Weakly Secure Symmetric Multilevel Diversity Coding

Tao Guo, Chao Tian, Tie Liu
Dept. of Electrical and Computer Engineering
Texas A&M University
{guotao,chao.tian,tieliu}@tamu.edu

Raymond W. Yeung
Dept. of Information Engineering
The Chinese University of Hong Kong
whyeung@ie.cuhk.edu.hk

***Abstract*—Multilevel diversity coding is a classical coding model where multiple mutually independent information messages are encoded, such that different reliability requirements can be afforded to different messages. It is well known that *superposition coding*, namely separately encoding the independent messages, is optimal for symmetric multilevel diversity coding (SMDC) (Yeung-Zhang 1999). In the current paper, we consider weakly secure SMDC where secrecy constraints are injected on each individual message, and provide a complete characterization of the conditions under which superposition coding is sum-rate optimal. Two joint coding strategies, which lead to rate savings compared to superposition coding, are proposed, where some coding components for one message can be used as the encryption key for another. By applying different variants of Han's inequality, we show that the lack of opportunities to apply these two coding strategies directly implies the optimality of superposition coding. It is further shown that under a particular security configuration, one of the proposed joint coding strategies can be used to achieve the optimal sum rate.**

I. Introduction

Symmetric multilevel diversity coding (SMDC) was introduced by Roche *et al.* [1] and Albanese *et al.* [2] for applications in distributed data storage and robust network communication. In a symmetric L-level diversity coding system, there are L independent messages $(M_1, M_2, \ldots, M_L)$, where the importance of messages decreases with the subscript l. The messages are encoded by L encoders. There are totally $2^L - 1$ decoders, each of which has access to the outputs of a distinct subset of the encoders. A decoder which can access any α encoders, called a Level-α decoder, is required to reconstruct the first α messages. The system is symmetric in the sense that the reconstruction requirements depend on the set of encoders only via its cardinality. It was shown that separately encoding these independent messages, referred to as *superposition coding*, is optimal in the sense that it can achieve not only the minimum sum rate [1], [2] but in fact the entire rate region [1], [3]. Later efforts extending and generalizing this result can be found in, for example, [4]–[8].

In this paper we consider a *weakly* secure setting of the classical SMDC problem, where the security level of each message is specified by a separate security parameter N_α. More specifically, for any $\alpha = 1, 2, \ldots, L$, we require the message M_α to be kept perfectly secure if the outputs of no more than N_α encoders are accessible by an eavesdropper. Such a security requirement is "weak" in the sense that the eavesdropper is only prevented from obtaining any information about the *individual* messages. By comparison, the security requirement of [5], [6] is strong in that it prevents the eavesdropper to obtain any information about the *entire* set of messages. The notion of weak security has been considered in various network coding settings [9]–[12] and multi-access wiretap channel settings [13] in the literature and is generally considered to be more practical for protecting individual messages regardless of their relations.

Note that on the one hand, the notion of weak security has significantly enriched the collection of secure SMDC problems: Unlike the strongly secure setting where it is necessary to consider a single security parameter for all messages, for the weakly secure setting each message can be associated with a different security parameter. On the other hand, the notion of weak security has also cast the optimality of superposition coding in much greater doubt, as asking the messages only to be protected *marginally* (instead of jointly) also significantly opens up the set of feasible coding strategies. Our main goal for this paper is: 1) to understand under what configurations of the security parameters $(N_1, N_2, \ldots, N_L)$ superposition coding remains to be optimal; and 2) to identify optimal coding strategies when superposition coding is suboptimal.

The main result of this paper is a precise classification of the cases, in terms of the parameter $(N_1, N_2, \ldots, N_L)$, where superposition is sum-rate optimal, i.e., a set of necessary and sufficient conditions on $(N_1, N_2, \ldots, N_L)$ for superposition to be sum-rate optimal. Two joint coding strategies, which are built on utilizing some coding components for one message as the encryption key for another and lead to rate savings compared to superposition coding, are proposed. Then we show that lacking the opportunity to utilize either of the two joint coding strategies implies that superposition coding is in fact optimal. We further provide a special security configuration where the proposed joint coding strategies directly lead to the optimal sum-rate. The essential proofs can be found in [14].

II. Problem Formulation and Preliminaries

A. Problem Formulation

Let $\mathcal{L} \triangleq \{1, 2, \cdots, L\}$, where $L \geq 2$. Let $M_1, M_2, \cdots, M_L$ be a collection of L mutually independent messages uniformly distributed over the direct product of certain finite sets. For

The work of C. Tian was supported in part by the National Science Foundation under Grants CCF-18-32309 and CCF-18-16546.

978-1-5386-6900-6/19/$31.00 ©2019 IEEE

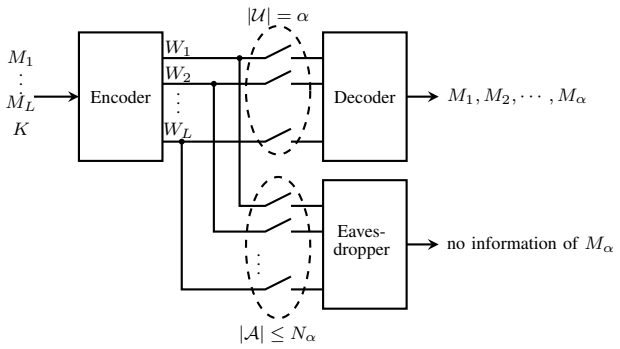

Fig. 1: The Weakly Secure SMDC Model

simplicity, we assume the message set to be $\mathbb{F}_{p^{m_1}} \times \mathbb{F}_{p^{m_2}} \times \cdots \times \mathbb{F}_{p^{m_L}}$, where $\mathbb{F}_{p^{m_1}}$ is a finite field of order p^{m_1} and p is an integer power of some prime number. We may also regard M_α ($\alpha \in \mathcal{L}$) as $M_\alpha = (M_\alpha^1, M_\alpha^2, \cdots, M_\alpha^{m_\alpha})$ where $M_\alpha^i \in \mathbb{F}_p$ for $i = 1, 2, \cdots, m_\alpha$.

The weakly secure SMDC problem is depicted in Fig. 1. There are L encoders, indexed by $\mathcal{L}$, each of which can access all the L information messages. There are also $2^L - 1$ decoders. For each $\mathcal{U} \subseteq \mathcal{L}$ such that $\mathcal{U} \neq \emptyset$, Decoder-$\mathcal{U}$ can access the outputs of the subset of encoders indexed by $\mathcal{U}$. For $\alpha \in \mathcal{L}$ and any $\mathcal{U}$ such that $|\mathcal{U}| = \alpha$, Decoder-$\mathcal{U}$ can completely recover the first α messages $M_1, M_2, \cdots, M_\alpha$. In addition, there is an eavesdropper who has access to the outputs of a subsets $\mathcal{A}$ of encoders. Let $\boldsymbol{N} = (N_1, N_2, \cdots, N_L)$ be L non-negative integers, where $N_\alpha < \alpha$ for $\alpha \in \mathcal{L}$. Weak security requires that each individual message M_α should be kept perfectly secure from the eavesdropper if $|\mathcal{A}| \le N_\alpha$.

Formally, an $(m_1, m_2, \cdots, m_L, R_1, R_2, \cdots, R_L)$ code is defined by the encoding functions

$$E_l : \prod_{i=1}^{L} \mathbb{F}_{p^{m_i}} \times \mathcal{K} \to \mathbb{F}_{p^{R_l}}, \text{ for } l \in \mathcal{L} \tag{1}$$

and decoding functions

$$D_\mathcal{U} : \prod_{l \in \mathcal{U}} \mathbb{F}_{p^{R_l}} \to \prod_{i=1}^{|\mathcal{U}|} \mathbb{F}_{p^{m_i}}, \text{ for } \mathcal{U} \subseteq \mathcal{L} \text{ and } \mathcal{U} \neq \emptyset. \tag{2}$$

Denote the shared key as K (accessible to all the encoders), which is uniformly distributed in the key space $\mathcal{K}$. Let $W_l = E_l(M_1, M_2, \cdots, M_L, K)$ be the output of Encoder-l and $W_\mathcal{U} = (W_i : i \in \mathcal{U})$ for $\mathcal{U} \subseteq \mathcal{L}$. Define the normalized message rates $\mathsf{m}_l \triangleq m_l / \sum_{l=1}^{L} m_l$, from which it follows that $\sum_l \mathsf{m}_l = 1$. A normalized non-negative rate tuple $(\mathsf{R}_1, \mathsf{R}_2, \cdots, \mathsf{R}_L)$ is *achievable* for the normalized message rates $(\mathsf{m}_1, \ldots, \mathsf{m}_L)$, if for any $\epsilon > 0$, there exist an integer a and an $(a\mathsf{m}_1, a\mathsf{m}_2, \cdots, a\mathsf{m}_L, R_1, R_2, \cdots, R_L)$ code such that

perfect reconstruction: $D_\mathcal{U}(W_\mathcal{U}) = (M_1, M_2, \cdots, M_\alpha)$,
$$\forall \alpha \in \mathcal{L} \text{ and } \mathcal{U} \subseteq \mathcal{L} \text{ s.t. } \mathcal{U} \neq \emptyset, \tag{3}$$

perfect secure: $H(M_\alpha | W_\mathcal{A}) = H(M_\alpha)$,
$$\forall \alpha \in \mathcal{L} \text{ and } \mathcal{A} \subseteq \mathcal{L} \text{ s.t. } |\mathcal{A}| \le N_\alpha, \tag{4}$$

and

$$\textbf{coding rate: } \mathsf{R}_l + \epsilon \ge a^{-1} R_l, \qquad l \in \mathcal{L}. \tag{5}$$

Remark 1. Here each message M_α can essentially be represented in $m_\alpha \log p$ bits, and each codeword W_l can be represented in $R_l \log p$ bits. Thus R_l can be viewed as the coding rate of encoder E_l when the definition of the entropy function uses logarithm in the base p, which will be assumed hereafter. The quantity R_l is then essentially the normalized R_l.

In this work we focus on the minimum achievable normalized sum rate $\mathsf{R}_{\text{sum}}^* \triangleq \min \sum_{i=1}^{L} \mathsf{R}_i$, and our main result is a necessary and sufficient condition for superposition coding to be sum-rate optimal.

B. An Achievable Sum Rate via Superposition Coding

Let M be a message encoded by n encoders. For any $0 \le c < k \le n$, the (c, k, n) ramp secret sharing problem [15], also known as the secure symmetrical single-level diversity coding (S-SSDC) problem in [5], requires that the outputs from any subset of no more than c encoders provide no information about the message, and the outputs from any subset of k encoders can completely recover the message. The minimum sum rate for this problem can be found in [5], [16], as stated in the following lemma.

Lemma 1. *The minimum sum rate of the* (c, k, n) *ramp secret sharing is* $\frac{n}{k-c} H(M)$.

In light of this result, a natural coding scheme (i.e., superposition coding) for the weakly secure SMDC problem formulated above is to separately encode each message M_α using an (N_α, α, L) ramp secret sharing code. The sum rate induced by superposition coding provides an upper bound $\bar{\mathsf{R}}_{\text{sum}}$ for $\mathsf{R}_{\text{sum}}^*$, and by Lemma 1, it can be written simply as,

$$\bar{\mathsf{R}}_{\text{sum}} \triangleq \sum_{\alpha=1}^{L} \frac{L\mathsf{m}_\alpha}{\alpha - N_\alpha}. \tag{6}$$

C. Properties of MDS Code for Secret Sharing

In this section, we describe in some details two (n, k) maximum distance separable (MDS) codes for ramp secret sharing that achieve the minimum sum rate in Lemma 1, and provide important properties that are instrumental to the joint coding strategy we later propose.

Let $M = (U_1, U_2, \cdots, U_{k-c})$ be a length-$(k-c)$ message where each symbol is chosen uniformly and independently from the finite field $\mathbb{F}_p$. Let $Z_1, Z_2, \cdots, Z_c$ be independent random keys chosen uniformly from the same finite field $\mathbb{F}_p$. For $i = 1, 2, \cdots, k$, define the following length-k vectors:

$$f_i = [\underbrace{0 \ \cdots \ 0}_{i-1} \ 1 \ 0 \ \cdots \ 0]^T. \tag{7}$$

Let $g_1, g_2, \cdots, g_n$ be length-k vectors such that any k vectors $\{h_{j_1}, h_{j_2}, \cdots, h_{j_k}\}$ chosen from the set $\{f_1, f_2, \cdots, f_k, g_1, g_2, \cdots, g_n\}$ satisfy the full rank condition

$$\text{rank}\,[h_{j_1} \ h_{j_2} \ \cdots \ h_{j_k}] = k. \tag{8}$$

It can be shown that as long as $p \ge n + k$, there exist such vectors $g_1, g_2, \cdots, g_n$, e.g., it can be chosen as the columns

from a Cauchy matrix. The generator matrices of the two MDS codes of interest are given, respectively, as

$$G^{(1)} = [f_{k-c+1} \ \cdots \ f_k \ g_1 \ g_2 \ \cdots \ g_{n-c}], \quad (9)$$

$$G^{(2)} = [g_1 \ g_2 \ \cdots \ g_n]. \quad (10)$$

Then the codewords of two MDS codes are, respectively,

$$[Y_1, Y_2, \cdots, Y_n] = \begin{bmatrix} M_1 & \cdots & M_{k-c} & Z_1 & \cdots & Z_c \end{bmatrix} G^{(1)}, \quad (11)$$

$$[Y_1, Y_2, \cdots, Y_n] = \begin{bmatrix} M_1 & \cdots & M_{k-c} & Z_1 & \cdots & Z_c \end{bmatrix} G^{(2)}. \quad (12)$$

We shall refer these two codes as MDS-A and MDS-B, respectively. By the definition of $f_{k-c+1}, \cdots, f_k$ in (7), MDS-A has the random keys explicitly as part of the coded message,

$$[Y_1, Y_2, \cdots, Y_c] = \begin{bmatrix} Z_1 & Z_2 & \cdots & Z_c \end{bmatrix}. \quad (13)$$

It is obvious that for both codes, M and $Z_1, Z_2, \cdots, Z_c$ can be perfectly recovered from any k coded symbols.

Since all the coded symbols are linear combinations of the messages and the random keys that are uniformly distributed, we have the following lemma.

Lemma 2. *Any k coded symbols of MDS-A and MDS-B are uniformly distributed over $\mathbb{F}_{p^k}$.*

The main difference between the two codes, which is the most relevant to this work, is given in the following two lemmas. The proofs follow directly from (8) and the proofs are omitted due to space constraint.

Lemma 3. *For any integer t such that $c \le t \le k$, let $\mathcal{E} \subseteq \{1, 2, \cdots, n\}$ where $|\mathcal{E}| = t$, and $\mathcal{A} \subseteq \{1, 2, \cdots, k-c\}$ where $|\mathcal{A}| = k - t$. The codewords of MDS-A have the following property:*

$$I(Y_{\mathcal{E}}; M_{\mathcal{A}}) = 0, \quad (14)$$

where $Y_{\mathcal{E}} \triangleq \{Y_i : i \in \mathcal{E}\}$ and $M_{\mathcal{A}} \triangleq \{M_i : i \in \mathcal{A}\}$.

Remark 2. For $t = c$, Lemma 3 reduces to the stated security constraint of parameter c; on the other hand, for $t > c$ (but $t \le k$), any t coded symbols reveal no information about any subset of $k - t$ message symbols.

Lemma 4. *For any integer t such that $0 \le t \le k$, let $\mathcal{E} \subseteq \{1, 2, \cdots, n\}$ where $|\mathcal{E}| = t$, and $\mathcal{A}_1 \subseteq \{1, 2, \cdots, k-c\}$, and $\mathcal{A}_2 \subseteq \{1, 2, \cdots, c\}$ where $|\mathcal{A}_1| + |\mathcal{A}_2| = k - t$. The codewords of MDS-B have the following property:*

$$I(Y_{\mathcal{E}}; M_{\mathcal{A}_1}, Z_{\mathcal{A}_2}) = 0, \quad (15)$$

where $Y_{\mathcal{E}} \triangleq \{Y_i : i \in \mathcal{E}\}$, $M_{\mathcal{A}_1} \triangleq \{M_i : i \in \mathcal{A}_1\}$, and $Z_{\mathcal{A}_2} \triangleq \{Z_i : i \in \mathcal{A}_2\}$.

In contrast to MDS-A, MDS-B has the additional advantage that part of the keys can also be made secure against some t eavesdroppers, at the expense of exposing some message symbols. This property becomes important to us in the sequel.

III. Main Results

The main question we seek to answer here is under what condition the equality $\mathsf{R}^*_{\text{sum}} = \bar{\mathsf{R}}_{\text{sum}}$ will hold, and the following theorem provides the exact answer to this question.

Theorem 1. *$\mathsf{R}^*_{sum} = \bar{\mathsf{R}}_{sum}$, if and only if for any $\alpha < \beta \in \mathcal{L}$ where $\mathsf{m}_\alpha > 0$ and $\mathsf{m}_\beta > 0$, we have*

$$\text{either } N_\alpha < \alpha \le N_\beta < \beta, \text{ or } N_\alpha = N_\beta = 0. \quad (16)$$

Remark 3. If the security constraints are given as

$$N_\alpha = 0, \text{ for all } \alpha \in \mathcal{L}, \quad (17)$$

then the problem reduces to the classical SMDC problem without security constraints, where superposition is known to be optimal [3], and Theorem 1 reduces correctly for this case.

Theorem 1 can be alternatively written in the following form, by taking the complement of the conditions in (16).

Theorem 1'. *$\mathsf{R}^*_{sum} < \bar{\mathsf{R}}_{sum}$, if and only if there exist $\alpha < \beta \in \mathcal{L}$ where $\mathsf{m}_\alpha > 0$ and $\mathsf{m}_\beta > 0$ such that*

$$\text{either } (N_\alpha < N_\beta < \alpha), \text{ or } (N_\beta \le N_\alpha \ \& \ N_\alpha > 0). \quad (18)$$

We prove Theorem 1 in two parts. In Section IV, we show that superposition is suboptimal under the security constraints in (18), by providing joint coding strategies that can reduce the coding rates. In Section V, the optimality of superposition coding is established by a matching sum rate lower bound.

When superposition is not optimal, the characterization of the minimum sum rate remains open in general. Our next result is that for a special security configuration, a joint coding strategy proposed in Section IV can be used to build an optimal coding scheme. Consider the case where $\boldsymbol{N}$ is given by

$$N_\alpha = \begin{cases} \alpha - 1, & \text{for } 1 \le \alpha \le r \\ 0, & \text{for } r+1 \le \alpha \le L, \end{cases} \quad (19)$$

for certain parameters (L, r). We refer to this system as the (L, r) differential-constant secure SMDC (DS-SMDC), where the more important messages (i.e., small α values) are maximally protected ($N_\alpha = \alpha - 1$) and the less important messages are not protected at all ($N_\alpha = 0$). For $r = 1$, the problem reduces to the classical SMDC.

Let M_{L+1} be an independent message uniformly distributed over $\mathbb{F}_{p^{m_{L+1}}}$ with

$$m_{L+1} = \left[\sum_{\alpha=1}^{r} (\alpha - 1) m_\alpha - \sum_{\alpha=r+1}^{L} m_\alpha \right]^+, \quad (20)$$

where for any $x \in \mathbb{R}$, $[x]^+ \triangleq \max(0, x)$. Let $\eta^* \in \{r+1, r+2, \cdots, L+1\}$ be the unique integer such that

$$\sum_{\alpha=r+1}^{\eta^*-1} m_\alpha < \sum_{\alpha=1}^{r} (\alpha - 1) m_\alpha \le \sum_{\alpha=r+1}^{\eta^*} m_\alpha. \quad (21)$$

Our main result on DS-SMDC is the following theorem, the proof of which is omitted due to space constraint.

Theorem 2. *The minimum sum rate of DS-SMDC is*

$$\mathsf{R}^*_{sum} = \sum_{\alpha=1}^{r} \left(1 - \frac{\alpha - 1}{\eta^*}\right) L\mathsf{m}_\alpha + \sum_{\alpha=r+1}^{\eta^*} \frac{L\mathsf{m}_\alpha}{\eta^*} + \sum_{\alpha=\eta^*+1}^{L} \frac{L\mathsf{m}_\alpha}{\alpha}. \quad (22)$$

IV. Achievability of Theorem 1: Joint Coding Strategies

In order to prove the necessity direction of Theorem 1, we instead prove the sufficiency direction of Theorem 1', in the two separate cases given in (18).

A. Low Security Level at Higher Diversity Level

In this section, we provide a joint coding strategy for the case $N_\alpha < N_\beta < \alpha$ which provides rate saving, compared to superposition coding. We first discuss a motivating example to illustrate the key insight on how such rate saving is obtained.

Example 1. Let $L = 3, (\alpha, \beta) = (2, 3), (m_2, m_3) = (2, 2), (N_2, N_3) = (0, 1)$, and $p = 5$. Let Z_3 be an independent random key uniformly chosen from $\mathbb{F}_p$. Let the two messages be encoded with generator matrices constructed using MDS-A, which induce the coded symbols as shown in Table I(a) through superposition. The important insight is that the coded message of M_2 can be used as the secret key to encode M_3, which reduces the coding rate. More precisely, we replace Z_3 by $Y_2^1 = Z_2^1 + Z_2^2$ to serve as the key for M_3, where Y_2^1 denotes the code symbol for M_α and Encoder-1 under the superposition strategy. The coded symbols for this joint coding strategy are shown in Table I(b). By comparing the two tables, it is seen that the sum rate is reduced since the coded symbol Z_3 is eliminated.

TABLE I: Coding strategy for Example 1

	W_1	W_2	W_3
$\alpha = 2$	$Y_2^1 = M_2^1 + M_2^2$	$2M_2^1 + M_2^2$	$M_2^1 + 2M_2^2$
$\beta = 3$	Z_3	$M_3^1 + 2M_3^2 + Z_3$	$2M_3^1 + M_3^2 + Z_3$

(a) Superposition coding strategy

	W_1	W_2	W_3
$\alpha = 2$	$Y_2^1 = M_2^1 + M_2^2$	$2M_2^1 + M_2^2$	$M_2^1 + 2M_2^2$
$\beta = 3$		$M_3^1 + 2M_3^2 + Y_2^1$	$2M_3^1 + M_3^2 + Y_2^1$

(b) Joint coding strategy

The reconstruction requirements of both M_2 and M_3 are straightforward. There is no security requirement on M_2. For M_3, it is seen that any one coded symbol W_l reveals no information about M_3. For instance, eavesdropping W_2 gives

$$H(M_3 | M_3^1 + 2M_3^2 + Y_2^1, M_2^1 + Y_2^1) = H(M_3). \tag{23}$$

Coding strategy for general parameters:

First encode separately M_α and M_β with generator matrices G_α and G_β using MDS-A in Section II-C. The coded symbols for superposition coding strategy are as given in Table II(a). The joint coding strategy we propose is then to replace the first $\theta = \min\{N_\beta, \alpha - N_\beta\}$ encryption key symbols $(Z_\beta^1\ Z_\beta^2\ \cdots\ Z_\beta^\theta)$ using the coded symbols $(Y_\alpha^1, Y_\alpha^2, \cdots, Y_\alpha^\theta)$. Denote the corresponding codewords for M_β thus obtained as $(Y_\beta^{1*}, Y_\beta^{2*}, \cdots, Y_\beta^{L*})$. The joint coding strategy of M_α and M_β is illustrated in Table II(b) and can be described as follows:

$$W_i = \begin{cases} Y_\alpha^i, & \text{for } 1 \le i \le \theta \\ [Y_\alpha^i, Y_\beta^{i*}], & \text{for } \theta < i \le L. \end{cases} \tag{24}$$

By comparing Table II(a) and Table II(b), it can be seen that the coding rate is reduced compared to superposition coding because $(Y_\beta^1, Y_\beta^2, \cdots, Y_\beta^\theta)$ are removed from the codeword, while the rates for all the others are kept the same as before. Next, we verify the reconstruction and security constraints for the two messages.

TABLE II: Coding strategy to replace encryption keys for M_β

	W_1	W_2	$\cdots$	W_θ	$W_{\theta+1}$	$\cdots$	W_L
α	Y_α^1	Y_α^2	$\cdots$	Y_α^θ	$Y_\alpha^{\theta+1}$	$\cdots$	Y_α^L
β	Y_β^1	Y_β^2	$\cdots$	Y_β^θ	$Y_\beta^{\theta+1}$	$\cdots$	Y_β^L

(a) Superposition coding strategy

	W_1	W_2	$\cdots$	W_θ	$W_{\theta+1}$	$\cdots$	W_L
α	Y_α^1	Y_α^2	$\cdots$	Y_α^θ	$Y_\alpha^{\theta+1}$	$\cdots$	Y_α^L
β					$Y_\beta^{(\theta+1)*}$	$\cdots$	Y_β^{L*}

(b) Joint coding strategy

Reconstruction: The reconstruction requirements of both M_α and M_β are straightforward.

Security: We consider the security requirements for the two levels separately.

1) Assume we can access N_α coded symbols $W_\mathcal{B}, |\mathcal{B}| = N_\alpha$. Partition $\mathcal{B}$ into $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B}_1 \subseteq \{1, 2, \cdots, \theta\}$ and $\mathcal{B}_2 \subseteq \{\theta + 1, \cdots, L\}$. Notice
$$\begin{aligned} H(Y_\beta^{*\mathcal{B}_2} | M_\alpha, Y_\alpha^{\mathcal{B}_1} Y_\alpha^{\mathcal{B}_2}) &= H(Y_\beta^{*\mathcal{B}_2} | M_\alpha, Y_\alpha^{1:\theta}, Y_\alpha^{\mathcal{B}_2}) \\ &= H(Y_\beta^{*\mathcal{B}_2} | Y_\alpha^{1:\theta}) = H(Y_\beta^{*\mathcal{B}_2}), \end{aligned} \tag{25}$$
where the second equality follows from the Markov chain $Y_\beta^{*\mathcal{B}_2} \leftrightarrow Y_\alpha^{1:\theta} \leftrightarrow (M_\alpha, Y_\alpha^{\mathcal{B}_2})$, and the last equality follows from Lemma 2 because
$$|\mathcal{B}_2| + \theta \le N_\alpha + \theta \tag{26}$$
$$= N_\alpha + \min\{\alpha - N_\beta, N_\beta\} \le \alpha < \beta, \tag{27}$$
where the second inequality follows from $N_\beta < N_\alpha$. It follows that
$$\begin{aligned} &I(W_\mathcal{B}; M_\alpha) = I(W_{\mathcal{B}_1} W_{\mathcal{B}_2}; M_\alpha) \\ &= I(Y_\alpha^{\mathcal{B}_1}\ Y_\alpha^{\mathcal{B}_2} Y_\beta^{*\mathcal{B}_2}; M_\alpha) && (28) \\ &= I(Y_\alpha^{\mathcal{B}_1} Y_\alpha^{\mathcal{B}_2}; M_\alpha) + I(Y_\beta^{*\mathcal{B}_2}; M_\alpha | Y_\alpha^{\mathcal{B}_1} Y_\alpha^{\mathcal{B}_2}) && (29) \\ &= I(Y_\beta^{*\mathcal{B}_2}; M_\alpha | Y_\alpha^{\mathcal{B}_1} Y_\alpha^{\mathcal{B}_2}) = 0, && (30) \end{aligned}$$
where the last but one equality follows from Lemma 3 and the fact that $|\mathcal{B}_1| + |\mathcal{B}_2| = N_\alpha$, and (30) follows from (25). Thus indeed $W_\mathcal{B}$ reveals nothing about M_α.
2) Assume we can access N_β coded symbols $W_\mathcal{B}, |\mathcal{B}| = N_\beta$. Partition $\mathcal{B}$ into $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B}_1 \subseteq \{1, 2, \cdots, \theta\}$ and $\mathcal{B}_2 \subseteq \{\theta + 1, \cdots, L\}$. We first consider
$$\begin{aligned} &H(Y_\alpha^{\mathcal{B}_2} | Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}) \\ &\ge H(Y_\alpha^{\mathcal{B}_2} | Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2} M_\beta) && (31) \\ &\ge H(Y_\alpha^{\mathcal{B}_2} | Y_\alpha^1 \cdots Y_\alpha^\theta, Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta Y_\beta^{*\mathcal{B}_2}) && (32) \\ &= H(Y_\alpha^{\mathcal{B}_2} | Y_\alpha^1 \cdots Y_\alpha^\theta, Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta) && (33) \\ &= H(Y_\alpha^{\mathcal{B}_2}, Y_\alpha^1 \cdots Y_\alpha^\theta | Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta) \\ &\quad - H(Y_\alpha^1 \cdots Y_\alpha^\theta | Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta) && (34) \\ &= H(Y_\alpha^{\mathcal{B}_2}, Y_\alpha^1 \cdots Y_\alpha^\theta) - H(Y_\alpha^1 \cdots Y_\alpha^\theta) && (35) \\ &= H(Y_\alpha^{\mathcal{B}_2}), && (36) \end{aligned}$$
where both (31) and (32) follow from the fact that conditioning does not increase entropy, (33) follows from that

$Y_\beta^{*\mathcal{B}_2}$ is a function of $(Y_\alpha^1 \cdots Y_\alpha^\theta, Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta)$, (35) follows from that $(Z_\beta^{\theta+1} \cdots Z_\beta^{N_\beta}, M_\beta)$ are independent of $(Y_\alpha^{\mathcal{B}_2}, Y_\alpha^1 \cdots Y_\alpha^\theta)$, and the last equality follows from Lemma 2, because $|\mathcal{B}_2| + \theta \le \alpha$ which is induced from $\theta \le \alpha - N_\beta$. Since conditioning does not increase entropy, in light of (36), we obtain

$$\begin{aligned} &H(Y_\alpha^{\mathcal{B}_2}|Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2} M_\beta) \\ &= H(Y_\alpha^{\mathcal{B}_2}|Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}) = H(Y_\alpha^{\mathcal{B}_2}). \end{aligned} \tag{37}$$

Then we have

$$I(W_\mathcal{B}; M_\beta) = I(W_{\mathcal{B}_1} W_{\mathcal{B}_2}; M_\beta) \tag{38}$$
$$= I(Y_\alpha^{\mathcal{B}_1}\ Y_\alpha^{\mathcal{B}_2} Y_\beta^{*\mathcal{B}_2}; M_\beta) \tag{39}$$
$$= I(Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}; M_\beta) + I(Y_\alpha^{\mathcal{B}_2}; M_\beta | Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}) \tag{40}$$
$$= I(Y_\alpha^{\mathcal{B}_2}; M_\beta | Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}) \tag{41}$$
$$= H(Y_\alpha^{\mathcal{B}_2}|Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2}) - H(Y_\alpha^{\mathcal{B}_2}|Y_\alpha^{\mathcal{B}_1} Y_\beta^{*\mathcal{B}_2} M_\beta) \tag{42}$$
$$= H(Y_\alpha^{\mathcal{B}_2}) - H(Y_\alpha^{\mathcal{B}_2}) \tag{43}$$
$$= 0, \tag{44}$$

where (41) follows from Lemma 3 and the fact that $|\mathcal{B}_1| + |\mathcal{B}_2| = N_\beta$, and (43) follows from (37). Now we obtain that $W_\mathcal{B}$ reveals nothing about M_β.

B. Reversed Security Level

We next provide a joint coding strategy for the case $N_\beta \le N_\alpha \ \&\ N_\alpha > 0$.

Let G_α be a generator matrix generated using MDS-B in Section II-C, which can be used to encode M_α separately with encryption keys $(Z_1, Z_2, \ldots, Z_{N_\alpha})$. The joint coding strategy is simply to use $\eta = \min\{N_\alpha, \alpha - N_\beta\}$ symbols of the message M_β (i.e., $M_\beta^1, M_\beta^2, \cdots, M_\beta^\eta$) to replace the encryption keys $(Z_1\ Z_2\ \cdots\ Z_\eta)$ to encrypt M_α. Denote the corresponding coded symbols for M_α after this replacement as $(Y_\alpha^{1*}, Y_\alpha^{2*}, \cdots, Y_\alpha^{L*})$. Since the η message symbols of M_β do not need to be separately encoded, rate saving is thus obtained. Next, we verify the reconstruction and security constraints.

Reconstruction: By the code construction in Section II-C, both the message M_α and the keys M_β can be losslessly recovered from any α coded symbols. Since $\alpha < \beta$, the reconstruction requirements of both M_α and M_β are satisfied immediately.

Security: The security constraint of M_α is straightforward, and thus let us consider M_β. For any $\mathcal{B} \subseteq \mathcal{L}$ such that $|\mathcal{B}| = N_\beta$, let $Y_\alpha^{*\mathcal{B}} = (Y_\alpha^{i*} : i \in \mathcal{B})$. By Lemma 4, we have

$$I(Y_\alpha^{*\mathcal{B}}; M_\beta^1, M_\beta^2, \cdots, M_\beta^\eta) = 0, \tag{45}$$

since $\eta \le \alpha - N_\beta$.

V. Optimality Proof of Theorem 1

To show the optimality in Theorem 1, we only need to prove the sum rate bound in (6), which is

$$\sum_{l=1}^{L} \mathsf{R}_l \ge \sum_{\alpha=1}^{L} \frac{L\mathsf{m}_\alpha}{\alpha - N_\alpha}. \tag{46}$$

For any $\alpha \in \mathcal{L}$, let $\mathbb{B}_\alpha$ be the set of *disjoint subset* pairs $(\mathcal{B}_\alpha^1, \mathcal{B}_\alpha^2)$ such that $\mathcal{B}_\alpha^1, \mathcal{B}_\alpha^2 \subseteq \mathcal{L}$,

$$|\mathcal{B}_\alpha^1| = \alpha - N_\alpha \text{ and } |\mathcal{B}_\alpha^2| = N_\alpha. \tag{47}$$

For $\alpha \in \mathcal{L}$, let $M_{1:\alpha} \triangleq (M_1, M_2, \cdots, M_\alpha)$. Define μ_α by

$$\mu_\alpha = \frac{L}{\alpha - N_\alpha} \frac{1}{\binom{L}{N_\alpha}\binom{L-N_\alpha}{\alpha - N_\alpha}} \cdot \sum_{(\mathcal{B}_\alpha^1, \mathcal{B}_\alpha^2) \in \mathbb{B}_\alpha} H(W_{\mathcal{B}_\alpha^1} | W_{\mathcal{B}_\alpha^2} M_{1:\alpha}). \tag{48}$$

We need the following lemma to proceed.

Lemma 5. *For any $\alpha \in \mathcal{L}$, we have*

$$\sum_{l=1}^{L} H(W_l) \ge \sum_{j=1}^{\alpha} \frac{Lm_j}{j - N_j} + \mu_\alpha. \tag{49}$$

For $\alpha = L$, in light of (49), we can continue bounding as

$$\sum_{l=1}^{L} R_l = \sum_{l=1}^{L} H(W_l) \ge \sum_{\alpha=1}^{L} \frac{Lm_\alpha}{\alpha - N_\alpha} + \mu_L \ge \sum_{\alpha=1}^{L} \frac{Lm_\alpha}{\alpha - N_\alpha}, \tag{50}$$

from which we can obtain, by normalization, the sum rate bound (46).

References

[1] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1059–1064, May 1997. (conference version: ISIT 1995).

[2] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1737–1744, Nov. 1996.

[3] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 45, pp. 609–621, Mar. 1999.

[4] T. Guo and R. W. Yeung, "The explicit coding rate region of symmetrical multilevel diversity coding," Jan. 2018. (full version). [Online]. Available: http://arxiv.org/abs/1801.02376.

[5] A. Balasubramanian, H. D. Ly, S. Li, T. Liu, and S. L. Miller, "Secure symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 59, pp. 3572–3581, Jun. 2013.

[6] J. Jiang, N. Marukala, and T. Liu, "Symmetrical multilevel diversity coding and subset entropy inequalities," *IEEE Trans. Inf. Theory*, vol. 60, pp. 84–103, Jan. 2014.

[7] C. Tian and T. Liu, "Multilevel diversity coding with regeneration," *IEEE Trans. Inf. Theory*, vol. 62, pp. 4833–4847, Sep. 2016.

[8] S. Mohajer, C. Tian, and S. N. Diggavi, "Asymmetric multilevel diversity coding and asymmetric gaussian multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4367–4387, Sep. 2010.

[9] K. Bhattad and K. R. Narayanan, "Weakly secure network coding," in *Proc. NetCod 2005*, (Riva del Garda, Italy), Apr. 2005.

[10] M. Yan and A. Sprintson, "Weakly secure network coding for wireless cooperative data exchange," in *IEEE Global Telecommunications Conference (GLOBECOM)*, (Kathmandu, Nepal), Dec. 2011.

[11] M. Yan and A. Sprintson, "Algorithms for weakly secure data exchange," in *Proc. NetCod 2013*, (Calgary, Alberta, Canada), Jun. 2013.

[12] M. Yan, A. Sprintson, and I. Zelenko, "Weakly secure data exchange with generalized Reed Solomon codes," in *IEEE International Symposium on Information Theory (ISIT)*, (Honolulu, HI, USA), Jun. 2014.

[13] Y. Chen, O. O. Koyluoglu, and A. J. H. Vinck, "On secure communication over the multiple access channel," in *International Symposium on Information Theory and Its Applications (ISITA)*, (Monterey, CA, USA), Oct. 2016.

[14] T. Guo, C. Tian, T. Liu, and R. W. Yeung, "Weakly secure symmetric multilevel diversity coding," Apr. 2019. (full version). Available: https://www.dropbox.com/s/h52232lpyjroevc/WS-SMDC.pdf?dl=0.

[15] H. Yamamoto, "Secret sharing system using (k, L, n) threshold scheme," *IEICE Trans. Fund.(Jpn. Edition)*, vol. J68-A, Sep. 1985. (English Translation: Scripta Technica, Inc., Electronics and Commun. in Japan, Part I, vol. 69, pp. 4654, 1986).

[16] W.-A. Jackson and K. M. Martin, "A combinatorial interpretation of ramp schemes," *Australasian Journal of Combinatorics*, vol. 14, pp. 51–60, Jan. 1996.