# SECURE SYMMETRICAL MULTILEVEL DIVERSITY CODING

A Thesis

by

SHUO LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2012

Major Subject: Electrical Engineering

# SECURE SYMMETRICAL MULTILEVEL DIVERSITY CODING

A Thesis

by

SHUO LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

| | |
|---|---|
| Chair of Committee, | Tie Liu |
| Committee Members, | Serap Savari |
| | Robert Balog |
| | Anxiao Jiang |
| Head of Department, | Costas Georghiades |

May 2012

Major Subject: Electrical Engineering

# ABSTRACT

Secure Symmetrical Multilevel Diversity Coding. (May 2012)

Shuo Li, B.S., Xi'an Jiaotong University

Chair of Advisory Committee: Dr. Tie Liu

Secure symmetrical multilevel diversity coding (S-SMDC) is a source coding problem, where a total of $L-N$ discrete memoryless sources $(S_1,\dots,S_{L-N})$ are to be encoded by a total of $L$ encoders. This thesis considers a natural generalization of SMDC to the secure communication setting with an additional eavesdropper. In a general S-SMDC system, a legitimate receiver and an eavesdropper have access to a subset $U$ and $A$ of the encoder outputs, respectively. Which subsets $U$ and $A$ will materialize are unknown a priori at the encoders. No matter which subsets $U$ and $A$ actually occur, the sources $(S_1,\dots,S_k)$ need to be perfectly reconstructable at the legitimate receiver whenever $|U|=N+k$, and all sources $(S_1,\dots,S_{L-N})$ need to be kept perfectly secure from the eavesdropper as long as $|A|\leq N$. A precise characterization of the entire admissible rate region is established via a connection to the problem of secure coding over a three-layer wiretap network and utilizing some properties of basic polyhedral structure of the admissible rate region. Building on this result, it is then shown that superposition coding remains optimal in terms of achieving the minimum sum rate for the general secure SMDC problem.

# ACKNOWLEDGMENTS

It is my pleasure to thank all the people who helped me during my stay at Texas A&M University.

First of all, I would like to thank my advisor, Dr. Tie Liu, for his generous support and guidance. By clearly explaining the fundamental knowledge and concepts to me, he led me into real scientific research. His patience and encouragement became the greatest support for me to finish my thesis. It has been an impressive experience to work with him.

I would like to thank all my committee members, Dr. Serap Savari, Dr. Robert Balog and Dr. Anxiao Jiang, for giving me the feedback and suggestions on the thesis. I would also like to thank Dr. Krishina Narayanna and Dr. Jean-Francois Chamberland for their help and support in my career planning period.

I have been really lucky to have such wonderful lab mates, Hung D. Ly, Jinjing Jiang, Jae Won Yoo and Neeharika Marukala. They all have been great workmates and friends who gave me useful suggestions anytime I was in need. Hung D. Ly, whom I wish to especially thank, has been the best workmate and guide during my thesis preparation.

Texas A&M University is such a wonderful place and I greatly enjoyed my three years here. I would like to thank all my friends who have made my life fuller and more satisfying.

Finally, and most importantly, I would like to express my deepest gratitude to my parents for their ceaseless love and support, which helped me go through these years to reach this point.

## TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION . . . . . 1

A. Symmetrical Multilevel Diversity Coding . . . . . 1
B. Secure Network Coding on Wiretap Network . . . . . 3

II SECURE SYMMETRICAL SINGLE-LEVEL DIVERSITY CODING . . . . . 5

A. Problem Statement . . . . . 5
B. Main Results . . . . . 6
C. Proof of Theorem 1 . . . . . 11

III SECURE SYMMETRICAL MULTILEVEL DIVERSITY CODING . . . . . 16

A. Problem Statement . . . . . 16
B. Main Results . . . . . 17
C. Proof of the Main Results . . . . . 19

IV CONCLUSION . . . . . 24

REFERENCES . . . . . 26

APPENDIX A . . . . . 28

APPENDIX B . . . . . 30

VITA . . . . . 33

# LIST OF FIGURES

FIGURE Page

1 Illustration of the rate region $\mathcal{R}(L, k, H)$ for (a) $L = 2$ and (b)$L = 2$. The $R_l = 0$ slices of $\mathcal{R}(3, 1, H)$ are empty sets, and the $R_l = 0$ slices of $\mathcal{R}(3, 2, H)$ are isomorphic to $R(2, 1, H)$. . . . . . . . . . . . . . 9

2 Illustration of the $(L, N, m, (R_1, \ldots, R_L))$ WN. . . . . . . . . . . . . . 12

# CHAPTER I

# INTRODUCTION

## A. Symmetrical Multilevel Diversity Coding

Symmetrical Multilevel Diversity Coding (SMDC) system consists with $L$ independent discrete memoryless sources $(S_1, \ldots, S_L)$, where the *importance* of the sources is assumed to decrease with the subscript $l$. The sources are to be encoded by a total of $L$ encoders, where the rate of the $l$th encoder output is $R_l$. The decoder can access a subset $U \subseteq \Omega_L := \{1, \ldots, L\}$ of the encoder outputs. Which subset of the encoder outputs is available at the decoder is *unknown* a priori at the encoders. However, no matter which subset $U$ actually realizes, the sources $(S_1, \ldots, S_m)$ need to be asymptotically perfectly reconstructed at the decoder whenever $|U| \geq m$. Note that the word "symmetrical" here refers to the fact that the sources that need to be reconstructed at the decoder depend on the available subset of the encoder outputs only via its cardinality. The rate allocations at different encoders, however, can be different and are not necessarily symmetrical.

The problem of Multilevel Diversity Coding (MDC) was introduced by Roche [1] and Yeung [2] in the early 1990s. In particular, [2] considered the simple coding strategy of separately encoding different sources at the encoders, subsequently referred to as *superposition coding*. The aforementioned SMDC problem was first systematically studied in [3], where it was shown that superposition coding can achieve the minimum sum rate for the general SMDC problem (with an arbitrary total number of encoders $L$) and the entire admissible rate region with $L = 3$ encoders. The problem regarding

This thesis follows the style of *IEEE Transactions on Automatic Control.*

whether superposition coding can achieve the entire admissible rate region for the general SMDC problem, however, remained open. Finally, in a very elegant (albeit highly technical) paper [4], Yeung and Zhang resolved the open problem by positive through the so-called $\alpha$-*resolution* method.

Recent years have seen a flurry of research on information-theoretic security. See [5] and [6] for surveys of recent progress in this field. Motivated by this renewed interest, in this thesis I consider the problem of *Secure* Symmetrical Multilevel Diversity Coding (S-SMDC) in the presence of an additional eavesdropper. Specifically, a collection of $L-N$ independent discrete memoryless sources $(S_1, \ldots, S_{L-N})$ are to be encoded by a total of $L$ encoders, where the rate of the $l$th encoder output is $R_l$. A legitimate receiver and an eavesdropper can access a subset $U \subseteq \Omega_L$ and $A \subseteq \Omega_L$ of the encoder outputs, respectively. Which subsets of the encoder outputs are available at the legitimate receiver and the eavesdropper are *unknown* a priori at the encoders. However, no matter which subsets $U$ and $A$ actually occur, the sources $(S_1, \ldots, S_k)$ need to be asymptotically perfectly reconstructed at the legitimate receiver whenever $|U| \geq N+k$, and the entire collection of the sources $(S_1, \ldots, S_{L-N})$ needs to be kept *perfectly* secret from the eavesdropper as long as $|A| \leq N$. As before, the word "symmetrical" here refers to the access structure at the legitimate receiver and the eavesdropper, but not to the rate allocations at different encoders. We envision that such a communication scenario is useful for designing distributed information storage systems [1] where information retrieval needs to be both robust and secure.

As mentioned previously, separate encoding of different sources (superposition coding) can achieve the entire admissible rate region for the general SMDC problem without any secrecy constraints [4]. It is thus natural to ask whether the same separate encoding strategy would remain optimal for the general S-SMDC problems. For the classical SMDC problems without any secrecy constraints, the problem of efficient encoding of individual sources is essentially to transmit the source over an *erasure* channel and is well understood based on the earlier work of Singleton [7]. For the S-SMDC problems, however, the problem of efficient encoding of individual sources is closely related to the problem of secure coding over a *Wiretap Network (WN)* [8], which, in its most general setting, is a very challenging problem in information-theoretic security.

### B. Secure Network Coding on Wiretap Network

Secure Network Coding on Wiretap Network is a information-theoretic security problem of network coding and was studied by N. Cai and R. W. Yeung in [8]. They proposed a way to construct a secure linear network code for wiretap network which was proved to be optimal for the case that the wiretapper may access any subset of channels of a fixed size.

A wiretap network, denoted as $(\mathcal{G}, s, \mathcal{U}, \mathcal{A})$, consists of a directed acyclic multigraph $\mathcal{G}$, a source node $s$, a set of user nodes $\mathcal{U}$, and a collection of sets of wiretapped edges $\mathcal{A}$. The multigraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is seen as a network where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the edge set. A wiretapper may access on member of $\mathcal{A}$ but no more than one. In order to protect the message generated by the source node from the wiretapper, the key has to be introduced in the network. A code is admissible for a wiretap network if the decodable condition and the secure condition are satisfied. A wiretap network is called $r$- Wiretap Network if all subsets in $\mathcal{A}$ have cardinalities no more than $r$. A wiretapper will get no information of the source by accessing no more than $r$ channels when the admissible $r$- secure network code applied.

The most common example of the $r$- Wiretap Network, studied in [9] and [10], is the so called 'secret sharing' problem. An $(r, n)-$threshold secret sharing scheme, which is equivalent to an $(r-1)-$threshold secret sharing scheme, consists of $n$ participants such that any $r$ participants can perfectly recover the source message while any $r-1$ or less participants can not acquired any information.

The rest of the thesis is organized as follows. In Chapter II, we focus on the problem of encoding individual sources, i.e., Secure Symmetrical Single-level Diversity Coding (S-SSDC). By leveraging the results of [8] on secure coding over a three-layer WN and utilizing some basic polyhedral structure of the admissible rate region, we provide a precise characterization of the entire admissible rate region for the general S-SSDC problem. Building on this result, in Chapter III we show that superposition coding can achieve the minimum sum rate for the general S-SMDC problem. Finally, in Chapter IV we conclude the thesis with some remarks.

# CHAPTER II

# SECURE SYMMETRICAL SINGLE-LEVEL DIVERSITY CODING

## A. Problem Statement

Let $\{S[t]\}_{t=1}^{\infty}$ be a discrete memoryless source with time index $t$ and let $S^n := (S[1], \ldots, S[n])$. An $(L, N, m)$ S-SSDC problem consists of a set of $L$ encoders, a legitimate receiver who has access to a subset $U \subseteq \Omega_L$ of the encoder outputs, and an eavesdropper who has access to a subset $A \subseteq \Omega_L$ of the encoder outputs. Which subsets of the encoder outputs are available at the legitimate receiver and the eavesdropper are *unknown* a priori at the encoders. However, no matter which subsets $U$ and $A$ actually occur, the legitimate receiver must be able to asymptotically perfectly reconstruct the source whenever $|U| \geq m$, and the source must be kept *perfectly* secret from the eavesdropper as long as $|A| \leq N$. Obviously, reliable and secure communication of the source is possible only when $m > N$.

Formally, an $(n, (M_1, \ldots, M_L))$ code is defined by a collection of $L$ encoding functions

$$e_l : \mathcal{S}^n \times \mathcal{K} \to \{1, \ldots, M_l\}, \quad \forall l = 1, \ldots, L \tag{2.1}$$

and decoding functions

$$d_U : \prod_{l \in U} \{1, \ldots, M_l\} \to \mathcal{S}^n, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq m. \tag{2.2}$$

Here, $\mathcal{K}$ denotes the key space accessible to all $L$ encoders. There are no limitations on the size of the key space $\mathcal{K}$. However, the secret key is only shared by the encoders, but *not* with the legitimate receiver or the eavesdropper. A nonnegative rate tuple $(R_1, \ldots, R_L)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block length $n$, an $(n, (M_1, \ldots, M_L))$ code such that:

- (Rate constraints)

$$\frac{1}{n}\log M_l \leq R_l + \epsilon, \quad \forall l = 1, \ldots, L; \tag{2.3}$$

- (Asymptotically perfect reconstruction at the legitimate receiver)

$$\Pr\{d_U(X_U) \neq S^n\} \leq \epsilon, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq m \tag{2.4}$$

  where $X_l := e_l(S^n, K)$ is the output of the $l$th encoder, $K$ is the secret key shared by all $L$ encoders, and $X_U := \{X_l : l \in U\}$; and

- (Perfect secrecy at the eavesdropper)

$$H(S^n|X_A) = H(S^n), \qquad \forall A \subseteq \Omega_L \text{ s.t. } |A| \leq N \tag{2.5}$$

  i.e., observing the encoder outputs $X_A$ does not provide *any* information regarding to the source sequence $S^n$.

The *admissible rate region* $\mathcal{R}$ is the collection of *all* admissible rate tuples $(R_1, \ldots, R_L)$. The *minimum sum rate* $R_{ms}$ is defined as

$$R_{ms} := \min_{(R_1,\ldots,R_L)\in\mathcal{R}} \sum_{l=1}^{L} R_l. \tag{2.6}$$

### B. Main Results

The following lemma provides a simple outer bound on the admissible rate region of the general S-SSDC problem. Let $\mathcal{R}(L, k, H)$ be the collection of all nonnegative rate tuples $(R_1, \ldots, R_L)$ satisfying

$$\sum_{l\in D} R_l \geq H, \quad \forall D \in \Omega_L^{(k)} \tag{2.7}$$

where $\Omega_L^{(k)}$ is the collection of all subsets of $\Omega_L$ of size $k$.

**Lemma 1.** *For any $(L, N, m)$ S-SSDC problem, the admissible rate region*

$$\mathcal{R} \subseteq \mathcal{R}(L, m - N, H(S)). \tag{2.8}$$

Lemma 1 can be proved using standard information-theoretic techniques. For completeness, a proof is included in Appendix A. The above outer bound is known to be tight in the following two special cases:

1) When $N = 0$, the $(L, N, m)$ S-SSDC problem reduces to the classical $(L, m)$ SSDC problem without any secrecy constraints, for which the admissible rate region is known [7] to be $\mathcal{R}(L, m, H(S))$.

2) With $N > 0$ but $m = N + 1$, a collection $D$ of the encoder outputs will either lead to an asymptotically perfect reconstruction of the source (whenever $|D| \geq N+1$), or provide zero information on the source (whenever $|D| \leq N$). In this case, the $(L, N, m)$ S-SSDC problem reduces to the classical $(L, N)$ *threshold secret sharing* problem, for which the admissible rate region is known [9, 10] to be $\mathcal{R}(L, 1, H(S))$.

The main result of this section is that the outer bound $\mathcal{R}(L, m - N, H(S))$ is in fact the admissible rate region for the *general* S-SSDC problem, as summarized in the following theorem.

**Theorem 1.** *For any $(L, N, m)$ S-SSDC problem, the admissible rate region*

$$\mathcal{R} = \mathcal{R}(L, m - N, H(S)). \tag{2.9}$$

A proof of the theorem is provided in Sec. C. To show that *every* rate tuple in $\mathcal{R}(L, m - N, H(S))$ is admissible, our proof proceeds in the following two steps. First, we show that for any $(L, N, m)$ S-SSDC problem, the symmetrical rate tuple

$(H(S)/(m-N),\ldots,H(S)/(m-N))$ is admissible. In our proof, this is accomplished by relating the S-SSDC problem to the problem of secure coding over a *three-layer* WN and using the result of [8, Th. 3] on an achievable secrecy rate for the generic WN. Building on the previous result, next we show that every rate tuple in $\mathcal{R}(L,m-N,H(S))$ is admissible via an induction argument (inducting on the total number of encoders $L$) and the following polyhedral structure of $\mathcal{R}(L,k,H)$.

**Proposition 1.** *$\mathcal{R}(L,k,H)$ is a pointed polyhedron in $\mathbb{R}^L$ with the following structural properties:*

1) *The characteristic cone of $\mathcal{R}(L,k,H)$ is given by $\{(R_1,\ldots,R_L) : R_l \geq 0,\ \forall l = 1,\ldots,L\}$.*
2) *Among all corner points (vertices) of $\mathcal{R}(L,k,H)$, $(H/k,\ldots,H/k)$ is the* only *one with all* strictly *positive entries (if there exists any).*
3) *For any $l=1,\ldots,L$, the $R_l=0$ slice of $\mathcal{R}(L,k,H)$ is* isomorphic *to $\mathcal{R}(L-1,k-1,H)$. In particular, the $R_L=0$ slice of $\mathcal{R}(L,k,H)$ is* identical *to $\mathcal{R}(L-1,k-1,H)$, i.e.,*

$$\{(R_1,\ldots,R_{L-1}) : (R_1,\ldots,R_{L-1},0) \in \mathcal{R}(L,k,H)\} = \mathcal{R}(L-1,k-1,H). \quad (2.10)$$

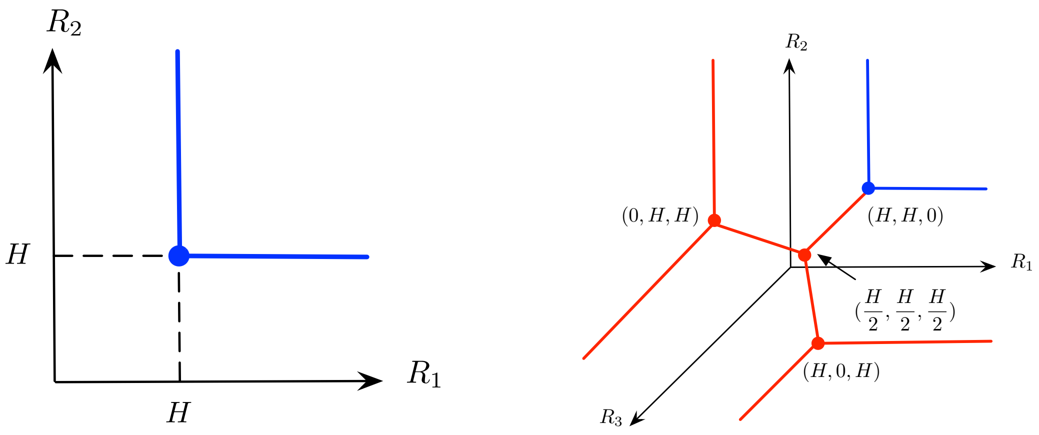
*Proof.* Property 1 follows directly from the definition of characteristic cone [11, Lec. 2]. Property 2 is due to the fact that

$$(R_1,\ldots,R_L) = (H/k,\ldots,H/k) \quad (2.11)$$

is a solution to the equations

$$\sum_{l\in D} R_l = H, \quad \forall D \in \Omega_L^{(k)}. \quad (2.12)$$

To see property 3, note that the $R_l=0$ slice of $\mathcal{R}(L,k,H)$ is given by all



(a) $\mathcal{R}(2,1,H)$

(b) $\mathcal{R}(3,1,H)$

Fig. 1. Illustration of the rate region $\mathcal{R}(L,k,H)$ for (a) $L=2$ and (b)$L=2$. The $R_l=0$ slices of $\mathcal{R}(3,1,H)$ are empty sets, and the $R_l=0$ slices of $\mathcal{R}(3,2,H)$ are isomorphic to $R(2,1,H)$.

nonnegative rate tuples $(R_1,\ldots,R_{l-1},R_{l+1},\ldots,R_L)$ satisfying

$$\sum_{d\in D} R_d \geq H, \quad \forall D \in \Omega^{(k-1)}_{L\setminus\{l\}} \cup \Omega^{(k)}_{L\setminus\{l\}} \tag{2.13}$$

where $\Omega^{(k)}_{L\setminus\{l\}}$ denotes all subsets of $\Omega_L \setminus \{l\}$ of size $k$. Since every inequality with $D \in \Omega^{(k)}_{L\setminus\{l\}}$ is dominated by every inequality with $D' \in \Omega^{(k-1)}_{L\setminus\{l\}}$ and such that $D' \subseteq D$, we have the desired property. □

Fig. 1 illustrates the above polyhedral structure of $\mathcal{R}(L,k,H)$ for $L=2$ and 3. The following corollary summarizes the minimum sum rate for the general S-SSDC problem.

**Corollary 2.** *For any $(L,N,m)$ S-SSDC problem, the minimum sum rate*

$$R_{ms} = \frac{L}{m-N} H(S). \tag{2.14}$$

*Proof.* Let us first verify that

$$\min_{(R_1,\ldots,R_L)\in\mathcal{R}(L,k,H)} \sum_{l=1}^{L} R_l = \frac{L}{k}H. \tag{2.15}$$

For any rate tuple $(R_1, \ldots, R_L) \in \mathcal{R}(L,k,H)$, we have

$$\sum_{l\in D} R_l \geq H, \quad \forall D \in \Omega_L^{(k)}. \tag{2.16}$$

Summing over all $D \in \Omega_L^{(k)}$ gives

$$\sum_{D\in\Omega_L^{(k)}} \sum_{l\in D} R_l = \begin{pmatrix} L-1 \\ k-1 \end{pmatrix} \sum_{l=1}^{L} R_l \geq \begin{pmatrix} L \\ k \end{pmatrix} H. \tag{2.17}$$

We thus have

$$\sum_{l=1}^{L} R_l \geq \frac{\begin{pmatrix} L \\ k \end{pmatrix}}{\begin{pmatrix} L-1 \\ k-1 \end{pmatrix}} H = \frac{L}{k}H \tag{2.18}$$

for any rate tuple $(R_1, \ldots, R_L) \in \mathcal{R}(L,k,H)$. On the other hand, note that the symmetrical rate tuple

$$(H/k, \ldots, H/k) \in \mathcal{R}(L,k,H) \tag{2.19}$$

so

$$\min_{(R_1,\ldots,R_L)\in\mathcal{R}(L,k,H)} \sum_{l=1}^{L} R_l \leq \frac{L}{k}H. \tag{2.20}$$

Combining (2.18) and (2.20) completes the proof of (2.15).

Now by Theorem 1,

$$R_{ms} = \min_{(R_1,\ldots,R_L)\in\mathcal{R}} \sum_{l=1}^{L} R_l = \min_{(R_1,\ldots,R_L)\in\mathcal{R}(L,m-N,H(S))} \sum_{l=1}^{L} R_l = \frac{L}{m-N}H(S). \tag{2.21}$$

This completes the proof of the corollary. □

### C. Proof of Theorem 1

Let us first show that the symmetrical rate tuple $(H(S)/(m-N),\ldots,H(S)/(m-N))$ is admissible by considering the following simple *source-channel separation* scheme for the $(L,N,m)$ S-SSDC problem:

- First compress the source sequence $S^n$ into a source message $W$ using a fixed-length lossless source code. It is well known [12, Ch. 3.2] that the rate $R$ of the source message $W$ can be made arbitrarily close to the entropy rate $H(S)$ for sufficiently large block length $n$.
- Next, the source message $W$ is delivered to the legitimate receiver using a secure
$$(L,N,m,(R_1,\ldots,R_L))$$
WN code.

The problem of secure coding over a WN was formally introduced in [8]. A generic WN $(\mathcal{G},s,\mathcal{U},\mathcal{A})$ consists of a directed acyclic network $\mathcal{G}$, a source node $s$, a set of user nodes $\mathcal{U}$, and a collection of sets of wiretapped edges $\mathcal{A}$. Each member of $\mathcal{A}$ may be fully accessed by an eavesdropper, but no eavesdropper may access more than one member of $\mathcal{A}$. The source node has access to a message $W$, which is intended for all user nodes in $\mathcal{U}$ but needs to be kept *perfectly* secret from the eavesdroppers. The maximum achievable secrecy rate for $W$ is called the *secrecy capacity* of the WN and is denoted by $C_s(\mathcal{G},s,\mathcal{U},\mathcal{A})$.

An $(L,N,m,(R_1,\ldots,R_L))$ WN is a special WN with three layers of nodes: top, middle, and bottom. As illustrated in Fig. 2, the only node in the top layer is the source node $s$. There are $L$ intermediate nodes in the middle layer, each corresponding to an encoder in the $(L,N,m)$ S-SSDC problem. For each $l=1,\ldots,L$, the source

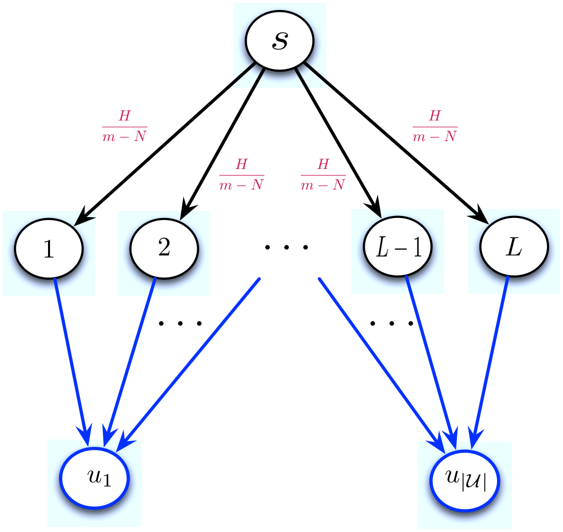


Fig. 2. Illustration of the $(L, N, m, (R_1, \ldots, R_L))$ WN.

node $s$ is connected to the intermediate node $l$ by a channel $(s, l)$ with capacity $R_l$. There are

$$|\mathcal{U}| = \begin{pmatrix} L \\ m \end{pmatrix} \tag{2.22}$$

user nodes in the bottom layer, each corresponding to a possible realization of the legitimate receiver in the $(L, N, m)$ S-SSDC problem and is connected to $m$ intermediate nodes through $m$ infinite-capacity channels. Finally, the collection of sets of wiretapped edges $\mathcal{A}$ is defined as

$$\mathcal{A} := \left\{ \{(s, l) | l \in A\} : A \in \Omega_L^{(N)} \right\} \tag{2.23}$$

where each set of wiretapped edges in $\mathcal{A}$ corresponds to a possible realization of the eavesdropper in the $(L, N, m)$ S-SSDC problem.

Based on the aforementioned connection between the $(L, N, m)$ S-SSDC problem and the problem of secure coding over the $(L, N, m, (R_1, \ldots, R_L))$ WN, we have the following simple lemma.

**Lemma 2.** *A nonnegative rate tuple* $(R_1, \ldots, R_L)$ *is admissible for the* $(L, N, m)$ *S-SSDC problem if the entropy rate of the source is less than or equal to the secrecy capacity of the* $(L, N, m, (R_1, \ldots, R_L))$ *WN, i.e.,*

$$H(S) \leq C_s(L, N, m, (R_1, \ldots, R_L)). \tag{2.24}$$

In general, characterizing the exact secrecy capacity of a WN can be very difficult. For a generic WN $(\mathcal{G}, s, \mathcal{U}, \mathcal{A})$, the following secrecy rate

$$R_s = \min_{u \in \mathcal{U}, A \in \mathcal{A}} [mincut(s, u) - mincut(s, A)] \tag{2.25}$$

is known [8] to be achievable. Here, $mincut(s, u)$ denotes the value of a minimum cut between the source node $s$ and the user node $u$, and $mincut(s, A)$ denotes the value of a minimum cut between the source node $s$ and the set of wiretapped edges $A$. For the $(L, N, m, (H(S)/(m-N), \ldots, H(S)/(m-N)))$ WN, it is straightforward to verify that

$$mincut(s, u) = \frac{m}{m-N} H(S), \quad \forall u \in \mathcal{U} \tag{2.26}$$

and

$$mincut(s, A) = \frac{N}{m-N} H(S), \quad \forall A \in \mathcal{A}. \tag{2.27}$$

Hence, the secrecy rate

$$R_s = \frac{m}{m-N} H(S) - \frac{N}{m-N} H(S) = H(S) \tag{2.28}$$

is achievable for the $(L, N, m, (H(S)/(m-N), \ldots, H(S)/(m-N)))$ WN. We summarize this result in the following lemma.

**Lemma 3.** *For any* $(L, N, m, (H(S)/(m-N), \ldots, H(S)/(m-N)))$ *WN, the secrecy* *capacity can be bounded from below as*

$$C_s(L, N, m, (H(S)/(m-N), \ldots, H(S)/(m-N))) \geq H(S). \tag{2.29}$$

Combining Lemmas 2 and 3 proves the admissibility of the symmetrical rate tuple

$$(H(S)/(m-N), \ldots, H(S)/(m-N)).$$

Building on the previous result, next we show that *every* rate tuple in $\mathcal{R}(L, m-N, H(S))$ is admissible. By Proposition 1, $\mathcal{R}(L, m-N, H(S))$ is a pointed polyhedron with the characteristic cone given by $\{(R_1, \ldots, R_L) : R_l \geq 0,\ \forall l = 1, \ldots, L\}$. Thus, to show that all rate tuples in $\mathcal{R}(L, m-N, H(S))$ are admissible, it is sufficient to show that all *corner points* of $\mathcal{R}(L, m-N, H(S))$ are admissible.

We shall consider proof by induction, where the induction is on the total number of encoders $L$. First consider the base case with $L = 2$. When $L = 2$, there is only one nontrivial $(L, N, m)$ S-SSDC problem: the $(2, 1, 2)$ S-SSDC problem. Note that the rate region $\mathcal{R}(2, 1, H(S))$ has only one corner point: the symmetrical rate pair $(H(S), H(S))$, whose admissibility has already been established. We thus conclude that every rate tuple in $\mathcal{R}(2, 1, H(S))$ is admissible for the $(2, 1, 2)$ S-SSDC problem.

Now assume that for every nontrivial $(L-1, N', m')$ S-SSDC problem, all rate tuples in $\mathcal{R}(L', m'-N', H(S))$ are admissible. Based on this assumption, next we show that all corner points of $\mathcal{R}(L, m-N, H(S))$ are admissible for the $(L, N, m)$ S-SSDC problem. We shall consider the corner points with all *strictly* positive entries (it they exist) and those with *at least one zero* entry separately:

1) By Proposition 1, the symmetrical rate tuple $(H(S)/(m-N), \ldots, H(S)/(m-N))$ is the *only* corner point of $\mathcal{R}(L, m-N, H(S))$ with all *strictly* positive entries (if it exists), whose admissibility has already been established.

2) To prove the admissibility of the corner points of $\mathcal{R}(L, m-N, H(S))$ with *at least one zero* entry, by the *symmetry* of the rate region $\mathcal{R}(L, m-N, H(S))$ we may consider without loss of generality those with $R_L = 0$. Note that if an $(n, (M_1, \ldots, M_{L-1}))$ code satisfies both asymptotically perfect reconstruction constraint (2.4) and the perfect secrecy constraint (2.5) for the $(L-1, N, m-1)$ S-SSDC problem, an $(n, (M_1, \ldots, M_{L-1}, 1))$ code with the *same* encoding functions for encoders 1 to $L-1$ (encoder $L$ uses a constant encoding function) also satisfies (trivially) both constraints for the $(L, N, m)$ S-SSDC problem. Thus, if $(R_1, \ldots, R_{L-1})$ is an admissible rate tuple for the $(L-1, N, m-1)$ S-SSDC problem, then $(R_1, \ldots, R_{L-1}, 0)$ is also an admissible rate tuple for the $(L, N, m)$ S-SSDC problem. By the induction assumption, all rate tuples in $\mathcal{R}(L-1, m-N-1, H(S))$ are admissible for the $(L-1, N, m-1)$ problem. Combined with Proposition 1, this implies that all rate tuples in

$$\begin{aligned} &\{(R_1, \ldots, R_{L-1}, 0) : (R_1, \ldots, R_{L-1}) \in \mathcal{R}(L-1, m-N-1, H(S))\} \\ &\qquad\qquad = \{(R_1, \ldots, R_L) \in \mathcal{R}(L, m-N, H(S)) : R_L = 0\} \end{aligned} \tag{2.30}$$

i.e., the $R_L = 0$ slice of $\mathcal{R}(L, m-N, H(S))$, are admissible for the $(L, N, m)$ S-SSDC problem. As a special case, all corner points of $\mathcal{R}(L, m-N, H(S))$ with $R_L = 0$ are admissible for the $(L, N, m)$ S-SSDC problem.

Combining Steps 1 and 2 proves that all corner points of $\mathcal{R}(L, m-N, H(S))$ are admissible. We thus conclude that all rate tuples in $\mathcal{R}(L, m-N, H(S))$ are admissible. This completes the induction step and hence the proof of the theorem.

# CHAPTER III

# SECURE SYMMETRICAL MULTILEVEL DIVERSITY CODING

## A. Problem Statement

Let $\{S_1[t], \ldots, S_{L-N}[t]\}_{t=1}^{\infty}$ be a collection of $L - N$ independent discrete memoryless sources with time index $t$ and let $S_k^n := (S_k[1], \ldots, S_k[n])$. An $(L, N)$ S-SMDC problem consists of a set of $L$ encoders, a legitimate receiver who has access to a subset $U$ of the encoder outputs, and an eavesdropper who has access to a subset $A$ of the encoder outputs. Which subsets of the encoder outputs are available at the legitimate receiver and the eavesdropper are *unknown* a priori at the encoders. However, no matter which subsets $U$ and $A$ actually occur, the legitimate receiver must be able to asymptotically perfectly reconstruct the sources $(S_1, \ldots, S_k)$ whenever $|U| = N + k$, and all sources $(S_1, \ldots, S_{L-N})$ must be kept perfectly secure from the eavesdropper as long as $|A| \leq N$.

Formally, an $(n, (M_1, \ldots, M_L))$ code is defined by a collection of $L$ encoding functions

$$e_l : \prod_{k=1}^{L-N} \mathcal{S}_k^n \times \mathcal{K} \to \{1, \ldots, M_l\}, \quad \forall l = 1, \ldots, L \tag{3.1}$$

and decoding functions

$$d_U : \prod_{l \in U} \{1, \ldots, M_l\} \to \prod_{k=1}^{|U|-N} \mathcal{S}_k^n, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq N + 1. \tag{3.2}$$

Here, $\mathcal{K}$ is the key space accessible to all $L$ encoders. A nonnegative rate tuple $(R_1, \ldots, R_L)$ is said to be *admissible* if for every $\epsilon > 0$, there exits, for sufficiently large block length $n$, an $(n, (M_1, \ldots, M_L))$ code such that:

- (Rate constraints)

$$\frac{1}{n}\log M_l \leq R_l + \epsilon, \quad \forall l = 1, \ldots, L; \tag{3.3}$$

- (Asymptotically perfect reconstruction at the legitimate receiver)

$$\Pr\{d_U(X_U) \neq (S_1^n, \ldots, S_{|U|-N}^n)\} \leq \epsilon, \quad \forall U \subseteq \Omega_L \text{ s.t. } |U| \geq N+1 \tag{3.4}$$

  where $X_l := e_l((S_1^n, \ldots, S_{L-N}^n), K)$ is the output of the $l$th encoder, $K$ is the secret key shared by all $L$ encoders, and $X_U := \{X_l : l \in U\}$; and

- (Perfect secrecy at the eavesdropper)

$$H(S_1^n, \ldots, S_{L-N}^n | X_A) = H(S_1^n, \ldots, S_{L-N}^n), \quad \forall A \subseteq \Omega_L \text{ s.t. } |A| \leq N \tag{3.5}$$

  i.e., observing the encoder outputs $X_A$ does not provide *any* information regarding to the source sequences $(S_1^n, \ldots, S_{L-N}^n)$.

## B. Main Results

Motivated by the success of [2–4] on the classical SMDC problem without any secrecy constraints, here we focus on superposition coding where the output of the $l$th encoder $X_l$ is given by

$$X_l = \left(X_l^{(1)}, \ldots, X_l^{(L-N)}\right) \tag{3.6}$$

and $X_l^{(k)}$ is the coded message for source $S_k$ at the $l$th encoder using an $(L, N, N+k)$ S-SSDC code. Note here that all sources are encoded *separately* at the encoders, and there is *no* coding across different sources. Thus, if $(R_1^{(k)}, \ldots, R_L^{(k)})$ is an admissible rate tuple for the $(L, N, N+k)$ S-SSDC problem with source $S_k$, then the rate tuple

$$(R_1, \ldots, R_L) = \left(\sum_{k=1}^{L-N} R_1^{(k)}, \ldots, \sum_{k=1}^{L-N} R_L^{(k)}\right) \tag{3.7}$$

is admissible for the $(L, N)$ S-SMDC problem.

By Corollary 2, the minimum sum rate for the $(L, N, N + k)$ S-SSDC problem with source $S_k$ is given by $(L/k)H(S_k)$. It follows that $\sum_{k=1}^{L-N}(L/k)H(S_k)$ is the minimum sum rate that can be achieved by superposition coding for the $(L, N)$ S-SMDC problem. The main result of this section is that $\sum_{k=1}^{L-N}(L/k)H(S_k)$ is in fact the minimum sum rate that can be achieved by *any* coding scheme for the $(L, N)$ S-SMDC problem. Thus, superposition coding is optimal in terms of achieving the minimum sum rate for the general S-SMDC problem. We summarize this result in the following theorem.

**Theorem 3.** *Superposition coding can achieve the minimum sum rate for the general $(L, N)$ S-SMDC problem, which is given by*

$$R_{ms} = \sum_{k=1}^{L-N} \frac{L}{k} H(S_k). \tag{3.8}$$

A proof of the theorem is provided in Sec C. The proof uses an induction argument and is built on the classical subset inequality of Han [12, Ch. 17.6] and the following key proposition.

**Proposition 2.** *For any $(n, (M_1, \dots, M_L))$ code that satisfies both asymptotically perfect reconstruction constraint* (3.4) *and the perfect secrecy constraint* (3.5)*, we have*

$$H(X_D|S_1^n, \dots, S_{k-1}^n, X_A) \geq nH(S_k) + H(X_D|S_1^n, \dots, S_k^n, X_A) - n\delta_k(n, \epsilon) \tag{3.9}$$

*where*

$$\delta_k(n, \epsilon) := 1/n + \epsilon \sum_{\alpha=1}^{k} \log |\mathcal{S}_\alpha| \tag{3.10}$$

*for any $A \in \Omega_L^{(N)}$ and $D \in \Omega_L^{(k)}$ such that $A \cap D = \emptyset$ and any $k = 1, \dots, L - N$.*

### C. Proof of the Main Results

Let us first prove Proposition 2. Since $|A| = N$, $|D| = k$, and $A \cap D = \emptyset$, we have $|D \cup A| = N + k$. For any $(n, (M_1, \dots, M_L))$ code that satisfies both asymptotically perfect reconstruction constraint (3.4) and the perfect secrecy constraint (3.5), we have by Fano's inequality

$$H(S_1^n, \dots, S_k^n | X_D, X_A) \leq n\delta_k(n, \epsilon) \tag{3.11}$$

and

$$H(S_1^n, \dots, S_k^n | X_A) = H(S_1^n, \dots, S_k^n). \tag{3.12}$$

Thus,

$$\begin{aligned}
&H(X_D | S_1^n, \dots, S_{k-1}^n, X_A) + n\delta_k(n, \epsilon) \\
&\quad \geq \; H(X_D | S_1^n, \dots, S_{k-1}^n, X_A) + H(S_1^n, \dots, S_k^n | X_D, X_A) && (3.13) \\
&\quad \geq \; H(X_D | S_1^n, \dots, S_{k-1}^n, X_A) + H(S_k^n | S_1^n, \dots, S_{k-1}^n, X_D, X_A) && (3.14) \\
&\quad = \; H(X_D, S_k^n | S_1^n, \dots, S_{k-1}^n, X_A) && (3.15) \\
&\quad = \; H(S_k^n | S_1^n, \dots, S_{k-1}^n, X_A) + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.16) \\
&\quad = \; H(S_1^n, \dots, S_k^n | X_A) - H(S_1^n, \dots, S_{k-1}^n | X_A) \\
&\qquad + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.17) \\
&\quad = \; H(S_1^n, \dots, S_k^n) - H(S_1^n, \dots, S_{k-1}^n | X_A) \\
&\qquad + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.18) \\
&\quad \geq \; H(S_1^n, \dots, S_k^n) - H(S_1^n, \dots, S_{k-1}^n) + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.19) \\
&\quad = \; H(S_k^n | S_1^n, \dots, S_{k-1}^n) + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.20) \\
&\quad = \; H(S_k^n) + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.21) \\
&\quad = \; nH(S_k) + H(X_D | S_1^n, \dots, S_k^n, X_A) && (3.22)
\end{aligned}$$

where (3.13) follows from (3.11), (3.18) follows from (3.12), (3.19) follows from the fact that conditioning reduces entropy, (3.21) follows from the fact that the sources $S_1, \ldots, S_k$ are mutually independent, and (3.22) follows from the fact that the source $S_k$ is memoryless. Moving $n\delta_k(n,\epsilon)$ to the right-hand side of the inequality completes the proof of Proposition 2.

Building on the result of Proposition 2, next let us show that for any $(n,(M_1,\ldots,M_L))$ code that satisfies both asymptotically perfect reconstruction constraint (3.4) and the perfect secrecy constraint (3.5) and any $\alpha = 1,\ldots,L-N$, we have

$$\sum_{l=1}^{L} H(X_l) \geq \sum_{k=1}^{\alpha} \frac{nL}{k} H(S_k) + \Delta_\alpha - \sum_{k=1}^{\alpha} nL\delta_k(n,\epsilon) \tag{3.23}$$

where

$$\Delta_\alpha := \frac{L}{\begin{pmatrix} L \\ N \end{pmatrix}\begin{pmatrix} L-N \\ \alpha \end{pmatrix}} \sum_{A\in\Omega_L^{(N)}} \sum_{D\in\Omega_{L\setminus A}^{(\alpha)}} \frac{H(X_D|S_1^n,\ldots,S_\alpha^n,X_A)}{\alpha}. \tag{3.24}$$

We shall consider proof by induction, where the induction is on $\alpha$. First consider the base case with $\alpha = 1$. Let $A \in \Omega_L^{(N)}$ and let $l \in \Omega_L \setminus A$. Applying Proposition 2 with $k = 1$, we have

$$H(X_l) \geq nH(S_1) + H(X_l|S_1^n, X_A) - n\delta_1(n,\epsilon). \tag{3.25}$$

Averaging (3.25) over all $l \in \Omega_L \setminus A$ and all $A \in \Omega_L^{(N)}$, we have

$$\frac{1}{\begin{pmatrix} L \\ N \end{pmatrix}\begin{pmatrix} L-N \\ 1 \end{pmatrix}} \sum_{A\in\Omega_L^{(N)}} \sum_{l\in\Omega_L\setminus A} H(X_l) \geq nH(S_1) + \frac{1}{L}\Delta_1 - n\delta_1(n,\epsilon). \tag{3.26}$$

Note that

$$\frac{1}{\begin{pmatrix} L \\ N \end{pmatrix}\begin{pmatrix} L-N \\ 1 \end{pmatrix}} \sum_{A \in \Omega_L^{(N)}} \sum_{l \in \Omega_L \backslash A} H(X_l) = \frac{1}{L}\sum_{l=1}^{L} H(X_l). \tag{3.27}$$

We thus have

$$\sum_{l=1}^{L} H(X_l) \geq nLH(S_1) + \Delta_1 - nL\delta_1(n,\epsilon) \tag{3.28}$$

which completes the proof of the base case.

Now assume that (3.23) holds for $\alpha - 1$ for some $2 \leq \alpha \leq L-N$. Based on this assumption, next we show that (3.23) also holds for $\alpha$. By the classical subset inequality of Han [12, Ch. 17.6], for any $A \in \Omega_L^{(N)}$ we have

$$\begin{aligned} &\frac{1}{\begin{pmatrix} L-N \\ \alpha-1 \end{pmatrix}} \textstyle\sum_{D \in \Omega_{L\backslash A}^{(\alpha-1)}} \frac{H(X_D|S_1^n,\ldots,S_{\alpha-1}^n,X_A)}{\alpha-1} \\ &\qquad\qquad \geq \frac{1}{\begin{pmatrix} L-N \\ \alpha \end{pmatrix}} \textstyle\sum_{D \in \Omega_{L\backslash A}^{(\alpha)}} \frac{H(X_D|S_1^n,\ldots,S_{\alpha-1}^n,X_A)}{\alpha}. \end{aligned} \tag{3.29}$$

It follows that

$$\Delta_{\alpha-1} \geq \frac{L}{\begin{pmatrix} L \\ N \end{pmatrix}\begin{pmatrix} L-N \\ \alpha \end{pmatrix}} \sum_{A \in \Omega_L^{(N)}} \sum_{D \in \Omega_{L\backslash A}^{(\alpha)}} \frac{H(X_D|S_1^n,\ldots,S_{\alpha-1}^n,X_A)}{\alpha}. \tag{3.30}$$

By Proposition 2, for any $A \in \Omega_L^{(N)}$ and any $D \in \Omega_{L\backslash A}^{(\alpha)}$ we have

$$H(X_D|S_1^n,\ldots,S_{\alpha-1}^n,X_A) \geq nH(S_\alpha) + H(X_D|S_1^n,\ldots,S_\alpha^n,X_A) - n\delta_n(\alpha,\epsilon). \tag{3.31}$$

Substituting (3.31) into (3.30) gives

$$\Delta_{\alpha-1} \geq \frac{L}{\begin{pmatrix} L \\ N \end{pmatrix}\begin{pmatrix} L-N \\ \alpha \end{pmatrix}} \sum_{A\in\Omega_L^{(N)}} \sum_{D\in\Omega_{L\setminus A}^{(\alpha)}} \frac{nH(S_\alpha)+H(X_D|S_1^n,\ldots,S_\alpha^n,X_A)-n\delta_\alpha(n,\epsilon)}{\alpha} \tag{3.32}$$

$$= \frac{nL}{\alpha}H(S_\alpha)+\Delta_\alpha-nL\delta_\alpha(n,\epsilon). \tag{3.33}$$

By the induction assumption,

$$\sum_{l=1}^{L} H(X_l) \geq \sum_{k=1}^{\alpha-1}\frac{nL}{k}H(S_k)+\Delta_{\alpha-1}-\sum_{k=1}^{\alpha-1}nL\delta_k(n,\epsilon) \tag{3.34}$$

$$\geq \sum_{k=1}^{\alpha-1}\frac{nL}{k}H(S_k)+\left(\frac{nL}{\alpha}H(S_\alpha)+\Delta_\alpha-nL\delta_\alpha(n,\epsilon)\right) -\sum_{k=1}^{\alpha-1}nL\delta_k(n,\epsilon) \tag{3.35}$$

$$= \sum_{k=1}^{\alpha}\frac{nL}{k}H(S_k)+\Delta_\alpha-\sum_{k=1}^{\alpha}nL\delta_k(n,\epsilon). \tag{3.36}$$

This completes the proof of the induction step and hence (3.23).

Finally, let $\alpha = L-N$ in (3.23). For any admissible rate tuple $(R_1,\ldots,R_L)$ and any $\epsilon > 0$, we have

$$n\sum_{l=1}^{L}(R_l+\epsilon) \geq \sum_{l=1}^{L}H(X_l) \tag{3.37}$$

$$\geq \sum_{k=1}^{L-N}\frac{nL}{k}H(S_k)+\Delta_{L-N}-\sum_{k=1}^{L-N}nL\delta_k(n,\epsilon) \tag{3.38}$$

$$\geq \sum_{k=1}^{L-N}\frac{nL}{k}H(S_k)-\sum_{k=1}^{L-N}nL\delta_k(n,\epsilon) \tag{3.39}$$

where (3.39) follows from the fact that $\Delta_{L-N} \geq 0$. Divide both sides of (3.39) by $n$

and let $n \to \infty$ and $\epsilon \to 0$. Note that $\delta_k(n, \epsilon) \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$ for all $k = 1, \ldots, L - N$. We thus have

$$\sum_{l=1}^{L} R_l \geq \sum_{k=1}^{L-N} \frac{L}{k} H(S_k) \tag{3.40}$$

for any admissible rate tuple $(R_1, \ldots, R_L)$. This completes the proof of Theorem 3.

# CHAPTER IV

# CONCLUSION

This thesis considered the problem of S-SMDC, which is a natural (perhaps also the simplest) extension of the classical SMDC problem [1–4] to the secrecy communication setting. First, the problem of encoding individual sources, i.e., the S-SSDC problem, was studied. A precise characterization of the entire admissible rate region was established via a connection to the problem of secure coding over a three-layer WN [8] and utilizing some basic polyhedral structure of the admissible rate region. Building on this result, it was then shown that the simple coding strategy of separately encoding individual sources at the encoders (superposition coding) can achieve the minimum sum rate for the general S-SMDC problem.

Based on the result of Theorem 3 (and the fact that superposition coding can achieve the entire admissible rate region for the classical SMDC problems without secrecy constraints), it is very tempting to conjecture that superposition coding can in fact achieve the entire admissible rate region for the general S-SMDC problem. In Appendix B, we verify that this is indeed the case for the simplest nontrivial S-SMDC problem: the $(3, 1)$ S-SMDC problem. Our proof relies on an *explicit* characterization of the superposition coding rate region via a Fourier-Motzkin elimination procedure. The optimality of superposition coding is then established by carefully using the results of Proposition 2.

Extending such a proof strategy to the general $(L, N)$ S-SMDC problem, however, faces a number of challenges. To begin with, the complexity of Fourier-Motzkin elimination procedure grows unboundedly as the total number of encoders $L$ increases. Thus, establishing an explicit characterization of the superposition coding rate region for the general $(L, N)$ S-SMDC problem appears to be very difficult. An alternative strategy is to look for an *implicit* characterization of the superposition coding rate region using *optimal $\alpha$-resolutions*, similar to that [4] for the classical SMDC problem without any secrecy constraints. In fact, note from Theorem 1 that the admissible rate region of an $(L, N, m)$ S-SSDC problem depends on the parameters $N$ and $m$ only via its difference $m - N$. As mentioned previously in Sec. II, when $N = 0$, the $(L, N, m)$ S-SSDC problem reduces to the classical $(L, m)$ SSDC problem without any secrecy constraints. Thus, the admissible rate region of the $(L, N, N + k)$ S-SSDC problem with source $S_k$ is *identical* to that of the classical $(L, k)$ SSDC problem with the same source. As a result, the superposition coding rate region of the $(L, N)$ S-SMDC problem with sources $(S_1, \ldots, S_{L-N})$ is *identical* to the superposition coding rate region of the classical SMDC problem with a total of $L$ encoders and sources $(S_1, \ldots, S_L)$ where the entropy rate of the source $H(S_l) = 0$ for $l = L-N+1, \ldots, L$. Based on this observation, the $\alpha$-resolution characterization of the superposition coding rate region for the general SMDC problem can be directly translated to the S-SMDC problem. It remains to see whether the properties provided in [4] on optimal $\alpha$-resolutions are sufficient for establishing the optimality of superposition coding for the general S-SMDC problem. This problem is currently under our investigations.

# REFERENCES

[1] J. R. Roche, “Distributed information storage,” Ph.D. Dissertation, Department of Statistics, Stanford University, Stanford, CA, Mar. 1992.

[2] R. W. Yeung, “Multilevel diversity coding with distortion,” *IEEE Trans. Inf. Theory*, vol. 41, pp. 412–422, Mar. 1995.

[3] 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.

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

[5] Y. Liang, H. V. Poor, and S. Shamai (Shitz), *Information Theoretic Security.* Dordrecht, The Netherlands: Now Publisher, 2009.

[6] R. Liu and W. Trappe, Eds, *Securing Wireless Communications at the Physical Layer.* New York: Springer Verlag, 2010.

[7] R. C. Singleton, “Maximum distance $q$-nary codes,” *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 116–118, Apr. 1964.

[8] N. Cai and R. W. Yeung, “Secure network coding on a wiretap network,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 424-435, Jan. 2011.

[9] A. Shamir, “How to share a secret,” *Comm. ACM*, vol. 22, pp. 612–613, Nov. 1979.

[10] G. R. Blakley, “Safeguarding cryptographic keys,” in *Proc. National Computer Conference*, New York, June 1979, vol. 48, pp. 313–317.

[11] C. Chekuri, *Lecture Notes on Combinatorial Optimization.* University of Illinois, Urbana-Champaign, IL. Available online at http://www.cs.illinois.edu/homes/chekuri/

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd ed.* Hoboken, NJ: John Wiley & Sons, 2006.

# APPENDIX A

## PROOF OF LEMMA 1

Let $D \in \Omega_L^{(m-N)}$ and let $A \in \Omega_{L\backslash D}^{(N)}$. Since $A \cap D = \emptyset$, we have $|D \cup A| = N + (m-N) = m$. For any $(n, (M_1, \ldots, M_L))$ code that satisfies both asymptotically perfect reconstruction constraint (2.4) and the perfect secrecy constraint (2.5), we have by Fano's inequality

$$H(S^n|X_D, X_A) \leq n\delta(n, \epsilon) \tag{A.1}$$

where

$$\delta(n, \epsilon) = 1/n + \epsilon \log |\mathcal{S}| \tag{A.2}$$

and

$$H(S^n|X_A) = H(S^n). \tag{A.3}$$

For any admissible rate tuple $(R_1, \ldots, R_L)$ and any $\epsilon > 0$, we have

$$\begin{aligned}
n\sum_{l\in D}(R_l + \epsilon) &\geq \sum_{l\in D} H(X_l) && \text{(A.4)}\\
&\geq H\left(X_D\right) && \text{(A.5)}\\
&\geq H\left(X_D|X_A\right) && \text{(A.6)}\\
&\geq H\left(X_D|X_A\right) + H\left(S^n|X_D, X_A\right) - n\delta(n, \epsilon) && \text{(A.7)}\\
&= H\left(X_D, S^n|X_A\right) - n\delta(n, \epsilon) && \text{(A.8)}\\
&= H\left(S^n|X_A\right) + H\left(X_D|S^n, X_A\right) - n\delta(n, \epsilon) && \text{(A.9)}\\
&\geq H\left(S^n|X_A\right) - n\delta(n, \epsilon) && \text{(A.10)}\\
&= H\left(S^n\right) - n\delta(n, \epsilon) && \text{(A.11)}\\
&= nH(S) - n\delta(n, \epsilon) && \text{(A.12)}
\end{aligned}$$

where (A.5) follows from the independence bound on entropy, (A.6) follows from the fact that conditioning reduces entropy, (A.7) follows from (A.1), (A.11) follows from (A.3), and (A.12) follows from the fact that the source $S$ is memoryless. Divide both sides of (A.12) by $n$ and let $n \to \infty$ and $\epsilon \to 0$. Note that $\delta(n, \epsilon) \to 0$ in the limit as $n \to \infty$ and $\epsilon \to 0$. We have from (A.12) that

$$\sum_{l \in D} R_l \geq H(S), \quad \forall D \in \Omega_L^{(m-N)}. \tag{A.13}$$

This completes the proof of Lemma 1.

# APPENDIX B

## THE ADMISSIBLE RATE REGION OF THE $(3,1)$ S-SMDC PROBLEM

In this appendix, we show that superposition coding can achieve the entire admissible rate region for the $(3,1)$ S-SMDC problem (the simplest nontrivial S-SMDC problem). The result is summarized in the following theorem.

**Theorem 4.** *Superposition coding can achieve the entire admissible rate region for the* $(3,1)$ *S-SMDC problem, which is given by the collection of all rate triples* $(R_1, R_2, R_3)$ *satisfying*

$$\begin{aligned} R_1 &\geq H(S_1) \\ R_2 &\geq H(S_1) \\ R_3 &\geq H(S_1) \\ R_1 + R_2 &\geq 2H(S_1) + H(S_2) \\ R_2 + R_3 &\geq 2H(S_1) + H(S_2) \\ R_3 + R_1 &\geq 2H(S_1) + H(S_2). \end{aligned} \tag{B.1}$$

*Proof. Achievability.* Consider the superposition coding scheme that separately encodes the sources $S_1$ and $S_2$ using the $(3,1,2)$ and $(3,1,3)$ S-SSDC codes, respectively. By Theorem 1, the admissible rate region for the $(3,1,2)$ S-SSDC problem is given by all rate triples $(R_1^{(1)}, R_2^{(1)}, R_3^{(1)})$ satisfying

$$\begin{aligned} R_1^{(1)} &\geq H(S_1) \\ R_2^{(1)} &\geq H(S_1) \\ R_3^{(1)} &\geq H(S_1) \end{aligned} \tag{B.2}$$

and the admissible rate region for the $(3,1,3)$ S-SSDC problem is given by all rate triples $(R_1^{(2)}, R_2^{(2)}, R_3^{(2)})$ satisfying

$$\begin{aligned} R_1^{(2)} &\geq 0 \\ R_2^{(2)} &\geq 0 \\ R_3^{(2)} &\geq 0 \\ R_1^{(2)} + R_2^{(2)} &\geq H(S_2) \\ R_2^{(2)} + R_3^{(2)} &\geq H(S_2) \\ R_3^{(2)} + R_1^{(2)} &\geq H(S_2). \end{aligned} \tag{B.3}$$

Following (3.7), all rate triples $(R_1, R_2, R_3)$ as given by

$$R_l = R_l^{(1)} + R_l^{(2)}, \quad \forall l = 1, 2, 3 \tag{B.4}$$

are admissible via superposition coding. Using Fourier-Motzkin elimination to eliminate $R_l^{(k)}$, $l = 1, 2, 3$ and $k = 1, 2$, from (B.2)–(B.4), we obtain the explicit characterization of the superposition coding rate region for the $(3, 1)$ S-SMDC problem as expressed by (B.1).

*The converse.* Next, we establish the optimality of superposition coding by proving that every inequality in (B.1) must hold for *all* admissible rate triples $(R_1, R_2, R_3)$ for the $(3, 1)$ S-SMDC problem. Let

$$a \oplus b := \begin{cases} a + b, & \text{if } a + b \leq 3 \\ a + b - 3, & \text{otherwise.} \end{cases} \tag{B.5}$$

For any admissible rate triple $(R_1, R_2, R_3)$, any $l = 1, 2, 3$, and any $\epsilon > 0$, we have

$$\begin{aligned} n(R_l + \epsilon) &\geq H(X_l) && \text{(B.6)} \\ &\geq H(X_l | X_{l\oplus 1}) && \text{(B.7)} \\ &\geq nH(S_1) + H(X_l | S_1^n, X_{l\oplus 1}) - n\delta_1(n, \epsilon) && \text{(B.8)} \\ &\geq nH(S_1) - n\delta_1(n, \epsilon) && \text{(B.9)} \end{aligned}$$

and

$$\begin{aligned} & n(R_l + R_{l\oplus 1} + 2\epsilon) \\ &\geq H(X_l) + H(X_{l\oplus 1}) && \text{(B.10)} \\ &\geq H(X_l | X_{l\oplus 1}) + H(X_{l\oplus 1} | X_{l\oplus 2}) && \text{(B.11)} \\ &\geq 2nH(S_1) + H(X_l | S_1^n, X_{l\oplus 1}) + H(X_{l\oplus 1} | S_1^n, X_{l\oplus 2}) - 2n\delta_1(n, \epsilon) && \text{(B.12)} \\ &\geq 2nH(S_1) + H(X_l | S_1^n, X_{l\oplus 1}, X_{l\oplus 2}) + H(X_{l\oplus 1} | S_1^n, X_{l\oplus 2}) \\ &\quad -2n\delta_1(n, \epsilon) && \text{(B.13)} \\ &= 2nH(S_1) + H(X_l, X_{l\oplus 1} | S_1^n, X_{l\oplus 2}) - 2n\delta_1(n, \epsilon) && \text{(B.14)} \\ &\geq 2nH(S_1) + (nH(S_2) + H(X_l, X_{l\oplus 1} | S_1^n, S_2^n, X_{l\oplus 2}) - n\delta_2(n, \epsilon)) \\ &\quad -2n\delta_1(n, \epsilon) && \text{(B.15)} \\ &\geq 2nH(S_1) + nH(S_2) - n\delta_2(n, \epsilon) - 2n\delta_1(n, \epsilon). && \text{(B.16)} \end{aligned}$$

Here, (B.7), (B.11) and (B.13) follow from the fact that conditioning reduces entropy, and (B.8), (B.12) and (B.15) follow from Proposition 2. Dividing both sides of (B.9) and (B.16) by $n$ and letting $n \to \infty$ and $\epsilon \to 0$ complete the proof of the converse part of the theorem. □

# VITA

Name: Shuo Li

Email: lishuoixa@gmail.com

Education: B.S. Electrical Engineering  
Xi’an Jiaotong University, Xi’an, Shaanxi, China (July 2009)  
M.S. Electrical Engineering  
Texas A&M University, College Station, TX (May 2012)

Address: 237-J Zachry Engineering Center,  
3128 TAMU, College Station, TX 77843