihara and H. Kuwakado, "Secure regenerating codes based on MSR codes for distributed storage systems," IEICE Trans. Fundamentals (Japanese Edition), to be published.
[11] M. Kurihara and H. Kuwakado, "Secret sharing schemes based on minimum bandwidth regenerating codes," Int. Symp. Inf. Theory and its Appl.(ISITA), Oct. 2012.

Appendix A: Proof of Theorem 1

We consider two cases of $l < m$ and $l \geq m$ for the parameters l and m. The following proof is similar to that of [2, Appendix A].

First, in the case of $l < m$, since $H(S|D_{i_1}, \ldots, D_{i_l}) = H(S) = H(S|C_{i_1}, \ldots, C_{i_m})$ from the two secrecy conditions of the secrecy capacity by Eq. (10), we have

$$\begin{aligned} H(S) &\overset{(a)}{=} H(S|D_{i_1}, \ldots, D_{i_l}) - H(S|C_{i_1}, \ldots, C_{i_k}) \\ &\overset{(b)}{=} H(S|C_{i_1}, \ldots, C_{i_m}) - H(S|C_{i_1}, \ldots, C_{i_k}) \\ &= I(S; C_{i_{m+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_m}) \\ &= H(C_{i_{m+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_m}) \\ &\quad - H(C_{i_{m+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_m}, S) \\ &\leq H(C_{i_{m+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_m}) \\ &\overset{(c)}{=} \sum_{j=m+1}^{k} H(C_{i_j}|C_{i_1}, \ldots, C_{i_{j-1}}) \\ &\overset{(d)}{\leq} \sum_{j=m+1}^{k} \min\{(d-j+1)\beta, \alpha\}, \end{aligned}$$

where $I(S; C_{i_{m+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_m})$ denotes the conditional mutual information of S and $C_{i_{m+1}}, \ldots, C_{i_k}$ given $C_{i_1}, \ldots, C_{i_m}$, equality (a) follows from $H(S|D_{i_1}, \ldots, D_{i_l}) = H(S)$ and $H(S|C_{i_1}, \ldots, C_{i_k}) = 0$, (b) follows from $H(S|D_{i_1}, \ldots, D_{i_l}) = H(S) = H(S|C_{i_1}, \ldots, C_{i_m})$, (c) follows from Chain rule of entropy, and (d) follows from the same reason of [2, Inequality (3) in Appendix A].

Next, in the case of $l \geq m$, since $H(S|D_{i_1}, \ldots, D_{i_l}) = H(S)$ implies $H(S|C_{i_1}, \ldots, C_{i_l}) = H(S)$ from Eq. (9), we have

$$\begin{aligned} H(S) &= H(S|D_{i_1}, \ldots, D_{i_l}) - H(S|C_{i_1}, \ldots, C_{i_k}) \\ &\overset{(e)}{=} H(S|C_{i_1}, \ldots, C_{i_l}) - H(S|C_{i_1}, \ldots, C_{i_k}) \\ &= I(S; C_{i_{l+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_l}) \\ &= H(C_{i_{l+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_l}) \\ &\quad - H(C_{i_{l+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_l}, S) \\ &\leq H(C_{i_{l+1}}, \ldots, C_{i_k}|C_{i_1}, \ldots, C_{i_l}) \\ &= \sum_{j=l+1}^{k} H(C_{i_j}|C_{i_1}, \ldots, C_{i_{j-1}}) \\ &\leq \sum_{j=l+1}^{k} \min\{(d-j+1)\beta, \alpha\}, \end{aligned}$$

where equality (e) follows from $H(S|C_{i_1}, \ldots, C_{i_l}) = H(S)$ by Eq. (9).

From the above argument about both cases and the definition of the secrecy capacity by Eq. (10), we have the following upper bound of the secrecy capacity:

$$C_S \leq \sum_{j=\max\{l,m\}+1}^{k} \min\{(d-j+1)\beta, \alpha\}.$$

□

Appendix B: Proof of Lemma 13

First, we show the relation between two shares in the following lemma.

Lemma 19: For any two shares $\underline{c}_i = [c_{i,1}, \ldots, c_{i,d}]^t$ and $\underline{c}_j = [c_{j,1}, \ldots, c_{j,d}]^t$,

$$\sum_{a=1}^{d} (x_j^{a-1} c_{i,a} - x_i^{a-1} c_{j,a}) = 0, \tag{A·1}$$

where x_i and x_j are elements assigned to node i and j, respectively.

Proof: From $c_{i,a} = \sum_{b=1}^{d} u_{a,b} x_i^{b-1}$, $c_{j,a} = \sum_{b=1}^{d} u_{a,b} x_j^{b-1}$, rewriting the left-hand side of Eq. (A·1) shows that

$$\sum_{a=1}^{d} \sum_{b=1}^{d} (x_j^{a-1} x_i^{b-1} - x_i^{a-1} x_j^{b-1}) u_{a,b}. \tag{A·2}$$

i) If $b = a$ then $(x_j^{a-1} x_i^{a-1} - x_i^{a-1} x_j^{a-1}) u_{a,a} = 0 \cdot u_{a,a} = 0$. ii) On the other hand, if $b \neq a$ then $(x_j^{a-1} x_i^{b-1} - x_i^{a-1} x_j^{b-1}) u_{a,b} + (x_j^{b-1} x_i^{a-1} - x_i^{b-1} x_j^{a-1}) u_{b,a} = 0$ because of $u_{a,b} = u_{b,a}$. Therefore, this lemma is proved. □

For an integer b such that $2 \leq b \leq t$, we define the two kinds of sets $A_{1,i}(b)$, $1 \leq j \leq b$, and $A_{2,i}(b)$, $2 \leq j \leq b$, and three sets $A_1(b)$, $A_2(b)$ and $A(b)$, which consist of elements of b shares $\underline{c}_1, \ldots, \underline{c}_b$ as follows:

$$A_{1,i}(b) = \{c_{i,j} | i \leq j \leq d\}, \ 1 \leq i \leq b, \tag{A·3}$$

$$A_{2,i}(b) = \{c_{i,j} | 1 \leq j \leq i-1\}, \ 2 \leq i \leq b, \tag{A·4}$$

$$A_1(b) = \cup_{i=1}^{b} A_{1,i}(b), \tag{A·5}$$

$$A_2(b) = \cup_{i=2}^{b} A_{2,i}(b), \tag{A·6}$$

$$A(b) = A_1(b) \cup A_2(b). \tag{A·7}$$

The set $A(b)$ is a direct sum set of $A_1(b)$ and $A_2(b)$ since the two sets $A_1(b)$ and $A_2(b)$ are disjoint from the above definition of the sets. Thus, the set $A(b)$ includes all elements of b shares $\underline{c}_1, \ldots, \underline{c}_b$.

Let i and b be indexes which satisfy that $1 \leq i < b \leq t$. For a pair (i, b) of such indexes i and b, we have the following $b-1$ linear equations, which are derived by substituting b for j in Eq. (A·1), from Lemma 19.

$$\sum_{a=1}^{d} (x_b^{a-1} c_{i,a} - x_i^{a-1} c_{b,a}) = 0, \ 1 \leq i \leq b-1. \tag{A·8}$$

Note that there are components of only two shares $\underline{c}_i$ and $\underline{c}_b$ in each equation. We rewrite each equation of the above equations so that $b-1$ terms $c_{b,1},\ldots,c_{b,b-1}$, which are elements of the share $\underline{c}_b$, are transposed to the left-hand side of the equation, and all the other terms are transposed to the right-hand side of the equation, that is,

$$\sum_{a=1}^{b-1} x_i^{a-1} c_{b,a} = \sum_{a=b}^{d} (x_b^{a-1} c_{i,a} - x_i^{a-1} c_{b,a}) + \sum_{a=1}^{b-1} x_b^{a-1} c_{i,a}, \quad 1 \le i \le b-1. \quad (\text{A}\cdot 9)$$

Note that in all bd elements of b shares, $b-1$ elements $c_{b,1},\ldots,c_{b,b-1}$ belong to the set $A_2(b)$, and all the other elements belong to the set $A_1(b)$.

To construct the proof of Lemma 13: "For any element in A_2, the element can be represented by a linear combination of elements in A_1 over $\mathbb{F}_q$", let $P(b)$ denote the proposition function for integers b, $2 \le b \le t$: "For any element in $A_2(b)$, the element can be represented by a linear combination of elements in $A_1(b)$ over $\mathbb{F}_q$".

Basis Step: When $b=2$, $A_2(2)=\{c_{2,1}\}$. $P(2)$ is true, since from Eq. (A· 9),

$$c_{2,1} = \sum_{a=2}^{d} (x_2^{a-1} c_{1,a} - x_1^{a-1} c_{2,a}) + c_{1,1}. \quad (\text{A}\cdot 10)$$

Inductive Step: Assume that $P(b-1)$ is true, that is, for any element in $A_2(b-1)$, the element can be represented by a linear combination of elements in $A_1(b-1)$ over $\mathbb{F}_q$.

We must show that $P(b)$ is true. That is, we must show that for any element in $A_2(b)$, the element can be represented by a linear combination of elements in $A_1(b)$ over $\mathbb{F}_q$. From the assumption of the inductive step, we only have to show that for any element in $A_2(b) \setminus A_2(b-1)$, the element can be represented by a linear combination of elements in $A_1(b)$ over $\mathbb{F}_q$. Note that $A_2(b) \setminus A_2(b-1) = \{c_{b,1},\ldots,c_{b,b-1}\}$.

For each pair (i,b) such that $1 \le i \le b-1$, from Eq. (A· 9), we have the following $b-1$ linear equations:

$$\begin{aligned}
\sum_{a=1}^{b-1} x_1^{a-1} c_{b,a} &= Z_{1,b},\\
\sum_{a=1}^{b-1} x_2^{a-1} c_{b,a} &= Z_{2,b},\\
&\vdots \quad \vdots\\
\sum_{a=1}^{b-1} x_{b-1}^{a-1} c_{b,a} &= Z_{b-1,b},
\end{aligned}$$

where for simplifying the notation of the right-hand side of Eq. (A· 9), we define that $Z_{i,b} = \sum_{a=b}^{d}(x_b^{a-1}c_{i,a} - x_i^{a-1}c_{b,a}) + \sum_{a=1}^{b-1} x_b^{a-1} c_{i,a}$ for $1 \le i \le b-1$. We rewrite the system of the $b-1$ linear equations in $b-1$ unknowns $c_{b,1},\ldots,c_{b,b-1}$ as follows:

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^{b-2} \\ 1 & x_2 & \cdots & x_2^{b-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{b-1} & \cdots & x_{b-1}^{b-2} \end{bmatrix} \begin{bmatrix} c_{b,1} \\ c_{b,2} \\ \vdots \\ c_{b,b-1} \end{bmatrix} = \begin{bmatrix} Z_{1,b} \\ Z_{2,b} \\ \vdots \\ Z_{b-1,b} \end{bmatrix} \quad (\text{A}\cdot 11)$$

The leftmost $(b-1)\times(b-1)$ square matrix of this equation is nonsingular because the determinant of the square matrix is the Vandermonde determinant from the conditions for x_i. Therefore, each element of the set $\{c_{b,1},\ldots,c_{b,b-1}\}$ can be represented by a linear combination of elements in $A_1(b)$ over $\mathbb{F}_q$. This induction step is complete.

Therefore, by the principle of mathematical induction, it has been shown that $P(b)$ is true for all integers b such that $2 \le b \le t$. Hence, this completes the proof of the lemma since $A_1 = A_1(t)$ and $A_2 = A_2(t)$ when $b = t$. □

Appendix C: Proof of Theorem 15

Recall three vectors $\underline{u}^{(1)}$, $\underline{u}^{(2)}$ and $\underline{u}^{(3)}$ defined by Eq. (33), (34) and (37), respectively. Let $[(\underline{u}^{(1)})^t, (\underline{u}^{(2)})^t, (\underline{u}^{(3)})^t]$ be the vector consisting of $\underline{u}^{(1)}$, $\underline{u}^{(2)}$ and $\underline{u}^{(3)}$. Moreover, recall three vectors $\underline{v}^{(1)}$, $\underline{v}^{(2)}$ and $\underline{v}$ defined by Eqs. (39), (40) and (41) , respectively, such that $\underline{v} = [(\underline{v}^{(1)})^t, (\underline{v}^{(2)})^t]^t$.

From the relation between the message symbols and the components of the shares defined by Eq. (16), we have the following system of linear equations for the two vectors $[(\underline{u}^{(1)})^t, (\underline{u}^{(2)})^t, (\underline{u}^{(3)})^t]^t$ and $\underline{v} = [(\underline{v}^{(1)})^t, (\underline{v}^{(2)})^t]^t$:

$$\begin{bmatrix} \underline{v}^{(1)} \\ \underline{v}^{(2)} \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,1} & X_{2,2} & X_{2,3} \end{bmatrix} \begin{bmatrix} \underline{u}^{(1)} \\ \underline{u}^{(2)} \\ \underline{u}^{(3)} \end{bmatrix} \quad (\text{A}\cdot 12)$$

$$= \begin{bmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{bmatrix} \begin{bmatrix} \underline{u}^{(1)} \\ \underline{u}^{(2)} \end{bmatrix} + \begin{bmatrix} X_{1,3} \\ X_{2,3} \end{bmatrix} \underline{u}^{(3)} \quad (\text{A}\cdot 13)$$

$$= \begin{bmatrix} X_{1,1} \\ X_{2,1} \end{bmatrix} \underline{u}^{(1)} + \begin{bmatrix} X_{1,2} \\ X_{2,2} \end{bmatrix} \underline{u}^{(2)} + \begin{bmatrix} X_{1,3} \\ X_{2,3} \end{bmatrix} \underline{u}^{(3)} \quad (\text{A}\cdot 14)$$

where all components of the $(L_1+L_2)\times B$ matrix $\left[\begin{smallmatrix} X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,1} & X_{2,2} & X_{2,3} \end{smallmatrix}\right]$ of the right-hand side of Eq. (A· 12) are belong to the set $\{0\} \cup \{x_i^{j-1} | 1 \le i \le t \text{ and } 1 \le j \le d\}$. The matrix $X_{1,1}$ is an $L_1 \times L_1$ square matrix corresponding to $\underline{u}^{(1)}$, and the matrix $X_{2,2}$ is an $L_2 \times L_2$ square matrix corresponding to $\underline{u}^{(2)}$. On the other hand, the matrix $X_{2,1}$ is an $L_2 \times L_1$ zero matrix of which all components are zeros, since each component of the vector $\underline{v}^{(2)}$ is represented by a linear combination of components of the vectors $\underline{u}^{(2)}$ and $\underline{u}^{(3)}$ that consist of entries of the submatrix M_2 of the message matrix M. The matrix $\left[\begin{smallmatrix} X_{1,3} \\ X_{2,3} \end{smallmatrix}\right]$ is an $(L_1+L_2)\times(B-L_1-L_2)$ matrix that corresponds with $\underline{u}^{(3)}$. (See Example 20)

To show the square matrix $\left[\begin{smallmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{smallmatrix}\right]$ corresponding to $[(\underline{u}^{(1)})^t, (\underline{u}^{(2)})^t]$ is nonsingular, we show that the matrix $X_{1,1}$ is nonsingular and the matrix $X_{2,2}$ is nonsingular, that is, $\det X_{1,1} \neq 0$ and $\det X_{2,2} \neq 0$.

First, we will show that $\det X_{1,1} \neq 0$. From the definitions of the vectors $[(\underline{u}^{(1)})^t, (\underline{u}^{(2)})^t]$ and $[(\underline{v}^{(1)})^t, (\underline{v}^{(2)})^t]$, the matrix $X_{1,1}$ is written as

$$X_{1,1} = \begin{bmatrix} X_1 & & \Delta \\ & \ddots & \\ \mathbf{O} & & X_k \end{bmatrix} \quad (\text{A}\cdot 15)$$

where $X_1, \ldots, X_k$ are square matrices such that

$$X_j = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{j-1} \\ 1 & x_2 & \cdots & x_2^{j-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_j & \cdots & x_j^{j-1} \end{bmatrix}, \ 1 \le j \le t-1, \quad (\text{A}\cdot 16)$$

$$X_j = \begin{bmatrix} x_1^{j-t} & x_1^{j-t+1} & \cdots & x_1^{j-1} \\ x_2^{j-t} & x_2^{j-t+1} & \cdots & x_2^{j-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_t^{j-t} & x_t^{j-t+1} & \cdots & x_t^{j-1} \end{bmatrix}, \ t \le j \le k. \quad (\text{A}\cdot 17)$$

And then, $X_1, \ldots, X_k$ are block diagonal components of $X_{1,1}$. The matrix $X_{1,1}$ is a block upper-triangular matrix, that is, entries in the strictly block lower-triangular half of the matrix equal to zero. From the structure of the matrix $X_{1,1}$, the determinant of the matrix $X_{1,1}$ is obtained as $\det X_{1,1} = \prod_{j=1}^{k} \det X_j$. For each j, $1 \le j \le k \le t-1$, $\det X_j \neq 0$ from the condition for x_i, since the determinant of the matrix X_j is the Vandermonde determinant. Similarly, for each j, $t \le j \le k$, $\det X_j \neq 0$ from the condition for x_i, since the determinant of the matrix X_j is the scalar multiple of the Vandermonde determinant, that is, $\det X_j = x_1^{j-t} \cdots x_t^{j-t} \prod_{1 \le b < a \le t} (x_a - x_b)$. Therefore, we have $\det X_{1,1} = \prod_{j=1}^{k} \det X_j \neq 0$.

Next, we will show that $\det X_{2,2} \neq 0$. In a way similar to the case of the matrix $X_{1,1}$, from the definitions of the vectors $[(\underline{u}^{(1)})^t, (\underline{u}^{(2)})^t]$ and $[(\underline{v}^{(1)})^t, (\underline{v}^{(2)})^t]$, the matrix $X_{2,2}$ is written as

$$X_{2,2} = \begin{bmatrix} X_{k+1} & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & X_d \end{bmatrix} \quad (\text{A}\cdot 18)$$

where $X_{k+1}, \ldots, X_d$ are $t \times t$ square matrices such that

$$X_j = \begin{bmatrix} x_1^{k-t} & x_1^{k-t+1} & \cdots & x_1^{k-1} \\ x_2^{k-t} & x_2^{k-t+1} & \cdots & x_2^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_t^{k-t} & x_t^{k-t+1} & \cdots & x_t^{k-1} \end{bmatrix}, \ k+1 \le j \le d. \quad (\text{A}\cdot 19)$$

And then, $X_{k+1}, \ldots, X_d$ are block diagonal components of $X_{2,2}$, that is, $X_{2,2}$ is a block diagonal matrix. From the structure of the matrix $X_{2,2}$, the determinant of the matrix $X_{2,2}$ is obtained as $\det X_{2,2} = \prod_{j=k+1}^{d} \det X_j$. For each j, $k+1 \le j \le d$, $\det X_j \neq 0$ from the condition for x_i, since the determinant of the matrix X_j is the scalar multiple of the Vandermonde determinant. Therefore, we have $\det X_{2,2} = \prod_{j=k+1}^{d} \det X_j \neq 0$.

From the above argument, we have $\det \left[\begin{smallmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{smallmatrix}\right] = \prod_{i=1}^{2} (\det X_{i,i}) \neq 0$, that is, the matrix $\left[\begin{smallmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{smallmatrix}\right]$ is nonsingular.

Finally, we show that Eq. (44) is valid by using the fact such that the matrix $\left[\begin{smallmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{smallmatrix}\right]$ is nonsingular.

Considering the L_S secret symbols and the L_R random symbols for the (n, k, d, m) secure regenerating code in the vectors $\underline{u}^{(1)}$, $\underline{u}^{(2)}$ and $\underline{u}^{(3)}$, we have the following facts for $m \le t \le k$. (See Example 21)

- All the L_R random symbols of the random vector $\underline{R}$ are components of the vectors $\underline{u}^{(1)}$ and $\underline{u}^{(2)}$ from the definitions of $\underline{u}^{(1)}$ and $\underline{u}^{(2)}$. Note that the number of all components of $\underline{u}^{(1)}$ and $\underline{u}^{(2)}$ is $L_1 + L_2$, and $L_1 + L_2 \ge L_R$ when $m \le t$. Let R denote a uniform random variable representing a random vector $\underline{R}$.
- The remaining $L_1 + L_2 - L_R$ components of $\underline{u}^{(1)}$ and $\underline{u}^{(2)}$ are the secret symbols of the secret $\underline{S}$. Let $\underline{S}_1$ be a vector consisting of the $L_1 + L_2 - L_R$ secret symbols, and then, let S_1 denote a random variable representing the vector $\underline{S}_1$.
- The remaining $B - L_1 - L_2$ secret symbols of the secure $\underline{S}$ are components of the vector $\underline{u}^{(3)}$. Let $\underline{S}_2$ be a vector consisting of the $B - L_1 - L_2$ secret symbols, and then, let S_2 denote a random variable representing the vector $\underline{S}_2$.

By using the random variables S_1, S_2, R and V, the equation (A· 12) is represented as

$$H(V|RS_1S_2) = 0. \quad (\text{A}\cdot 20)$$

Furthermore, because of the nonsingular matrix $\left[\begin{smallmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{smallmatrix}\right]$, the vector $[\underline{u}^{(1)}, \underline{u}^{(2)}]$ is uniquely determined from the vectors $[\underline{v}^{(1)}, \underline{v}^{(2)}]$ and $\underline{u}^{(3)}$ as follows:

$$\begin{bmatrix} \underline{u}^{(1)} \\ \underline{u}^{(2)} \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} \\ \mathbf{O} & X_{2,2} \end{bmatrix}^{-1} \left(\begin{bmatrix} \underline{v}^{(1)} \\ \underline{v}^{(2)} \end{bmatrix} - \begin{bmatrix} X_{1,3} \\ X_{2,3} \end{bmatrix} \underline{u}^{(3)} \right). \quad (\text{A}\cdot 21)$$

The above equation is represented as

$$H(RS_1|VS_2) = 0. \quad (\text{A}\cdot 22)$$

From Eqs. (A· 20) and (A· 22), we have

$$\begin{aligned} 0 &\le H(S_2) - H(S_2|V) = I(S_2; V) \\ &= H(V) - H(V|S_2) \\ &\overset{(a)}{=} H(V) - H(V|S_2) + H(V|RS_1S_2) \\ &= H(V) - I(V; RS_1|S_2) \\ &= H(V) - H(RS_1|S_2) + H(RS_1|VS_2) \\ &\overset{(b)}{=} H(V) - H(RS_1|S_2) \\ &= H(V) - H(S_1|S_2) - H(R|S_1S_2) \\ &\overset{(c)}{=} H(V) - H(S_1S_2) + H(S_2) - H(R) \\ &= H(V) - L_S + B - L_1 - L_2 - L_R \\ &\overset{(d)}{\le} L_1 + L_2 - L_1 - L_2 \\ &= 0, \end{aligned}$$

where equality (a) follows from Eq. (A· 20), equality (b) follows from Eq. (A· 22), equality (c) follows from $H(RS_1|S_2V) = H(R)$ because R is independent of S_1S_2, and

inequality (d) follows from $H(V) \leq L_1 + L_2$ and $B = L_S + L_R$. The above logic is a generalized version of that of Shah et al. [9, Eqs. (14)–(20)]. When $t = m$, that is, $L_1 + L_2 = L_R$, our logic is identical to that of Shah et al. as a special case. Therefore, for $m \leq t \leq k$, we have $H(S_2|V) = H(S_2)$ and

$$\begin{aligned} H(S|V) &= H(S_1 S_2|V) \\ &= H(S_2|V) + H(S_1|S_2 V) \\ &\overset{(a)}{=} H(S_2|V) = H(S_2) \\ &= B - L_1 - L_2 = \frac{1}{2}(t-k)(t-(2d-k+1)) \\ &= g(t), \end{aligned} \tag{A·23}$$

where equality (a) follows from Eq. (A·22).

1. In the case of $0 \leq t \leq m$: i) For $t = m$, we have $H(S|V) = g(m) = L_S = H(S)$ from Eq. (A·23). ii) For $0 \leq t \leq m-1$, we have $H(S|V) \geq H(S)$ from the above result i). Note that V is characterized by the parameter t in Eq. (41). On the other hand, in general, we have $H(S|V) \leq H(S)$. Hence, $H(S|V) = H(S)$. Thus, from both cases of i) and ii), we have $H(S|V) = H(S) = L_S$ for $0 \leq t \leq m$.
2. In the case of $m+1 \leq t \leq k-1$: We have $H(S|V) = g(t)$ for $m+1 \leq t \leq k-1$ from Eq. (A·23).
3. In the case of $k \leq t \leq n$: i) For $t = k$, we have $H(S|V) = g(k) = 0$ from Eq. (A·23). ii) For $k+1 \leq t \leq n$, we have $H(S|V) \leq 0$ from the above result i). On the other hand, in general, we have $H(S|V) \geq 0$. Hence, $H(S|V) = 0$. Thus, from both cases of i) and ii), we have $H(S|V) = 0$ for $k \leq t \leq n$.

This completes the proof of Theorem 15. □

Example 20: (Example 12 continued) Let $(k, d, t) = (4, 6, 3)$. We have then the vectors and the matrices in Eq. (A·13) as follows:

$$\begin{bmatrix} \underline{v}^{(1)} \\ \hline \underline{v}^{(2)} \end{bmatrix} = \left[\begin{array}{c} c_{1,1} \\ c_{1,2} \\ c_{2,2} \\ c_{1,3} \\ c_{2,3} \\ c_{3,3} \\ c_{1,4} \\ c_{2,4} \\ \hline c_{3,4} \\ c_{1,5} \\ c_{2,5} \\ c_{3,5} \\ c_{1,6} \\ c_{2,6} \\ c_{3,6} \end{array}\right], \quad \begin{bmatrix} \underline{u}^{(1)} \\ \hline \underline{u}^{(2)} \end{bmatrix} = \left[\begin{array}{c} u_{1,1} \\ u_{1,2} \\ u_{2,2} \\ u_{1,3} \\ u_{2,3} \\ u_{3,3} \\ u_{2,4} \\ u_{3,4} \\ \hline u_{4,4} \\ u_{2,5} \\ u_{3,5} \\ u_{4,5} \\ u_{2,6} \\ u_{3,6} \\ u_{4,6} \end{array}\right], \quad \underline{u}^{(3)} = \begin{bmatrix} u_{1,4} \\ u_{1,5} \\ u_{1,6} \end{bmatrix},$$

$$\left[\begin{array}{c|c} X_{1,1} & X_{1,2} \\ \hline \mathbf{O} & X_{2,2} \end{array}\right]$$

$$= \left[\begin{array}{cccccccc|cccccc} 1 & x_1 & 0 & x_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & x_1 & 0 & x_1^2 & 0 & x_1^3 & 0 & x_1^4 & 0 & 0 & x_1^5 & 0 & 0 \\ 0 & 1 & x_2 & 0 & x_2^2 & 0 & x_2^3 & 0 & x_2^4 & 0 & 0 & x_2^5 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_1 & x_1^2 & 0 & x_1^3 & 0 & x_1^4 & 0 & 0 & x_1^5 & 0 \\ 0 & 0 & 0 & 1 & x_2 & x_2^2 & 0 & x_2^3 & 0 & x_2^4 & 0 & 0 & x_2^5 & 0 \\ 0 & 0 & 0 & 1 & x_3 & x_3^2 & 0 & x_3^3 & 0 & x_3^4 & 0 & 0 & x_3^5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_1 & x_1^2 & x_1^3 & 0 & 0 & x_1^4 & 0 & 0 & x_1^5 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_2 & x_2^2 & x_2^3 & 0 & 0 & x_2^4 & 0 & 0 & x_2^5 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_3 & x_3^2 & x_3^3 & 0 & 0 & x_3^4 & 0 & 0 & x_3^5 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_1 & x_1^2 & x_1^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_2 & x_2^2 & x_2^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_3 & x_3^2 & x_3^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_1 & x_1^2 & x_1^3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_2 & x_2^2 & x_2^3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_3 & x_3^2 & x_3^3 \end{array}\right],$$

$$\begin{bmatrix} X_{1,3} \\ \hline X_{2,3} \end{bmatrix} = \left[\begin{array}{ccc} x_1^3 & x_1^4 & x_1^5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\right].$$

Example 21: (Examples 6 and 11 continued) Let $(k, d, m, t) = (4, 6, 2, 3)$. We have $B = 18, L_S = 7, L_R = 11, L_1 = 9$ and $L_2 = 6$. We consider the relation between the seven secret symbols $s_1, \ldots, s_6$ and the 11 random symbols $r_1, \ldots, r_{11}$ and the vectors $\underline{u}^{(1)}$, $\underline{u}^{(2)}$ and $\underline{u}^{(3)}$. The vectors $\underline{u}_j^{(1)}$, $\underline{u}_j^{(2)}$ and $\underline{u}_j^{(3)}$ are as follows:

$$\begin{aligned} &\underline{u}_1^{(1)} = [u_{1,1}] = [r_1],\ \underline{u}_2^{(1)} = [u_{1,2}, u_{2,2}]^t = [r_2, r_3]^t, \\ &\underline{u}_3^{(1)} = [u_{1,3}, u_{2,3}, u_{3,3}]^t = [s_1, r_4, r_5]^t, \\ &\underline{u}_4^{(3)} = [u_{1,4}] = [s_2],\ \underline{u}_4^{(1)} = [u_{2,4}, u_{3,4}, u_{4,4}]^t = [s_3, r_6, r_7]^t, \\ &\underline{u}_5^{(3)} = [u_{1,5}] = [s_4],\ \underline{u}_5^{(2)} = [u_{2,5}, u_{3,5}, u_{4,5}]^t = [s_5, r_8, r_9]^t, \\ &\underline{u}_6^{(3)} = [u_{1,6}] = [s_6],\ \underline{u}_6^{(2)} = [u_{2,6}, u_{3,6}, u_{4,6}]^t = [s_7, r_{10}, r_{11}]^t, \end{aligned}$$

and

$$\begin{aligned} &\underline{u}^{(1)} = [r_1, r_2, r_3, s_1, r_4, r_5, s_3, r_6, r_7]^t \in \mathbb{F}_q^9, \\ &\underline{u}^{(2)} = [s_5, r_8, r_9, s_7, r_{10}, r_{11}]^t \in \mathbb{F}_q^6, \\ &\underline{u}^{(3)} = [s_2, s_4, s_6]^t \in \mathbb{F}_q^3. \end{aligned}$$

Furthermore, the four secret symbols s_1, s_3, s_5, s_7 are components of the vector $\underline{S}_1$ and the remaining three secret symbols s_2, s_4, s_6 are components of the vector $\underline{S}_2$.

Masazumi Kurihara received the B.S. degree in Mathematics from Tokyo Metropolitan University in 1990, and the M.E. and Ph.D. degrees in Computer Science and Information Mathematics from the University of Electro-Communications (UEC) in 1992 and 2002, respectively. From 1992 to 2007, he was a research associate in the Faculty of Electro-Communications at UEC. From 2007 to 2010, he was an assistant professor in the Faculty of Electro-Communications at UEC. Since 2010, he was an assistant professor in the Graduate School of Informatics and Engineering. His research interests are in algebraic coding theory.

Hidenori Kuwakado received the B.E., M.E. and D.E. degrees from Kobe University in 1990, 1992, and 1999 respectively. He worked for Nippon Telegraph and Telephone Corporation from 1992 to 1996. From 1996 to 2002 he was a Research Associate in the Faculty of Engineering, Kobe University. From 2002 to 2007, he was an Associate Professor in the Faculty of Engineering, Kobe University. Since 2007, he has been an Associate Professor in Graduate School of Engineering, Kobe University. His research interests are in cryptography and information security.