The Pennsylvania State University

The Graduate School

College of Engineering

INFORMATION THEORETIC SECURITY GUARANTEES AGAINST MORE CAPABLE ADVERSARIES

A Dissertation in

Electrical Engineering

by

Mohamed Nafea

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2018

The dissertation of Mohamed Nafea was reviewed and approved* by the following:

Aylin Yener
Professor of Electrical Engineering
Dissertation Adviser
Chair of Committee

David Miller
Professor of Electrical Engineering

Viveck Cadambe
Assistant Professor of Electrical Engineering

Jan Reimann
Associate Professor of Mathematics

Kultegin Aydin
Professor of Electrical Engineering
Head of the Department of Electrical Engineering

*Signatures are on file in the Graduate School.

Abstract

With the recent growth in network connectivity, the issue of securing information against unauthentic and adversarial parties has become essential. By exploiting the intrinsic noise in physical communication channels, provable –information theoretic– security guarantees can be provided even against adversaries with *unlimited computation capabilities*. Earlier approaches in physical layer security typically assume the adversary is weaker than the legitimate terminals, and is a passive entity which does not design attacks. This thesis aims at addressing these issues, with the goal of bringing the information theoretic security vision closer to reality. Our approach includes introducing stronger adversaries in existing models, as well as introducing new adversarial models in emerging communication systems.

We begin with considering the multiple antenna Gaussian wiretap channel. An adversary which has resources to deploy more antennas than the legitimate terminals can deteriorate, or even preclude, secure communication. Enlisting a multiantenna terminal with the specific goal of diminishing the adversary's reception can re-enable secure communication. Towards that end, we characterize the secrecy capacity pre-log factor, i.e., the secure degrees of freedom (s.d.o.f.), of the multiple antenna wiretap channel with a multiantenna cooperative jammer, for all possible antenna configurations. Achievability is established by developing a variety of signaling, beamforming, and alignment schemes which vary according to the relative number of antennas at each terminal. Among these schemes, we devise a *projection and cancellation* decoding scheme that is optimal for

certain antenna configurations. Our results show that positive s.d.o.f. is attainable as long as the combined number of transmit antennas is greater than the number of the adversary's antennas.

We next turn our focus to the wiretap II channel, introduced back in 1984. The wiretap II channel provides an adversary model that is well received mainly due to its elegance in featuring a *designed* attack. Additionally, the model has an impact on several problems such as secure network coding, linear codes for secrecy, and adversarial erasure channels. The original communication theoretic version of the model has remained linked to the noiseless legitimate channel assumption for almost *three decades*. We introduce a noisy, i.e., discrete memoryless, legitimate channel to the model, and derive inner and outer bounds on its capacity-equivocation region, which match for certain cases. The achievability relies on combining random coding and random partitioning.

Subsequently, we introduce a *generalized wiretap channel model* which subsumes both the classical wiretap and wiretap II with a noisy main channel models as special cases. In this model, the adversary chooses a subset of the codeword symbols to tap into, while observing the remainder through a noisy channel. We identify the strong secrecy capacity of the model, and show that it is identical to the secrecy capacity when the subset is randomly chosen by nature. Achievability is established by utilizing source-channel duality, and concentration of measures to prove a super-exponential convergence rate for the security measure, which dominates the exponentially many possible strategies for the adversary. Converse is derived by using Sanov's theorem in method of types.

Next, we investigate several multi-terminal extensions of our generalized wiretap model, including the multiple access and broadcast wiretap channels, and the interference

and broadcast channels with confidential messages. These extensions require non trivial generalizations of the tool set we develop for the single user case. Our results for the generalized wiretap model and its multiterminal extensions demonstrate the robustness of *stochastic wiretap encoding* against a powerful adversary which chooses where to tap.

Finally, we introduce a new model, namely the notion of *cache tapping* into the information theoretic models of coded caching, in which the adversary taps into a subset of symbols of its choice either from the cache placement, delivery, or both phases. The legitimate parties know neither the relative fractions of tapped symbols in each phase, nor their positions. We identify the strong secrecy capacity for the instance of two users and two library files. We derive lower and upper bounds on the strong secrecy file rate for a library size of three or larger. Achievability is established by a code design which integrates wiretap coding, security embedding codes, one-time pad keys, and coded caching techniques. Our results establish that strong information theoretic security is possible against a powerful adversary which optimizes its chosen attack over both phases of communication in a cache-aided system.

Table of Contents

List of Figures

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my academic advisor Prof. Aylin Yener for her unceasing support, patience, and encouragement throughout the years of my PhD. It is not doubtful that this thesis would not have been possible without her invaluable guidance, passionate dedication, and the uncountable hours she spent in helping me to improve my intellectual, technical, and presentation skills. Her philosophy and strength in dealing with the technical challenges and general problems not only influenced me to become a much better researcher, but also helped me to grow wiser, for which I will always be grateful.

I would like to thank Prof. David Miller, Prof. Jan Reimann, and Prof. Viveck Cadambe, for serving in my PhD Dissertation committee and for their valuable comments. I sincerely thank Prof. Jan Reimann for his unwavering help during my Masters studies in Mathematics and for introducing me to the fields of mathematical logic and large graphs, which I find extremely exciting.

I would like to thank all my former and current Lab-mates at WCAN; Kaya Tütüncüoğlu, Başak Guler, Burak Varan, Ahmed Zewail, Abdelrahman Ibrahim, and Shiyang Leng, for their help and support during my PhD study and for many interesting technical and non-technical discussions. Special thanks are due to Prof. Rémi Chou with whom I thoroughly enjoyed conducting research and from whom I have learnt a lot, throughout the two years he spent as a post-doc at WCAN. Special thanks are due also to Dr. Kaya Tütüncüoğlu for all the times he was there when I needed a friend,

a colleague, or a brain. Special thanks as well are due to Prof. Raef Bassily for his valuable help during the time he spent at Penn state as Post-doc.

I would like to thank all my friends at Penn State; Mahmoud Ashour and Mohamed Tarek, among many others, who helped me a lot during these years.

I would also like to thank my close friends Ahmed Sobhy, Mohamed Abdelmonem, Abdelrahman Baknina, Islam Bakoury, Yahya Ezzeldin, Ghada Hatem, and Noha Helal, who shared many of my difficult moments as a grad student. Special thanks are due to Abelrahman Baknina for the numerous times he was there when I needed a sincere advice. Special thanks are also due to Yahya Ezzeldin for many insightful mathematical and engineering discussions.

I am very grateful to meet Waleed Khan at Penn State, whose genuine friendship and unwavering support made this town feel like home.

I am as well very grateful to have my three wonderful siblings, Fatma, Ahmed, and Mostafa, whose love and support have always been and continue to be invaluable. Thank you for sharing all my moments of happiness and difficulties.

Last but certainly not least, I am, and always have been, incalculably fortunate and eternally grateful to have my parents Saeed Nafea and Amal Abdelhafiz, whose love and devotion is beyond what words can ever describe.

Funding Acknowledgment and Disclaimer: This material is based upon work supported by the National Science Foundation (NSF) under Awards No. CCF 13-19338 and CNS 13-14719. Any opinions, findings, and conclusions or recommendations expressed in this thesis are those of the author and do not necessarily reflect the views of the National Science Foundation

Chapter 1

Introduction

Back in 1949, Shannon has established the fundamental principles of secure communications. In particular, he has introduced the notion of *perfect secrecy* that is achieved when the enemy crypt-analyst's equivocation about the message is equal to the apriori uncertainty of the message [105]. For a noiseless communication channel and an external adversary which has noiseless access to the communication, Shannon has shown that perfect secrecy requires a *one-time pad* key whose length is larger than or equal to the length of the message. Both cryptographic and information theoretic security have emerged from Shannon's work. The mathematical cryptographic schemes on one hand focus on the practical aspect of a computationally-limited crypt-analyst and rely on the computational hardness of solving certain problems such as prime factorization of large integers. Information theoretic secrecy on the other hand focuses on the practical assumption of *noisy communication channels* and rely on exploiting the intrinsic noise in the channel to secure the legitimate communication against a wiretapper with unbounded computational power.

Wyner in [121] has introduced the wiretap channel and the notion of secrecy capacity, where a legitimate transmitter and receiver are communicating over a discrete memoryless channel, referred to as the legitimate (main) channel, in the presence of a wiretapper which overhears the legitimate communication through a cascaded second

discrete memoryless channel, referred to as the wiretapper channel. Wyner's wiretap model is extended to a general discrete memoryless setting in [23] and to the Gaussian channel in [61]. Subsequently, an extensive body of work has been devoted to study a variety of network information theoretic extensions of Wyner's wiretap model, such as the broadcast channel [31,67,92], the multiple access channel [64,112,113], the fading channel [34, 57], the interference channel [41], the relay channel [28, 40, 60, 95], the compound channel [56, 62], and two-way communications [42].

The majority of research in information theoretic security literature assume an adversary model in which the adversary is (i) weaker than the legitimate terminals, and (ii) a passive entity which *only listens* to the legitimate communication through a noisy channel, and does not perform an active or chosen attacks. The adversary however is a malicious entity which potentially attempts to turn the communication scenario into its favor, and can have a resource advantage over the legitimate terminals. It is thus essential to address these challenges in order for the research in information theoretic security to move forward toward real systems. In this thesis, we aim at addressing more capable models for the adversary, and we show that information theoretic security guarantees against such adversaries are possible. We first consider an adversary which potentially has more resources (antennas) than the legitimate terminals. We next introduce adversarial models in which the adversary performs chosen codeword (cipher-text) attacks, and study multi-terminal extensions of these models. Finally, we study the impact of the chosen adversarial attacks in multiple phase communication setups, such as the cache-aided communication system.

With regard to the secrecy measure in information theoretic secrecy literature, the majority of works assume a *weak secrecy metric* which requires the mutual information between the wiretapper's observation and the message, normalized with respect to the block-length n, to vanish asymptotically with n. With this weak constraint, the wiretapper is able to access a substantial amount of information about the message, which increases at a rate that is strictly less than n. References [22, 75, 77] have addressed this drawback of the weak secrecy metric, and introduced a stronger metric, termed as strong secrecy, in which the overall mutual information between the wiretapper's signal and the message is required to vanish with n. Unlike for the secrecy results under the weak metric, there has not been a standard approach in the literature for proving secrecy with respect to this stronger metric. Instead, several approaches were considered for strong secrecy proofs such as privacy amplification [11] and channel resolvability [37], see for example [16, 22, 43, 75]. With the exception of Chapters 3 and 4, we consider the strong secrecy metric for the models presented in this thesis, and develop the necessary tools to prove strong security.

1.1 Adversarial Resource Advantage

Coding design for the wiretap channel relies on adding randomness in mapping each message to a codeword, i.e., stochastic encoding, such that the wiretapper is kept ignorant of the message, while the legitimate receiver is able to decode the message. This can be easily achieved when there is a physical advantage for the legitimate channel over the wiretapper channel and results in positive secure communication rates. When that is not the case, the features that the wireless medium brings can serve as tools to

create an overall channel that is advantageous for the legitimate parties. These include (i) utilizing multiple transmit and/or receive antennas in order to creat an advantage for the legitimate channel [58], or (ii) employing one or more helper nodes, which can be legitimate nodes in the network or external nodes augmented to the network, to transmit intentional interference with the specific goal of diminishing the reception capability of the wiretapper, known as *cooperative jamming* [111, 113].

In Chapter 3, we study the multiple antenna Gaussian wiretap channel, and introduce a multi-antenna cooperative jammer to the model in order to provide a secrecy rate that scales with transmit power.[1] We characterize the pre-log factor of the secrecy capacity, i.e., the secure degrees of freedom (s.d.o.f.), of the channel for all possible antenna configurations [81–84,89]. The achievability is based on a variety of signaling, beamforming, and alignment schemes in order to handle different antenna configurations. We show that whenever the s.d.o.f. is integer valued, Gaussian signaling for both transmission and cooperative jamming, linear precoding, and linear receiver processing, are sufficient for achieving the s.d.o.f. of the channel. By contrast, when the s.d.o.f. is not an integer, the achievable schemes need to rely on *structured*, i.e., discrete, signaling at the transmitter and the cooperative jammer. For the latter case, we devise a *projection and cancellation* decoding scheme which involves joint signal space and signal scale alignment in the complex plane [59, 71]. The converse is established by combining an upper bound which allows for full cooperation between the transmitter and the cooperative jammer, with another upper bound which exploits the secrecy and reliability constraints. Overall, our

[1] For a multiple antenna Gaussian wiretap channel, the secrecy capacity of the channel does not scale with the transmit power whenever the wiretapper has more antennas than the legitimate transmitter [58].

results show that positive s.d.o.f. for the channel is attainable as long as the combined number of transmit antennas, i.e., the combined number of antennas at the transmitter and the cooperative jammer, is greater than the number of antennas at the wiretapper.

1.2 Adversarially Designed Attacks and Multi-terminal Extensions

Ozarow and Wyner in [96] have introduced the wiretap II channel which models a noiseless legitimate channel, and a wiretapper which taps into a subset of its choosing of the transmitted symbols. The wiretap II model has received relatively less attention, mainly due to technical challenges in generalizing the model outside of this special communication setting. Yet, the notion of providing the wiretpper with this additional capability of choosing what to observe is appealing and represents a positive step towards providing confidentiality guarantees in stronger attack models.

In Chapter 4, we introduce a discrete memoryless legitimate channel to the wiretap II channel model and derive inner and outer bounds on its capacity equivocation regions [85]. The derived bounds match for the special instance of the maximizing input distribution being uniform. The achievability is established by identifying a class of good codes for which there exists a good partition that achieves the required level of equivocation no matter what subset the wiretapper chooses, and showing that the probability of this class of good codes goes to one with increasing the block-length.

Further, in Chapter 5, we introduce a *generalized wiretap channel* model which generalizes both the classical wiretap channel [23, 121] and the wiretap II model with a noisy legitimate channel [85]. In this model, the legitimate channel is discrete memroyless and the wiretapper chooses a subset of the codeword symbols to noiselessly observe

and observes the remainder of the codeword through a noisy channel. We characterize the strong secrecy capacity of the model and quantify the secrecy penalty of the additional capability at the wiretapper with respect to the previous wiretap models [88, 91]. We as well show that the secrecy capacity of the generalized wiretap model is identical to the secrecy capacity when the subset of noiseless observations is randomly chosen by nature. We establish the achievability by solving a dual secret key agreement problem in the source model [2, 76], and converting the solution to the original channel model using probability distribution approximation arguments. In the dual problem, a source encoder and decoder, which observe random sequences independent and identically distributed according to the input and output distributions of the legitimate channel in the original problem, communicate a confidential key over a public error-free channel using a single forward transmission, in the presence of a compound wiretapping source which has perfect access to the public discussion. The security of the key is guaranteed for the exponentially many possibilities of the subset chosen at the wiretapper by deriving a lemma which provides a *doubly-exponential* convergence rate for the probability that, for a fixed choice of the subset, the key is uniform and independent from the public discussion and the wiretapping source's observation. The converse is derived by using Sanov's theorem in the method of types [21, Theorem 11.4.1] to upper bound the secrecy capacity of the generalized wiretap channel by the secrecy capacity when the tapped subset is randomly chosen by nature.

Much like what has happened with Wyner's wiretap channel [128], exploring the multi-terminal extensions of our generalized wiretap model is the natural next step. In Chapter 6, we first introduce the multiple access wiretap II model with a noisy legitimate

channel, under three different wiretapping scenarios [86, 87]. The wiretapper, as in the classical wiretap channel II model, chooses a fixed-length subset of the channel uses on which it obtains noise-free observations of (i) one of the codewords, (ii) a superposition of the two codewords, or (iii) each of the two codewords. These thus extend the wiretap II channel with a noisy legitimate channel, introduced in Chapter 4, to a multiple access setting with a variety of attack models for the wiretapper. Next, we propose a generalized multiple access wiretap channel model, which further generalizes the multiple access wiretap channel II under the third wiretapping scenario, i.e., that which features the strongest adversarial model. In this model, the wiretapper, besides choosing a subset of the channel uses to noiselessly observe the transmitted codeword symbols of both users, observes the remainder of the two codewords through a discrete memoryless multiple access channel [90]. Achievable strong secrecy rate regions for all the proposed models are derived. As for the single user case, achievability is established using source-channel duality, i.e., solving the problem in the dual *multi-terminal* secret key agreement source model and inferring the design for the encoders and decoder of the original multiple access channel from the solution of the dual problem. This requires extending the tool set we derived for the single source case in Chapter 5 to the case of two independent sources. The derived achievable rate regions quantify the secrecy cost due to the additional capabilities of the wiretapper with respect to the previous multiple access wiretap models.

In Chapter 7, we further examine secure communication in multi-receiver models when the eavesdropping terminal is capable of tapping into a subset of its choosing of the transmitted codeword(s) symbols. This generalizes the classical multi-receiver wiretap models [9, 26, 34, 67] into those with chosen adversarial attacks. In particular, we first

propose a generalized two-user broadcast wiretap channel in the presence of a wiretapper which has perfect access to a fixed-size subset of its choosing of the transmitted symbols, while observing the remainder through a noisy channel. Additionally, we introduce generalized models for the two-user broadcast and interference channels with confidential messages. In these models, each receiver, besides its noisy observations, is provided with a subset of noiseless observations for the transmitted codeword(s) symbols. Achievable strong secrecy rate regions for these three models are derived. Achievability is established by source-channel duality, and requires extending the tools we develop in Chapter 5 to *multiple correlated sources*. For the generalized broadcast wiretap channel, the optimality of the proposed achievable scheme is established for two special cases. In general, the derived rate region for this model extends Marton's inner bound [30, Theorem 8.4] for the two user broadcast channel with a common message to the proposed setting, and indicates the secrecy cost due to the additional capability at the wiretapper. For the generalized broadcast and interference channels with confidential messages, the size of the subset at each receiver induces a trade-off between their rates. We highlight the special instance of the generalized broadcast channel with confidential messages, when one receiver is degraded with respect to the other and only the degraded receiver is provided with the subset of noiseless observations. In this case, the degraded receiver has a positive rate after a certain threshold of its noiseless observations, i.e., with the aid of these symbols.

1.3 Attacks in a Cache-aided Communication System

Caching is a technique proposed to efficiently reduce network traffic congestion by storing popular information contents at the cache memories of end users earlier during off-peak times [4,25]. Reference [72] has shown that a careful design of cache contents at the end users in a multi-receiver setting allows the transmitter to send delivery transmissions that are simultaneously useful for many users, termed as *coded caching*. Coded caching with security requirements has recently been studied in [8, 53, 98, 103, 130–132]. These references assume secure cache placement. At the other extreme, if cache placement were to be perfectly accessed by an adversary, the presence of cache memories cannot increase the secrecy capacity [2,106]. One might hence think of an intermediate scenario in which the adversary may have *partial access* to cache placement. The wiretap II model we have studied in Chapters 4, 5, 6, and 7, provides a model for an adversary with partial access to the legitimate communication in terms of a threshold on the fraction the adversary is able to tap into the communications.

In Chapter 8, we introduce the wiretap II channel model in the presence of multiple receivers equipped with fixed-size cache memories, and an adversary which is able to choose symbols to tap into from cache placement, in addition to or in lieu of, delivery transmission. The model is hence termed the caching broadcast channel with a *wire and cache tapping adversary* of type II. The legitimate parties know neither whether cache placement, delivery, or both phases are tapped, nor the positions in which they are tapped. Only the size of the overall tapped set is known. For the instance of two receivers and two library files, we identify the strong secrecy capacity of the model, i.e.,

the maximum achievable file rate while keeping the overall library strongly secure. Lower and upper bounds on the achievable strong secrecy file rate are derived when the library has more than two files. In order to establish the achievability, we propose a code design which combines wiretap coding, security embedding codes, one-time pad keys, and coded caching. A genie-aided upper bound, in which a genie provides the transmitter with user demands before cache placement, establishes the converse for the two files instance. For the library of more than two files, the upper bound is constructed by three successive channel transformations. Our results establish that strong information theoretic security is possible against a powerful adversary which optimizes its attack over both phases of communication in a cache-aided system.

Chapter 2

Notation and Preliminaries

2.1 Notation

We first provide the notation we use throughout the remainder of the thesis. Scalars are denoted by lower-case letters while random variables are denoted using upper-case letters. Vectors are denoted by bold lower-case super-scripted letters, while their components are denoted by lower-case sub-scripted letter. Similar convention, but with upper-case letters, is used for random vectors and their components. The vector's super-script is dropped when its dimension is clear from the context. Matrices are also denoted by bold upper-case super-scripted letters. The distinction between matrices and random vectors should be clear from the context. Sets are denoted using calligraphic fonts.

We use $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$, to denote the sets of natural, rational, integer, real, and complex numbers, respectively. $\mathbb{Z}_{\mathbb{C}}$ denotes the set of complex integers, i.e., $\mathbb{Z}_{\mathbb{C}} \triangleq \{n + jm : n, m \in \mathbb{Z}\}$. For $Q \in \mathbb{Z}$, the set of integers $\{-Q, \cdots, Q\}$ is denoted by $(-Q, Q)_{\mathbb{Z}}$. For $a, b \in \mathbb{R}$, $[a : b]$ denotes the integers between a and b, i.e., $[a : b] \triangleq \{i \in \mathbb{Z} : \ a \leq i \leq b\}$.

For $S \subseteq \mathbb{N}$, $\mathbf{X}_S$ denotes the sequence $\{X_i\}_{i \in S}$. We use $A_{[1:n]}$ to denote the sequence of variables $\{A_1, A_2, \cdots, A_n\}$. For matrix $\mathbf{A}$, $\mathcal{N}(\mathbf{A})$ denotes its null space, $\det(\mathbf{A})$ denotes its determinant, and $||\mathbf{A}||$ denotes its *induced* norm. For vector $\mathbf{V}$, $||\mathbf{V}||$ denotes its Euclidean norm, and $\mathbf{V}_i^j$ denotes the ith to jth components in $\mathbf{V}$. We use $|.|$ to denote the cardinality or the absolute value, when used for a set or a real (complex)

number, respectively. We use $\mathbf{V}^n$ to denote the n-letter extension of the random vector $\mathbf{V}$, i.e., $\mathbf{V}^n = [\mathbf{V}(1)\,\mathbf{V}(2)\,\cdots\mathbf{V}(n)]$. The operators ^{T}, ^{H}, and † denote the transpose, Hermitian, and pseudo inverse operations. A circularly symmetric Gaussian random vector with zero mean and covariance matrix $\mathbf{K}$ is denoted by $\mathcal{CN}(\mathbf{0}, \mathbf{K})$.

All logarithms are taken to be base 2. We use $j = \sqrt{-1}$ to denote the imaginary unit in a complex number. $\mathbf{0}_{m\times n}$ denotes an $m \times n$ matrix of zeros, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. For two sets $\mathcal{A}_1$ and $\mathcal{A}_2$, we use $\mathcal{A}_1 \times \mathcal{A}_2$ to denote their Cartesian product. $\mathcal{A}^T$ denotes the T-fold Cartesian product of the set $\mathcal{A}$. For $W_1, W_2 \in [1 : M]$, $W_1 \oplus W_2$ denotes the bit-wise XOR on the binary bit strings that correspond to W_1 and W_2.

We use $\mathbb{1}\{\mathcal{A}\}$ to denote the indicator function of the event $\mathcal{A}$. We use upper-case letters to denote random probability distributions, e.g., P_X, and lower-case letters to denote deterministic probability distributions, e.g., p_X. We use p_X^U to denote a uniform distribution over the random variable X. The argument of the probability distribution is omitted when it is clear from its subscript, and vice versa. $\mathbb{V}(p_X, q_X)$ and $\mathbb{D}(p_X||q_X)$ denote the total variation distance and the Kullback-Leibler (K-L) divergence between the distributions p_X and q_X. We use $\{\epsilon_n\}_{n\geq 1}$ to denote a sequence of positive real numbers such that $\epsilon_n \to 0$ as $n \to \infty$.

2.2 Preliminaries

In this section, we provide the definitions and preliminary results we utilize throughout the remainder of the thesis.

2.2.1 Definitions and Results from Transcendental Number Theory

Transcendental number theory deals with transcendental numbers which are not solutions of any polynomial equation with integer coefficients. One corner result in the field of classification of transcendental complex numbers provides a bound on the absolute value of a complex algebraic number with rational coefficients in terms of its height, i.e., the maximum coefficient [59, 109, 110]. We utilize this result in Chapter 3 in order to extend real interference alignment [78] to complex channels.

In order to present the desired result, let us first define the Diophantine exponent of a length-n complex vector.

Definition 1. *[59]* The Diophantine exponent $\omega(\mathbf{z})$ of $\mathbf{z} \in \mathbb{C}^n$ is defined as

$$\omega(\mathbf{z}) \triangleq \sup \left\{ v : |p + \mathbf{z}.\mathbf{q}| \leq (||\mathbf{q}||_\infty)^{-v} \text{ for infinetly many } \mathbf{q} \in \mathbb{Z}^n, p \in \mathbb{Z} \right\}, \tag{2.1}$$

where $\mathbf{q} = [q_1 \; q_2 \; \cdots \; q_n]^T$ and $||\mathbf{q}||_\infty = \max_i |q_i|$.

Next, we state the following lemma, which determines the Diophantine exponent for almost all length-n complex vectors.

Lemma 1. *[59]* For almost all $\mathbf{z} \in \mathbb{C}^n$, the Diophantine exponent $\omega(\mathbf{z})$ is equal to $\frac{n-1}{2}$.

2.2.2 Distance Measures and Concentration Inequalities

We first present the following two measures, which we extensively utilize throughout the thesis.

Definition 2. *The total variation distance between two probability distributions p_X and q_X, defined on the same probability space, is given by*

$$\mathbb{V}(p_X, q_X) \triangleq \frac{1}{2} \sum_{x \in \mathcal{X}} |p(x) - q(x)|$$

$$= \sum_{x \in \mathcal{X}:\, p(x) > q(x)} [p(x) - q(x)]. \tag{2.2}$$

Definition 3. *The Kullback-Leibler divergence, or relative entropy, between the two distributions p_X and q_X, defined on the same probability space, is given by*

$$\mathbb{D}(p_X || q_X) \triangleq \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}. \tag{2.3}$$

The following Lemma states two properties for the total variation distance between two distributions defined on the same probability space.

Lemma 2. *(Properties of Total Variation Distance) [24, Lemmas V.1 and V.2]: Consider the joint distributions $p_{X,Y} = p_X p_{Y|X}$ and $q_{X,Y} = q_X q_{Y|X}$, defined on the same alphabet $\mathcal{X} \times \mathcal{Y}$. Then, we have,*

$$\mathbb{V}(p_X, q_X) \leq \mathbb{V}(p_{X,Y}, q_{X,Y}) \tag{2.4}$$

$$\mathbb{V}(p_X p_{Y|X}, q_X p_{Y|X}) = \mathbb{V}(p_X, q_X). \tag{2.5}$$

Next, we present *Hoeffding's inequality* and a variation on *Chernoff's inequality*. The two inequalities provide exponentially decaying upper bounds, with different degrees of strength, on the tail distribution of the sum of independent random variables.

We first state Hoeffding's inequality in the following lemma.

Lemma 3. *(Hoeffding's Inequality) [46, Theorem 2]:*

Let $U_1, U_2, \cdots, U_n$ be independent random variables with $U_i \in [0, b]$ for all $i \in [1 : n]$, and let $\bar{m} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(U_i)$. Then, for $\epsilon > 0$, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} U_i \leq (1-\epsilon)\bar{m}\right) \leq \exp\left(\frac{-2\epsilon^2 \bar{m}^2}{b^2} n\right). \tag{2.6}$$

Next, we state the variation on *Chernoff's bound*.

Lemma 4. *(A variation on Chernoff bound:) Let $U_1, U_2, \cdots, U_n$ be a sequence of non-negative independent random variables with respective means $\mathbb{E}(U_i) = \bar{m}_i$. If $U_i \in [0, b]$, for all $i \in [1 : n]$, and $\sum_{i=1}^{n} \bar{m}_i \leq \bar{m}$, then, for every $\epsilon \in [0, 1]$, we have*

$$\mathbb{P}\left(\sum_{i=1}^{n} U_i \geq (1+\epsilon)\bar{m}\right) \leq \exp\left(-\epsilon^2 \frac{\bar{m}}{3b}\right). \tag{2.7}$$

Proof: The proof is adapted from [33, Appendix C]. The details are provided in Appendix A. ■

2.2.3 Selection Lemma

The selection lemma below is utilized in the achievability proofs in Chapters 5, 6, and 7, for channel (source) coding problems, in order to show the existence of a codebook (a source binning realization) which simultaneously satisfies multiple constraints.

Lemma 5. *(Selection Lemma) [14, Lemma 2.2]:*

Let $A_1, A_2, \cdots, A_n$ be a sequence of random variables where $A_n \in \mathcal{A}^n$, and let $\mathcal{F}_n =$

$\{f_{1,n}, \cdots, f_{M,n}\}$ be a finite set of bounded functions $f_{i,n} : \mathcal{A}^n \mapsto \mathbb{R}^+$, $i \in [1:M]$, such that $|\mathcal{F}_n| = M$ does not depend on n, and

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{A}^n}\left(f_{i,n}(A_n)\right) = 0 \quad \textit{for all } i \in [1:M]. \tag{2.8}$$

Then, there exists a specific realization $\{a_n^*\}$ of the sequence $\{A_n\}$ such that

$$\lim_{n\to\infty} f_{i,n}(a_n^*) = 0 \quad \textit{for all } i \in [1:M]. \tag{2.9}$$

Chapter 3

Multiple Antenna Wiretap Channel with a Multi-antenna Cooperative Jammer

3.1 Introduction

The secrecy capacity region for most of multi-terminal models remain open despite significant progress on bounds and associated insights. Recent work thus includes efforts that concentrate on characterizing the more tractable high signal-to-noise ratio (SNR) scaling behavior of secrecy capacity region for Gaussian multi-terminal models [39, 45, 56, 122–124]. Among the multi-transmitter models studied, a recurrent theme in achievability is enlisting one or more terminals to transmit intentional interference with the specific goal of diminishing the reception capability of the eavesdropper, known as *cooperative jamming* [111]. For the Gaussian wiretap channel, adding a cooperative jammer terminal transmitting Gaussian noise can improve the secrecy rate considerably [113], albeit not the scaling of the secrecy capacity with power at high SNR. Recently, reference [45] has shown that, for the Gaussian wiretap channel, adding a cooperative jammer and utilizing structured codes for message transmission and cooperative jamming, i.e., transmitted signals from discrete constellations, provide an achievable secrecy rate scalable with power, i.e., a positive secure degrees of freedom (s.d.o.f.), an improvement from the zero degrees of freedom of the Gaussian wiretap channel. More recently,

reference [122] has proved that, for this channel, the s.d.o.f. $\frac{1}{2}$, achievable by codebooks constructed from integer lattices along with real interference alignment, is tight. References [123, 124] have subsequently identified the s.d.o.f. region for multi-terminal Gaussian wiretap channel models.

While the above development is for single-antenna terminals, multiple antennas have also been utilized to improve secrecy rates and s.d.o.f. for several channel models [27, 38, 44, 55, 56, 58, 68, 69, 94, 104]. The secrecy capacity of the multi-antenna (MIMO) wiretap channel, identified in [58], scales with power only when the legitimate transmitter has an advantage over the eavesdropper in the number of antennas. It then follows naturally to utilize a cooperative jamming terminal to improve the secrecy rate and scaling for multi-antenna wiretap channels as well which is the focus of this chapter.

In this chapter, we study the multi-antenna wiretap channel with a multi-antenna cooperative jammer. We characterize the high SNR scaling of the secrecy capacity, i.e., the secure degrees of freedom (s.d.o.f.), of the channel with N_c antennas at the cooperative jammer, N_t antennas at the transmitter, N_r antennas at the receiver, N_e antennas at the eavesdropper, under the assumption of known channel state information at all terminals. The achievability and converse techniques both are methodologically developed for ranges of the parameters, i.e., the number of antennas at each terminal. The upper and lower bounds for all parameter values are shown to match one another.

We remark that secure degrees of freedom for single and multiple antenna wiretap channels have recently been investigated under the assumption of unknown eavesdropper channel state information at the legitimate terminals. The secure degrees of freedom for

the single-antenna wiretap channel with multiple helpers, multiple-access wiretap channel, and interference wiretap channel, with unknown and static eavesdropper channel, have been derived in [79]. The strongly secure degrees of freedom of the multiple antenna wiretap channel with unknown and *varying* eavesdropper channel is established in [44] by showing the existence of a universal scheme that can counter any eavesdropper state. [44] thus quantifies the reduction in degrees of freedom that results from universal immunity to eavesdropping. This work, by contrast, addresses the improvement provided by adding a *multi-antenna helper* in the benchmark case that is the static and known channel state information for the MIMO wiretap channel.

The proposed achievable schemes for different ranges of the values for N_c, N_t, N_r, and N_e all involve linear precoding and linear receiver processing. The common goal to all these schemes is to perfectly align the cooperative jamming signals over the information signals observed at the eavesdropper while simultaneously enabling information and cooperative jamming signal separation at the legitimate receiver. We show that whenever the s.d.o.f. of the channel is integer valued, Gaussian signaling both at the transmitter and the cooperative jammer suffices to achieve the s.d.o.f. By contrast, non-integer s.d.o.f. requires structured signaling, i.e., signals from discrete constellations, along with joint signal space and signal scale alignment in the complex plane [59,71]. The necessity of structured signaling follows from the fact that fractional s.d.o.f. indicates sharing at least one spatial dimension between information and cooperative jamming signals at the receiver's signal space. In this case, sharing the same spatial dimension between Gaussian information and jamming signals, which have similar power scaling, does not provide positive degrees of freedom, and we need for structured signals that can

be separated over this single dimension at high SNR. The tools that enable the signal scale alignment are available in the field of transcendental number theory [59, 109, 110], which we utilize.

Overall, this study determines the value in jointly utilizing signal scale and spatial interference alignment techniques for secrecy and quantifies the impact of a multi-antenna helper for the MIMO wiretap channel by settling the question of the secrecy prelog for the $(N_t \times N_r \times N_e)$ MIMO wiretap channel in the presence of an N_c-antenna cooperative jammer, for all possible values of N_c. In contrast with the single antenna case, where integer lattice codes and real interference alignment suffice to achieve the s.d.o.f. of the channel, in the MIMO setting, one needs to utilize a variety of signaling, beam-forming, and alignment techniques, in order to coordinate the transmitted and received signals for different values of N_t, N_r, N_e, and N_c.

The remainder of the chapter is organized as follows. Section 3.2 introduces the channel model, and Section 3.3 provides the main results. For clarity of exposition, we first present the converse and achievability for the MIMO wiretap channel with $N_t = N_r = N$ in Sections 3.4 and 3.5. Section 3.6 then extends the converse and achievability proofs for the case $N_t \neq N_r$. Section 3.7 discusses the results of this work and Section 3.8 concludes the chapter.

3.2 Channel Model

We consider the MIMO wiretap channel with an N_t-antenna transmitter, N_r-antenna receiver, N_e-antenna eavesdropper, and an N_c-antenna cooperative jammer as depicted in Fig. 3.1.

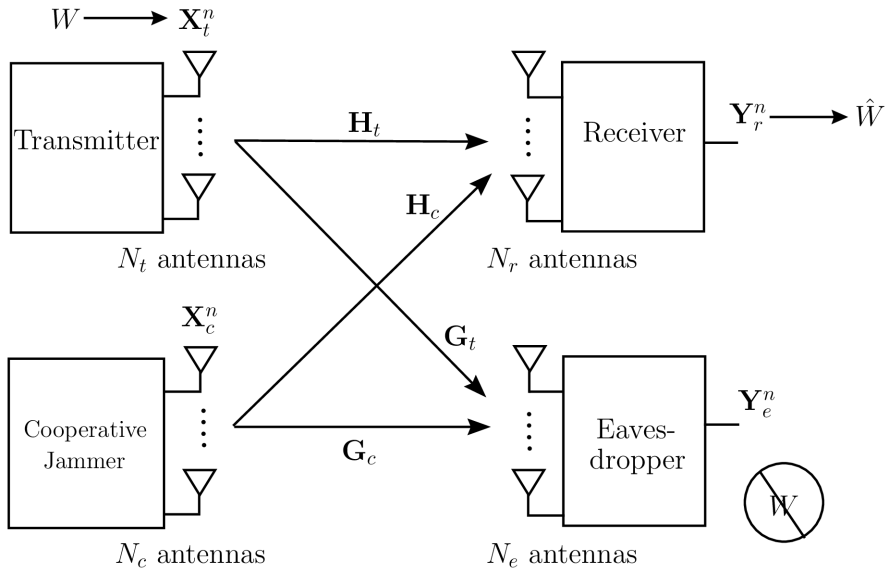


Fig. 3.1. $(N_t \times N_r \times N_e)$ multiple antenna wiretap channel with an N_c-antenna cooperative jammer.

The received signals at the receiver and eavesdropper, at the nth channel use, are given by

$$\mathbf{Y}_r(n) = \mathbf{H}_t\mathbf{X}_t(n) + \mathbf{H}_c\mathbf{X}_c(n) + \mathbf{Z}_r(n) \tag{3.1}$$

$$\mathbf{Y}_e(n) = \mathbf{G}_t\mathbf{X}_t(n) + \mathbf{G}_c\mathbf{X}_c(n) + \mathbf{Z}_e(n), \tag{3.2}$$

where $\mathbf{X}_t(n)$ and $\mathbf{X}_c(n)$ are the transmitted signals from the transmitter and the cooperative jammer at the nth channel use. $\mathbf{H}_t \in \mathbb{C}^{N_r \times N_t}$, $\mathbf{H}_c \in \mathbb{C}^{N_r \times N_c}$ are the channel gain matrices from the transmitter and the cooperative jammer to the receiver, while $\mathbf{G}_t \in \mathbb{C}^{N_e \times N_t}$, $\mathbf{G}_c \in \mathbb{C}^{N_e \times N_c}$ are the channel gain matrices from the transmitter and the cooperative jammer to the eavesdropper. It is assumed that the channel gains are drawn independently from a *complex-valued* continuous distribution. All channel gains

are assumed to be known at all terminals. $\mathbf{Z}_r(n)$ and $\mathbf{Z}_e(n)$ are the complex Gaussian noise at the receiver and eavesdropper at the nth channel use, where $\mathbf{Z}_r(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$ and $\mathbf{Z}_e(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_e})$ for all n. $\mathbf{Z}_r(n)$ is independent from $\mathbf{Z}_e(n)$ and both are independent and identically distributed (i.i.d.) across the time index[1] n. The power constraints on the transmitted signals at the transmitter and the cooperative jammer are $\mathbb{E}\left(\mathbf{X}_t^H \mathbf{X}_t\right), \mathbb{E}\left(\mathbf{X}_c^H \mathbf{X}_c\right) \leq P$.

The transmitter aims to send a message W to the receiver, and keep it secret from the external eavesdropper. A stochastic encoder, which maps the message W to the transmitted signal $\mathbf{X}_t^n \in \mathcal{X}_t^n$, is used at the transmitter. The receiver uses its observation, $\mathbf{Y}_r^n \in \mathcal{Y}_r^n$, to obtain an estimate $\hat{W}$ of the transmitted message. Secrecy rate R_s is achievable if for any $\epsilon > 0$, there is a channel code $(2^{nR_s}, n)$ satisfying[2]

$$P_e \triangleq \mathbb{P}\left(\hat{W} \neq W\right) \leq \epsilon, \tag{3.3}$$

$$\frac{1}{n} H(W|\mathbf{Y}_e^n) \geq \frac{1}{n} H(W) - \epsilon. \tag{3.4}$$

The secrecy capacity of a channel, $C_s(P)$, is defined as the closure of all its achievable secrecy rates. For a channel with complex-valued coefficients, the maximum secure degrees of freedom (s.d.o.f.) is defined as

$$D_s \triangleq \lim_{P \to \infty} \frac{C_s(P)}{\log P}. \tag{3.5}$$

[1] Throughout the chapter, we omit the index n whenever possible.

[2] We consider weak secrecy throughout this chapter.

The cooperative jammer transmits the signal $\mathbf{X}_c^n \in \mathcal{X}_c^n$ in order to reduce the reception capability of the eavesdropper. However, this transmission affects the receiver as well, as interference. The jamming signal, $\mathbf{X}_c^n$, does not carry any information. Additionally, there is no shared secret between the transmitter and the cooperative jammer.

3.3 Main Result

We first state the s.d.o.f. results for $N_t = N_r = N$.

Theorem 1. The s.d.o.f. of the MIMO wiretap channel with an N_c-antenna cooperative jammer, N antennas at each of the transmitter and receiver, and N_e antennas at the eavesdropper, is almost surely[3] (a.s.)

$$D_s = \begin{cases} [N + N_c - N_e]^+, & \text{if } 0 \leq N_c \leq N_e - \frac{\min\{N,N_e\}}{2} \\ N - \frac{\min\{N,N_e\}}{2}, & \text{if } N_e - \frac{\min\{N,N_e\}}{2} < N_c \leq \max\{N, N_e\} \\ \frac{N+N_c-N_e}{2}, & \text{if } \max\{N, N_e\} < N_c \leq N + N_e, \\ N, & \text{if } N_c > N + N_e \end{cases} \tag{3.6}$$

Proof: The proof for Theorem 1 is provided in Sections 3.4 and 3.5. ■

Next, in Theorem 2 below, we generalize the result in Theorem 1 to $N_t \neq N_r$.

Theorem 2. The s.d.o.f. of the MIMO wiretap channel with an N_c-antenna cooperative jammer, N_t-antenna transmitter, N_r-antenna receiver, and N_e-antenna eavesdropper, is

[3]The subset of the channel gains for which the result does not hold has a Lebesgue measure zero.

a.s.

$$D_s = \begin{cases} \min\left\{N_r, [N_c + N_t - N_e]^+\right\}, & \text{if } 0 \le N_c \le N_1 \\ \min\left\{N_t, N_r, \frac{N_r + [N_t - N_e]^+}{2}\right\}, & \text{if } N_1 < N_c \le N_2 \\ \min\left\{N_t, N_r, \frac{N_c + N_t - N_e}{2}\right\}, & \text{if } N_2 < N_c \le N_3, \\ \min\left\{N_t, N_r\right\}, & \text{if } N_c > N_3, \end{cases} \tag{3.7}$$

where,

$$N_1 = \min\left\{N_e, \left[\frac{N_r}{2} + \frac{N_e - N_t}{2 - 1_{N_e > N_t}}\right]^+\right\}, \qquad 1_{N_e > N_t} = \begin{cases} 1, & \text{if } N_e > N_t \\ 0, & \text{if } N_e \le N_t \end{cases}$$

$$N_2 = N_r + [N_e - N_t]^+, \qquad N_3 = \max\left\{N_2, 2\min\left\{N_t, N_r\right\} + N_e - N_t\right\}.$$

Proof: The proof for Theorem 2 is provided in Section 3.6. ■

Remark 1. *Theorem 2 provides a complete characterization for the s.d.o.f. of the channel. The s.d.o.f. at $N_c = N_3$ is equal to $\min\{N_t, N_r\}$, which is equal to the d.o.f of the $(N_t \times N_r)$ point-to-point MIMO Gaussian channel. Thus, increasing the number of antennas at the cooperative jammer, N_c, over N_3 cannot increase the s.d.o.f. over $\min\{N_t, N_r\}$.*

Remark 2. *For $N_t \ge N_r + N_e$, the s.d.o.f. of the channel is equal to N_r at $N_c = 0$, i.e., the maximum s.d.o.f. of the channel is achieved without the help of the cooperative jammer.*

Remark 3. *The converse proof for Theorem 2 involves combining two upper bounds for the s.d.o.f. derived for two different ranges of N_c. These two bounds are a straightforward generalization of those derived for the symmetric case in Theorem 1. However, combining them is more tedious since more cases of the number of antennas at the different terminals should be handled carefully. Achievability for Theorem 2 utilizes similar techniques to those used for Theorem 1 as well, where handling more cases is required. For clarity of exposition, we derive the s.d.o.f. for the symmetric case first in order to present the main ideas, and then utilize these ideas and generalize the result to the asymmetric case of Theorem 2.*

For illustration purposes, the s.d.o.f. for $N_t = N_r = N_e = N$, and N_c varying from 0 to $2N$, is depicted in Fig. 3.2. The s.d.o.f. curves with N even and odd are shown in Fig. 3.2a and Fig. 3.2b, respectively.

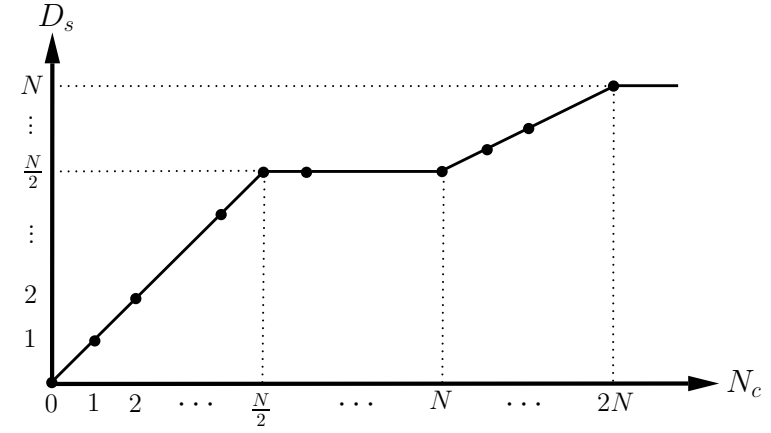
We provide the discussion of the results of this work in Section 3.7.

3.4 Converse for $N_t = N_r = N$

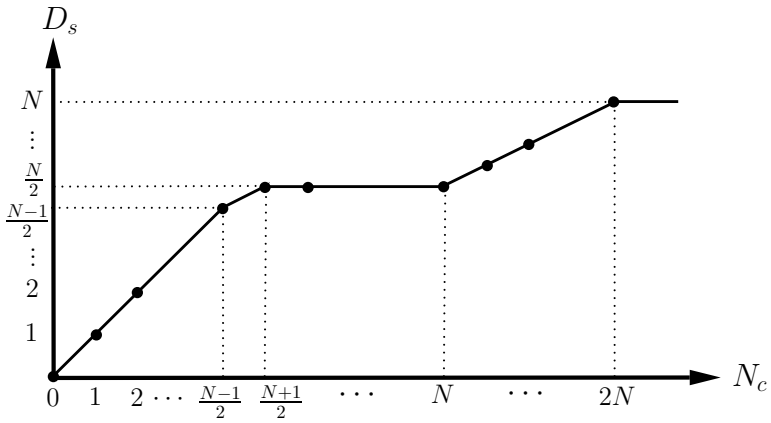
In Section 3.4.1, we derive the upper bound for the s.d.o.f. for $0 \leq N_c \leq N_e$. In Section 3.4.2, we derive the upper bound for $\max\{N, N_e\} \leq N_c \leq N + N_e$. The two bounds are combined in Section 3.4.3 to provide the desired upper bound in (3.6).

3.4.1 $0 \leq N_c \leq N_e$

Allow for full cooperation between the transmitter and the cooperative jammer. This cooperation cannot decrease the s.d.o.f. of the channel, and yields a MIMO wiretap



N even.



N odd.

Fig. 3.2. Secure degrees of freedom for a MIMO wiretap channel, with N antennas at each of its nodes, and a cooperative jammer with N_c antennas, where N_c varies from 0 to $2N$.

channel with $N + N_c$-antenna transmitter, N-antenna receiver, and N_e-antenna eavesdropper. It has been shown in [58] that, at high SNR, i.e., $P \to \infty$, the secrecy capacity of this channel, $C_s(P)$, takes the asymptotic form

$$C_s(P) = \log\det\left(\mathbf{I}_N + \frac{P}{p}\mathbf{H}\mathbf{G}^{\sharp}\mathbf{H}^H\right) + o(\log P), \tag{3.8}$$

where $\lim_{P\to\infty} \frac{o(\log P)}{\log P} = 0$, $\mathbf{H} \in \mathbb{C}^{N\times(N+N_c)}$ and $\mathbf{G} \in \mathbb{C}^{N_e\times(N+N_c)}$ are the channel gains from the combined transmitter to the receiver and eavesdropper, and $\mathbf{G}^{\sharp}$ is the projection matrix onto the null space of $\mathbf{G}$, $\mathcal{N}(\mathbf{G})$. $p \triangleq \dim\left\{\mathcal{N}(\mathbf{H})^{\perp} \cap \mathcal{N}(\mathbf{G})\right\}$, where $\mathcal{N}(\mathbf{H})^{\perp}$ is the space orthogonal to the null space of $\mathbf{H}$. Due to the randomly generated channel gains, if a vector $\mathbf{x} \in \mathcal{N}(\mathbf{G})$, then $\mathbf{x} \in \mathcal{N}(\mathbf{H})^{\perp}$ a.s., for all $0 \le N_c \le N_e$. Thus, $p = \dim(\mathcal{N}(\mathbf{G})) = [N + N_c - N_e]^+$.

$\mathbf{H}\mathbf{G}^{\sharp}\mathbf{H}^H$ can be decomposed as

$$\mathbf{H}\mathbf{G}^{\sharp}\mathbf{H}^H = \mathbf{\Psi}\begin{bmatrix} \mathbf{0}_{(N-p)\times(N-p)} & \mathbf{0}_{(N-p)\times p} \\ \mathbf{0}_{p\times(N-p)} & \mathbf{\Omega} \end{bmatrix}\mathbf{\Psi}^H, \tag{3.9}$$

where $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ is a unitary matrix and $\mathbf{\Omega} \in \mathbb{C}^{p\times p}$ is a non-singular matrix [58]. Let $\mathbf{\Psi} = [\mathbf{\Psi}_1\ \mathbf{\Psi}_2]$, where $\mathbf{\Psi}_1 \in \mathbb{C}^{N\times(N-p)}$ and $\mathbf{\Psi}_2 \in \mathbb{C}^{N\times p}$. Substituting (3.9) in (3.8) yields

$$C_s(P) = \log\det\left(\mathbf{I}_N + \frac{P}{p}\mathbf{\Psi}_2\mathbf{\Omega}\mathbf{\Psi}_2^H\right) + o(\log P) \tag{3.10}$$

$$= \log\det\left(\mathbf{I}_p + \frac{P}{p}\mathbf{\Omega}\mathbf{\Psi}_2^H\mathbf{\Psi}_2\right) + o(\log P) \tag{3.11}$$

$$= \log P^p \det\left(\frac{1}{P}\mathbf{I}_p + \frac{1}{p}\mathbf{\Omega}\right) + o(\log P) \tag{3.12}$$

$$= p \log P + o(\log P), \tag{3.13}$$

where (3.11) follows from Sylvester's determinant identity and (3.12) follows from $\mathbf{\Psi}$ being unitary.

Thus, the s.d.o.f. of the original channel, for $0 \leq N_c \leq N_e$, is upper bounded as

$$D_s \leq \lim_{P\to\infty} \frac{C_s(P)}{\log P} = \lim_{P\to\infty} \frac{p \log P + o(\log P)}{\log P} \tag{3.14}$$

$$= [N + N_c - N_e]^+. \tag{3.15}$$

3.4.2 $\max\{N, N_e\} < N_c \leq N + N_e$

The upper bound we derive here is inspired by the converse of the single antenna Gaussian wiretap channel with a single antenna cooperative jammer derived in [122], though as we will see shortly, the vector channel extension resulting from multiple antennas does require care. Let ϕ_i, for $i = 1, 2, \cdots, 10$, denote constants which do not depend on the power P.

The secrecy rate R_s can be upper bounded as follows

$$nR_s = H(W) \tag{3.16}$$

$$= H(W) - H(W|\mathbf{Y}_e^n) + H(W|\mathbf{Y}_e^n) - H(W|\mathbf{Y}_r^n) + H(W|\mathbf{Y}_r^n) \tag{3.17}$$

$$\leq n\epsilon + H(W|\mathbf{Y}_e^n) - H(W|\mathbf{Y}_r^n, \mathbf{Y}_e^n) + n\delta \tag{3.18}$$

$$= I(W; \mathbf{Y}_r^n|\mathbf{Y}_e^n) + n\phi_1 \tag{3.19}$$

$$= h(\mathbf{Y}_r^n|\mathbf{Y}_e^n) - h(\mathbf{Y}_r^n|W, \mathbf{Y}_e^n) + n\phi_1 \tag{3.20}$$

$$\leq h(\mathbf{Y}_r^n|\mathbf{Y}_e^n) - h(\mathbf{Y}_r^n|W, \mathbf{Y}_e^n, \mathbf{X}_t^n, \mathbf{X}_c^n) + n\phi_1 \tag{3.21}$$

$$= h(\mathbf{Y}_r^n, \mathbf{Y}_e^n) - h(\mathbf{Y}_e^n) - h(\mathbf{Z}_r^n) + n\phi_1, \tag{3.22}$$

where (3.18) follows since $H(W) - H(W|\mathbf{Y}_e^n) \leq n\epsilon$ by the secrecy constraint in (3.4), $H(W|\mathbf{Y}_r^n) \leq n\delta$ by Fano's inequality, and $H(W|\mathbf{Y}_r^n) \geq H(W|\mathbf{Y}_r^n, \mathbf{Y}_e^n)$ by the fact that conditioning does not increase entropy, (3.22) follows since $\mathbf{Z}_r^n$ is independent from $\{W, \mathbf{Y}_e^n, \mathbf{X}_t^n, \mathbf{X}_c^n\}$, and $\phi_1 = \epsilon + \delta$.

Let $\tilde{\mathbf{X}}_t = \mathbf{X}_t + \tilde{\mathbf{Z}}_t$ and $\tilde{\mathbf{X}}_c = \mathbf{X}_c + \tilde{\mathbf{Z}}_c$, where $\tilde{\mathbf{Z}}_t \sim \mathcal{CN}(\mathbf{z}, \mathbf{K}_t)$ and $\tilde{\mathbf{Z}}_c \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_c)$. The covariance matrices, $\mathbf{K}_t$ and $\mathbf{K}_c$, are chosen as $\mathbf{K}_t = \rho^2\mathbf{I}_N$ and $\mathbf{K}_c = \rho^2\mathbf{I}_{N_c}$, where $0 < \rho \leq 1/\max\left\{||\mathbf{H}_c^H||, \sqrt{||\mathbf{G}_t^H||^2 + ||\mathbf{G}_c^H||^2}\right\}$. Note that $\tilde{\mathbf{X}}_t$ and $\tilde{\mathbf{X}}_c$ are noisy versions of the transmitted signals $\mathbf{X}_t$ and $\mathbf{X}_c$, respectively. $\tilde{\mathbf{Z}}_t$ is independent from $\tilde{\mathbf{Z}}_c$ and both are independent from $\{\mathbf{X}_t, \mathbf{X}_c, \mathbf{Z}_r, \mathbf{Z}_e\}$. $\tilde{\mathbf{Z}}_t^n$ and $\tilde{\mathbf{Z}}_c^n$ are i.i.d. sequences of the random vectors $\tilde{\mathbf{Z}}_t$ and $\tilde{\mathbf{Z}}_c$. In addition, let $\tilde{\mathbf{Z}}_1 = -\mathbf{H}_t\tilde{\mathbf{Z}}_t - \mathbf{H}_c\tilde{\mathbf{Z}}_c + \mathbf{Z}_r$ and $\tilde{\mathbf{Z}}_2 = -\mathbf{G}_t\tilde{\mathbf{Z}}_t - \mathbf{G}_c\tilde{\mathbf{Z}}_c + \mathbf{Z}_e$. Note that $\tilde{\mathbf{Z}}_1 \sim \mathcal{CN}\left(\mathbf{z}, \mathbf{\Sigma}_{\tilde{\mathbf{Z}}_1}\right)$ and $\tilde{\mathbf{Z}}_2 \sim \mathcal{CN}\left(\mathbf{z}, \mathbf{\Sigma}_{\tilde{\mathbf{Z}}_2}\right)$, where $\mathbf{\Sigma}_{\tilde{\mathbf{Z}}_1} = \mathbf{H}_t\mathbf{K}_t\mathbf{H}_t^H + \mathbf{H}_c\mathbf{K}_c\mathbf{H}_c^H + \mathbf{I}_N$ and $\mathbf{\Sigma}_{\tilde{\mathbf{Z}}_2} = \mathbf{G}_t\mathbf{K}_t\mathbf{G}_t^H + \mathbf{G}_c\mathbf{K}_c\mathbf{G}_c^H + \mathbf{I}_{N_e}$. $\tilde{\mathbf{Z}}_1^n$ and $\tilde{\mathbf{Z}}_2^n$ are i.i.d. sequences of $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$, since each of $\mathbf{Z}_r^n, \mathbf{Z}_e^n, \tilde{\mathbf{Z}}_t^n, \tilde{\mathbf{Z}}_c^n$ is i.i.d. across time. The choice of $\mathbf{K}_t$ and $\mathbf{K}_c$ above guarantees the finiteness of $h(\tilde{\mathbf{Z}}_t), h(\tilde{\mathbf{Z}}_c), h(\tilde{\mathbf{Z}}_1)$, and $h(\tilde{\mathbf{Z}}_2)$ as shown in Appendix B. Starting from (3.22), we have

$$nR_s \leq h(\mathbf{Y}_r^n, \mathbf{Y}_e^n) - h(\mathbf{Y}_e^n) + n\phi_2 \tag{3.23}$$

$$= h(\mathbf{Y}_r^n, \mathbf{Y}_e^n, \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) - h(\tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n|\mathbf{Y}_r^n, \mathbf{Y}_e^n) - h(\mathbf{Y}_e^n) + n\phi_2 \tag{3.24}$$

$$\leq h(\tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) + h(\mathbf{Y}_r^n, \mathbf{Y}_e^n | \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) - h(\tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n | \mathbf{Y}_r^n, \mathbf{Y}_e^n, \mathbf{X}_t^n, \mathbf{X}_c^n) - h(\mathbf{Y}_e^n) + n\phi_2 \tag{3.25}$$

$$\leq h(\tilde{\mathbf{X}}_t^n) + h(\tilde{\mathbf{X}}_c^n) + h(\mathbf{Y}_r^n | \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) + h(\mathbf{Y}_e^n | \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) - h(\tilde{\mathbf{Z}}_t^n, \tilde{\mathbf{Z}}_c^n) - h(\mathbf{Y}_e^n) + n\phi_2 \tag{3.26}$$

$$= h(\tilde{\mathbf{X}}_t^n) + h(\tilde{\mathbf{X}}_c^n) + h(\tilde{\mathbf{Z}}_1^n | \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) + h(\tilde{\mathbf{Z}}_2^n | \tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_c^n) - h(\mathbf{Y}_e^n) + n\phi_3 \tag{3.27}$$

$$\leq h(\tilde{\mathbf{X}}_t^n) + h(\tilde{\mathbf{X}}_c^n) + h(\tilde{\mathbf{Z}}_1^n) + h(\tilde{\mathbf{Z}}_2^n) - h(\mathbf{Y}_e^n) + n\phi_3 \tag{3.28}$$

$$= h(\tilde{\mathbf{X}}_t^n) + h(\tilde{\mathbf{X}}_c^n) - h(\mathbf{Y}_e^n) + n\phi_4, \tag{3.29}$$

where (3.26) follows since $\tilde{\mathbf{Z}}_t^n$ and $\tilde{\mathbf{Z}}_c^n$ are independent from $\{\mathbf{X}_t^n, \mathbf{X}_c^n, \mathbf{Y}_r^n, \mathbf{Y}_e^n\}$, $\phi_2 = \phi_1 - h(\mathbf{Z}_r)$, $\phi_3 = \phi_2 - h(\tilde{\mathbf{Z}}_t) - h(\tilde{\mathbf{Z}}_c)$, and $\phi_4 = \phi_3 + h(\tilde{\mathbf{Z}}_1) + h(\tilde{\mathbf{Z}}_2)$.

We have utilized the noisy versions $\tilde{\mathbf{X}}_t = \mathbf{X}_t + \tilde{\mathbf{Z}}_t$ and $\tilde{\mathbf{X}}_c = \mathbf{X}_c + \tilde{\mathbf{Z}}_c$ instead of $\mathbf{X}_t$, $\mathbf{X}_c$ so that (3.24)-(3.29) hold whether $\mathbf{X}_t, \mathbf{X}_c$ are continuous or discrete random vectors. This requires continuing the analysis with stochastically equivalent versions of $\mathbf{Y}_r, \mathbf{Y}_e$ in which they are expressed as functions of $\tilde{\mathbf{X}}_t$ and/or $\tilde{\mathbf{X}}_c$. To do so, we divide the Gaussian noise $\mathbf{Z}_r, \mathbf{Z}_e$ into sums of other independent Gaussian noise variables. The infinite divisibility of the Gaussian distribution ensures such division of $\mathbf{Z}_r, \mathbf{Z}_e$. We now consider the following two cases.

Case 1: $N_e \leq N$

We first lower bound $h(\mathbf{Y}_e^n)$ in (3.29) as follows. Using the infinite divisibility of Gaussian distribution, we can express a stochastically equivalent form of $\mathbf{Z}_e$, denoted by

$\mathbf{Z}'_e$, as

$$\mathbf{Z}'_e = \mathbf{G}_t \tilde{\mathbf{Z}}_t + \tilde{\mathbf{Z}}_e. \tag{3.30}$$

where[4] $\tilde{\mathbf{Z}}_e \sim \mathcal{CN}(\mathbf{z}, \mathbf{I}_{N_e} - \mathbf{G}_t \mathbf{K}_t \mathbf{G}_t^H)$ is independent from $\{\tilde{\mathbf{Z}}_t, \tilde{\mathbf{Z}}_c, \mathbf{X}_t, \mathbf{X}_c, \mathbf{Z}_r\}$. $\tilde{\mathbf{Z}}_e^n$ is an i.i.d. sequence of the random vectors $\tilde{\mathbf{Z}}_e$. Using (3.30), a stochastically equivalent form of $\mathbf{Y}_e^n$ is

$$\mathbf{Y}_e'^n = \mathbf{G}_t \tilde{\mathbf{X}}_t^n + \mathbf{G}_c \mathbf{X}_c^n + \tilde{\mathbf{Z}}_e^n. \tag{3.31}$$

Let $\mathbf{X}_t = \left[X_{t,1} \cdots X_{t,N}\right]^T$, $\tilde{\mathbf{Z}}_t = [\tilde{Z}_{t,1} \cdots \tilde{Z}_{t,N}]^T$, and $\tilde{\mathbf{X}}_t = [\tilde{\mathbf{X}}_{t_1}^T \ \tilde{\mathbf{X}}_{t_2}^T]^T$, where $\tilde{\mathbf{X}}_{t_1} = [\tilde{X}_{t,1} \cdots \tilde{X}_{t,N_e}]^T$, $\tilde{\mathbf{X}}_{t_2} = [\tilde{X}_{t,N_e+1} \cdots \tilde{X}_{t,N}]^T$, and $\tilde{X}_{t,k} = X_{t,k} + \tilde{Z}_{t,k}$, $k = 1, 2, \cdots, N$. In addition, let $\mathbf{G}_t = \left[\mathbf{G}_{t_1} \ \mathbf{G}_{t_2}\right]$, where $\mathbf{G}_{t_1} \in \mathbb{C}^{N_e \times N_e}$ and[5] $\mathbf{G}_{t_2} \in \mathbb{C}^{N_e \times (N - N_e)}$. Using (3.31), we have

$$h(\mathbf{Y}_e^n) = h(\mathbf{Y}_e'^n) = h(\mathbf{G}_t \tilde{\mathbf{X}}_t^n + \mathbf{G}_c \mathbf{X}_c^n + \tilde{\mathbf{Z}}_e^n) \tag{3.32}$$

$$\geq h(\mathbf{G}_t \tilde{\mathbf{X}}_t^n) = h(\mathbf{G}_{t_1} \tilde{\mathbf{X}}_{t_1}^n + \mathbf{G}_{t_2} \tilde{\mathbf{X}}_{t_2}^n) \tag{3.33}$$

$$\geq h(\mathbf{G}_{t_1} \tilde{\mathbf{X}}_{t_1}^n + \mathbf{G}_{t_2} \tilde{\mathbf{X}}_{t_2}^n | \tilde{\mathbf{X}}_{t_2}^n) = h(\mathbf{G}_{t_1} \tilde{\mathbf{X}}_{t_1}^n | \tilde{\mathbf{X}}_{t_2}^n) \tag{3.34}$$

$$= h(\tilde{\mathbf{X}}_{t_1}^n | \tilde{\mathbf{X}}_{t_2}^n) + n \log |\det(\mathbf{G}_{t_1})|. \tag{3.35}$$

[4]The choice of $\mathbf{K}_t$ guarantees that $\mathbf{I}_{N_e} - \mathbf{G}_t \mathbf{K}_t \mathbf{G}_t^H$ is a valid covariance matrix.

[5]Note that when $N_e = N$, the vector $\tilde{\mathbf{X}}_{t_2}$ and the matrix $\mathbf{G}_{t_2}$ vanish and the analysis below holds in the same manner, by discarding $\tilde{\mathbf{X}}_{t_2}^n$ and $\mathbf{G}_{t_2}$.

where the inequality in (3.33) follows since $\{\mathbf{G}_t\tilde{\mathbf{X}}_t^n\}$ and $\{\mathbf{G}_c\mathbf{X}_c^n+\tilde{\mathbf{Z}}_e^n\}$ are independent, as for two independent random vectors $\mathbf{X}$ and $\mathbf{Y}$, we have $h(\mathbf{X}+\mathbf{Y})\geq h(\mathbf{X})$.

Substituting (3.35) in (3.29) results in

$$nR_s \leq h(\tilde{\mathbf{X}}_{t_1}^n,\tilde{\mathbf{X}}_{t_2}^n)+h(\tilde{\mathbf{X}}_c^n)-h(\tilde{\mathbf{X}}_{t_1}^n|\tilde{\mathbf{X}}_{t_2}^n)-n\log|\det(\mathbf{G}_{t_1})|+n\phi_4 \tag{3.36}$$

$$= h(\tilde{\mathbf{X}}_{t_2}^n)+h(\tilde{\mathbf{X}}_c^n)+n\phi_5, \tag{3.37}$$

where $\phi_5=\phi_4-\log|\det(\mathbf{G}_{t_1})|$.

We now exploit the reliability constraint in (3.3) to derive another upper bound for R_s, which we combine with the bound in (3.37) in order to obtain the desired bound for the s.d.o.f. when $N_e\leq N$ and $N<N_c\leq N+N_e$. The reliability constraint in (3.3) can be achieved only if [21]

$$nR_s \leq I(\mathbf{X}_t^n;\mathbf{Y}_r^n)=h(\mathbf{Y}_r^n)-h(\mathbf{Y}_r^n|\mathbf{X}_t^n) \tag{3.38}$$

$$= h(\mathbf{Y}_r^n)-h(\mathbf{H}_c\mathbf{X}_c^n+\mathbf{Z}_r^n). \tag{3.39}$$

Similar to (3.30), a stochastically equivalent form of $\mathbf{Z}_r$ is given by

$$\mathbf{Z}_r'=\mathbf{H}_c\tilde{\mathbf{Z}}_c+\tilde{\mathbf{Z}}_r, \tag{3.40}$$

where[6] $\tilde{\mathbf{Z}}_r\sim\mathcal{CN}(\mathbf{z},\mathbf{I}_N-\mathbf{H}_c\mathbf{K}_c\mathbf{H}_c^H)$ is independent from $\{\tilde{\mathbf{Z}}_t,\tilde{\mathbf{Z}}_c,\mathbf{X}_t,\mathbf{X}_c,\mathbf{Z}_e\}$. $\tilde{\mathbf{Z}}_r^n$ is an i.i.d. sequence of the random vectors $\tilde{\mathbf{Z}}_r$.

[6]The choice of $\mathbf{K}_c$ guarantees that $\mathbf{I}_N-\mathbf{H}_c\mathbf{K}_c\mathbf{H}_c^H$ is a valid covariance matrix.

Let $\mathbf{X}_c = \left[X_{c,1} \cdots X_{c,N_c}\right]^T$, $\tilde{\mathbf{Z}}_c = [\tilde{Z}_{c,1} \cdots \tilde{Z}_{c,N_c}]^T$, and $\tilde{\mathbf{X}}_c = [\tilde{\mathbf{X}}_{c_1}^T \ \tilde{\mathbf{X}}_{c_2}^T]^T$, where $\tilde{\mathbf{X}}_{c_1} = [\tilde{X}_{c,1} \ \ldots \ \tilde{X}_{c,N}]^T$, $\tilde{\mathbf{X}}_{c_2} = [\tilde{X}_{c,N+1} \ \cdots \ \tilde{X}_{c,N_c}]^T$, and $\tilde{X}_{c,k} = X_{c,k} + \tilde{Z}_{c,k}$, $k = 1, 2, \cdots, N_c$. In addition, let $\mathbf{H}_c = \left[\mathbf{H}_{c_1} \ \mathbf{H}_{c_2}\right]$, where $\mathbf{H}_{c_1} \in \mathbb{C}^{N \times N}$ and $\mathbf{H}_{c_2} \in \mathbb{C}^{N \times (N_c - N)}$. Using (3.40), we have

$$h(\mathbf{H}_c\mathbf{X}_c^n + \mathbf{Z}_r^n) = h(\mathbf{H}_c\mathbf{X}_c^n + \mathbf{Z}_r'^n) = h(\mathbf{H}_c\tilde{\mathbf{X}}_c^n + \tilde{\mathbf{Z}}_r^n) \tag{3.41}$$

$$\geq h(\mathbf{H}_c\tilde{\mathbf{X}}_c^n) = h(\mathbf{H}_{c_1}\tilde{\mathbf{X}}_{c_1}^n + \mathbf{H}_{c_2}\tilde{\mathbf{X}}_{c_2}^n) \tag{3.42}$$

$$\geq h(\mathbf{H}_{c_1}\tilde{\mathbf{X}}_{c_1}^n|\tilde{\mathbf{X}}_{c_2}^n) \tag{3.43}$$

$$= h(\tilde{\mathbf{X}}_{c_1}^n|\tilde{\mathbf{X}}_{c_2}^n) + n\log|\det(\mathbf{H}_{c_1})|. \tag{3.44}$$

Substituting (3.44) in (3.39) yields

$$nR_s \leq h(\mathbf{Y}_r^n) - h(\tilde{\mathbf{X}}_{c_1}^n|\tilde{\mathbf{X}}_{c_2}^n) - n\log|\det(\mathbf{H}_{c_1})|. \tag{3.45}$$

Let $\mathbf{Y}_r = \left[Y_{r,1} \ \cdots \ Y_{r,N}\right]^T$. Summing (3.37) and (3.45) results in

$$nR_s \leq \frac{1}{2}\left\{h(\mathbf{Y}_r^n) + h(\tilde{\mathbf{X}}_{t_2}^n) + h(\tilde{\mathbf{X}}_{c_2}^n)\right\} + n\phi_6 \tag{3.46}$$

$$\leq \frac{1}{2}\sum_{i=1}^{n}\left\{\sum_{k=1}^{N} h(Y_{r,k}(i)) + \sum_{k=N_e+1}^{N} h(\tilde{X}_{t,k}(i)) + \sum_{k=N+1}^{N_c} h(\tilde{X}_{c,k}(i))\right\} + n\phi_6, \tag{3.47}$$

where $\phi_6 = \frac{1}{2}\left(\phi_5 - \log|\det(\mathbf{H}_{c_1})|\right)$.

In Appendix C, we show, for $i = 1, \cdots, n$, $k = 1, \cdots, N$, and $m = 1, \cdots, N_c$, that

$$h(Y_{r,k}(i)) \leq \log 2\pi e + \log(1 + h^2 P) \tag{3.48}$$

$$h(\tilde{X}_{t,k}(i)), h(\tilde{X}_{c,m}(i)) \leq \log 2\pi e + \log(\rho^2 + P), \tag{3.49}$$

where $h^2 = \max_k \left(||\mathbf{h}_{t,k}^r||^2 + ||\mathbf{h}_{c,k}^r||^2 \right)$; $\mathbf{h}_{t,k}^r$ and $\mathbf{h}_{c,k}^r$ denote the transpose of the kth row vectors of $\mathbf{H}_t$ and $\mathbf{H}_c$, respectively. Using (3.47), (3.48), and (3.49), we have

$$R_s \leq \frac{N}{2} \log(1 + h^2 P) + \frac{N_c - N_e}{2} \log(\rho^2 + P) + \phi_7, \tag{3.50}$$

where $\phi_7 = \phi_6 + \frac{N+N_c-N_e}{2} \log 2\pi e$. Using (3.5), we get

$$D_s \leq \lim_{P \to \infty} \frac{\frac{N}{2} \log(1 + h^2 P) + \frac{N_c - N_e}{2} \log(\rho^2 + P) + \phi_7}{\log P} \tag{3.51}$$

$$= \frac{N + N_c - N_e}{2}. \tag{3.52}$$

Thus, the s.d.o.f. for $N_e \leq N$ and $N < N_c \leq N + N_e$, is upper bounded by $\frac{N+N_c-N_e}{2}$.

Case 2: $N_e > N$

Another stochastically equivalent form of $\mathbf{Z}_e$ is

$$\mathbf{Z}_e'' = \mathbf{G}_t \tilde{\mathbf{Z}}_t + \mathbf{G}_c \tilde{\mathbf{Z}}_c + \tilde{\mathbf{Z}}_e'. \tag{3.53}$$

where[7] $\tilde{\mathbf{Z}}_e' \sim \mathcal{CN}(\mathbf{z}, \mathbf{I}_{N_e} - \mathbf{G}_t \mathbf{K}_t \mathbf{G}_t^H - \mathbf{G}_c \mathbf{K}_c \mathbf{G}_c^H)$ is independent from $\{\tilde{\mathbf{Z}}_t, \tilde{\mathbf{Z}}_c, \mathbf{X}_t, \mathbf{X}_c, \mathbf{Z}_r\}$. $\tilde{\mathbf{Z}}_e'^n$ is an i.i.d. sequence of the random vectors $\tilde{\mathbf{Z}}_e'$. Using (3.53), another stochastically equivalent form of $\mathbf{Y}_e^n$ is given by

$$\mathbf{Y}_e''^n = \mathbf{G}_t \tilde{\mathbf{X}}_t + \mathbf{G}_c \tilde{\mathbf{X}}_c^n + \tilde{\mathbf{Z}}_e'^n. \tag{3.54}$$

[7]The choice of $\mathbf{K}_t$ and $\mathbf{K}_c$ guarantees that $\mathbf{I}_{N_e} - \mathbf{G}_t \mathbf{K}_t \mathbf{G}_t^H - \mathbf{G}_c \mathbf{K}_c \mathbf{G}_c^H$ is a valid covariance matrix.

Let us rewrite $\tilde{\mathbf{X}}_c$ and $\mathbf{H}_c$ as follows. $\tilde{\mathbf{X}}_c = [\tilde{\mathbf{X}}_{c_1}^{\prime T}\ \tilde{\mathbf{X}}_{c_2}^{\prime T}]^T$, where $\tilde{\mathbf{X}}_{c_1}^{\prime} = [\tilde{X}_{c,1} \cdots \tilde{X}_{c,N_e-N}]^T$, $\tilde{\mathbf{X}}_{c_2}^{\prime} = [\tilde{\mathbf{X}}_{c_{21}}^{\prime T}\ \tilde{\mathbf{X}}_{c_{22}}^{\prime T}]^T$, $\tilde{\mathbf{X}}_{c_{21}}^{\prime} = [\tilde{X}_{c,N_e-N+1} \cdots \tilde{X}_{c,N_e}]^T$, and $\tilde{\mathbf{X}}_{c_{22}}^{\prime} = [\tilde{X}_{c,N_e+1} \cdots \tilde{X}_{c,N_c}]^T$. $\mathbf{H}_c = [\mathbf{H}_{c_1}^{\prime}\ \mathbf{H}_{c_2}^{\prime}]$, where $\mathbf{H}_{c_1}^{\prime} \in \mathbb{C}^{N\times(N_e-N)}$, $\mathbf{H}_{c_2}^{\prime} = [\mathbf{H}_{c_{21}}^{\prime}\ \mathbf{H}_{c_{22}}]$, $\mathbf{H}_{c_{21}}^{\prime} \in \mathbb{C}^{N\times N}$, and $\mathbf{H}_{c_{22}} \in \mathbb{C}^{N\times(N_c-N_e)}$. Let $\mathbf{G}_c = \big[\mathbf{G}_{c_1}\ \mathbf{G}_{c_2}\big]$, where $\mathbf{G}_{c_1} \in \mathbb{C}^{N_e\times(N_e-N)}$ and $\mathbf{G}_{c_2} \in \mathbb{C}^{N_e\times(N+N_c-N_e)}$. Using (3.54), we have

$$h(\mathbf{Y}_e^n) = h(\mathbf{Y}_e^{\prime\prime n}) = h\Big([\mathbf{G}_t\ \mathbf{G}_{c_1}] \begin{bmatrix} \tilde{\mathbf{X}}_t^n \\ \tilde{\mathbf{X}}_{c_1}^{\prime n} \end{bmatrix} + \mathbf{G}_{c_2}\tilde{\mathbf{X}}_{c_2}^{\prime n} + \tilde{\mathbf{Z}}_e^{\prime n}\Big) \tag{3.55}$$

$$\geq h(\tilde{\mathbf{X}}_t^n, \tilde{\mathbf{X}}_{c_1}^{\prime n} | \tilde{\mathbf{X}}_{c_2}^{\prime n}) + n\log|\det[\mathbf{G}_t\ \mathbf{G}_{c_1}]| \tag{3.56}$$

$$= h(\tilde{\mathbf{X}}_t^n) + h(\tilde{\mathbf{X}}_{c_1}^{\prime n} | \tilde{\mathbf{X}}_{c_2}^{\prime n}) + n\log|\det[\mathbf{G}_t\ \mathbf{G}_{c_1}]|, \tag{3.57}$$

where (3.57) follows since $\tilde{\mathbf{X}}_t^n$ and $\tilde{\mathbf{X}}_c^n$ are independent. Substituting (3.57) in (3.29) gives

$$nR_s \leq h(\tilde{\mathbf{X}}_{c_2}^{\prime n}) + n\phi_8, \tag{3.58}$$

where $\phi_8 = \phi_4 - \log|\det[\mathbf{G}_t\ \mathbf{G}_{c_1}]|$.

In order to obtain another upper bound for R_s, which we combine with (3.58) to obtain the desired bound for $N_e > N$ and $N_e < N_c \leq N + N_e$, we proceed as follows. Consider a modified channel where the first $N_e - N$ antennas at the cooperative jammer are removed, i.e., the cooperative jammer uses only the last $N + N_c - N_e$ out of its N_c antennas. The transmitted signals in the modified channel are $\mathbf{X}_t^n$ and $\mathbf{X}_{c_2}^{\prime n}$, and hence,

the legitimate receiver receives

$$\bar{\mathbf{Y}}_r^n = \mathbf{H}_t \mathbf{X}_t^n + \mathbf{H}'_{c_2} \mathbf{X}'^n_{c_2} + \mathbf{Z}_r^n. \tag{3.59}$$

Let R and $\bar{R}$ denote reliable communication rates, i.e., the achievable rates without the secrecy constraint, for the original and the modified channels, respectively. Since the cooperative jamming signal is additive interference for the legitimate receiver, the reliable communication rate of the modified channel, $\bar{R}$, is an upper bound for that of the original channel, R. Since R_s satisfies the reliability and secrecy constraints in (3.3) and (3.4), we have that

$$\begin{aligned} nR_s &\le nR \le n\bar{R} \\ &\le I(\mathbf{X}_t^n; \bar{\mathbf{Y}}_r^n) = h(\bar{\mathbf{Y}}_r^n) - h(\mathbf{H}'_{c_2} \mathbf{X}'^n_{c_2} + \mathbf{Z}_r^n). \end{aligned} \tag{3.60}$$

Let $\tilde{\mathbf{Z}}_{c_2} = [\tilde{Z}_{c,N_e-N+1} \cdots \tilde{Z}_{c,N_c}]^T \sim \mathcal{CN}(\mathbf{z}, \mathbf{K}'_c)$, where $\mathbf{K}'_c = \rho^2 \mathbf{I}_{N+N_c-N_e}$. Another stochastically equivalent form of $\mathbf{Z}_r$ is $\mathbf{Z}''_r = \mathbf{H}'_{c_2} \tilde{\mathbf{Z}}_{c_2} + \tilde{\mathbf{Z}}'_r$, where[8] $\tilde{\mathbf{Z}}'_r \sim \mathcal{CN}(\mathbf{z}, \mathbf{I}_N - \mathbf{H}'_{c_2} \mathbf{K}'_c \mathbf{H}'^H_{c_2})$ is independent from $\{\tilde{\mathbf{Z}}_t, \tilde{\mathbf{Z}}_c, \mathbf{X}_t, \mathbf{X}_c, \mathbf{Z}_e\}$, and $\tilde{\mathbf{Z}}'^n_r$ is an i.i.d. sequence of $\tilde{\mathbf{Z}}'_r$. Thus, using (3.60), we have

$$nR_s \le h(\bar{\mathbf{Y}}_r^n) - h(\mathbf{H}'_{c_2} \tilde{\mathbf{X}}'^n_{c_2} + \tilde{\mathbf{Z}}'^n_r) \tag{3.61}$$

$$\le h(\bar{\mathbf{Y}}_r^n) - h(\mathbf{H}'_{c_2} \tilde{\mathbf{X}}'^n_{c_2}) \tag{3.62}$$

$$\le h(\bar{\mathbf{Y}}_r^n) - h(\tilde{\mathbf{X}}'^n_{c_{21}} | \tilde{\mathbf{X}}'^n_{c_{22}}) - n \log |\det(\mathbf{H}'_{c_{21}})|. \tag{3.63}$$

[8]The choice of $\mathbf{K}_c$ guarantees that $\mathbf{I}_N - \mathbf{H}'_{c_2} \mathbf{K}'_c \mathbf{H}'^H_{c_2}$ is a valid covariance matrix.

Let $\bar{\mathbf{Y}}_r = [\bar{Y}_{r,1} \cdots \bar{Y}_{r,N}]^T$. Summing (3.58) and (3.63) yields

$$nR_s \leq \frac{1}{2}\left\{h(\bar{\mathbf{Y}}_r^n) + h(\tilde{\mathbf{X}}_{c_{22}}^{\prime n})\right\} + n\phi_9 \tag{3.64}$$

$$\leq \frac{1}{2}\sum_{i=1}^{n}\left\{\sum_{k=1}^{N} h(\bar{Y}_{r,k}(i)) + \sum_{k=N_e+1}^{N_c} h(\tilde{X}_{c,k}(i))\right\} + n\phi_9, \tag{3.65}$$

where $\phi_9 = \frac{1}{2}\{\phi_8 - \log|\det(\mathbf{H}'_{c_{21}})|\}$. In Appendix C, we also show that

$$h(\bar{Y}_{r,k}(i)) \leq \log 2\pi e + \log(1 + \bar{h}^2 P), \tag{3.66}$$

where $\bar{h}^2 = \max_k \left(||\mathbf{h}_{t,k}^r||^2 + ||\mathbf{h}_{c,k}^{\prime r}||^2\right)$; $\mathbf{h}_{c,k}^{\prime r}$ denotes the transpose of the kth row vector of $\mathbf{H}'_{c_2}$.

Similar to case 1, using (3.65), (3.66), and (3.49), the secrecy rate is bounded as

$$R_s \leq \frac{N}{2}\log(1 + \bar{h}^2 P) + \frac{N_c - N_e}{2}\log(\rho^2 + P) + n\phi_{10}, \tag{3.67}$$

where $\phi_{10} = \phi_9 + \frac{N+N_c-N_e}{2}\log 2\pi e$. Thus, the s.d.o.f., for $N_e > N$ and $N_e < N_c \leq N+N_e$, is upper bounded as

$$D_s \leq \frac{N + N_c - N_e}{2}. \tag{3.68}$$

3.4.3 Obtaining the Upper Bound

For $N_e \leq N$, the upper bound for the s.d.o.f. derived in Section 3.4.1 is equal to $N + N_c - N_e$ for all $0 \leq N_c \leq N_e$. In addition, the upper bound derived in Section 3.4.2, at $N_c = N$, is equal to $N - \frac{N_e}{2}$, c.f. equations (3.15) and (3.52). As the former upper

bound is greater than the latter for all $\frac{N_e}{2} < N_c \leq N$, the s.d.o.f. is upper bounded by $N - \frac{N_e}{2}$ for all $\frac{N_e}{2} < N_c \leq N$. Combining these statements, we have the following upper bound for the s.d.o.f. for $N_e \leq N$:

$$D_s \leq \begin{cases} N + N_c - N_e, & \text{if } 0 \leq N_c \leq \frac{N_e}{2} \\ N - \frac{N_e}{2}, & \text{if } \frac{N_e}{2} < N_c \leq N \\ \frac{N+N_c-N_e}{2}, & \text{if } N < N_c \leq N + N_e, \\ N, & \text{if } N_c > N + N_e. \end{cases} \tag{3.69}$$

Similarly, when $N_e > N$ and for all $N_e - \frac{N}{2} < N_c \leq N_e$, the upper bound derived for $0 \leq N_c \leq N_e$ in Section 3.4.1 is greater than the upper bound derived in Section 3.4.2 at $N_c = N_e$. Thus, the s.d.o.f. for $N_e - \frac{N}{2} < N_c \leq N_e$ is upper bounded by $\frac{N}{2}$. In addition, the upper bound in (3.15) is equal to zero for all $0 \leq N_c \leq N_e - N$. Thus, the s.d.o.f. for $N_e > N$ is upper bounded as:

$$D_s \leq \begin{cases} 0, & \text{if } 0 \leq N_c \leq N_e - N \\ N + N_c - N_e, & \text{if } N_e - N < N_c \leq N_e - \frac{N}{2} \\ \frac{N}{2}, & \text{if } N_e - \frac{N}{2} < N_c \leq N_e \\ \frac{N+N_c-N_e}{2}, & \text{if } N_e < N_c \leq N + N_e, \\ N, & \text{if } N_c > N + N_e. \end{cases} \tag{3.70}$$

By combining the bounds for $N_e \leq N$ in (3.69) and for $N_e > N$ in (3.70), we obtain the upper bound for the s.d.o.f. in (3.6). In the next section, we will show the achievability of (3.6).

3.5 Achievablility for $N_t = N_r = N$

In this section, we provide the achievability proof for Theorem 1 by showing the achievability of (3.69) when $N_e \leq N$, and the achievability of (3.70) when $N_e > N$. For both $N_e \leq N$ and $N_e > N$, we divide the range of the number of antennas at the cooperative jammer, N_c, into five ranges and propose an achievable scheme for each range. For all the achievable schemes in this section, we have the n-letter signals, $\mathbf{X}_t^n$ and $\mathbf{X}_c^n$, as i.i.d. sequences. Since $\mathbf{X}_c^n$ is independent from $\mathbf{X}_t^n$, and each of them is i.i.d. across time, we have in effect a memoryless wiretap channel and the secrecy rate

$$R_s = [I(\mathbf{X}_t; \mathbf{Y}_r) - I(\mathbf{X}_t; \mathbf{Y}_e)]^+, \tag{3.71}$$

is achievable by *stochastic encoding* at the transmitter [23].

The transmitted signals at the transmitter and the cooperative jammer, for each of the following schemes, are

$$\mathbf{X}_t = \mathbf{P}_t \mathbf{U}_t, \quad \mathbf{X}_c = \mathbf{P}_c \mathbf{V}_c, \tag{3.72}$$

where $\mathbf{U}_t = [U_1 \cdots U_d]^T$ and $\mathbf{V}_c = [V_1 \cdots V_l]^T$ are the information and cooperative jamming streams, respectively. $\mathbf{P}_t = \left[\mathbf{p}_{t,1} \cdots \mathbf{p}_{t,d}\right] \in \mathbb{C}^{N \times d}$ and $\mathbf{P}_c = \left[\mathbf{p}_{c,1} \cdots \mathbf{p}_{c,l}\right] \in \mathbb{C}^{N_c \times l}$ are the precoding matrices at the transmitter and the cooperative jammer.

Signaling, precoding, and decoding techniques utilized in this proof vary according to the relative number of antennas at the different terminals and whether the s.d.o.f. of the channel is integer valued or not an integer. In particular, we show that Gaussian signaling both for transmission and cooperative jamming is sufficient to achieve the integer valued s.d.o.f., while achieving non-integer s.d.o.f. requires structured signaling and cooperative jamming, i.e., signals from discrete constellations, along with a combination of linear receiver processing, and the complex field equivalent of real interference alignment [59, 71]. Additionally, the linear precoding at the transmitter and the cooperative jammer depends on whether N_e is equal to, smaller than, or larger than N, and whether the number of antennas at the cooperative jammer, N_c, results in a s.d.o.f. for the channel that is before, after, or at the flat s.d.o.f. range in the s.d.o.f. plot versus N_c. This leads to an achievability proof that involves 10 distinct achievable schemes, which differ from each other in the type of signals used (Gaussian or structured), and/or precoding at the transmitter and cooperative jammer, and/or decoding at the legitimate receiver.

Remark 4. *Note that integer-valued s.d.o.f. can also be achieved using structured signals. However, Gaussian signaling often outperforms structured signaling for finite SNR; see for example [127, Fig. 2]. Although our focus in this chapter is on characterizing the s.d.o.f., i.e., secrecy rate scaling at high SNR, for the channel, we use Gaussian signaling whenever possible for the achievability for this reason.*

In order to extend real interference alignment to complex channels, we need to utilize different results than those used for real channels. For real channels, to analyze the decoder performance, reference [78] proposed utilizing the convergence part of

Khintchine-Groshev theorem in the field of Diophantine approximation [102], which deals with the approximation of real numbers with rational numbers. For complex channels, transforming the channel into a real channel with twice the dimensions, as is usually the convention, is not sufficient here, since real interference alignment relies on the linear independence over rational numbers of the channel gains, which does not continue to hold after such channel transformation. Luckily, we can utilize the result stated in Lemma 1. For complex channel coefficients, this result ends up playing the same role of the Khintchine-Groshev theorem for real coefficients.

Before continuing with the achievability proof for the different cases, we state the following lemma, which is utilized to show the linear independence between the directions of the received streams at the legitimate receiver.

Lemma 6. Consider two matrices $\mathbf{E}_1 \in \mathbb{C}^{N\times K}$ and $\mathbf{E}_2 \in \mathbb{C}^{K\times M}$, where $N, M < K$. If the matrix $\mathbf{E}_2$ is full column rank and the matrix $\mathbf{E}_1$ has all of its entries independently and randomly drawn according to a continuous distribution, then $\text{rank}(\mathbf{E}_1\mathbf{E}_2) = \min(N, M)$ a.s.

Proof: The proof of Lemma 6 is given in Appendix D. ■

3.5.1 Case 1: $N_e \leq N$ and $0 \leq N_c \leq \frac{N_e}{2}$

The s.d.o.f. for this case is equal to $N + N_c - N_e$, i.e., integer valued, for which we utilize Gaussian signaling and cooperative jamming. Since $N_e \leq N$, the transmitter exploits this advantage by sending a part of its signal invisible to the eavesdropper.

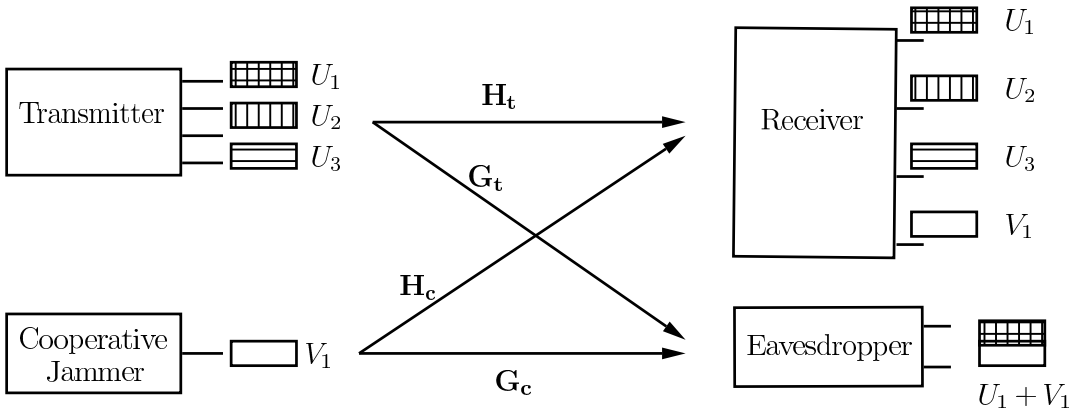


Fig. 3.3. An example for the achievability scheme for Case 1, when $N = 4$, $N_e = 2$, and $N_c = 1$.

There is no need for linear precoding at the cooperative jammer for this case. Increasing the number of the cooperative jammer antennas, N_c, increases the s.d.o.f. of the channel.

The transmitted signals, $\mathbf{X}_t$ and $\mathbf{X}_c$, are given by (3.72) with $d = N + N_c - N_e$, $l = N_c$, $\mathbf{U}_t \sim \mathcal{CN}(\mathbf{0}, \bar{P}\mathbf{I}_d)$, $\mathbf{V}_c \sim \mathcal{CN}(\mathbf{0}, \bar{P}\mathbf{I}_l)$, and $\bar{P} = \frac{1}{\alpha}P$, in accordance with the power constraints on the transmitted signals at the transmitter and the cooperative jammer, where $\alpha = \max\left\{l, \sum_{i=1}^{d} ||\mathbf{p}_{t,i}||^2\right\}$ is a constant which does not depend on the power P. The precoders $\mathbf{P}_c$ and $\mathbf{P}_t$ are given by $\mathbf{P}_c = \mathbf{I}_l$, and

$$\mathbf{P}_t = \begin{bmatrix}\mathbf{P}_{t,a} & \mathbf{P}_{t,n}\end{bmatrix} \in \mathbb{C}^{N \times d}, \tag{3.73}$$

where $\mathbf{P}_{t,a} = \mathbf{G}_t^{\dagger}\mathbf{G}_c$ in order to align the information streams over the cooperative jamming streams at the eavesdropper, and the $N - N_e$ columns of $\mathbf{P}_{t,n}$ are chosen to span $\mathcal{N}(\mathbf{G}_t)$. The achievability scheme for this case, when $N = 4$, $N_e = 2$, and $N_c = 1$, is depicted in Fig. 3.3.

Since $N_c \leq \frac{N_e}{2}$, the total number of superposed received streams at the receiver, $2N_c + N - N_e$, is less than or equal to the number of its available spatial dimensions, N. Thus, the receiver can decode all the information and cooperative jamming streams at high SNR. Using (3.1), (3.2), and (3.72), the received signals at the receiver and the eavesdropper are

$$\mathbf{Y}_r = \begin{bmatrix} \mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c \end{bmatrix} \begin{bmatrix} \mathbf{U}_t \\ \mathbf{V}_c \end{bmatrix} + \mathbf{Z}_r, \tag{3.74}$$

$$\mathbf{Y}_e = \begin{bmatrix} \mathbf{G}_t\mathbf{G}_t^{\dagger}\mathbf{G}_c & \mathbf{z}_{N_e\times(N-N_e)} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{t1}^{\,l} \\ \mathbf{U}_{tl+1}^{\,d} \end{bmatrix} + \mathbf{G}_c\mathbf{V}_c + \mathbf{Z}_e \tag{3.75}$$

$$= \mathbf{G}_c(\mathbf{U}_{t1}^{\,l} + \mathbf{V}_c) + \mathbf{Z}_e. \tag{3.76}$$

We lower bound the secrecy rate in (3.71) as follows. First, in order to compute $I(\mathbf{X}_t; \mathbf{Y}_r)$, we show that the matrix $[\mathbf{H}_t\mathbf{P}_t \;\; \mathbf{H}_c] \in \mathbb{C}^{N\times(d+l)}$ in (3.74) is full column-rank a.s.

The columns of $\mathbf{P}_{t,a} = \mathbf{G}_t^{\dagger}\mathbf{G}_c$ are linearly independent a.s. due to the randomly generated channel gains, and the $N - N_e$ columns of $\mathbf{P}_{t,n}$ are linearly independent as well, since they span an $N - N_e$-dimensional subspace. In addition, each of the columns of $\mathbf{P}_{t,a}$ is linearly independent from the columns of $\mathbf{P}_{t,n}$ a.s. since $\mathbf{G}_t\mathbf{P}_{t,a} = \mathbf{G}_c$, and hence $\mathbf{G}_t\mathbf{p}_{t_i} \neq \mathbf{z}$ for all $i = 1, 2, \cdots, l$. Thus $\mathbf{P}_t = [\mathbf{P}_{t,a}\; \mathbf{P}_{t,n}]$ is full column rank a.s. The

matrix $[\mathbf{H}_t\mathbf{P}_t \;\; \mathbf{H}_c]$ can be written as

$$\begin{bmatrix}\mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\end{bmatrix} = \begin{bmatrix}\mathbf{H}_t & \mathbf{H}_c\end{bmatrix} \begin{bmatrix}\mathbf{P}_t & \mathbf{z}_{N\times l} \\ \mathbf{z}_{l\times d} & \mathbf{I}_l\end{bmatrix}. \tag{3.77}$$

The matrix $[\mathbf{H}_t \;\; \mathbf{H}_c]$ has all of its entries independently and randomly drawn according to a continuous distribution, while the second matrix on the right hand side (RHS) of (3.77) is full column rank a.s. By applying Lemma 6 to (3.77), we have that the matrix $[\mathbf{H}_t\mathbf{P}_t \;\; \mathbf{H}_c]$ is full column rank a.s. Thus, using (3.74), we obtain the lower bound

$$I(\mathbf{X}_t;\mathbf{Y}_r) \geq d\log P + o(\log P). \tag{3.78}$$

Next, using (3.76), we upper bound $I(\mathbf{X}_t;\mathbf{Y}_e)$ as follows:

$$I(\mathbf{X}_t;\mathbf{Y}_e) = h(\mathbf{Y}_e) - h(\mathbf{Y}_e|\mathbf{X}_t) \tag{3.79}$$

$$= h(\mathbf{G}_c(\mathbf{U}_{t1}^{l} + \mathbf{V}_c) + \mathbf{Z}_e) - h(\mathbf{G}_c\mathbf{V}_c + \mathbf{Z}_e) \tag{3.80}$$

$$= \log\frac{\det(\mathbf{I}_{N_e} + 2\bar{P}\mathbf{G}_c\mathbf{G}_c^H)}{\det(\mathbf{I}_{N_e} + \bar{P}\mathbf{G}_c\mathbf{G}_c^H)} \tag{3.81}$$

$$= \log\frac{\det(\mathbf{I}_l + 2\bar{P}\mathbf{G}_c^H\mathbf{G}_c)}{\det(\mathbf{I}_l + \bar{P}\mathbf{G}_c^H\mathbf{G}_c)} \tag{3.82}$$

$$= \log\frac{2^l\det(\frac{1}{2}\mathbf{I}_l + \bar{P}\mathbf{G}_c^H\mathbf{G}_c)}{\det(\mathbf{I}_l + \bar{P}\mathbf{G}_c^H\mathbf{G}_c)} \tag{3.83}$$

$$\leq l. \tag{3.84}$$

Substituting (3.78) and (3.84) in (3.71), we have

$$R_s \geq d \log P + o(\log P) - l \tag{3.85}$$

$$= (N + N_c - N_e) \log P + o(\log P) - N_c, \tag{3.86}$$

and hence, using (3.5), we conclude that the achievable s.d.o.f. is $D_s \geq N + N_c - N_e$.

3.5.2 Case 2: $N_e \leq N, \frac{N_e}{2} < N_c \leq N$, and N_e is even

Unlike case 1, the s.d.o.f. for this case does not increase by increasing N_c. For all N_c in this case, the transmitter sends the same number of information streams, while the cooperative jammer utilizes a linear precoder which allows for discarding any unnecessary antennas. The s.d.o.f. here is integer valued, and we use Gaussian signaling for transmission and cooperative jamming.

In particular, for N_e is even, $N_c = \frac{N_e}{2}$, and $N_e \leq N$, the achievable s.d.o.f., using the scheme in Section 3.5.1, is equal to $N - \frac{N_e}{2}$. However, from (3.69), we observe that the s.d.o.f. is upper bounded by $N - \frac{N_e}{2}$ for all $\frac{N_e}{2} < N_c \leq N$. Thus, when $N_e \leq N$ and N_e is even, the scheme for $N_c = \frac{N_e}{2}$ in Section 3.5.1 can be used to achieve the s.d.o.f. for all $\frac{N_e}{2} < N_c \leq N$ by discarding the remaining $N_c - \frac{N_e}{2}$ antennas. That is, the cooperative jammer uses the precoder

$$\mathbf{P}_c = \begin{bmatrix} \mathbf{I}_l \\ \\ \mathbf{z}_{(N_c - l) \times l} \end{bmatrix}, \tag{3.87}$$

with $l = \frac{N_e}{2}$, to utilize only $\frac{N_e}{2}$ out of its N_c antennas, and the transmitter utilizes

$$\mathbf{P}_t = \left[\mathbf{P}_{t,a}\, \mathbf{P}_{t,n}\right], \tag{3.88}$$

$\mathbf{P}_{t,a} = \mathbf{G}_t^{\dagger}\mathbf{G}_c\mathbf{P}_c \in \mathbb{C}^{N\times l}$, $\mathbf{P}_{t,n} \in \mathbb{C}^{N\times(N-N_e)}$ is defined as in (3.73), in order to send $d = N - \frac{N_e}{2}$ Gaussian information streams. Following the same analysis as in the previous case, the achievable s.d.o.f. is $N - \frac{N_e}{2}$ for all $\frac{N_e}{2} < N_c \leq N$, where N_e is even and $N_e \leq N$.

3.5.3 Case 3: $N_e \leq N$, $\frac{N_e}{2} < N_c \leq N$, and N_e is odd

The s.d.o.f. for this case is equal to $N - \frac{N_e}{2}$, which is not an integer. As Gaussian signaling cannot achieve fractional s.d.o.f. for the channel, we utilize structured signaling both for transmission and cooperative jamming for this case. In particular, we propose utilizing *joint* signal space alignment and the complex field equivalent of real interference alignment [59, 71].

The decoding scheme at the receiver is as follows. The receiver projects its received signal over a direction that is orthogonal to all but one information and one cooperative jamming streams. Then, the receiver decodes these two streams from the projection using complex field analogy of real interference alignment. Finally, the receiver removes the decoded information and cooperative jamming streams from its received signal, leaving $N - 1$ spatial dimensions for the other $N - \frac{N_e+1}{2}$ information and $\frac{N_e-1}{2}$ cooperative jamming streams.

Before continuing with the details for the achievability scheme for this case, we provide the following example, which illustrates the ideas utilized for this case.

Example 1. Consider a multi-antenna wiretap channel with 4-antenna transmitter, 4-antenna receiver, 3-antenna eavesdropper, and 2-antenna cooperative jammer as shown in Fig. 3.4.

The transmitter sends 3 structured information streams, U_1, U_2, U_3, and the cooperative jammer sends 2 structured jamming streams, V_1, V_2. The streams U_1, V_1 are integer valued, while the streams U_2, U_3, V_2, are complex integers. That is, $U_2 = U_{2,\mathrm{Re}} + jU_{2,\mathrm{Im}}, U_3 = U_{3,\mathrm{Re}} + jU_{3,\mathrm{Im}}$, and $V_2 = V_{2,\mathrm{Re}} + jV_{2,\mathrm{Im}}$, where $\{U_1, U_{2,\mathrm{Re}}, U_{2,\mathrm{Im}}, U_{3,\mathrm{Re}}, U_{3,\mathrm{Im}}, V_1, V_{2,\mathrm{Re}}, V_{2,\mathrm{Im}}\}$ are i.i.d. random variables uniform over a set of integer that scales with the transmit power as it will be explained later in (3.89). The transmitter chooses its precoder as in (3.88) so that U_3 is sent over $\mathcal{N}(\mathbf{G}_t)$, and hence U_3 is invisible to the eavesdropper, and that U_1, V_1 and U_2, V_2 are perfectly aligned at the eavesdropper. The cooperative jammer chooses its precoder as in (3.87) so that it utilizes only 2 out of its 3 antennas to send V_1, V_2. The legitimate receiver projects its received signal over a single dimension that is orthogonal to $\{U_2, U_3, V_2\}$, and hence, only U_1 and V_1 remain in this dimension. The received signal after projection is of the form $f_1U_1 + f_2V_1 + Z$, where f_1, f_2 are the coefficients resulting from multiplying the channel gains with the projection matrix, and Z is the projection of the Gaussian noise over the single dimension. The receiver utilizes a hard decision decoder which maps $f_1U_1 + f_2V_1 + Z$ to the nearest point in the constellation of $f_1U_1 + f_2V_1$. It has been shown in [78] that U_1, V_1 can be uniquely decoded from $f_1U_1 + f_2V_1$. Thus, the receiver decodes U_1, V_1, subtracts them from its original received signal, and then utilizes the remaining 3 dimensions in its signal space

to decode U_2, U_3, V_2. Thus, the receiver utilizes 2.5 dimensions to decode the information streams, i.e., 2.5 useful dimensions, where each of U_2 and U_3 is decoded from a separate dimension while both U_1 and V_1 are decoded from a single dimension (each occupies half of that dimension), leading to 2.5 achievable s.d.o.f. for the channel.

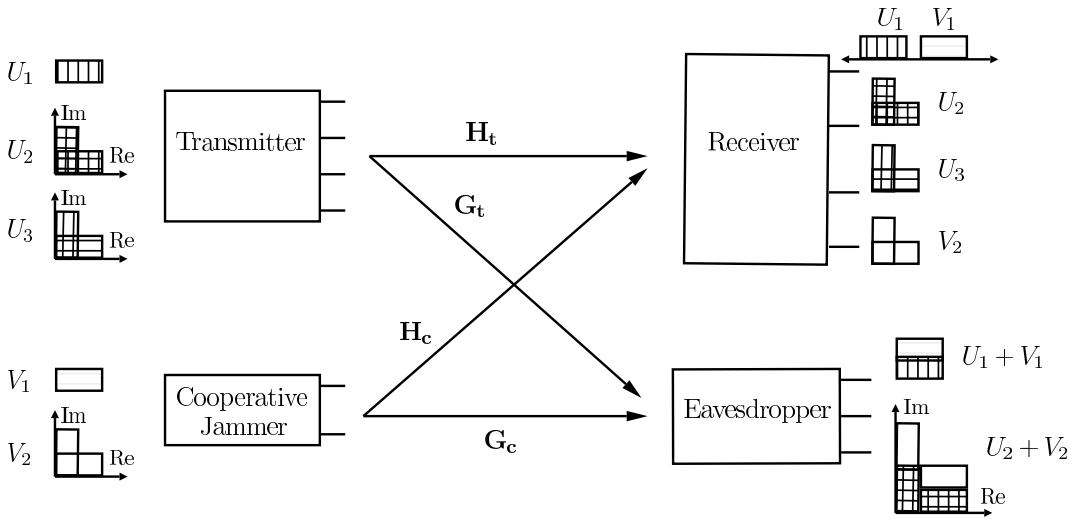


Fig. 3.4. An example for the achievability scheme for Case 3, when $N = 4$, $N_e = 3$, and $N_c = 2$.

Now, we continue with the detailed explanation for the achievability scheme for this case. The transmitted signals are given by (3.72), with $d = N - \frac{N_e-1}{2}$, $l = \frac{N_e+1}{2}$, $\mathbf{P}_c, \mathbf{P}_t$ are defined as in (3.87) and (3.88), and $U_i = U_{i,\mathrm{Re}} + jU_{i,\mathrm{Im}}$, $V_k = V_{k,\mathrm{Re}} + jV_{k,\mathrm{Im}}$, $i = 2, 3, \cdots, d$ and $k = 2, 3, \cdots, l$. The random variables U_1, V_1, $\{U_{i,\mathrm{Re}}\}_{i=2}^{d}$, $\{U_{i,\mathrm{Im}}\}_{i=2}^{d}$, $\{V_{i,\mathrm{Re}}\}_{i=2}^{l}$, and $\{V_{i,\mathrm{Im}}\}_{i=2}^{l}$ are i.i.d. uniform over the set $\{a(-Q, Q)_{\mathbb{Z}}\}$. The values for a and the integer Q are chosen as

$$Q = \left\lfloor P^{\frac{1-\epsilon}{2+\epsilon}} \right\rfloor = P^{\frac{1-\epsilon}{2+\epsilon}} - \nu \tag{3.89}$$

$$a = \gamma P^{\frac{3\epsilon}{2(2+\epsilon)}}, \tag{3.90}$$

in order to satisfy the power constraints, where ϵ is an arbitrarily small positive number, and ν, γ are constants that do not depend on the power P. Justification for the choice of a and Q is provided in Appendix E.

The received signal at the eavesdropper is

$$\mathbf{Y}_e = \tilde{\mathbf{G}}_c(\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c) + \mathbf{Z}_e, \tag{3.91}$$

where $\tilde{\mathbf{G}}_c = \mathbf{G}_c\mathbf{P}_c$. We upper bound the second term in (3.71), $I(\mathbf{X}_t; \mathbf{Y}_e)$, as follows:

$$I(\mathbf{X}_t; \mathbf{Y}_e) \leq I(\mathbf{X}_t; \mathbf{Y}_e, \mathbf{Z}_e) \tag{3.92}$$

$$= I(\mathbf{X}_t; \mathbf{Y}_e|\mathbf{Z}_e) \tag{3.93}$$

$$= H(\mathbf{Y}_e|\mathbf{Z}_e) - H(\mathbf{Y}_e|\mathbf{Z}_e, \mathbf{X}_t) \tag{3.94}$$

$$= H\left(\tilde{\mathbf{G}}_c(\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c)\right) - H\left(\tilde{\mathbf{G}}_c\mathbf{V}_c\right) \tag{3.95}$$

$$= H(\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c) - H(\mathbf{V}_c) \tag{3.96}$$

$$= H\left(U_1 + V_1,\ U_{2,\mathrm{Re}} + V_{2,\mathrm{Re}},\ U_{2,\mathrm{Im}} + V_{2,\mathrm{Im}}, \cdots,\ U_{l,\mathrm{Re}} + V_{l,\mathrm{Re}},\ U_{l,\mathrm{Im}} + V_{l,\mathrm{Im}}\right)$$

$$- H\left(V_1,\ V_{2,\mathrm{Re}},\ V_{2,\mathrm{Im}},\ \cdots,\ V_{l,\mathrm{Re}},\ V_{l,\mathrm{Im}}\right) \tag{3.97}$$

$$\leq \log(4Q+1)^{2l-1} - \log(2Q+1)^{2l-1} \tag{3.98}$$

$$= (2l-1)\log\frac{4Q+1}{2Q+1} \tag{3.99}$$

$$\leq 2l - 1, \tag{3.100}$$

where (3.93) follows since $\mathbf{X}_t$ and $\mathbf{Z}_e$ are independent, and (3.98) follows since the entropy of a uniform random variable over the set $\{a(-2Q, 2Q)_{\mathbb{Z}}\}$ upper bounds the entropy of each of $U_1 + V_1, U_{2,\mathrm{Re}} + V_{2,\mathrm{Re}}, U_{2,\mathrm{Im}} + V_{2,\mathrm{Im}}, \cdots, U_{l,\mathrm{Im}} + V_{l,\mathrm{Im}}$. Equation (3.96) follows since the mappings $\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c \mapsto \tilde{\mathbf{G}}_c(\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c)$ and $\mathbf{V}_c \mapsto \tilde{\mathbf{G}}_c\mathbf{V}_c$ are bijective. The reason for this is that the entries of $\tilde{\mathbf{G}}_c$ are *rationally independent*, as illustrated in Definition 4 below, and that $(\mathbf{U}_{t1}^{\ l} + \mathbf{V}_c)$ and $\mathbf{V}_c$ belong to $\mathbb{Z}_{\mathbb{C}}^{l}$.

Definition 4. A set of complex numbers $\{c_1, c_2, \cdots, c_L\}$ are rationally independent, i.e., linearly independent over $\mathbb{Q}$, if there is no set of rational numbers $\{r_i\}$, $r_i \neq 0$ for all $i = 1, 2, \cdots, L$, such that $\sum_{i=1}^{L} r_i c_i = 0$.

Next, we derive a lower bound for $I(\mathbf{X}_t; \mathbf{Y}_r)$. The received signal at the legitimate receiver is given by

$$\mathbf{Y}_r = \mathbf{A}\mathbf{U}_t + \mathbf{H}_c'\mathbf{V}_c + \mathbf{Z}_r, \tag{3.101}$$

where $\mathbf{A} = \mathbf{H}_t\mathbf{P}_t = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_d]$ and $\mathbf{H}_c' = \mathbf{H}_c\mathbf{P}_c = \left[\mathbf{h}_{c,1}\ \mathbf{h}_{c,2}\ \cdots\ \mathbf{h}_{c,l}\right]$. The receiver chooses $\mathbf{b} \in \mathbb{C}^N$ such that $\mathbf{b} \perp \mathrm{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c,2}, \cdots, \mathbf{h}_{c,l}\}$ and obtains

$$\widetilde{\mathbf{Y}}_r = \mathbf{D}\mathbf{Y}_r \tag{3.102}$$

where

$$\mathbf{D} \triangleq \begin{bmatrix} & \mathbf{b}^H & \\ \mathbf{0}_{(N-1)\times 1} & & \mathbf{I}_{N-1} \end{bmatrix}. \tag{3.103}$$

Note that $(d-1)+(l-1) = N - \frac{N_e-1}{2} + \frac{N_e+1}{2} - 2 = N-1$, and hence the dimension of $\text{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c,2}, \cdots, \mathbf{h}_{c,l}\}$ is at most $N-1$. This shows the existence of a vector $\mathbf{b} \in \mathbb{C}^N$ such that $\mathbf{b} \perp \text{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c,2}, \cdots, \mathbf{h}_{c,l}\}$.

Due to the fact that channel gains are continuous and randomly generated, $\mathbf{a}_1$ and $\mathbf{h}_{c,1}$ are linearly independent from $\text{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c,2}, \cdots, \mathbf{h}_{c,l}\}$, and hence, $\mathbf{b}$ is not orthogonal to $\mathbf{a}_1$ and $\mathbf{h}_{c,1}$ a.s. Thus, we have

$$\widetilde{\mathbf{Y}}_r = \begin{bmatrix} \widetilde{Y}_{r_1} \\ \widetilde{\mathbf{Y}}_{r_2}^{n_t} \end{bmatrix} = \begin{bmatrix} \mathbf{b}^H\mathbf{a}_1 & \mathbf{z}_{1\times(d-1)} \\ \multicolumn{2}{c}{\widetilde{\mathbf{A}}} \end{bmatrix} \begin{bmatrix} U_1 \\ \mathbf{U}_{t2}^{d} \end{bmatrix} + \begin{bmatrix} \mathbf{b}^H\mathbf{h}_{c,1} & \mathbf{z}_{1\times(l-1)} \\ \multicolumn{2}{c}{\widetilde{\mathbf{H}}_c} \end{bmatrix} \begin{bmatrix} V_1 \\ \mathbf{V}_{c2}^{l} \end{bmatrix} + \begin{bmatrix} \mathbf{b}^H\mathbf{Z}_r \\ \mathbf{Z}_{r2}^{n_t} \end{bmatrix}, \tag{3.104}$$

where $\widetilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ \cdots\ \tilde{\mathbf{a}}_d] \in \mathbb{C}^{(N-1)\times d}$, $\tilde{\mathbf{a}}_i = \mathbf{a}_{i2}^{N}$ for all $i = 1, 2, \cdots, d$. Similarly, $\widetilde{\mathbf{H}}_c = [\tilde{\mathbf{h}}_{c,1}\ \tilde{\mathbf{h}}_{c,2}\ \cdots\ \tilde{\mathbf{h}}_{c,l}] \in \mathbb{C}^{(N-1)\times l}$, where $\tilde{\mathbf{h}}_{c,i} = \mathbf{h}_{c,i2}^{N}$ for all $i = 1, 2, \cdots, l$.

Next, the receiver uses $\widetilde{Y}_{r_1}$ to decode the information stream U_1 and the cooperative jamming stream V_1 as follows. Let $Z' = \mathbf{b}^H\mathbf{Z}_r \sim \mathcal{CN}(0, ||\mathbf{b}||^2)$, $f_1 = \mathbf{b}^H\mathbf{a}_1$, and $f_2 = \mathbf{b}^H\mathbf{h}_{c,1}$. Thus, $\widetilde{Y}_{r_1}$ is given by

$$\widetilde{Y}_{r_1} = f_1U_1 + f_2V_1 + Z'. \tag{3.105}$$

Once again, with randomly generated channel gains, $f_1 = \mathbf{b}^H\mathbf{a}_1$ and $f_2 = \mathbf{b}^H\mathbf{h}_{c,1}$ are rationally independent a.s. Thus, the mapping $(U_1, V_1) \mapsto f_1U_1 + f_2V_1$ is invertible [78]. The receiver employs a hard decision decoder which maps $\widetilde{Y}_{r_1} \in \tilde{\mathcal{Y}}_{r_1}$ to the nearest point in the constellation $\mathcal{R}_1 = f_1\mathcal{U}_1 + f_2\mathcal{V}_1$, where $\mathcal{U}_1, \mathcal{V}_1 = \{a(-Q, Q)_{\mathbb{Z}}\}$. Then, the

receiver passes the output of the hard decision decoder through the bijective mapping $f_1U_1 + f_2V_1 \mapsto (U_1, V_1)$ in order to decode both U_1 and V_1.

The receiver can now use

$$\bar{\tilde{\mathbf{Y}}}_r = \widetilde{\mathbf{Y}}_{r_2}^{n_t} - \tilde{\mathbf{a}}_1 U_1 - \tilde{\mathbf{h}}_{c,1} V_1 \tag{3.106}$$

$$= \begin{bmatrix} \tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_d \end{bmatrix} \mathbf{U}_{t2}^{\ d} + \begin{bmatrix} \tilde{\mathbf{h}}_{c,2} & \cdots & \tilde{\mathbf{h}}_{c,l} \end{bmatrix} \mathbf{V}_{c2}^{\ l} + \mathbf{Z}_{r2}^{\ n_t} \tag{3.107}$$

$$= \mathbf{B} \begin{bmatrix} \mathbf{U}_{t2}^{\ d} \\ \mathbf{V}_{c2}^{\ l} \end{bmatrix} + \mathbf{Z}_{r2}^{\ n_t}, \tag{3.108}$$

to decode $U_2, \cdots, U_d$, where,

$$\mathbf{B} \triangleq \begin{bmatrix} \tilde{\mathbf{a}}_2 & \cdots & \tilde{\mathbf{a}}_d & \tilde{\mathbf{h}}_{c,2} & \cdots & \tilde{\mathbf{h}}_{c,l} \end{bmatrix} \in \mathbb{C}^{(N-1)\times(N-1)}, \tag{3.109}$$

is full rank a.s. To show that $\mathbf{B}$ is full rank a.s., let $\bar{\mathbf{H}}_t$ and $\bar{\mathbf{H}}_c$ be generated by removing the first row from $\mathbf{H}_t$ and $\mathbf{H}_c$, and let $\bar{\mathbf{P}}_t$ and $\bar{\mathbf{P}}_c$ be generated by removing the first column from $\mathbf{P}_t$ and $\mathbf{P}_c$, respectively. $\mathbf{B}$ can be rewritten as

$$\mathbf{B} = \begin{bmatrix} \bar{\mathbf{H}}_t & \bar{\mathbf{H}}_c \end{bmatrix} \begin{bmatrix} \bar{\mathbf{P}}_t & \mathbf{z}_{N\times(l-1)} \\ \mathbf{z}_{N_c\times(d-1)} & \bar{\mathbf{P}}_c \end{bmatrix}. \tag{3.110}$$

Note that $\begin{bmatrix} \bar{\mathbf{H}}_t & \bar{\mathbf{H}}_c \end{bmatrix}$ has all of its entries independently and randomly drawn from a continuous distribution, and the second matrix in the RHS of (3.110) is full column rank. Using Lemma 6, the matrix $\mathbf{B}$ is full rank a.s.

Hence, by zero forcing, the receiver obtains

$$\widehat{\mathbf{Y}}_r = \mathbf{B}^{-1}\bar{\bar{\mathbf{Y}}}_r = \begin{bmatrix} \mathbf{U}_{t2}^{\ d} \\ \mathbf{V}_{c2}^{\ l} \end{bmatrix} + \bar{\mathbf{Z}}_r, \tag{3.111}$$

where $\bar{\mathbf{Z}}_r = \mathbf{B}^{-1}\mathbf{Z}_{r2}^{\ n_t} \sim \mathcal{CN}\left(\mathbf{z}, \mathbf{B}^{-1}\mathbf{B}^{-H}\right)$. Thus, at high SNR, the receiver can decode the other information streams, $U_2, \cdots, U_d$, from $\widehat{\mathbf{Y}}_r$.

The mutual information between the transmitter and receiver is lower bounded as follows:

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq I(\mathbf{U}_t; \widetilde{\mathbf{Y}}_r) \tag{3.112}$$

$$= I(U_1, \mathbf{U}_{t2}^{\ d}; \widetilde{Y}_{r_1}, \widetilde{\mathbf{Y}}_{r_2}^{n_t}) \tag{3.113}$$

$$= I(U_1, \mathbf{U}_{t2}^{\ d}; \widetilde{Y}_{r_1}) + I(U_1, \mathbf{U}_{t2}^{\ d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | \widetilde{Y}_{r_1}) \tag{3.114}$$

$$= I(U_1; \widetilde{Y}_{r_1}) + I(\mathbf{U}_{t2}^{\ d}; \widetilde{Y}_{r_1} | U_1) + I(U_1; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | \widetilde{Y}_{r_1}) + I(\mathbf{U}_{t2}^{\ d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | U_1, \widetilde{Y}_{r_1}) \tag{3.115}$$

$$\geq I(U_1; \widetilde{Y}_{r_1}) + I(\mathbf{U}_{t2}^{\ d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | U_1, \widetilde{Y}_{r_1}), \tag{3.116}$$

where (3.112) follows since $\mathbf{U}_t - \mathbf{X}_t - \mathbf{Y}_r - \widetilde{\mathbf{Y}}_r$ forms a Markov chain. We next lower bound each term in the RHS of (3.116).

We lower bound the first term, $I(U_1; \widetilde{Y}_{r_1})$ as follows, see also [71, 78]. Let P_{e_1} denote the probability of error in decoding U_1 at the receiver, i.e., $P_{e_1} \triangleq \mathbb{P}\left(\hat{U}_1 \neq U_1\right)$, where $\hat{U}_i$, $i = 1, 2, \cdots, d$, is the estimate of U_i at the legitimate receiver. Thus, using

Fano's inequality, we have

$$I(U_1;\widetilde{Y}_{r_1}) = H(U_1) - H(U_1|\widetilde{Y}_{r_1}) \tag{3.117}$$

$$\geq H(U_1) - 1 - P_{e_1}\log|\mathcal{U}_1| \tag{3.118}$$

$$= \big(1 - P_{e_1}\big)\log(2Q+1) - 1. \tag{3.119}$$

From (3.105), since the mapping $(U_1, V_1) \mapsto f_1U_1 + f_2V_1$ is invertible, the only source of error in decoding U_1 from $\widetilde{Y}_{r_1}$ is the additive Gaussian noise Z'. Note that, since $Z' \sim \mathcal{CN}(0, ||\mathbf{b}||^2)$, $\mathrm{Re}\{Z'\}$ and $\mathrm{Im}\{Z'\}$ are i.i.d. with $\mathcal{N}\left(0, \frac{||\mathbf{b}||^2}{2}\right)$ distribution, and $|Z'| \sim \text{Rayleigh}\left(\frac{||\mathbf{b}||}{\sqrt{2}}\right)$. Thus, we have

$$P_{e_1} \triangleq \mathbb{P}\left(\hat{U}_1 \neq U_1\right) \tag{3.120}$$

$$\leq \mathbb{P}\left((\hat{U}_1, \hat{V}_1) \neq (U_1, V_1)\right) \tag{3.121}$$

$$\leq \mathbb{P}\left(|Z'| \geq \frac{d_{\min}}{2}\right) \tag{3.122}$$

$$= \exp\left(\frac{-d_{\min}^2}{4||\mathbf{b}||^2}\right), \tag{3.123}$$

where $d_{\min}$ is the minimum distance between the points in the constellation $\mathcal{R}_1 = f_1\mathcal{U}_1 + f_2\mathcal{V}_1$.

In order to upper bound P_{e_1}, we lower bound $d_{\min}$. To do so, similar to [71], we extend real interference alignment [78] to complex channels. In particular, we utilize Lemma 1 in Section 2. Lemma 1 implies the following:

Corollary 1. For almost all $\mathbf{z} \in \mathbb{C}^n$ and for all $\epsilon > 0$,

$$|p + \mathbf{z}.\mathbf{q}| > (\max_i |q_i|)^{-\frac{(n-1+\epsilon)}{2}}, \tag{3.124}$$

holds for all $\mathbf{q} \in \mathbb{Z}^n$ and $p \in \mathbb{Z}$ except for finitely many of them.

Since the number of integers that violate the inequality in (3.124) is finite, there exists a constant κ such that, for almost all $\mathbf{z} \in \mathbb{C}^n$ and all $\epsilon > 0$, the inequality

$$|p + \mathbf{z}.\mathbf{q}| > \kappa(\max_i |q_i|)^{-\frac{(n-1+\epsilon)}{2}}, \tag{3.125}$$

holds for all $\mathbf{q} \in \mathbb{Z}^n$ and $p \in \mathbb{Z}$.

Thus, for almost all channel gains, the minimum distance $d_{\min}$ is lower bounded as follows:

$$d_{\min} = \inf_{Y'_{r_1}, Y''_{r_1} \in \mathcal{R}_1} |Y'_{r_1} - Y''_{r_1}| \tag{3.126}$$

$$= \inf_{U_1, V_1 \in \{a(-2Q, 2Q)_{\mathbb{Z}}\}} |f_1 U_1 + f_2 V_1| \tag{3.127}$$

$$= \inf_{U_1, V_1 \in (-2Q, 2Q)_{\mathbb{Z}}} a|f_1| \left| U_1 + \frac{f_2}{f_1} V_1 \right| \tag{3.128}$$

$$\geq \kappa \frac{a|f_1|}{(2Q)^{\frac{\epsilon}{2}}} \tag{3.129}$$

$$\geq \kappa\gamma|f_1| 2^{-\frac{\epsilon}{2}} P^{\frac{\epsilon}{2}}, \tag{3.130}$$

where (3.129) follows from (3.125), and (3.130) follows by substituting (3.89) and (3.90) in (3.129). Substituting (3.130) in (3.123) gives the following bound on P_{e_1},

$$P_{e_1} \leq \exp(-\mu P^{\epsilon}), \tag{3.131}$$

where $\mu = \frac{\kappa^2\gamma^2|f_1|^2 2^{-\epsilon}}{4||\mathbf{b}||^2}$ is a constant which does not depend on the power P. Thus, using (3.119) and (3.131), we have

$$I(U_1; \widetilde{Y}_{r_1}) \geq \left(1 - \exp(-\mu P^{\epsilon})\right) \log(2Q+1) - 1. \tag{3.132}$$

Next, we lower bound the second term in the RHS of (3.116), $I(\mathbf{U}_{t2}^{d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | U_1, \widetilde{Y}_{r_1})$. Let $\widetilde{\mathbf{B}} = \begin{bmatrix} \mathbf{z}_{(N-1)\times 1} & \mathbf{I}_{N-1} \end{bmatrix} - \frac{1}{f_2}\tilde{\mathbf{h}}_{c,1}\mathbf{b}^H$, and

$$\bar{\widetilde{\mathbf{Y}}}'_r = \mathbf{B} \begin{bmatrix} \mathbf{U}_{t2}^{d} \\ \\ \mathbf{V}_{c2}^{l} \end{bmatrix} + \widetilde{\mathbf{B}}\mathbf{Z}_r \tag{3.133}$$

$$\widehat{\mathbf{Y}}'_r = \mathbf{B}^{-1}\bar{\widetilde{\mathbf{Y}}}'_r = \begin{bmatrix} \mathbf{U}_{t2}^{d} \\ \\ \mathbf{V}_{c2}^{l} \end{bmatrix} + \mathbf{B}^{-1}\widetilde{\mathbf{B}}\mathbf{Z}_r, \tag{3.134}$$

where $\mathbf{B}$ is defined as in (3.109). Thus, we have

$$I\left(\mathbf{U}_{t2}^{d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | U_1, \widetilde{Y}_{r_1}\right) = I\left(\mathbf{U}_{t2}^{d}; \widetilde{\mathbf{A}}\mathbf{U}_t + \widetilde{\mathbf{H}}_c\mathbf{V}_c + \mathbf{Z}_{r2}^{\ n_t} | U_1, f_2V_1 + Z'\right) \tag{3.135}$$

$$= I\left(\mathbf{U}_{t2}^{d}; \mathbf{B} \begin{bmatrix} \mathbf{U}_{t2}^{d} \\ \\ \mathbf{V}_{c2}^{l} \end{bmatrix} + \mathbf{Z}_{r2}^{\ n_t} - \frac{1}{f_2}\tilde{\mathbf{h}}_{c,1}\mathbf{b}^H\mathbf{Z}_r \Big| f_2V_1 + Z'\right) \tag{3.136}$$

$$= I(\mathbf{U}_{t2}^{d}; \bar{\widetilde{\mathbf{Y}}}'_r | f_2V_1 + Z') \tag{3.137}$$

$$\geq I(\mathbf{U}_{t2}^{d};\bar{\bar{\mathbf{Y}}}'_r) \tag{3.138}$$

$$\geq I(\mathbf{U}_{t2}^{d};\widehat{\mathbf{Y}}'_r) \tag{3.139}$$

$$= H(\mathbf{U}_{t2}^{d}) - H(\mathbf{U}_{t2}^{d}|\widehat{\mathbf{Y}}'_r) \tag{3.140}$$

$$\geq H(\mathbf{U}_{t2}^{d}) - P_{e_2}^{d}\log(2Q+1)^{2(d-1)} - 1 \tag{3.141}$$

$$= 2(d-1)\left(1-P_{e_2}^{d}\right)\log(2Q+1) - 1, \tag{3.142}$$

where $P_{e_2}^{d} \triangleq \mathbb{P}\left((\hat{U}_2,\hat{U}_3,\cdots,\hat{U}_d) \neq (U_2,U_3,\cdots,U_d)\right)$, (3.135) follows from (3.104), (3.138) follows since $\mathbf{U}_{t2}^{d}$ and f_2V_1+Z' are independent, (3.139) follows since $\mathbf{U}_{t2}^{d} - \bar{\bar{\mathbf{Y}}}'_r - \widehat{\mathbf{Y}}'_r$ forms a Markov chain, and (3.141) follows from Fano's inequality.

Let $\widehat{\mathbf{Z}}_r \triangleq \mathbf{\Theta}\mathbf{Z}_r = [\hat{Z}_{r_2} \ \cdots \ \hat{Z}_{r_N}]^T$, where $\mathbf{\Theta} = \mathbf{B}^{-1}\widetilde{\mathbf{B}}$. Thus, $\widehat{\mathbf{Z}}_r \sim \mathcal{CN}(\mathbf{z},\mathbf{\Theta}\mathbf{\Theta}^H)$ and $|\hat{Z}_{r_i}| \sim$ Rayleigh(σ_i), where $\sigma_i^2 = \mathbf{\Theta}\mathbf{\Theta}^H(i,i)$, $i = 2,3,\cdots,N$. Using the union bound, we have

$$P_{e_2}^{d} = \mathbb{P}\left((\hat{U}_2,\hat{U}_3,\cdots,\hat{U}_d) \neq (U_2,U_3,\cdots,U_d)\right) \tag{3.143}$$

$$\leq \sum_{i=2}^{d}\mathbb{P}\left(\hat{U}_i \neq U_i\right) \tag{3.144}$$

$$\leq \sum_{i=2}^{d}\mathbb{P}\left(|\hat{Z}_{r_i}| \geq \frac{a}{2}\right) \tag{3.145}$$

$$= \sum_{i=2}^{d}\exp\left(-\frac{a^2}{8\sigma_i^2}\right) \tag{3.146}$$

$$\leq (d-1)\exp\left(-\frac{\gamma^2}{8\sigma_{\max}^2}P^{\frac{3\epsilon}{2+\epsilon}}\right) \tag{3.147}$$

$$= (d-1)\exp(-\mu' P^{\epsilon'}), \tag{3.148}$$

where $\sigma_{\max} = \max_i \sigma_i$, $\mu' = \frac{\gamma^2}{8\sigma_{\max}^2}$, $\epsilon' = \frac{3\epsilon}{2+\epsilon}$, and (3.147) follows by substituting (3.90) in (3.146).

Substituting (3.148) in (3.142) yields

$$I\left(\mathbf{U}_{t2}^{d}; \widetilde{\mathbf{Y}}_{r_2}^{n_t} | U_1, \widetilde{Y}_{r_1}\right) \geq \left(2d - 2 - 2(d-1)^2 \exp(-\mu' P^{\epsilon'})\right) \log(2Q+1) - 1. \tag{3.149}$$

Using (3.89), (3.116), (3.132), and (3.149), we have

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq \left[2d - 1 - \exp(-\mu P^{\epsilon}) - 2(d-1)^2 \exp(-\mu' P^{\epsilon'})\right] \log\left(2P^{\frac{1-\epsilon}{2+\epsilon}} - 2\nu + 1\right) - 2 \tag{3.150}$$

$$= \frac{1-\epsilon}{2+\epsilon}\left[2d - 1 - \exp(-\mu P^{\epsilon}) - 2(d-1)^2 \exp(-\mu' P^{\epsilon'})\right] \log P + o(\log P). \tag{3.151}$$

Using the upper bound in (3.100) and the lower bound in (3.151), we get

$$R_s \geq \frac{1-\epsilon}{2+\epsilon}\left[2d - 1 - \exp(-\mu P^{\epsilon}) - 2(d-1)^2 \exp(-\mu' P^{\epsilon'})\right] \log P + o(\log P) - (2l-1) \tag{3.152}$$

$$= \frac{1-\epsilon}{2+\epsilon}\left[2N - N_e - \exp(-\mu P^{\epsilon}) - \frac{1}{2}(2N - N_e - 1)^2 \exp(-\mu' P^{\epsilon'})\right] \log P + o(\log P) - N_e. \tag{3.153}$$

Thus, it follows that the s.d.o.f. is lower bounded as

$$D_s \geq \frac{(1-\epsilon)(2N - N_e)}{2+\epsilon}. \tag{3.154}$$

Since $\epsilon > 0$ can be chosen arbitrarily small, we can achieve s.d.o.f. of $N - \frac{N_e}{2}$.

3.5.4 Case 4: $N_e \leq N$, $N < N_c \leq N + N_e$, and $N + N_c - N_e$ is even

Since $N_c > N$ for this case, the cooperative jammer, unlike the previous three cases, chooses its precoder such that $N_c - N$ of its jamming streams are sent invisible to the receiver, in order to allow for more space for the information streams at the receiver. The s.d.o.f. for this case is integer valued, which we can achieve using Gaussian information and cooperative jamming streams.

The transmitted signals are given by (3.72), with $d = \frac{N+N_c-N_e}{2}$, $l = \frac{N_c+N_e-N}{2}$, $\mathbf{U}_t \sim \mathcal{CN}\left(\mathbf{z}, \bar{P}\mathbf{I}_d\right)$, $\mathbf{V}_c \sim \mathcal{CN}\left(\mathbf{z}, \bar{P}\mathbf{I}_l\right)$,

$$\mathbf{P}_c = [\mathbf{P}_{c,\mathrm{I}}\, \mathbf{P}_{c,n}], \tag{3.155}$$

where $\mathbf{P}_{c,\mathrm{I}}$ is given by

$$\mathbf{P}_{c,\mathrm{I}} = \begin{bmatrix} \mathbf{I}_g \\ \mathbf{z}_{(N_c-g)\times g} \end{bmatrix}, \tag{3.156}$$

$g = \frac{N_e+N-N_c}{2}$, and $\mathbf{P}_{c,n} \in \mathbb{C}^{N_c \times (N_c-N)}$ is a matrix whose columns span $\mathcal{N}(\mathbf{H}_c)$, $\mathbf{P}_t$ is defined as in Section 3.5.2, and $\bar{P} = \frac{1}{\alpha'}P$, where $\alpha' = \max\left\{\sum_{i=1}^{d} ||\mathbf{p}_{t,i}||^2, g + \sum_{i=g+1}^{l} ||\mathbf{p}_{c,i}||^2\right\}$. At high SNR, the receiver can decode the d information and the g cooperative jamming streams, where $d + g = N$.

The received signals at the legitimate receiver and the eavesdropper are given by

$$\mathbf{Y}_r = \mathbf{H}_t\mathbf{P}_t\mathbf{U}_t + \begin{bmatrix} \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} & \mathbf{z}_{N\times(N_c-N)} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{c_1}^{\,g} \\ \mathbf{V}_{c_{g+1}}^{\,l} \end{bmatrix} + \mathbf{Z}_r \tag{3.157}$$

$$= \begin{bmatrix} \mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_t \\ \mathbf{V}_{c1}^{g} \end{bmatrix} + \mathbf{Z}_r \tag{3.158}$$

$$\mathbf{Y}_e = \tilde{\mathbf{G}}_c(\mathbf{U}_{t1}^{l} + \mathbf{V}_c) + \mathbf{Z}_e, \tag{3.159}$$

where $\tilde{\mathbf{G}}_c = \mathbf{G}_c\mathbf{P}_c$.

The matrix $\begin{bmatrix} \mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} \end{bmatrix} \in \mathbb{C}^{N\times N}$ in (3.158) can be rewritten as

$$\begin{bmatrix} \mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_t & \mathbf{H}_c \end{bmatrix} \begin{bmatrix} \mathbf{P}_t & \mathbf{z}_{N\times g} \\ \mathbf{z}_{N_c\times d} & \mathbf{P}_{c,\mathrm{I}} \end{bmatrix}. \tag{3.160}$$

By applying Lemma 6 on (3.160), the matrix $\begin{bmatrix} \mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} \end{bmatrix}$ is full rank a.s. Thus,

$$I(\mathbf{X}_t;\mathbf{Y}_r) \geq d\log P + o(\log P). \tag{3.161}$$

Using similar steps as from (3.79) to (3.84), we can show that

$$I(\mathbf{X}_t;\mathbf{Y}_e) = \log\frac{\det(\mathbf{I}_l + 2\bar{P}\tilde{\mathbf{G}}_c^H\tilde{\mathbf{G}}_c)}{\det(\mathbf{I}_l + \bar{P}\tilde{\mathbf{G}}_c^H\tilde{\mathbf{G}}_c)} \leq l. \tag{3.162}$$

Thus, the achievable secrecy rate in (3.71) is lower bounded as

$$R_s \geq d\log P + o(\log P) - l \tag{3.163}$$

$$= \frac{N + N_c - N_e}{2}\log P + o(\log P) - \frac{N_c + N_e - N}{2}, \tag{3.164}$$

and, using (3.5), the s.d.o.f. is lower bounded as

$$D_s \geq \frac{N + N_c - N_e}{2}. \tag{3.165}$$

3.5.5 Case 5: $N_e \leq N$, $N < N_c \leq N + N_e$, and $N + N_c - N_e$ is odd

As in case 3, the s.d.o.f. for this case is not an integer, and as in case 4, we have $N_c > N$, which allows the cooperative jammer to send some signals invisible to the receiver. Consequently, the achievable scheme for this case combines the techniques used in Sections 3.5.3 and 3.5.4.

The transmitted signals are given by (3.72) with $d = \frac{N+N_c-N_e+1}{2}$, $l = \frac{N_c+N_e-N+1}{2}$, $\mathbf{P}_t$ and $\mathbf{P}_c$ are defined as in Section 3.5.4 with $g = \frac{N_e+N-N_c+1}{2}$, and $\mathbf{U}_t$, $\mathbf{V}_c$ are defined as in Section 3.5.3. Similar to the proof in Appendix E, the values of Q and a are chosen as in (3.89) and (3.90), with

$$\gamma = \frac{1}{\sqrt{\max\left\{||\mathbf{p}_{t,1}||^2 + 2\sum_{i=2}^{d}||\mathbf{p}_{t,i}||^2, 2g - 1 + 2\sum_{i=g+1}^{l}||\mathbf{p}_{c,i}||^2\right\}}}, \tag{3.166}$$

and ν are constants that do not depend on the power P.

The legitimate receiver uses the projection and cancellation technique described in Section 3.5.3 in order to decode the information streams. The received signal at the eavesdropper is the same as in (3.159), with $l = \frac{N_c+N_e-N+1}{2}$. Similar to the derivation from (3.92) to (3.100), we have

$$I(\mathbf{X}_t; \mathbf{Y}_e) \leq 2l - 1. \tag{3.167}$$

Let $\mathbf{A} = \mathbf{H}_t\mathbf{P}_t = [\mathbf{a}_1 \cdots \mathbf{a}_d]$, and $\mathbf{H}_c' = \mathbf{H}_c\mathbf{P}_{c,\mathrm{I}} = [\mathbf{h}_{c,1} \cdots \mathbf{h}_{c,g}]$. The received signal at the legitimate receiver is

$$\mathbf{Y}_r = \begin{bmatrix} \mathbf{A} & \mathbf{H}_c' \end{bmatrix} \begin{bmatrix} \mathbf{U}_t \\ \mathbf{V}_{c1}^{\,g} \end{bmatrix} + \mathbf{Z}_r. \tag{3.168}$$

As in case 3, we have that $d+g-2 = N-1$, and hence the dimension of $\mathrm{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c_2}, \cdots, \mathbf{h}_{c_g}\}$ is at most $N-1$, and there exists $\mathbf{b} \in \mathbb{C}^N$ such that $\mathbf{b}$ is orthogonal to $\mathrm{span}\{\mathbf{a}_2, \cdots, \mathbf{a}_d, \mathbf{h}_{c_2}, \cdots, \mathbf{h}_{c_g}\}$. The receiver chooses such $\mathbf{b}$ and multiplies its received signal by the matrix $\mathbf{D}$ given in (3.103) to obtain $\widetilde{\mathbf{Y}}_r = \left[\widetilde{Y}_{r_1} \ (\widetilde{\mathbf{Y}}_{r_2}^{n_t})^T\right]^T$, where

$$\widetilde{Y}_{r_1} = f_1 U_1 + f_2 V_1 + Z', \tag{3.169}$$

$$\widetilde{\mathbf{Y}}_{r_2}^{n_t} = \widetilde{\mathbf{A}}\mathbf{U}_t + \widetilde{\mathbf{H}}_c \mathbf{V}_{c1}^{\,g} + \mathbf{Z}_{r2}^{\,n_t}, \tag{3.170}$$

f_1, f_2, Z', $\widetilde{\mathbf{A}}$, and $\widetilde{\mathbf{H}}_c$, are defined as in Section 3.5.3. In order to decode U_1 and V_1, the receiver passes $\widetilde{Y}_{r_1}$ through a hard decision decoder, $\widetilde{Y}_{r_1} \mapsto f_1 U_1 + f_2 V_1$, and passes the output of the hard decision decoder through the bijective map $f_1 U_1 + f_2 V_1 \mapsto (U_1, V_1)$, where f_1 and f_2 are rationally independent.

Using similar steps to the derivation from (3.112) to (3.151) in Section 3.5.3, we obtain

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq \frac{1-\epsilon}{2+\epsilon}\Big[2d - 1 - \exp\left(-\mu P^{\epsilon}\right) - 2(d-1)^2 \exp(-\mu' P^{\epsilon'})\Big] \log P + o(\log P), \tag{3.171}$$

where $\epsilon > 0$ is arbitrarily small, $\epsilon' = \frac{3\epsilon}{2+\epsilon}$, and μ, μ' are constants which do not depend on P.

Thus, the achievable secrecy rate in (3.71) is lower bounded as

$$R_s \geq \frac{1-\epsilon}{2+\epsilon}\left[2d - 1 - \exp(-\mu P^{\epsilon}) - (d-1)^2 \exp(-\mu' P^{\epsilon'})\right] \log P + o(\log P) - (2l-1) \tag{3.172}$$

$$= \frac{1-\epsilon}{2+\epsilon}\left[N + N_c - N_e - \exp(-\mu P^{\epsilon}) - \frac{1}{2}(N + N_c - N_e - 1)^2 \exp(-\mu' P^{\epsilon'})\right] \log P$$

$$+ o(\log P) - (N_c + N_e - N), \tag{3.173}$$

and hence the s.d.o.f is lower bounded as

$$D_s \geq \frac{(1-\epsilon)(N + N_c - N_e)}{2+\epsilon}. \tag{3.174}$$

Since $\epsilon > 0$ can be chosen arbitrarily small, $D_s = \frac{N+N_c-N_e}{2}$ is achievable for this case, which completes the achievability of (3.69). Next, we show the achievability of (3.70), where $N_e > N$, i.e., the eavesdropper has more antennas than the legitimate receiver.

3.5.6 Case 6: $N_e > N$ and $N_e - N < N_c \leq N_e - \frac{N}{2}$

Unlike the previous five cases, since $N_e > N$, no information streams can be sent invisible to the eavesdropper. In fact, the precoder at the transmitter is not adequate for achieving the alignment of the information and cooperative jamming streams at the eavesdropper. We need to design both precoders at the transmitter and the cooperative

jammer to take part in achieving the alignment condition. The s.d.o.f. here is integer valued, and hence we can utilize Gaussian streams.

The transmitted signals are given by (3.72), with $d = l = N + N_c - N_e$, and $\mathbf{U}_t, \mathbf{V}_c \sim \mathcal{CN}\left(\mathbf{z}, \bar{P}\mathbf{I}_d\right)$. The matrices $\mathbf{P}_t$ and $\mathbf{P}_c$ are chosen as follows. Let $\mathbf{G} = [\mathbf{G}_t \ \ -\mathbf{G}_c] \in \mathbb{C}^{N_e \times (N+N_c)}$, and let $\mathbf{Q} \in \mathbb{C}^{(N+N_c)\times d}$ be a matrix whose columns are randomly[9] chosen to span $\mathcal{N}(\mathbf{G})$. Write the matrix $\mathbf{Q}$ as $\mathbf{Q} = \left[\mathbf{Q}_1^T \ \ \mathbf{Q}_2^T\right]^T$, where $\mathbf{Q}_1 \in \mathbb{C}^{N\times d}$ and $\mathbf{Q}_2 \in \mathbb{C}^{N_c \times d}$. Set $\mathbf{P}_t = \mathbf{Q}_1$ and $\mathbf{P}_c = \mathbf{Q}_2$. $\bar{P} = \frac{1}{\alpha''}P$, where $\alpha'' = \max\left\{\sum_{i=1}^d ||\mathbf{p}_{t,i}||^2, \sum_{i=1}^d ||\mathbf{p}_{c,i}||^2\right\}$.

The choice of $\mathbf{P}_t$ and $\mathbf{P}_c$ results in $\mathbf{G}_t\mathbf{P}_t = \mathbf{G}_c\mathbf{P}_c$. Thus, the eavesdropper receives

$$\mathbf{Y}_e = \mathbf{G}_c\mathbf{P}_c(\mathbf{U}_t + \mathbf{V}_c) + \mathbf{Z}_e. \tag{3.175}$$

Similar to going from (3.79) to (3.84), it follows that we have

$$I(\mathbf{X}_t; \mathbf{Y}_e) \leq N + N_c - N_e. \tag{3.176}$$

The received signal at the receiver in turn is given by

$$\mathbf{Y}_r = \begin{bmatrix}\mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_c\end{bmatrix}\begin{bmatrix}\mathbf{U}_t \\ \\ \mathbf{V}_c\end{bmatrix} + \mathbf{Z}_r. \tag{3.177}$$

Note that, without conditioning on $\mathbf{G}_t$ and $\mathbf{G}_c$, the matrix $\mathbf{Q}$ has all of its entries independently and randomly drawn according to a continuous distribution. Thus, each

[9] Out of all possible sets of $d = N + N_c - N_e$ linearly independent vectors which span $\mathcal{N}(\mathbf{G})$, the columns of $\mathbf{Q}$ are the elements of one randomly chosen set.

of $\mathbf{P}_t$ and $\mathbf{P}_c$ is full column rank a.s. Thus, by using Lemma 6, we can show that the matrix $[\mathbf{H}_t\mathbf{P}_t \;\; \mathbf{H}_c\mathbf{P}_c]$ is full column rank a.s. Using (3.177), we have

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq (N + N_c - N_e) \log P + o(\log P). \tag{3.178}$$

Hence, using (3.176), (3.178), (3.71), and (3.5), the s.d.o.f. is lower bounded as $D_s \geq N + N_c - N_e$.

3.5.7 Case 7: $N_e > N$, $N_e - \frac{N}{2} < N_c \leq N_e$, and N is even

The s.d.o.f. for this case does not increase by increasing N_c. The scheme in Section 3.5.6 for $N_c = N_e - \frac{N}{2}$, i.e., $d = \frac{N}{2}$, can be used to achieve the s.d.o.f. for all $N_e - \frac{N}{2} < N_c \leq N_e$, when $N_e > N$ and N is even. However, since $\dim(\mathcal{N}(\mathbf{G})) = N + N_c - N_e > \frac{N}{2}$, the $d = \frac{N}{2}$ columns of the matrix $\mathbf{Q}$ are randomly chosen as linearly independent vectors from $\mathcal{N}(\mathbf{G})$. Following the same analysis as in Section 3.5.6, we can show that the s.d.o.f. is lower bounded as $D_s \geq \frac{N}{2}$.

3.5.8 Case 8: $N_e > N$, $N_e - \frac{N}{2} < N_c \leq N_e$, and N is odd

The difference here from Section 3.5.7 is that the s.d.o.f. is not an integer, and hence, structured signaling for transmission and cooperative jamming is needed, and the difference from 3.5.3 is that $N_e > N$, and hence both the precoders at the transmitter and cooperative jammer have to participate in achieving the alignment condition at the eavesdropper.

The transmitted signals are given by (3.72), with $d = l = \frac{N+1}{2}$, $\mathbf{U}_t$ and $\mathbf{V}_c$ are defined as in Section 3.5.3, and the values for Q and a are chosen as in (3.89) and (3.90),

with

$$\gamma = \frac{1}{\sqrt{\max\left\{|||\mathbf{p}_{t,1}||^2 + 2\sum_{i=2}^{d}||\mathbf{p}_{t,i}||^2, ||\mathbf{p}_{c,1}||^2 + 2\sum_{i=2}^{d}||\mathbf{p}_{c,i}||^2\right\}}}, \tag{3.179}$$

and ν are constants which do not depend P. $\mathbf{P}_t, \mathbf{P}_c$ are chosen as in Section 3.5.7, with $d = \frac{N+1}{2}$. The eavesdropper's received signal is the same as in (3.175). Similar to (3.92)-(3.100), we have

$$I(\mathbf{X}_t; \mathbf{Y}_e) \leq N. \tag{3.180}$$

The receiver employs the decoding scheme in Sections 3.5.3 and 3.5.5. Following similar steps as in Sections 3.5.3 and 3.5.5, we have

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq \frac{(1-\epsilon)N}{2+\epsilon}\log P + o(\log P). \tag{3.181}$$

Using (3.180), (3.181), (3.71), and (3.5), the s.d.o.f. is lower bounded as $D_s \geq \frac{(1-\epsilon)N}{2+\epsilon}$, and since $\epsilon > 0$ is arbitrarily small, the s.d.o.f. of $\frac{N}{2}$ is achievable for this case.

3.5.9 Case 9: $N_e > N$, $N_e < N_c \leq N + N_e$, and $N + N_c - N_e$ is even

In Sections 3.5.7 and 3.5.8, we observe that the flat s.d.o.f. range extends to $N_c = N_e$, and not $N_c = N$ as in Sections 3.5.2 and 3.5.3. Achieving the alignment of information and cooperative jamming at the eavesdropper requires that $N_c > N_e$ in order for the cooperative jammer to begin sending some jamming signals invisible to the legitimate receiver. For this case, in addition to choosing its precoding matrix jointly

with the transmitter to satisfy the alignment condition, the cooperative jammer chooses its precoder to send $N_c - N_e$ jamming streams invisible to the receiver. The s.d.o.f. here is integer valued, for which we utilize Gaussian streams.

The transmitted signals are given by (3.72) with $d = l = \frac{N+N_c-N_e}{2}$, and $\mathbf{U}_t, \mathbf{V}_c$ are defined as in Section 3.5.6. Let $\mathbf{P}_t = \begin{bmatrix}\mathbf{P}_{t,1}\, \mathbf{P}_{t,2}\end{bmatrix}$, and $\mathbf{P}_c = \begin{bmatrix}\mathbf{P}_{c,1}\, \mathbf{P}_{c,2}\end{bmatrix}$, where $\mathbf{P}_{t,1} \in \mathbb{C}^{N\times g}$, $\mathbf{P}_{t,2} \in \mathbb{C}^{N\times(N_c-N_e)}$, $\mathbf{P}_{c,1} \in \mathbb{C}^{N_c\times g}$, $\mathbf{P}_{c,2} \in \mathbb{C}^{N_c\times(N_c-N_e)}$, and $g = \frac{N_e+N-N_c}{2}$. The matrices $\mathbf{P}_t$ and $\mathbf{P}_c$ are chosen as follows. Let $\mathbf{G} = [\mathbf{G}_t \ -\mathbf{G}_c] \in \mathbb{C}^{N_e\times(N+N_c)}$, and let $\mathbf{G}' \in \mathbb{C}^{(N_e+N)\times(N+N_c)}$ be expressed as

$$\mathbf{G}' = \begin{bmatrix} \mathbf{G}_t & -\mathbf{G}_c \\ \mathbf{z}_{N\times N} & \mathbf{H}_c \end{bmatrix}. \tag{3.182}$$

Let $\mathbf{Q}' \in \mathbb{C}^{(N+N_c)\times(N_c-N_e)}$ be randomly chosen such that its columns span $\mathcal{N}(\mathbf{G}')$, and let the columns of the matrix $\mathbf{Q} \in \mathbb{C}^{(N+N_c)\times g}$ be randomly chosen as linearly independent vectors in $\mathcal{N}(\mathbf{G})$, and not in $\mathcal{N}(\mathbf{G}')$. Write the matrix $\mathbf{Q}$ as $\mathbf{Q} = \begin{bmatrix}\mathbf{Q}_1^T\ \mathbf{Q}_2^T\end{bmatrix}^T$, and the matrix $\mathbf{Q}'$ as $\mathbf{Q}' = \begin{bmatrix}\mathbf{Q}_1'^T\ \mathbf{Q}_2'^T\end{bmatrix}^T$, where $\mathbf{Q}_1 \in \mathbb{C}^{N\times g}$, $\mathbf{Q}_2 \in \mathbb{C}^{N_c\times g}$, $\mathbf{Q}_1' \in \mathbb{C}^{N\times(N_c-N_e)}$, and $\mathbf{Q}_2' \in \mathbb{C}^{N_c\times(N_c-N_e)}$. Set $\mathbf{P}_{t,1} = \mathbf{Q}_1$, $\mathbf{P}_{t,2} = \mathbf{Q}_1'$, $\mathbf{P}_{c,1} = \mathbf{Q}_2$, and $\mathbf{P}_{c,2} = \mathbf{Q}_2'$.

This choice of $\mathbf{P}_t$ and $\mathbf{P}_c$ results in $\mathbf{G}_t\mathbf{P}_t = \mathbf{G}_c\mathbf{P}_c$ and $\mathbf{H}_c\mathbf{P}_{c,2} = \mathbf{z}_{N\times(N_c-N_e)}$. Thus, the received signals at the receiver and eavesdropper are given by

$$\mathbf{Y}_r = \begin{bmatrix}\mathbf{H}_t\mathbf{P}_t & \mathbf{H}_c\mathbf{P}_{c,1}\end{bmatrix} \begin{bmatrix}\mathbf{U}_t \\ \mathbf{V}_{c1}^g\end{bmatrix} + \mathbf{Z}_r \tag{3.183}$$

$$\mathbf{Y}_e = \mathbf{G}_c\mathbf{P}_c(\mathbf{U}_t + \mathbf{V}_c) + \mathbf{Z}_e. \tag{3.184}$$

Using (3.184), and similar to going from (3.79) to (3.84), we have

$$I(\mathbf{X}_t; \mathbf{Y}_e) \leq \frac{N + N_c - N_e}{2}. \tag{3.185}$$

Because of the assumption of randomly generated channel gains, each of $\mathbf{P}_t$ and $\mathbf{P}_c$ is full column rank a.s. Using Lemma 6, we have the matrix $\left[\mathbf{H}_t\mathbf{P}_t \;\; \mathbf{H}_c\mathbf{P}_{c,1}\right]$ is full column rank a.s., and hence, using (3.183), we have

$$I(\mathbf{X}_t; \mathbf{Y}_r) \geq \frac{N + N_c - N_e}{2} \log P + o(\log P). \tag{3.186}$$

Thus, using (3.185), (3.186), (3.71), and (3.5), the s.d.o.f. is lower bounded as $D_s \geq \frac{N+N_c-N_e}{2}$.

3.5.10 Case 10: $N_e > N$, $N_e < N_c \leq N + N_e$, and $N + N_c - N_e$ is odd

The s.d.o.f. for this case is not an integer, and we have $N_c > N_e$, and hence, we utilize here precoding as in Section 3.5.9, and signaling and decoding scheme as in Section 3.5.8; $\mathbf{U}_t, \mathbf{V}_c$ are defined as in Section 3.5.8, and $\mathbf{P}_t, \mathbf{P}_c$ are chosen as in Section 3.5.9, with $d = \frac{N+N_c-N_e+1}{2}$ and $g = \frac{N_e+N-N_c+1}{2}$. Using the same decoding scheme as in Section 3.5.8, we obtain that the s.d.o.f. is lower bounded as $D_s \geq \frac{N+N_c-N_e}{2}$ for this case, which completes the achievability proof of (3.70). Thus, we have completed the proof for Theorem 1.

3.6 Extending to the General Case: Theorem 2

The converse and achievability proofs for Theorem 2 involve the same techniques as those utilized for Theorem 1. However, one needs to carefully handle the antenna configurations when $N_t \neq N_r$. In the following, we summarize how to extend the main ideas presented in Sections 3.4 and 3.5 in order to prove Theorem 2.

3.6.1 Converse

The converse proof for Theorem 2 follows similar steps as in Section 3.4. In particular, we derive the following two upper bounds which hold for two different ranges of N_c.

3.6.1.1 $0 \leq N_c \leq N_e$

When $N_t \neq N_r$, the range of N_c for which the first upper bound holds is the same as in the case when $N_t = N_r = N$ in Section 3.4.1. However, unlike in Section 3.4.1, when $N_t \neq N_r$, this range of N_c is further subdivided into two ranges. The first upper bound on the s.d.o.f. we derive here is again $D_s \leq [N_t + N_c - N_e]^+$, yet, the maximum s.d.o.f. for the channel is equal to $\min\{N_t, N_r\}$. Hence, for the case $N_r < N_t + N_c - N_e$, the maximum s.d.o.f., N_r, is reached at an N_c that is smaller than N_e. In particular, using similar analysis as in Section 3.4.1, we have

$$R_s \leq C_s(P) = \rho \log P + o(\log P), \tag{3.187}$$

where, for $0 \leq N_c \leq \left[N_e - [N_t - N_r]^+\right]^+$, $\rho = [N_c + N_t - N_e]^+$. Since $[N_c + N_t - N_e]^+ \leq N_r$ for $\left[N_e - [N_t - N_r]^+\right]^+ \leq N_c \leq N_e$, we have, for $0 \leq N_c \leq N_e$,

$$D_s \leq \min\{N_r, [N_c + N_t - N_e]^+\}. \tag{3.188}$$

3.6.1.2 $N_r + [N_e - N_t]^+ < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$

Following similar steps as in Section 3.4.2, where the two cases we consider here are $N_e \leq N_t$ and $N_e > N_t$, the s.d.o.f. for this range of N_c is upper bounded as

$$D_s \leq \frac{N_c + N_t - N_e}{2}. \tag{3.189}$$

It easy to see that, when $N_t = N_r = N$, the range of N_c for which the second upper bound in (3.189) holds is reduced to the range $\max\{N, N_e\} < N_c \leq N + N_e$ in Section 3.4.2. However, when $N_t \neq N_r$, the range of N_c is different. In particular, we have that $N_c > N_r + [N_e - N_t]^+$ because, when $N_e > N_t$, (3.189) holds only when $N_c > N_r + N_e - N_t$ so that the number of antennas at the cooperative jammer in the modified channel, c.f. (3.59), is greater than N_r. We also have that $N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$. This this because, when $N_t < N_r$, we have $\frac{N_c + N_t - N_e}{2} = N_t$ at $N_c = N_t + N_e$, and when $N_t > N_r$, we have $\frac{N_c + N_t - N_e}{2} = N_r$ at $N_c = 2N_r + N_e - N_t$.

3.6.1.3 Obtaining the Upper Bound

For each of the following cases, we use the two bounds in (3.188) and (3.189) to obtain the upper bound for the s.d.o.f.

i) $N_t \geq N_r + N_e$

For this case, we use the trivial bound for the s.d.o.f., $D_s \leq N_r$ for all the values of N_c.

ii) $N_r \geq N_t \geq N_e$ and $N_r \geq N_t + N_e$

Using the bound in (3.188), we have

$$D_s \leq N_c + N_t - N_e, \text{ for } 0 \leq N_c \leq N_e,$$

where at $N_c = N_e$, we have $D_s \leq N_t$, which is the maximum achievable s.d.o.f. for this case.

iii) $N_t \geq N_e$ and $N_t - N_e < N_r < N_t + N_e$

Combining the bounds in (3.188) and (3.189), as in Section 3.4.3, yields

$$D_s \leq \begin{cases} N_c + N_t - N_e, & \text{if } 0 \leq N_c \leq \frac{N_r+N_e-N_t}{2} \\ \frac{N_r+N_t-N_e}{2}, & \text{if } \frac{N_r+N_e-N_t}{2} < N_c \leq N_r \\ \frac{N_c+N_t-N_e}{2}, & \text{if } N_r < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t. \end{cases} \tag{3.190}$$

iv) $N_e > N_t$ and $N_r \geq 2N_t$

Using the bound in (3.188), we have

$$D_s \leq [N_c + N_t - N_e]^+, \text{ for } 0 \leq N_c \leq N_e.$$

v) $N_e > N_t$ and $N_r < 2N_t$

By combining the bounds in (3.188) and (3.189), we have

$$D_s \leq \begin{cases} [N_c + N_t - N_e]^+, & \text{if } 0 \leq N_c \leq \frac{N_r}{2} + N_e - N_t \\ \frac{N_r}{2}, & \text{if } \frac{N_r}{2} + N_e - N_t < N_c \leq N_r + N_e - N_t \\ \frac{N_c + N_t - N_e}{2}, & \text{if } N_r + N_e - N_t < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t. \end{cases} \tag{3.191}$$

One can easily verify that the cases cited above cover all possible combinations of number of antennas at various terminals. By merging the upper bounds for these cases in one expression, we obtain (3.7) as the upper bound for the s.d.o.f. of the channel.

3.6.2 Achievability

The s.d.o.f. for the channel when N_t is not equal to N_r, given in (3.7), is achieved using techniques similar to what we presented in Section 3.5. There are few cases, of the number of antennas, where the achievability is straightforward. One such case is when $N_t \geq N_r + N_e$, where the transmitter can send N_r Gaussian information streams invisible to the eavesdropper, and the maximum possible s.d.o.f. of the channel, i.e., N_r, is achieved without the help of the cooperative jammer, i.e., $N_c = 0$. Another case is when $N_r \geq N_t + \min\{N_t, N_e\}$, where the receiver's signal space is sufficient for decoding the information and jamming streams, at high SNR, for all $0 \leq N_c \leq N_e$, arriving at the s.d.o.f. of N_t (the maximum possible s.d.o.f.) at $N_c = N_e$. Thus, there is no constant period in the s.d.o.f. characterization for this case where the s.d.o.f. keeps increasing by

increasing N_c, and Gaussian signaling and cooperative jamming are sufficient to achieve the s.d.o.f. of the channel.

We consider the five cases of the number of antennas at the different terminals listed in Section 3.6.1.3. In the following, we summarize the achievable schemes for these cases. Let d and l denote the number of information and cooperative jamming streams. $\mathbf{P}_t, \mathbf{P}_c$ are the precoding matrices at the transmitter and the cooperative jammer.

i) $N_t \geq N_r + N_e$

The transmitter sends N_r Gaussian information streams over $\mathcal{N}(\mathbf{G}_t)$. $D_s = N_r$ is achievable at $N_c = 0$.

ii) $N_r \geq N_t \geq N_e$ and $N_r \geq N_t + N_e$

For $0 \leq N_c \leq N_e$, $d = N_c + N_t - N_e$ and $l = N_c$ Gaussian streams are transmitted. Choose $\mathbf{P}_t$ to send $N_t - N_e$ information streams over $\mathcal{N}(\mathbf{G}_t)$ and align the remaining information streams over cooperative jamming streams at the eavesdropper. $D_s = N_c + N_t - N_e$.

iii) $N_t \geq N_e$ and $N_t - N_e < N_r < N_t + N_e$:

1) For $0 \leq N_c \leq \frac{N_r+N_e-N_t}{2}$:

The same scheme as in case (ii) is utilized. $D_s = N_c + N_t - N_e$.

2) For $\frac{N_r+N_e-N_t}{2} < N_c \leq N_r$ and $N_r + N_t - N_e$ is even:

The same scheme as in case (iii-1), with $d = \frac{N_r+N_t-N_e}{2}$ and $l = \frac{N_r+N_e-N_t}{2}$, is utilized. The cooperative jammer uses only $\frac{N_r+N_e-N_t}{2}$ of its N_c antennas. $D_s = \frac{N_r+N_t-N_e}{2}$.

3) For $\frac{N_r+N_e-N_t}{2} < N_c \leq N_r$ and $N_r + N_t - N_e$ is odd:

$d = \frac{N_r+N_t-N_e+1}{2}$ and $l = \frac{N_r+N_e-N_t+1}{2}$ structured streams, as defined in Section 3.5.3, are transmitted. The cooperative jammer uses only $\frac{N_r+N_e-N_t+1}{2}$ of its N_c antennas. $\mathbf{P}_t$ is chosen as in case (ii). The legitimate receiver uses the projection and cancellation technique, as in Section 3.5.3. $D_s = \frac{N_r+N_t-N_e}{2}$.

4) For $N_r < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$ and $N_c + N_t - N_e$ is even:

$d = \frac{N_c+N_t-N_e}{2}$ and $l = \frac{N_c+N_e-N_t}{2}$ Gaussian streams are transmitted. The cooperative jammer chooses $\mathbf{P}_c$ to send $N_c - N_r$ cooperative jamming streams over $\mathcal{N}(\mathbf{H}_c)$. $\mathbf{P}_t$ is chosen as in case (ii). $D_s = \frac{N_c+N_t-N_e}{2}$.

5) For $N_r < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$ and $N_c + N_t - N_e$ is odd:

$d = \frac{N_c+N_t-N_e+1}{2}$ and $l = \frac{N_c+N_e-N_t+1}{2}$ structured streams are transmitted. $\mathbf{P}_t, \mathbf{P}_c$ are chosen as in case (iii-4). The legitimate receiver uses the projection and cancellation technique. $D_s = \frac{N_c+N_t-N_e}{2}$.

iv) $N_e > N_t$ and $N_r \geq 2N_t$

For $0 \leq N_c \leq N_e$, $d = l = [N_c + N_t - N_e]^+$ Gaussian streams are transmitted. Both $\mathbf{P}_t, \mathbf{P}_c$ are chosen to align the information streams over the cooperative jamming streams at the eavesdropper as in Section 3.5.6. $D_s = [N_c + N_t - N_e]^+$.

v) $N_e > N_t$ and $N_r < 2N_t$:

1) For $0 \leq N_c \leq \frac{N_r}{2} + N_e - N_t$:

The same scheme as in case (iv) is utilized. $D_s = [N_c + N_t - N_e]^+$.

2) For $\frac{N_r}{2} + N_e - N_t < N_c \leq N_r + N_e - N_t$ and N_r is even:

   $d = l = \frac{N_r}{2}$ Gaussian streams are transmitted. $\mathbf{P}_t, \mathbf{P}_c$ are chosen as in case (iv). $D_s = \frac{N_r}{2}$.

3) For $\frac{N_r}{2} + N_e - N_t < N_c \leq N_r + N_e - N_t$ and N_r is odd:

   $d = l = \frac{N_r+1}{2}$ structured streams are transmitted. $\mathbf{P}_t, \mathbf{P}_c$ are as in case (iv). The legitimate receiver uses the projection and cancellation technique. $D_s = \frac{N_r}{2}$.

4) For $N_r + N_e - N_t < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$ and $N_c + N_t - N_e$ is even:

   $d = l = \frac{N_c+N_t-N_e}{2}$ Gaussian streams are transmitted. Both $\mathbf{P}_t, \mathbf{P}_c$ are chosen to align the information and the cooperative jamming streams at the eavesdropper. $\mathbf{P}_c$ is also chosen to send $N_c - N_r$ cooperative jamming streams over $\mathcal{N}(\mathbf{H}_c)$ as in Section 3.5.9. $N_c > N_r + N_e - N_t$ achieves the above two conditions. $D_s = \frac{N_c+N_t-N_e}{2}$.

5) For $N_r + N_e - N_t < N_c \leq 2\min\{N_t, N_r\} + N_e - N_t$ and $N_c + N_t - N_e$ is odd:

   $d = l = \frac{N_c+N_t-N_e+1}{2}$ structured streams are transmitted. $\mathbf{P}_t, \mathbf{P}_c$ are chosen as in case (v-4). The receiver uses the projection and cancellation technique. $D_s = \frac{N_c+N_t-N_e}{2}$.

Using the achievable schemes described above for the different cases of the number of antennas, and their analysis as in Section 3.5, we have (3.7) as the achievable s.d.o.f., which completes the proof for Theorem 2.

3.7 Discussion

At this point, it is useful to discuss the results and the implications of this work. Theorem 1, c.f. (3.6), shows the behavior of the s.d.o.f., for an $(N \times N \times N_e)$ multi-antenna Gaussian wiretap channel with an N_c-antenna cooperative jammer, associated with increasing N_c form 0 to $N+N_e$. The s.d.o.f. first increases linearly by increasing N_c from 0 to $N_e - \lceil \frac{\min\{N,N_e\}}{2} \rceil$, that is to say adding one antenna at the cooperative jammer provided the system to have one additional degrees of freedom. The s.d.o.f. remains constant in the N_c range of $N_e - \lfloor \frac{\min\{N,N_e\}}{2} \rfloor$ to $\max\{N, N_e\}$, and starts to increase again for N_c from $\max\{N, N_e\}$ to $N + N_e$, until the s.d.o.f. arrives at its maximum value, N, at $N_c = N + N_e$. This behavior transpires both when the eavesdropper antennas are fewer or more than that of the legitimate receiver.

The reason for the flat s.d.o.f. range is as follows: At high SNR, achieving the secrecy constraint requires i) the entropy of the cooperative jamming signal, $\mathbf{X}_c^n$, to be greater than or equal to that of the information signal visible to the eavesdropper, and ii) $\mathbf{X}_c^n$ to completely cover the information signal, $\mathbf{X}_t^n$, at the eavesdropper. For $N_e \leq N$, part of $\mathbf{X}_t^n$ can be sent invisible to the eavesdropper, and the information signal visible to the eavesdropper can be covered by jamming for all N_c. For $0 \leq N_c \leq \frac{N_e}{2}$, the spatial resources at the receiver are sufficient, at high SNR, for decoding information and jamming signals which satisfy the above constraints. Thus, increasing the possible entropy of $\mathbf{X}_c^n$ by increasing N_c from 0 to $\left\lfloor \frac{N_e}{2} \right\rfloor$ allows for increasing the entropy of $\mathbf{X}_t^n$, and hence, the achievable secrecy rate and the s.d.o.f. increase. At $N_c = \left\lceil \frac{N_e}{2} \right\rceil$, the possible entropy of $\mathbf{X}_c^n$ and, correspondingly, the maximum possible entropy of $\mathbf{X}_t^n$,

result in information and jamming signals which completely occupy the receiver's signal space. Thus, increasing the possible uncertainty of $\mathbf{X}_c^n$ by increasing N_c from $\left\lceil \frac{N_e}{2} \right\rceil$ to N is useless, since, in this range, $\mathbf{X}_c^n$ is totally observed by the receiver which has its signal space already full at $N_c = \left\lceil \frac{N_e}{2} \right\rceil$.

Increasing N_c over N increases the possible entropy of $\mathbf{X}_c^n$ and simultaneously increases the part of $\mathbf{X}_c^n$ that can be transmitted invisible to the receiver, leaving more space for $\mathbf{X}_t^n$ at the receiver. This allows for increasing the secrecy rate, and hence, the s.d.o.f. starts to increase again. For $N_e > N$, the s.d.o.f. is equal to zero for all $0 \leq N_c \leq N_e - N$, where $\mathbf{X}_c^n$ cannot cover the information at the eavesdropper for this case. The s.d.o.f. starts to increase again, after the flat range, at $N_c > N_e$, since sending jamming signals invisible to the receiver while satisfying the covering condition at the eavesdropper requires that $N_c > N_e$.

The difference in the slope for the increase in the s.d.o.f. in the ranges before and after the flat range, for both $N_e \leq N$ and $N_e > N$, can be explained as follows. For $0 \leq N_c \leq N_e - \frac{\min\{N,N_e\}}{2}$, each additional antenna at the cooperative jammer allows for utilizing two more spatial dimensions at the receiver; one spatial dimension is used for the jamming signal and the other is used for the information signal. By contrast, for $\max\{N, N_e\} < N_c \leq N + N_e$, each additional antenna at the cooperative jammer sets one spatial dimension at the receiver free from jamming, and this spatial dimension is shared between the extra cooperative jamming and information streams.

It is important to note that the result that suggests that increasing the cooperative jammer antennas is not useful in the range $N_e - \frac{\min\{N,N_e\}}{2} < N_c \leq \max\{N, N_e\}$ applies only to the prelog of the secrecy capacity, i.e., is specific to the high SNR behavior. This

should not be taken to mean that additional antennas do not improve secrecy rate, but only the secrecy rate scaling with power in the high SNR.

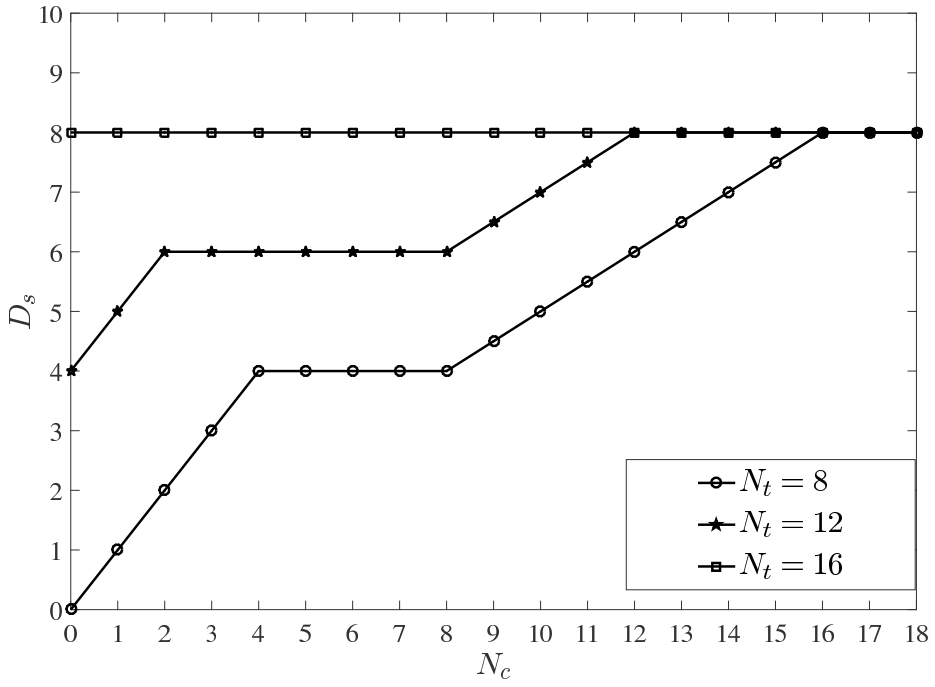


Fig. 3.5. D_s versus N_c when $N_r = N_e = 8$ and N_t increases from N_r to $N_r + N_e$.

Theorem 2 generalizes the results above to the case where the number of transmit antennas at the transmitter, N_t, is not equal to the number of receive antennas at the legitimate receiver, N_r. Although the maximum possible s.d.o.f. of the channel for this case is limited to $\min\{N_t, N_r\} = N_d$, increasing N_t over N_r, or increasing N_r over N_t, do change the behavior of the s.d.o.f. associated with increasing N_c until the maximum possible s.d.o.f. is reached. Let us start at $N_t = N_r = N_d$. For $N_t \geq N_e$, increasing N_t over $N_d = N_r$ increases the number of the information streams that can be sent invisible to the eavesdropper, and hence the s.d.o.f. without the help of the cooperative

jammer, i.e., $N_c = 0$, increases. This results in increasing the range of N_c for which the s.d.o.f. remains constant by increasing N_c, since the receiver's signal space gets full at a smaller N_c and remains full until N_c is larger than $N_d = N_r$. In addition, increasing N_t over N_d, when $N_t \geq N_e$, results in decreasing the value of N_c at which the maximum s.d.o.f. of the channel, N_d, is achievable, arriving at $N_t \geq N_r + N_e$, where the s.d.o.f. of N_d is achievable without the help of the cooperative jammer. Fig. 3.5 illustrates this behavior. When $N_e > N_t$, increasing N_t over N_d decreases the value of N_c at which the s.d.o.f. is positive, and decreases the value of N_c at which the s.d.o.f. of N_d is achievable, arriving at $N_t > N_e$, where the channel renders itself to the previous case. This behavior is demonstrated in Fig. 3.6. For both the cases $N_t \geq N_e$ and $N_t < N_e$, increasing N_r over $N_d = N_t$, results in increasing the available space at the receiver's signal space, and hence the constant period decreases, arriving at $N_r \geq N_t + N_e$ when $N_t \geq N_e$, or at $N_r \geq 2N_t$ when $N_e > N_t$, where the constant period vanishes. Fig. 3.7 illustrates the behavior of the s.d.o.f. curve associated with increasing N_r over N_t.

3.8 Conclusion

In this chapter, we have studied the multi-antenna wiretap channel with a N_c-antenna cooperative jammer, N_t-antenna transmitter, N_r-antenna receiver, and N_e-antenna eavesdropper. We have completely characterized the s.d.o.f. for this channel for all possible values of the number of antennas at the cooperative jammer, N_c. We have shown that when the s.d.o.f. of the channel is integer valued, it can be achieved by linear precoding at the transmitter and cooperative jammer, Gaussian signaling both for transmission and jamming, and linear processing at the legitimate receiver. By contrast,

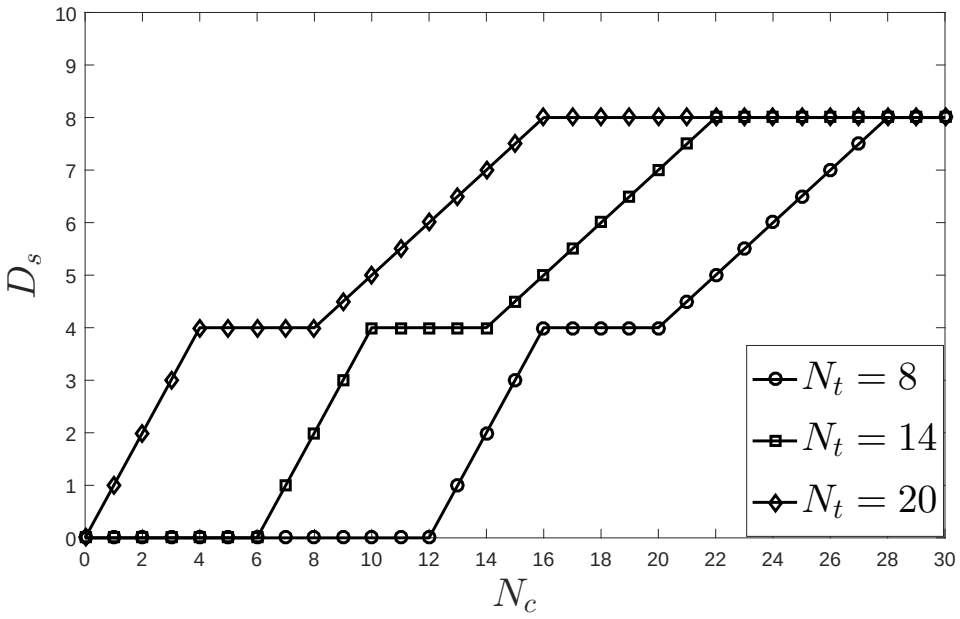


Fig. 3.6. D_s versus N_c when $N_r = 8$, $N_e = 20$ and N_t increases from N_r to N_e.

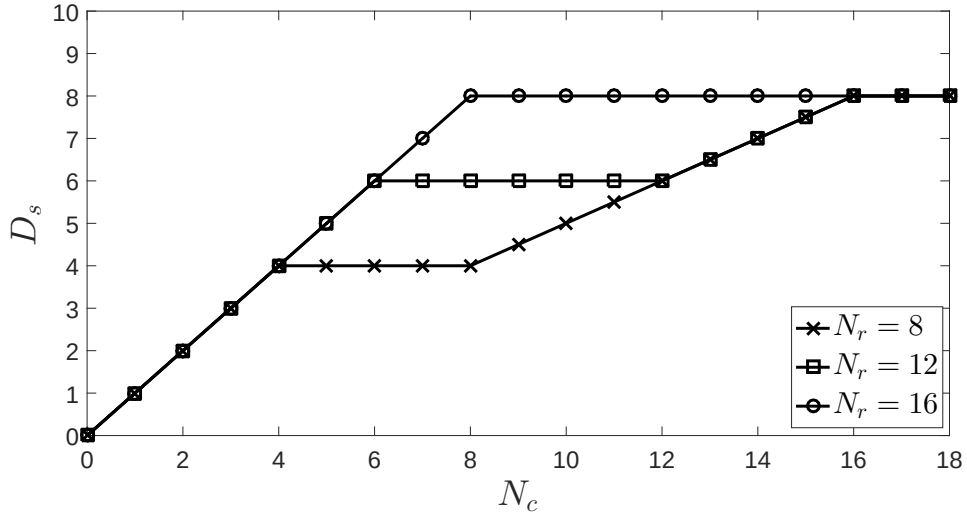


Fig. 3.7. D_s versus N_c when $N_t = N_e = 8$ and N_r increases from N_t to $N_t + N_e$.

when the s.d.o.f. is not an integer, we have shown that a scheme which employs structured signaling both at the transmitter and the cooperative jammer, along with joint signal space and signal scale alignment achieves the s.d.o.f. of the channel. We have seen that, when $N_t \geq N_e$, the transmitter uses its precoder to send a part of its information signal invisible to the eavesdropper, and to align the remaining part over jamming at the eavesdropper, while the cooperative jammer uses its precoder to send a part of its jamming signal invisible to the receiver, whenever possible. When $N_e > N_t$, more intricate precoding at the transmitter and cooperative jammer is required, where both the transmitter and cooperative jammer choose their precoders to achieve the alignment of information and jamming at the eavesdropper, and simultaneously, the cooperative jammer designs its precoder, whenever possible, to send a part of the jamming signal invisible to the receiver.

The converse was established by allowing for full cooperation between the transmitter and cooperative jammer for a certain range of N_c, and by incorporating both the secrecy and reliability constraints, for the other values of N_c. We note that while this work settles the degrees of freedom of this channel, its secrecy capacity is still open. Additionally, while the model considered here assumes channels to be known, universal secrecy as in [44] should be considered in the future.

Chapter 4

The Wiretap Channel II with a Noisy Main Channel

4.1 Introduction

Back to 1984, Ozarow and Wyner in reference [96] introduced the wiretap channel II model, which generalizes the special instance of a wiretap channel with a noiseless main channel and a binary erasure wiretapper channel, to a wiretapper which is able to select the positions of erasures. Authors in [96] derived an outer bound for the rate-equivocation region of the channel and proved its tightness by using random partitioning and combinatorial arguments, concluding that the secrecy capacity for the channel does not deteriorate despite this additional capability of the wiretapper.

Besides deriving the capacity-equivocation region for the wiretap channel II model, reference [96] proposed a randomized coset coding scheme, where a group code and its cosets were used as the sub-codebooks of the wiretap code, and showed that it achieves the capacity-equivocation region. This result has spurred a considerable amount of research on practical coding design for secure communication, see for example [1, 15, 66, 114, 118]. In [66], the wiretap channel with a noiseless main channel and binary-input symmetric-output memoryless wiretapper channel, is considered as type-II wiretap channel. Reference [1] studied a variation of the wiretap channel II model studied in [96], where the wiretapper not only noiselessly overhears a subset of the transmitted bits, but

also modifies (or corrupts) the bits, so that the legitimate receiver receives a corrupted version of the transmitted codeword.

In this chapter, we consider a wiretap channel with a finite input alphabet, a discrete memoryless main channel, and a wiretapper which noiselessly observes μ symbols of its choosing of the length-n transmitted codeword, where $\mu \leq n$ and $\alpha = \frac{\mu}{n}$. We first derive an outer bound for the rate-equivocation region of the channel as a function of α. Next, we propose an achievable scheme which extends the random partitioning argument in [96] constructed for the one codebook $\mathcal{C}_0$ that contains all possible codewords and has all of its components independently and identically distributed, to a random coding argument, which is exploited to guarantee reliable communication over the discrete memoryless main channel. In particular, we define a class of good codebooks for which we show, using random partitioning and combinatorial arguments, the existence of a partition which achieves the required level of equivocation. We then show that under the requirement of the claimed achievable rate, the probability of the class of the good codebooks goes to one as the block-length n increases. Note that the wiretapper's capability of choosing the positions of the symbols it observes results in a wiretapper channel with memory, and hence the results of the classical wiretap channel in [121] do not specialize to the performance of the model at hand.

The remainder of the chapter is organized as follows. Section 4.2 describes the channel model. Section 4.3 presents the main result in this section. Sections 4.4 and 4.5 provide outer and inner bounds for the rate-equivocation region of the channel. Section 4.6 concludes the chapter. Section 4.7 provides a discussion about the secrecy capacity of the model presented in this chapter.

4.2 Channel Model

We consider a wiretap channel II with a discrete memoryless main channel as illustrated in Fig. 4.1. The legitimate transmitter wishes to reliably transmit a message W to the legitimate receiver, and to keep it secret from the wiretapper. The message W is uniformly drawn from $\mathcal{W} = \{1, 2, \cdots, 2^{nR}\}$. The encoder at the transmitter $f_n : \mathcal{W} \mapsto \mathcal{X}^n$ maps the message $W \in \mathcal{W}$ to the transmitted codeword $\mathbf{X}^n \in \mathcal{X}^n$. The mapping f_n is allowed to be stochastic. The legitimate channel from the transmitter to the receiver is a discrete memoryless channel with a finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$, and probability distribution $p_{Y|X}(y|x)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The decoder at legitimate receiver, $g_n : \mathcal{Y}^n \mapsto \mathcal{W}$, which observes $\mathbf{Y}^n \in \mathcal{Y}^n$ and outputs an estimate $\hat{W}$ of the transmitted message, is parametrized by P_e, where

$$P_e = \mathbb{P}(\hat{W} \neq W) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \mathbb{P}(\hat{W} \neq w | W = w). \tag{4.1}$$

The wiretapper can noiselessly observe a subset, of its own choice, of the n transmitted symbols, $\mathbf{X}^n$. In particular, the wiretapper chooses $S \in \mathcal{S}$, with

$$\mathcal{S} = \left\{ S : \ S \subseteq \{1, 2, \cdots, n\}, \ |S| = \mu \leq n, \ \alpha = \frac{\mu}{n} \right\}, \tag{4.2}$$

and observes $\mathbf{Z}_S^n = [Z_{S,1} \ Z_{S,2} \ \cdots \ Z_{S,n}] \in \mathcal{Z}^n$, where

$$Z_{S,i} = \begin{cases} X_i, & \text{if } i \in S \\ ?, & \text{otherwise,} \end{cases} \tag{4.3}$$

and $\mathcal{Z} = \mathcal{X} \cup \{?\}$. Given the wiretapper's choice of the subset S, the equivocation at the wiretapper is measured by $H(W|\mathbf{Z}_S^n)$. In order to assure the required level of secrecy at the wiretapper, the encoding scheme has to be designed to maximize the equivocation

$$\Delta = \min_{S \in \mathcal{S}} H(W|\mathbf{Z}_S^n), \tag{4.4}$$

so that the equivocation at the wiretapper is at least Δ, no matter what subset S the wiretapper picks.

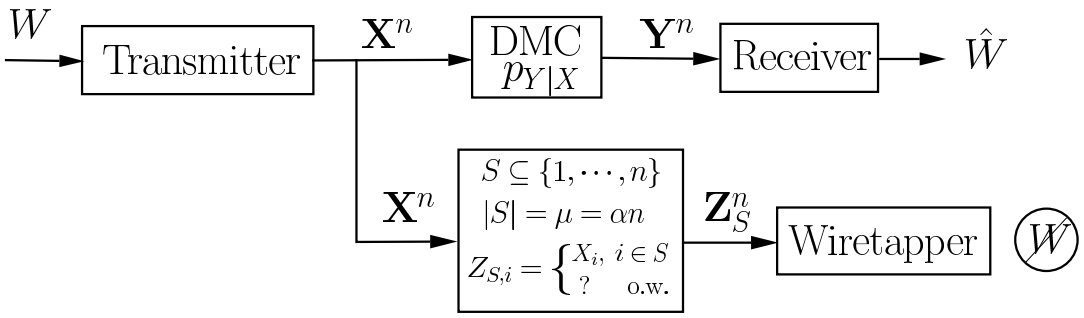


Fig. 4.1. Wiretap channel II with a discrete memoryless main channel.

We study the tradeoff between the rate of reliable transmission ($R = \frac{\log|\mathcal{W}|}{n}$ such that $\lim_{n\to\infty} P_e = 0$), the fraction of the transmitted symbols tapped by the wiretapper ($\alpha = \frac{\mu}{n}, 0 \le \alpha \le 1$), and the normalized equivocation at the wiretapper ($\delta = \frac{\Delta}{H(W)}, H(W) = nR$).

Definition 5. *The triple (R, α, δ) is said to be achievable if for every $\epsilon > 0$, there exists a sequence of encoder-decoder pairs, $\{f_n, g_n\}_{n\ge1}$, and $n_0 \ge 1$ such that $\frac{\mu}{n} \ge \alpha - \epsilon$, $\frac{\Delta}{H(W)} \ge \delta - \epsilon$, and $P_e \le \epsilon$, for all $n \ge n_0$.*

Definition 6. *For a fixed α, where $0 \leq \alpha \leq 1$, the secrecy rate R_s is achievable if the triple $(R_s, \alpha, 1)$ is achievable, i.e., the secrecy rate R_s, for some fixed α, is achievable if $P_e \to 0$, and $\frac{\Delta}{H(W)} \to 1$ as $n \to \infty$.*

4.3 Main Results

The following theorems provide outer and inner bounds for the set of achievable triples (R, α, δ), $\mathcal{R}$. Let C_m denote the capacity of the main channel $p_{Y|X}$, i.e., $C_m = \max_{p_X} I(X;Y)$. For the channel $p_{Y|X}$, let

$$C_u = I(X;Y) \text{ when } p_X(x) = \frac{1}{|\mathcal{X}|} \text{ for all } x \in \mathcal{X}. \tag{4.5}$$

Theorem 3. *The set $\mathcal{R} \subseteq \overline{\mathcal{R}}$, where*

$$\overline{\mathcal{R}} = \left((R, \alpha, \delta) : 0 \leq \alpha, \delta \leq 1, \;\; 0 \leq R \leq C_m, \; \delta \leq \begin{cases} 1, & \text{if } 0 \leq \alpha \leq 1 - \frac{R}{C_m} \\ (1-\alpha)\frac{C_m}{R}, & \text{if } 1 - \frac{R}{C_m} \leq \alpha \leq 1. \end{cases} \right). \tag{4.6}$$

Theorem 4. *The set $\mathcal{R} \supseteq \underline{\mathcal{R}}$, where*

$$\underline{\mathcal{R}} = \left((R, \alpha, \delta) : 0 \leq \alpha, \delta \leq 1, \;\; 0 \leq R \leq C_u, \; \delta \leq \begin{cases} 1, & \text{if } 0 \leq \alpha \leq \frac{C_u - R}{\log |\mathcal{X}|} \\ \left[\frac{C_u - \alpha \log |\mathcal{X}|}{R} \right]^+, & \text{if } \frac{C_u - R}{\log |\mathcal{X}|} < \alpha \leq 1. \end{cases} \right). \tag{4.7}$$

For a fixed rate R, the inner and outer bounds for the pair (α, δ) are shown in Fig. 4.2. As the achievable secrecy rate, R_s, for the model in question is of a particular interest, the following corollary, which directly follows from Theorems 3 and 4 by setting $\delta = 1$, gives lower and upper bounds for R_s.

Corollary 2. *For a fixed α, where $0 \leq \alpha \leq 1$, the achievable secrecy rate R_s satisfies*

$$[C_u - \alpha \log |\mathcal{X}|]^+ \leq R_s \leq (1-\alpha) C_m. \tag{4.8}$$

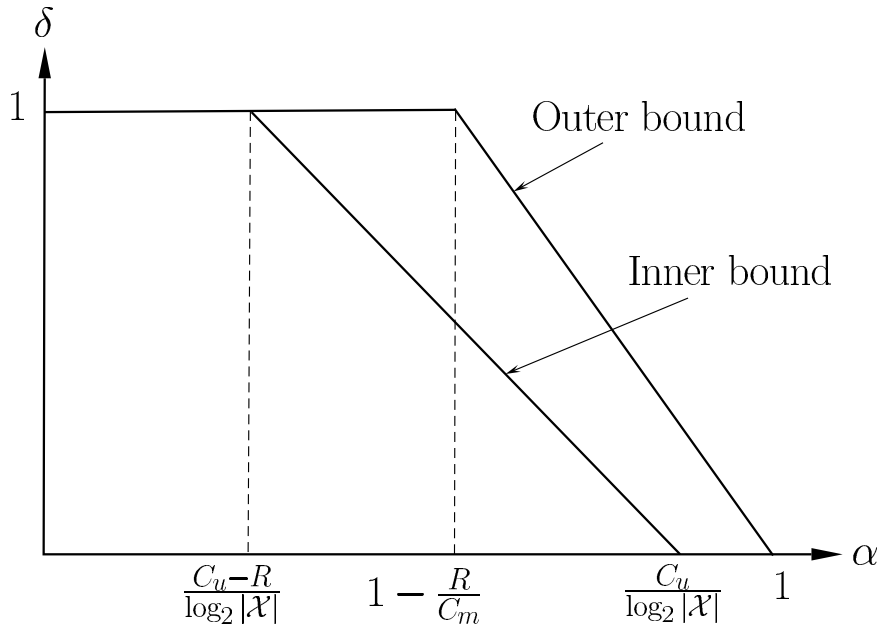


Fig. 4.2. Inner and outer bounds for (α, δ), for a fixed R.

Remark 5. *The lower bound for R_s in (4.8), computed for a binary main channel whose capacity is achieved by a uniform input distribution, is equal to the secrecy capacity of a*

wiretap channel with the same binary main channel and an erasure wiretapper channel with erasure probability $1-\alpha$.

4.4 Outer Bound

In order to prove Theorem 3, we show that any achievable triple, $(R,\alpha,\delta)\in\mathcal{R}$, satisfies $(R,\alpha,\delta)\in\overline{\mathcal{R}}$, where $\overline{\mathcal{R}}$ is given in (4.6). That $R\le C_m$ follows from the regular converse to the channel coding theorem [21]. That $\alpha\le 1$ follows since $\alpha\le\frac{\mu}{n}+\epsilon\le 1+\epsilon$ for every $\epsilon>0$. That $\delta\le 1$ follows since, for any $\epsilon>0$, $\delta\le\frac{\Delta}{H(W)}+\epsilon=\frac{\min_S H(W|\mathbf{Z}_S^n)}{H(W)}+\epsilon\le 1+\epsilon$. Thus, it remains to show the last inequality in (4.6).

Consider an arbitrary encoder-decoder pair, (f_n,g_n), and a fixed selection S at the wiretapper. Let $W,\mathbf{X}^n,\mathbf{Y}^n,\mathbf{Z}_S^n,\hat{W}$ correspond to the pair (f_n,g_n) and the selection S. Thus,

$$\Delta = H(W|\mathbf{Z}_S^n)\le H(W|\mathbf{Z}_S^n)-H(W|\mathbf{Y})+n\eta(P_e) \tag{4.9}$$

$$= I(W;\mathbf{Y})-I(W;\mathbf{Z}_S^n)+n\eta(P_e) \tag{4.10}$$

$$\le I(W;\mathbf{Y}\mathbf{Z}_S^n)-I(W;\mathbf{Z}_S^n)+n\eta(P_e) \tag{4.11}$$

$$= I(W;\mathbf{Y}|\mathbf{Z}_S^n)+n\eta(Pe), \tag{4.12}$$

where (4.9) follows from Fano's inequality, $\lim_{P_e\to 0}\eta(P_e)=0$. Let $S^c=\{1,2,\cdots,n\}\backslash S$, $\mathbf{X}_S=\{X_i\}_{i\in S}$, and $\mathbf{X}_{S^c}=\{X_i\}_{i\in S^c}$, and let $\mathbf{Y}_S$ and $\mathbf{Y}_{S^c}$ be defined similarly. Due to the Markov Chain $W-\mathbf{X}^n\mathbf{Z}_S^n-\mathbf{Y}^n$, (4.12) is bounded as

$$\Delta\le I(\mathbf{X};\mathbf{Y}|\mathbf{Z}_S^n)+n\eta(P_e) \tag{4.13}$$

$$= H(\mathbf{X}_S, \mathbf{X}_{S^c}|\mathbf{Z}_S) - H(\mathbf{X}_S, \mathbf{X}_{S^c}|\mathbf{Y}_S\mathbf{Y}_{S^c}\mathbf{Z}_S) + n\eta(P_e) \tag{4.14}$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S) - H(\mathbf{X}_{S^c}|\mathbf{Y}_S\mathbf{Y}_{S^c}\mathbf{X}_S) + n\eta(P_e) \tag{4.15}$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S) - H(\mathbf{X}_{S^c}|\mathbf{Y}_{S^c}\mathbf{X}_S) + n\eta(P_e) \tag{4.16}$$

$$= I(\mathbf{X}_{S^c}; \mathbf{Y}_{S^c}|\mathbf{X}_S) + n\eta(P_e) \tag{4.17}$$

$$\leq I(\mathbf{X}_{S^c}; \mathbf{Y}_{S^c}) + n\eta(P_e) \tag{4.18}$$

$$\leq \sum_{i \in S^c} I(X_i; Y_i) + n\eta(P_e) \tag{4.19}$$

$$\leq (n - \mu) \max_{p_X} I(X; Y) + n\eta(P_e) \tag{4.20}$$

$$= (n - \mu)C_m + n\eta(P_e), \tag{4.21}$$

where (4.16) follows from the Markov chain $\mathbf{X}_{S^c} - \mathbf{X}_S\mathbf{Y}_{S^c} - \mathbf{Y}_S$, (4.18) follows from the Markov chain $\mathbf{X}_S - \mathbf{X}_{S^c} - \mathbf{Y}_{S^c}$, and (4.19) follows from the memoryless channel assumption.

Thus, we have, for any selection of S, $\Delta \leq (n - \mu)C_m + n\eta(P_e)$. But, Δ is also upper bounded as $\Delta = \min_S H(W|\mathbf{Z}_S^n) \leq H(W) = nR$, and hence, we have, for every encoder-decoder pair,

$$\Delta \leq \min\left\{nR, (n - \mu)C_m + n\eta(P_e)\right\}. \tag{4.22}$$

Since $(R, \alpha, \delta) \in \mathcal{R}$, then for every $\epsilon > 0$, there exists an encoder-decoder pair with $\frac{\mu}{n} \geq \alpha - \epsilon$, $\frac{\Delta}{H(W)} \geq \delta - \epsilon$, and $P_e \leq \epsilon$. Thus, for every $\epsilon > 0$, by applying (4.22) to this

encoder-decoder pair, we obtain

$$\delta \leq \begin{cases} 1+\epsilon, & 0 \leq \alpha \leq 1 - \frac{R}{C_m} + O(\epsilon) \\ \frac{(1-\alpha)C_m}{R-\epsilon} + O(\epsilon), & 1 - \frac{R}{C_m} \leq \alpha + O(\epsilon) \leq 1. \end{cases} \tag{4.23}$$

Taking $\epsilon \to 0$, the last inequality in (4.6) is proved, which completes the proof for Theorem 3.

4.5 Inner Bound

The enabler for the achievability result in [96] is the assumption of a noiseless main channel over which a codebook $\mathcal{C}_0$, that contains all possible codewords, can be reliably communicated. The enabling property, satisfied by $\mathcal{C}_0$, is that for every possible observation at the wiretapper, $\mathbf{z}^n$ (which results from a transmitted codeword, $\mathbf{x}^n$, and a selection S), the number of codewords in $\mathcal{C}_0$ that can generate $\mathbf{z}^n$ is the same. Relying on this property, the existence of a good partition of $\mathcal{C}_0$ (of equal size subsets) that distributes the codewords among the subsets of the partition in a harmony that asymptotically (with the codeword length, n) achieves the secrecy constraint for every possible selection of S, is evident by [96]. When the main channel is a discrete memoryless channel, our achievability scheme relies on defining a class of good codebooks which possess a similar property to that of $\mathcal{C}_0$. We restrict the input distribution to be uniform, and α to be such that $\alpha < \frac{C_u}{\log|\mathcal{X}|}$, in order to show, using a random coding argument, that the probability of the class of good codes approaches 1 as $n \to \infty$. Thus,

the existence of a codebook, that achieves the reliability constraint, within the class of good codes follows. In the following, the achievable scheme is described in detail.

Codebook Generation: Let p_X be a uniform distribution over $\mathcal{X}$, i.e., $p_X(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$, and for the main channel $p_{Y|X}$, let $C_u = I(X;Y)$. Generate 2^{nC_u} length-n codewords randomly and independently, each with i.i.d. components according to p_X. Let $\mathcal{C}$ denote the random variable which represents the generated codebook. For the generated codebook $\mathcal{C} = C$, let $\{\mathcal{A}_w\}_{w=1}^{2^{nR}}$ be a partition of C into 2^{nR} disjoint subsets, $\mathcal{A}_w$s, each containing $2^{n(C_u - R)}$ codewords.

Encoder at the transmitter: In order to send the message W, the encoder randomly selects a codeword, $\mathbf{X}^n$, from $\mathcal{A}_W$, and transmits $\mathbf{X}^n$. For this encoder, let $\mu = \alpha n$.

Decoder at the legitimate receiver: Since the rate of the codebook C is equal to $(1/n)\log 2^{nC_u} = I(X;Y)$, it can be shown [21], using a typical set decoder at the legitimate receiver, that for every $\epsilon > 0$, there exists a sufficiently large n_1 such that $\mathbb{E}_{\mathcal{C}}(P_e) \leq \frac{\epsilon}{3}$ for all $n \geq n_1$.

Equivocation Analysis: For an arbitrary codebook C, an encoder defined as above, an arbitrary selection S, and $\mathbf{Z}_S^n$ which corresponds to S (as defined in (4.3)), we have

$$\Delta = H(W|\mathbf{Z}_S^n) = H(W, \mathbf{Z}_S^n) - H(\mathbf{Z}_S^n) \tag{4.24}$$

$$= H(W, \mathbf{X}, \mathbf{Z}_S^n) - H(\mathbf{X}|W, \mathbf{Z}_S^n) - H(\mathbf{Z}_S^n) \tag{4.25}$$

$$= H(\mathbf{X}|\mathbf{Z}_S^n) + H(W|\mathbf{X}, \mathbf{Z}_S^n) - H(\mathbf{X}|W, \mathbf{Z}_S^n). \tag{4.26}$$

By the construction of the encoding scheme, for every selection of the subset S, we have

$$H(W|\mathbf{X}, \mathbf{Z}_S^n) = 0. \tag{4.27}$$

Definition 7. *The codeword* $\mathbf{x}^n$ *is said to be consistent with the wiretapper's observation* $\mathbf{z}^n$ *if* $\mathbf{z}^n$ *can be obtained from* $\mathbf{x}^n$ *by switching* $(n-\mu)$ *components of* $\mathbf{x}^n$ *to* $'?'$.

For an arbitrary codebook C, $j = 1, 2, \cdots, 2^{nC_u}$, and $S \subseteq \{1, 2, \cdots, n\}$ with $|S| = \mu$, let $\mathbf{x}(j) = [x_1(j) \ \cdots \ x_n(j)]$ denote the jth codeword in C, and $\mathbf{z}(j, S)$ denote the length-n vector with $x_i(j)$ in the positions $i \in S$, and $'?'$ in the remaining positions, and define $\mathcal{Q}_C(j, S)$ and $m_C(j, S)$ as

$$\mathcal{Q}_C(j,S) = \Big\{ \mathbf{x}(i) \in C : \text{ for } \mathbf{x}(j) \in C, \ \mathbf{x}(i) \text{ is consistent with } \mathbf{z}(j,S), \ i = 1, \cdots, 2^{nC_u}, i \neq j \Big\}, \tag{4.28}$$

$$\text{and} \qquad m_C(j,S) = |\mathcal{Q}_C(j,S)|. \tag{4.29}$$

Now, let us define a set of good codebooks, $\mathcal{C}^*$, as

$$\mathcal{C}^* = \Big\{ C : \forall j = 1, 2, \cdots, 2^{nC_u}, \text{ and } \forall S \subseteq \{1, 2, \cdots, n\}, \text{ with } |S| = \mu, \ \ |m_C(j,S) - m| \leq t \Big\}, \tag{4.30}$$

where $m = 2^{n(C_u - \alpha \log |\mathcal{X}|)}$, $t = \beta \sqrt{n} 2^{\frac{n}{2}(C_u - \alpha \log |\mathcal{X}|)} + |\mathcal{X}|^{-\alpha n}$, and β is a constant which does not depend on n. We assume that 2^{nC_u}, and all such quantities, are integers. If not, a straight forward modification of the sequel is necessary.

Now, we consider a good codebook, $C \in \mathcal{C}^*$. Since the message W is uniformly distributed, and the encoder randomly selects a codeword from $\mathcal{A}_W$, we have $\mathbf{X}^n$ is uniformly distributed over C. Thus, given the wiretapper's observation, $\mathbf{Z}_S^n = \mathbf{z}^n$, $\mathbf{X}^n$ is uniformly distributed over the codewords in C consistent with $\mathbf{z}^n$. Using (4.30), we have, for all S, that

$$H(\mathbf{X}|\mathbf{Z}_S^n) \geq \log(m-t). \tag{4.31}$$

Next, we show the existence of a partition $\{\mathcal{A}_w\}_{w=1}^{2^{nR}}$ of the good codebook C, which satisfies that, for all S, $H(\mathbf{X}|W, \mathbf{Z}_S^n)$ is upper bounded by a constant which does not depend on n.

Definition 8. *A partition $\{\mathcal{A}_w\}_{w=1}^{2^{nR}}$ of the codebook C is said to be good, if there exists an integer $l_0 \geq 1$ such that for all $w = 1, \cdots, 2^{nR}$, $j = 1, \cdots, 2^{nC_u}$, and $S \subseteq \{1, 2, \cdots, n\}$ with $|S| = \mu$, we have*

$$\big|\, \{\mathbf{x} \in \mathcal{A}_w : \mathbf{x} \textit{ is consistent with } \mathbf{z}(j, S)\} \,\big| < l_0. \tag{4.32}$$

If a partition $\{\mathcal{A}_w\}$ of $C \in \mathcal{C}^*$ is good, we have, for all S,

$$H(\mathbf{X}|W, \mathbf{Z}_S^n) < \log l_0. \tag{4.33}$$

We now choose the partition $\{\mathcal{A}_w\}$ of $C \in \mathcal{C}^*$ uniformly at random from the set of all partitions of C into 2^{nR} equal size subsets. Define the functions $\psi(\{\mathcal{A}_w\})$ and

$\phi(\mathcal{A}_w, \mathbf{z}(j,S))$ as

$$\psi\left(\{\mathcal{A}_w\}\right) = \begin{cases} 0, \text{ if the partition } \{\mathcal{A}_w\} \text{ is good} \\ 1, \text{ otherwise.} \end{cases} \tag{4.34}$$

$$\phi(\mathcal{A}_w, \mathbf{z}) = \begin{cases} 0, \text{ if } \left|\{\mathbf{x} \in \mathcal{A}_w : \mathbf{x} \text{ is consistent with } \mathbf{z}\}\right| < l_0 \\ 1, \text{ otherwise.} \end{cases} \tag{4.35}$$

Note that, we have $\psi\left(\{\mathcal{A}_w\}\right) \leq \sum_{w=1}^{2^{nR}} \sum_{j,S} \phi(\mathcal{A}_w, \mathbf{z}(j,S))$, and hence, by linearity and monotonicity of expectation, we have

$$\mathbb{E}\left(\psi\left(\{\mathcal{A}_w\}\right)\right) \leq \sum_{w=1}^{2^{nR}} \sum_{j,S} \mathbb{E}\left(\phi(\mathcal{A}_w, \mathbf{z}(j,S))\right). \tag{4.36}$$

We now upper bound $\mathbb{E}\left(\phi(\mathcal{A}_w, \mathbf{z}(j,S))\right)$. For fixed w, j, and S, let L_r denote a random variable which represents the number of codewords $\mathbf{x} \in \mathcal{A}_w$, that are consistent with $\mathbf{z}(j,S)$. Let $n_C = |C| = 2^{nC_u}$, $m_C = |\mathcal{Q}_C(j,S)|$, and $n_r = |\mathcal{A}_w| = 2^{n(C_u - R)}$. Using a similar analysis as in [96], we have

$$\mathbb{P}(L_r = l) \leq \left(\frac{m_C n_r}{n_C}\right)^l \frac{2^l}{l!}. \tag{4.37}$$

Since we consider a good codebook $C \in \mathcal{C}^*$, we have $m_C \leq m + t = 2^{n(C_u - \alpha \log|\mathcal{X}|)}(1 + \frac{t}{m})$. For $\alpha < \frac{C_u}{\log|\mathcal{X}|}$, we have $\lim_{n\to\infty} \frac{t}{m} = 0$, i.e., $\frac{t}{m} \in o(n)$. Thus, using (4.37), we have

$$\mathbb{P}(L_r = l) \leq 2^{n(C_u - \alpha \log|\mathcal{X}| - R)l} \frac{\left(2(1 + \frac{t}{m})\right)^l}{l!}. \tag{4.38}$$

Thus, whenever $R > C_u - \alpha \log |\mathcal{X}|$, we have

$$\mathbb{E}\left(\phi(\mathcal{A}_w, \mathbf{z}(j,S))\right) = \mathbb{P}\left(\phi(\mathcal{A}_w, \mathbf{z}(j,S)) = 1\right) \tag{4.39}$$

$$\leq \sum_{l=l_0}^{n_r} 2^{n(C_u - \alpha \log |\mathcal{X}| - R)l} \frac{\left(2(1+\frac{t}{m})\right)^l}{l!} \tag{4.40}$$

$$\leq 2^{n(C_u - \alpha \log |\mathcal{X}| - R)l_0 + 2(1+\frac{t}{m}) \log e}, \tag{4.41}$$

and hence, using (4.36), we have

$$\mathbb{E}\left(\psi(\{\mathcal{A}_w\})\right) \leq 2^{nR} 2^{nC_u} \binom{n}{\alpha n} \times 2^{n(C_u - \alpha \log |\mathcal{X}| - R)l_0 + 2(1+\frac{t}{m}) \log e} \tag{4.42}$$

$$\leq 2^{n(R + C_u + H(\alpha) + o(n)) + 2(1+\frac{t}{m}) \log e + n(C_u - \alpha \log |\mathcal{X}| - R)l_0}, \tag{4.43}$$

where (4.43) follows from Stirling's approximation. Thus, for

$$l_0 > \frac{R + C_u + H(\alpha) + o(n) + \frac{2(1+\frac{t}{m}) \log e}{n}}{R - C_u + \alpha \log |\mathcal{X}|}, \tag{4.44}$$

we have $\mathbb{E}\left(\psi(\{\mathcal{A}_w\})\right) < 1$, and hence there must exist a good partition $\{\mathcal{A}_w\}$ of the codebook $C \in \mathcal{C}^*$. Since $\frac{t}{m} \in o(n)$, we have, for sufficiently large n_2, $o(n) \leq \epsilon_1$ and $\frac{t}{m} \leq \epsilon_2$, for all $n \geq n_2$. Thus, whenever $\frac{C_u - R}{\log |\mathcal{X}|} < \alpha < \frac{C_u}{\log |\mathcal{X}|}$, by setting

$$l_0 = \frac{R + C_u + H(\alpha) + \epsilon_1 + \frac{2(1+\epsilon_2) \log e}{n_2}}{R - C_u + \alpha \log |\mathcal{X}|} + 1 = B, \tag{4.45}$$

the existence of a good partition $\{\mathcal{A}_w\}$ is guaranteed. This is the partition chosen by the encoder at the transmitter.

Using (4.26), (4.27), (4.31), (4.33), and (4.45), we have, for $C \in \mathcal{C}^*$, $\frac{C_u - R}{\log|\mathcal{X}|} < \alpha < \frac{C_u}{\log|\mathcal{X}|}$, and sufficiently large n, that

$$\frac{\Delta}{H(W)} = \min_{S \in \mathcal{S}} \frac{H(W|\mathbf{Z}_S^n)}{H(W)} \tag{4.46}$$

$$\geq \frac{1}{nR}(\log(m-t) - \log B) \tag{4.47}$$

$$= \frac{\log m}{nR} - o(n) \tag{4.48}$$

$$\geq \frac{C_u - \alpha \log|\mathcal{X}|}{R} - \epsilon_3, \tag{4.49}$$

where, for sufficiently large n_3, $o(n) \leq \epsilon_3$ for all $n \geq n_3$.

We now show that with high probability, $C \in \mathcal{C}^*$, i.e., $\lim_{n\to\infty} \mathbb{P}\left(C \in \mathcal{C}^*\right) = 1$. Let $\mathbf{X}(j)$, $j = 1, \cdots, 2^{nC_u}$, represent the jth codeword of the random code $\mathcal{C}$, and let $\mathbf{Z}_S^n(\mathbf{X}(j))$ represent the codeword $\mathbf{X}(j)$ with $'?'$ for $i \notin S$. For $j = 1, \cdots, 2^{nC_u}$, and all S, define the event $\mathcal{E}(j, S)$ as

$$\mathcal{E}(j,S)$$

$$= \Big\{ \Big| \text{card}\big\{ i : \mathbf{X}(i) \text{ is consistent with } \mathbf{Z}_S^n(\mathbf{X}(j)), \text{ where } i = 1, 2, \cdots, 2^{nC_u}, i \neq j \big\} - m \Big| \leq t \Big\}, \tag{4.50}$$

and let $\mathcal{E}^c(j, S)$ denote the complement of (4.50). Using the definition of $\mathcal{C}^*$ in (4.30), we have

$$\mathbb{P}\left(\mathcal{C}^*\right) = \mathbb{P}\Big(\mathcal{E}(j,S), \forall j = 1, 2, \cdots, 2^{nC_u} \text{ and } \forall S \subseteq \{1, 2, \cdots, n\} \text{ with } |S| = \mu\Big) \tag{4.51}$$

$$= 1 - \mathbb{P}\left(\mathcal{E}^c(j,S) \text{ for some } j, \text{ or some } S\right) \tag{4.52}$$

$$\geq 1 - \sum_{j=1}^{2^{nC_u}} \sum_{S} \mathbb{P}\left(\mathcal{E}^c(j,S)\right) \tag{4.53}$$

$$= 1 - 2^{nC_u} \binom{n}{\alpha n} \mathbb{P}\left(\mathcal{E}^c(1,S^*)\right), \tag{4.54}$$

where S^* is some set such that $S^* \subseteq \{1,2,\cdots,n\}$ with $|S^*| = \mu$, (4.53) follows from the union bound, and (4.54) follows because of the symmetry in the codebook, $\mathcal{C}$, construction; $\mathbb{P}\{\mathcal{E}^c(j,S)\}$ is the same for all j and all S.

Let $\mathcal{E}_{\mathbf{x}(1)}(1,S^*)$, and $\mathbf{z}_{S^*}(\mathbf{x}(1))$ denote the event $\mathcal{E}(1,S^*)$, and the random vector $\mathbf{Z}^n_{S^*}(\mathbf{X}(1))$ when $\mathbf{X}(1) = \mathbf{x}(1)$, respectively. Using (4.50), we have,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}_{\mathbf{x}(1)}(1,S^*)\right) \\
&= \mathbb{P}\Big(\Big| \text{card}\left\{ i : \mathbf{X}(i) \text{ is consistent with } \mathbf{z}_{S^*}(\mathbf{x}(1)), i = 2,3,\cdots,2^{nC_u} \right\} - m \Big| \leq t \Big).
\end{aligned} \tag{4.55}$$

For $i = 2,3,\cdots,2^{nC_u}$, define the random variable V_i as

$$V_i = \begin{cases} 1, \text{ if } \mathbf{X}(i) \text{ is consistent with } \mathbf{z}_{S^*}(\mathbf{x}(1)) \\ 0, \text{ otherwise.} \end{cases} \tag{4.56}$$

Note that the random variables $\{V_i\}_{i=2}^{2^{nC_u}}$ are i.i.d. In addition, we have, for each value of $\mathbf{x}(1)$, and for $i = 2,\cdots,2^{nC_u}$,

$$\mathbb{E}\left(V_i\right) = \mathbb{P}(V_i = 1) = |\mathcal{X}|^{-\alpha n}. \tag{4.57}$$

Let $V = \sum_{i=2}^{2^{nC_u}} V_i$. We can rewrite $\mathbb{P}\{\mathcal{E}_{\mathbf{x}(1)}(1, S^*)\}$ as

$$\mathbb{P}\left(\mathcal{E}_{\mathbf{x}(1)}(1, S^*)\right) = \mathbb{P}\left(|V - m| \leq t\right). \tag{4.58}$$

Since $\mathbb{E}(V) = (2^{nC_u} - 1)|\mathcal{X}|^{-\alpha n} = m - |\mathcal{X}|^{-\alpha n}$, by setting $t' = t - |\mathcal{X}|^{-\alpha n}$, we have

$$\mathbb{P}\left(\mathcal{E}_{\mathbf{x}(1)}(1, S^*)\right) \geq \mathbb{P}\left(|V - \mathbb{E}(V)| \leq t'\right). \tag{4.59}$$

By applying Hoeffding bound in Lemma 3 to (4.59), we obtain

$$\mathbb{P}\left(\mathcal{E}_{\mathbf{x}(1)}(1, S^*)\right) \geq 1 - 2\exp\left(\frac{-t'^2}{(2+\tau)\mathbb{E}(V)}\right), \tag{4.60}$$

where $\tau = \frac{t'}{\mathbb{E}(V)}$. Using (4.60), we have

$$\mathbb{P}\left(\mathcal{E}(1, S^*)\right) = \sum_{\mathbf{x}(1)} \mathbb{P}(\mathbf{X}(1) = \mathbf{x}(1))\mathbb{P}\left(\mathcal{E}_{\mathbf{x}(1)}(1, S^*)\right) \tag{4.61}$$

$$\geq 1 - 2\exp\left(\frac{-t'^2}{(2+\tau)\mathbb{E}(V)}\right). \tag{4.62}$$

By substituting $\mathbb{P}\left(\mathcal{E}^c(1, S^*)\right) \leq 2\exp\left(\frac{-t'^2}{(2+\tau)\mathbb{E}(V)}\right)$ in (4.54), and choosing t' as $t' = \beta\sqrt{n}2^{\frac{n}{2}(C_u - \alpha\log|\mathcal{X}|)}$,

$$\mathbb{P}(\mathcal{C}^*) \geq 1 - 2^{nC_u}\binom{n}{\alpha n}2\exp\left(\frac{-t'^2}{(2+\tau)\mathbb{E}(V)}\right) \tag{4.63}$$

$$\geq 1 - 2^{n(C_u + H(\alpha) + o(n) - \frac{\beta^2}{(2+\tau)}\log e)}. \tag{4.64}$$

For sufficiently large n_4, we have $o(n) \leq \epsilon_4$ and $\tau \leq \epsilon_5$ for all $n \geq n_4$, where $\tau = \frac{t'}{\mathbb{E}(V)} \in o(n)$. By setting $\beta \geq \sqrt{\frac{(\gamma + C_u + H(\alpha) + \epsilon_4)(2+\epsilon_5)}{\log e}}$, where $\gamma > 0$, we have, for $n \geq n_4$,

$$\mathbb{P}(\mathcal{C}^*) \geq 1 - 2^{-\gamma n}. \tag{4.65}$$

Using (4.49) and (4.65), we have, for $\frac{C_u - R}{\log |\mathcal{X}|} < \alpha < \frac{C_u}{\log |\mathcal{X}|}$, and sufficiently large n,

$$\mathbb{E}_{\mathcal{C}} \left(\frac{\Delta}{H(W)} \right) \geq \left(\frac{C_u - \alpha \log |\mathcal{X}|}{R} - \epsilon_3 \right) (1 - 2^{-\gamma n}) \tag{4.66}$$

$$\geq \frac{C_u - \alpha \log |\mathcal{X}|}{R} - \frac{\epsilon}{3}, \tag{4.67}$$

where, for sufficiently large n_5, $3(\epsilon_3 + \frac{2^{-\gamma n}(C_u - \alpha \log |\mathcal{X}|)}{R}) \leq \epsilon$ for all $n \geq n_5$. Let $n_0 = \max_{i=1:5} n_i$. For $n \geq n_0$ and $\frac{C_u - R}{\log |\mathcal{X}|} < \alpha < \frac{C_u}{\log |\mathcal{X}|}$, using (4.67) and that $\mathbb{E}_{\mathcal{C}}(P_e) \leq \frac{\epsilon}{3}$, we have

$$\mathbb{P} \left(P_e \leq \epsilon, \frac{\Delta}{H(W)} \geq \frac{C_u - \alpha \log |\mathcal{X}|}{R} - \epsilon \right)$$

$$\geq 1 - \mathbb{P}(P_e \geq \epsilon) - \mathbb{P} \left(\frac{\Delta}{H(W)} \leq \frac{C_u - \alpha \log |\mathcal{X}|}{R} - \epsilon \right) \tag{4.68}$$

$$\geq \frac{1}{3}, \tag{4.69}$$

where (4.68) follows from the union bound and (4.69) follows from Markov inequality and (4.67). Thus, for a fixed R, $0 \leq R \leq C_u$, a fixed α, $\frac{C_u - R}{\log |\mathcal{X}|} < \alpha \leq 1$, and for every $\epsilon > 0$, there exists a sufficiently large n_0, and an encoder-decoder pair with $\mu = \alpha n$, $\frac{\Delta}{H(W)} \geq \frac{[C_u - \alpha \log |\mathcal{X}|]^+}{R} - \epsilon$, and $P_e \leq \epsilon$, for all $n \geq n_0$. This completes the proof for Theorem 4.

4.6 Conclusion

In this chapter, we have derived outer and inner bounds for the rate-equivocation region of a wiretap channel with a discrete memoryless main channel, and a wiretapper which has a noiseless access to μ symbols of its own choice of the n transmitted symbols. The characterization of the derived inner and outer bounds has provided a trade-off between the rate of reliable transmission, R, the level of equivocation at the wiretapper, δ, and the ratio of the tapped symbols by the wiretapper, $\alpha = \frac{\mu}{n}$. The achievability has been established by random coding and random partitioning arguments, where, for a uniform input distribution and a certain range of α, the existence of a good codebook which achieves the reliability constraint, and for which there exists a good partition that achieves the required level of equivocation no matter what subset of μ symbols the wiretapper chooses, is guaranteed. The inner and outer bounds proposed in this chapter provide insights for understanding the fundamental limits of the model, and count as a step towards characterizing its capacity-equivocation region, as well as towards understanding the impact of more powerful adversary than passive observers.

4.7 Discussion

The secrecy capacity of the wiretap II with a noisy main channel model presnted in this chapter and introduced in [85], was later identified in [33]. For $0 \leq \alpha \leq 1$, the secrecy capacity $C_s(\alpha)$ is given by

$$C_s(\alpha) = \max_{p_{UX}} [I(U;Y) - \alpha I(U;X)]^+, \tag{4.70}$$

where the maximization is taken over all the distributions p_{UX} which satisfy the Markov chain $U - X - Y$, and the cardinality of the auxiliary random variable U can be restricted to $|\mathcal{U}| \leq |\mathcal{X}|$.

The achievability of (4.70) is established by utilizing a regular wiretap code, where a stronger version of Wyner's soft covering lemma [120, Theorem 3], which provides a doubly exponential decay rate for the probability of not achieving the secrecy constraint for a fixed choice of the subset S, is used, along with the union bound, in order to guarantee secrecy for the exponentially many possibilities of the subset S. Note that, by setting $U = X$ and p_X uniform over $\mathcal{X}$, the secrecy capacity of the channel in (4.70) reduces to the achievable rate in (4.8) which establishes the optimality of our achievable scheme for this special instance. In the next section, we shall see a generalization of the wiretap channel II with a noisy main channel to a model with the wiretapper receiving noisy observations instead of the erasures.

Chapter 5

A Generalized Wiretap Channel Model and its Strong Secrecy Capacity

5.1 Introduction

In the previous chapter, we introduced a discrete memoryless (noisy) main channel to the wiretap channel II model, and derived outer and inner bounds for the capacity-equivocation region of the model, where the proposed achievability scheme is optimal for the special case of the maximizing input distribution being uniform. The secrecy capacity for this model is identified in [33], and shown to be equal to that of the case when the wiretapper channel is replaced with a discrete memoryless erasure channel.

In this chapter, we go one step further and introduce a *generalized wiretap channel model* with a discrete memoryless main channel and a wiretapper which observes a subset of the transmitted codeword symbols of its choosing perfectly, as well as observing the remaining symbols through a second discrete memroyless channel. This model includes as special cases both the classical wiretap channel in [23] by setting the subset size to zero, and the wiretap channel II with a noisy main channel in Chapter 4 by setting the wiretapper's discrete memroyless channel to an erasure channel with erasure probability one, and is termed as the *generalized wiretap channel* for that reason. We characterize the *strong* secrecy capacity for the proposed wiretap channel model, quantifying precisely the

cost in secrecy capacity due to the additional capability at the wiretapper, with respect to the previous wiretap models.

We first present the achievability. The achievability is established by using a framework similar to the output statistics of random binning framework in [126]. In particular, we solve a dual secret key agreement problem in the source model sense [2,76], and infer the design for the encoder and decoder of the original channel model from the solution of the dual problem. The difference between our achievability proof and the framework presented in [126] is that we measure the statistical dependence between the transmitted message and the wiretapper's observation in terms of the Kullback-Leibler (K-L) divergence instead of total variation distance, which requires establishing a convergence result, with a rate strictly faster than $\frac{1}{n}$, for the probability that the two induced distributions from the original and the dual models are close in the total variation distance sense. In addition, in the source model, we guarantee the secrecy of the confidential key for the exponentially many possibilities of the subset chosen at the wiretapper by deriving a *one-shot* result which provides a *doubly-exponential convergence rate* for the probability that the key is uniform and independent from the wiretapper's observation.

The converse is derived by identifying a channel model whose secrecy capacity is identical to that of the proposed channel model, and is easier to establish the converse of. This is done by means of upper bounding its secrecy capacity with that of a discrete memoryless channel[1] whose secrecy capacity is tractable.

[1] A similar approach was considered to derive [33, Proposition 1].

The remainder of the chapter is organized as follows. Section 5.2 describes the generalized wiretap channel model. Section 5.3 provides the main result of this chapter, i.e., the strong secrecy capacity for the generalized wiretap channel. Sections 5.4 and 5.5 provide the achievability and converse proofs. Section 5.6 provides a discussion about the main result and the adopted achievability approach. Section 5.7 concludes the chapter.

5.2 Channel Model

We consider the channel model illustrated in Figure 5.1. The main channel $\left\{\mathcal{X}, \mathcal{Y}, p_{Y|X}\right\}$ is a discrete memoryless channel which consists of a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a transition probability $p_{Y|X}$. The transmitter wishes to transmit a message W, uniformly distributed over $\mathcal{W} = [1 : 2^{nR_s}]$, to the legitimate receiver reliably, and to keep the message secret from the wiretapper. To do so, the transmitter maps the message W to the transmitted codeword $\mathbf{X}^n \in \mathcal{X}^n$ using a stochastic encoder. The legitimate receiver observes $\mathbf{Y}^n \in \mathcal{Y}^n$ and maps its observation to the estimate $\hat{W}$ of the message W. The wiretapper chooses a subset $S \in \mathcal{S}$ where the set $\mathcal{S}$ is defined as

$$\mathcal{S} \triangleq \left\{S : S \subseteq [1:n], \;\; |S| = \mu \leq n, \; \alpha = \frac{\mu}{n}\right\}. \tag{5.1}$$

Then, the wiretapper observes the sequence $\mathbf{Z}_S^n \triangleq [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$, with

$$Z_{S,i} = \begin{cases} X_i, & i \in S \\ V_i, & \text{otherwise,} \end{cases} \tag{5.2}$$

where $\mathbf{V}^n \triangleq [V_1, V_2, \cdots, V_n] \in \mathcal{V}^n$ is the output of the discrete memroyless channel $p_{V|X}$ when $\mathbf{X}^n$ is the input, and the alphabet $\mathcal{Z}$ is given by $\mathcal{Z} = \{\mathcal{X} \cup \mathcal{V}\}$.

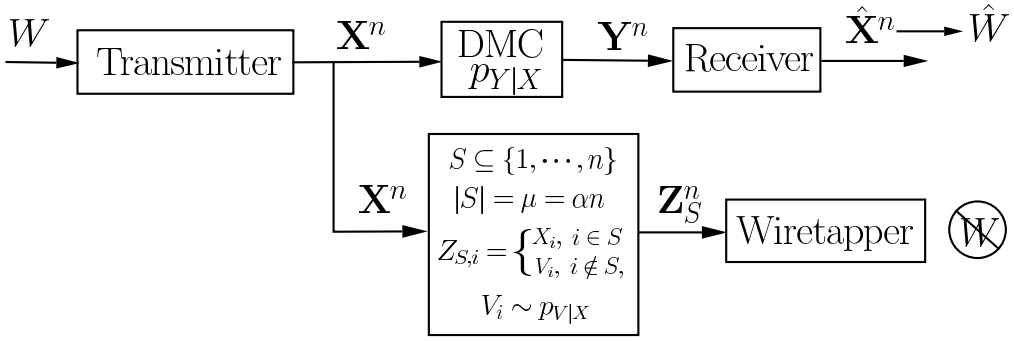


Fig. 5.1. The generalized wiretap channel model.

An $(n, 2^{nR_s})$ code $\mathcal{C}_n$ for the channel model in Figure 5.1 consists of

(i) the message set $\mathcal{W} = [1 : 2^{nR_s}]$,

(ii) the stochastic encoder $P_{\mathbf{X}^n|W}^{(\mathcal{C}_n)}$ at the transmitter, and

(iii) the decoder at the legitimate receiver.

We consider the strong secrecy constraint at the wiretapper [22,75]. Rate R_s is an achievable strong secrecy rate if there exists a sequence of $(n, 2^{nR_s})$ channel codes, $\{\mathcal{C}_n\}_{n \geq 1}$, such that

$$\lim_{n \to \infty} \mathbb{P}^{(\mathcal{C}_n)} \left(\hat{W} \neq W \right) = 0 \qquad \textbf{Reliability}, \tag{5.3}$$

$$\text{and} \quad \lim_{n \to \infty} \max_{S \in \mathcal{S}} I^{(\mathcal{C}_n)} \left(W; \mathbf{Z}_S^n \right) = 0 \qquad \textbf{Strong Secrecy}, \tag{5.4}$$

where $\mathcal{S}$ is defined as in (5.1). $\mathbb{P}^{(\mathcal{C}_n)}\left(\hat{W} \neq W\right)$ and $I^{(\mathcal{C}_n)}\left(W; \mathbf{Z}_S^n\right)$ are the error probability and the mutual information between the message W and the wiretapper's observation $\mathbf{Z}_S^n$, with respect to the joint distribution that corresponds to the code $\mathcal{C}_n$.

The strong secrecy capacity, C_s, is the supremum of all achievable strong secrecy rates.

5.3 Main Result

The main result of this chapter is stated in the following theorem.

Theorem 5. *For $0 \leq \alpha \leq 1$, the strong secrecy capacity of the generalized wiretap channel in Figure 5.1 is given by*

$$C_s(\alpha) = \max_{U-X-YV} \left[I(U;Y) - I(U;V) - \alpha I(U;X|V)\right]^+, \tag{5.5}$$

where the maximization is over all the distributions p_{UX} which satisfy the Markov chain $U - X - YV$, and the cardinality of U can be restricted as $|\mathcal{U}| \leq |\mathcal{X}|$.

Proof: The achievability and converse proofs for Theorem 5 are provided in Sections 5.4 and 5.5, respectively. ■

Remark 6. *An equivalent characterization for the strong secrecy capacity of the generalized wiretap channel is given by*

$$C_s(\alpha) = \max_{U-X-YV} \left[I(U;Y) - \alpha I(U;X) - (1-\alpha) I(U;V)\right]^+, \tag{5.6}$$

since $I(U;X|V)$ in (5.5) can be written as

$$I(U;X|V) = H(U|V) - H(U|X) \tag{5.7}$$

$$= H(U) - I(U;V) - H(U|X) \tag{5.8}$$

$$= I(U;X) - I(U;V), \tag{5.9}$$

where (5.7) follows from the Markov chain $U - X - V$.

Corollary 3. *By setting the tapped subset by the wiretapper, S, to the null set, or equivalently $\alpha = 0$, the secrecy capacity in (5.5) is equal to the secrecy capacity of the discrete memoryless wiretap channel in [23, Corollary 2], i.e.,*

$$C_s(0) = \max_{U-X-YV} \left[I(U;Y) - I(U;V)\right]^+. \tag{5.10}$$

Remark 7. *Comparing (5.5) and (5.10), we observe that the secrecy cost, with respect to the classical wiretap channel, of the additional capability of the wiretapper to choose a subset of size αn of the codewords to access perfectly, is equal to $\alpha I(U;X|V)$.*

Corollary 4. *By setting the wiretapper's discrete memroyless channel through which it observes the $(1-\alpha)n$ symbols it does not choose, $p_{V|X}$, to be an erasure channel with erasure probability one, the secrecy capacity in (5.5) is equal to the secrecy capacity of the wiretap channel II with a noisy main channel in [33, Theorem 2], i.e.,*

$$C_s(\alpha) = \max_{U-X-Y} \left[I(U;Y) - \alpha I(U;X)\right]^+. \tag{5.11}$$

Remark 8. *Comparing (5.6) and (5.11), the secrecy cost, with respect to the wiretap channel II with a noisy main channel, of the additional capability of the wiretapper of observing* $(1-\alpha)$ *fraction of the codeword through the discrete memroyless channel* $p_{V|X}$*, is equal to* $(1-\alpha)I(U;V)$*.*

5.4 Achievability

We establish the achievability for Theorem 5 using an indirect approach as in [24, 99, 126]. We first assume the availability of a certain common randomness at all terminals of the original channel model. We then define a *dual* secret key agreement problem in the source model which introduces a set of random variables similar to the set of variables introduced by the original problem with the assumed common randomness. The alphabets of the random variables in the original and dual problems are identical. In addition, a subset of the marginal and conditional distributions for these random variables in the original and dual problems are considered to be identical. Yet, the joint distribution of the random variables in the dual problem can differ from that of the original problem due to the different dynamics in the two problems. The main trick is to search for conditions such that the joint distributions of the random variables in the two problems are almost identical in the total variation distance sense. This enables converting the solution, i.e., finding an encoder and decoder which satisfy certain reliability and secrecy conditions, for the dual problem, which is more tractable, to a solution of the original problem. We finally eliminate the assumed common randomness from the original channel model by conditioning on a certain instance of it. Duality here

is an *operational* duality [35] in which the solution for the dual problem is converted to a solution for the original problem.

We first prove the achievability for the case $U = X$. We fix the input distribution p_X and define two protocols; each of these protocols introduces a set of random variables and random vectors and induces a joint distribution over them. The first protocol, protocol A, describes a dual secret key agreement problem in which a source encoder and decoder observe random sequences i.i.d. according to the input and output distributions of the original channel model. The source encoder and decoder intend to communicate a confidential key via transmitting a public message over an error-free channel, in the presence of a *compound* wiretapping source which has perfect access to the public message and observes another random sequence whose distribution belongs to a finite class of distributions, with no prior distribution over the class. The second protocol, protocol B, describes the original channel model in Figure 5.1, with the addition of assuming a common randomness that is available at all terminals. In the following, we describe the two protocols in detail.

Protocol A (Secret key agreement in source model): The protocol is illustrated in Figure 5.2. The random vectors $\mathbf{X}^n, \mathbf{Y}^n$ are i.i.d. according to $p_{XY} = p_X p_{Y|X}$, where $p_{Y|X}$ is the transition probability of the main channel in Figure 5.1. The source encoder observes the sequence $\mathbf{X}^n$ and randomly assigns (bins) it into the two bin indices $W = \mathcal{B}_{1,n}(\mathbf{X}^n)$ and $F = \mathcal{B}_{2,n}(\mathbf{X}^n)$, where $\mathcal{B}_{1,n}$ and $\mathcal{B}_{2,n}$ are uniformly distributed over $[1 : 2^{nR_s}]$ and $[1 : 2^{n\tilde{R}_s}]$, respectively. That is, each $\mathbf{x}^n \in \mathcal{X}^n$ is randomly and independently assigned to the indices $w \in [1 : 2^{nR_s}]$ and $f \in [1 : 2^{n\tilde{R}_s}]$. The bin index F represents the public message which is transmitted over a noiseless channel to the decoder and

perfectly accessed by the wiretapper. The bin index W represents the confidential key to be generated at the encoder and reconstructed at the decoder. The source decoder observes F and the i.i.d. sequence $\mathbf{Y}^n$, and outputs the estimate $\hat{\mathbf{X}}^n$ of $\mathbf{X}^n$, which in turn generates the estimate $\hat{W}$ of W. For any $S \in \mathcal{S}$, where $\mathcal{S}$ is defined as in (5.1), the wiretapper source node observes F and the sequence $\mathbf{Z}_S^n$ in (5.2). The subset S is selected by the wiretapper and its selection is unknown to the legitimate parties. Thus, the wiretapper can be represented as a compound source $\mathbf{Z}_S^n \triangleq \left\{\mathcal{Z}, p_{\mathbf{Z}_S^n}, S \in \mathcal{S}\right\}$ whose distribution is only known to belong to the finite class $\{p_{\mathbf{Z}_S^n}\}_{S\in\mathcal{S}}$ with no prior distribution over the class, with $|\mathcal{S}| = \binom{n}{\alpha n} \leq 2^n$. For $S \in \mathcal{S}$, the induced joint distribution for this protocol is

$$\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) = p_{\mathbf{XYZ}_S}(\mathbf{x}, \mathbf{y}, \mathbf{z})\tilde{P}_{WF|\mathbf{X}}(w, f|\mathbf{x})\tilde{P}_{\hat{\mathbf{X}}|\mathbf{YF}}(\hat{\mathbf{x}}|\mathbf{y}, f) \tag{5.12}$$

$$= p_{\mathbf{XYZ}_S}(\mathbf{x}, \mathbf{y}, \mathbf{z})\mathbb{1}\{\mathcal{B}_{1,n}(\mathbf{X}) = W\}\mathbb{1}\{\mathcal{B}_{2,n}(\mathbf{X}) = F\}\tilde{P}_{\hat{\mathbf{X}}|\mathbf{YF}}(\hat{\mathbf{x}}|\mathbf{y}, f) \tag{5.13}$$

$$= \tilde{P}_{WF}(w, f)\tilde{P}_{\mathbf{X}|WF}(\mathbf{x}|w, f)\; p_{\mathbf{YZ}_S|\mathbf{X}}(\mathbf{y}, \mathbf{z}|\mathbf{x})\tilde{P}_{\hat{\mathbf{X}}|\mathbf{YF}}(\hat{\mathbf{x}}|\mathbf{y}, f). \tag{5.14}$$

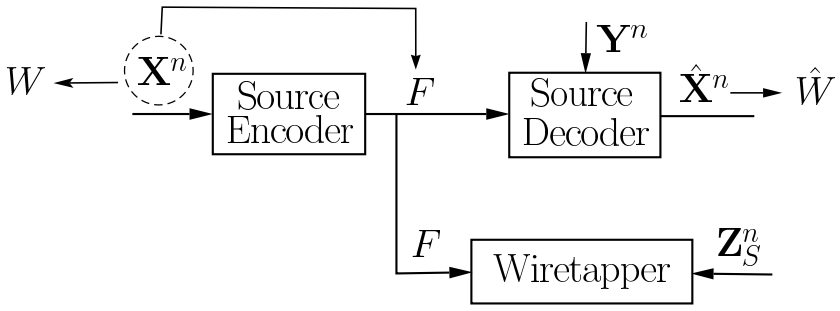


Fig. 5.2. Protocol A: Secret key agreement in the source model.

Protocol B (Main problem assisted with common randomness): This protocol is defined as the channel model in Figure 5.1, with an addition of a common randomness F that is uniformly distributed over $[1 : 2^{n\tilde{R}_s}]$, independent from all other variables, and known at all terminals. In fact, the assumed common randomness represents the random nature in generating the codebook, which is known at all nodes. At the end of the proof, we eliminate the assumed common randomness from the channel model in this protocol by conditioning on a certain instance of it. The encoder and decoder in this protocol are defined as in (5.14), i.e., $P_{\mathbf{X}|WF} = \tilde{P}_{\mathbf{X}|WF}$ and $P_{\hat{\mathbf{X}}|\mathbf{Y}F} = \tilde{P}_{\hat{\mathbf{X}}|\mathbf{Y}F}$. The induced joint distribution for this protocol is given by

$$P_{WF\mathbf{X}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{X}}}(w, f, \mathbf{x}, \mathbf{y}, \mathbf{z}, \hat{\mathbf{x}}) = p_W^U p_F^U \tilde{P}_{\mathbf{X}|WF}(\mathbf{x}|w, f)\; p_{\mathbf{Y}\mathbf{Z}_S|\mathbf{X}}(\mathbf{y}, \mathbf{z}|\mathbf{x}) \tilde{P}_{\hat{\mathbf{X}}|\mathbf{Y}F}(\hat{\mathbf{x}}|\mathbf{y}, f). \tag{5.15}$$

The induced joint distributions in (5.14) and (5.15) are random due to the random binning of $\mathbf{X}^n$. Note that we have ignored the random variables $\hat{W}$ from the induced joint distributions at this stage. We will introduce them later to the joint distributions as deterministic functions of the random vectors $\hat{\mathbf{X}}^n$, after fixing the binning functions.

The remaining steps of the proof are outlined as follows:

(i) We derive a condition on the rates R_s and $\tilde{R}_s$ such that the two induced joint distributions (5.14) and (5.15) are close in the total variation distance sense, when averaged over the random binning.

(ii) We then use Slepian-Wolf source coding theorem [20, 108] to derive a condition on the rate $\tilde{R}_s$ such that the decoding of $\hat{\mathbf{X}}$ in protocol A is reliable, i.e., the communication of the key W is reliable.

(iii) Next, for protocol A, we derive another condition on the rates R_s and $\tilde{R}_s$ such that the probability, with respect to the random binning, that for all $S \in \mathcal{S}$, the key W and the public message F are uniformly distributed, independent, and both independent from the wiretapper's observation $\mathbf{Z}_S^n$, goes to one as n goes to infinity, i.e., protocol A is secure.

(iv) We use the closeness of the two induced distributions for the two protocols to show that, under the same rate conditions for protocol B, the aforementioned reliability and secrecy properties in (ii) and (iii) hold for protocol B as well.

(v) The reliability and secrecy properties in (ii) and (iii), after being converted to the channel model in protocol B, are averaged over the random binning of the dual source model[2] in protocol A. We show the existence of a fixed binning realization such that both properties still hold for protocol B.

(vi) Finally, we eliminate the common randomness F from the channel model in protocol B by showing that the reliability and secrecy constraints still hold when we condition on a certain instance of F, i.e., $F = f^*$.

Note that, for the secrecy constraint, we have required the independence of the assumed common random F from both W and $\mathbf{Z}_S^n$ so that when we condition over an

[2]Note that the probability with respect to the random binning in (iii) is equivalent to an average over the random binning of an indicator function.

instance of F, the independence of W and $\mathbf{Z}_S^n$ is not affected. That is, the secrecy (independence) property in (iii) for protocol B, after fixing the binning function and removing the common randomness, results in an achievable strong secrecy rate for the original channel model.

Before continuing with the proof, we state the following lemmas.

5.4.1 Useful Lemmas

Lemma 7 is a *one-shot* result, which provides an exponential decay rate for the average, over the random binning, of the total variation distance between the two induced distributions from the two protocols. We utilize this lemma to show a convergence in probability result that allows converting the secrecy property from protocol A to protocol B. A result similar to Lemma 7 was derived in [126, Appendix A] which does not provide the required convergence rate, hence the need for Lemma 7.

Lemma 7. *Let the source $X \triangleq \{\mathcal{X}, p_X\}$ be randomly binned into $W = \mathcal{B}_1(X)$ and $F = \mathcal{B}_2(X)$, where $\mathcal{B}_1$ and $\mathcal{B}_2$ are uniform over $[1 : \tilde{W}]$ and $[1 : \tilde{F}]$, respectively. Let $\mathcal{B} \triangleq \{\mathcal{B}_1(x), \mathcal{B}_2(x)\}_{x\in\mathcal{X}}$, and for $\gamma > 0$, define*

$$\mathcal{D}_\gamma \triangleq \left\{ x \in \mathcal{X} : \log \frac{1}{p_X(x)} > \gamma \right\}. \tag{5.16}$$

Then, we have

$$\mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{WF}, p_W^U p_F^U \right) \right) \leq \mathbb{P} \left(X \notin \mathcal{D}_\gamma \right) + \frac{1}{2} \sqrt{\tilde{W}\tilde{F} 2^{-\gamma}}, \tag{5.17}$$

where P is the induced distribution over W and F.

Proof: The proof is provided in Appendix F. ■

Lemma 8 below is again a one-shot result which provides a *doubly-exponential* decay rate for the probability of failure of achieving the secrecy property for protocol A, for a fixed choice of the subset S. This lemma is needed, along with the union bound, to guarantee secrecy against the exponentially many possibilities of the tapped subset S.

Lemma 8. *Let* $X \triangleq \{\mathcal{X}, p_X\}$ *and* $\{Z_S\} \triangleq \left\{\mathcal{Z}, p_{Z_S}, S \in \mathcal{S}\right\}$ *be two correlated sources with* $|\mathcal{X}|, |\mathcal{Z}|$, *and* $|\mathcal{S}| < \infty$, *where* $\{Z_S\}_{S\in\mathcal{S}}$ *is a compound source whose distribution is known to belong to the finite class* $\{p_{Z_S}\}_{S\in\mathcal{S}}$. *Let* X *be randomly binned into the bin indices* W *and* F *as in Lemma 7. For* $\gamma > 0$ *and any* $S \in \mathcal{S}$, *define*

$$\mathcal{D}_\gamma^S \triangleq \left\{(x,z) \in \mathcal{X}\times\mathcal{Z} : \log\frac{1}{p_{X|Z_S}(x|z)} > \gamma\right\}. \tag{5.18}$$

If there exists $\delta \in \left(0, \frac{1}{2}\right)$ *such that for all* $S \in \mathcal{S}$, $\mathbb{P}_{p_{XZ_S}}\left((X, Z_S) \in \mathcal{D}_\gamma^S\right) \geq 1 - \delta^2$, *then, we have, for every* $\epsilon_1 \in [0,1]$, *that*

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(P_{WFZ_S} || p_W^U p_F^U p_{Z_S}\right) \geq \tilde{\epsilon}\right) \leq |\mathcal{S}||\mathcal{Z}| \exp\left(\frac{-\epsilon_1^2(1-\delta)2^\gamma}{3\tilde{W}\tilde{F}}\right), \tag{5.19}$$

where $\tilde{\epsilon} = \epsilon_1 + (\delta + \delta^2)\log(\tilde{W}\tilde{F}) + H_b(\delta^2)$, H_b *is the binary entropy function, and* P *is the induced distribution over* W, F, *and* Z_S.

Proof: The proof of Lemma 8 is given in Appendix G. ■

5.4.2 Proof

First, we apply Lemma 7 to protocol A. In Lemma 7, set $X = \mathbf{X}^n$, $\tilde{W} = 2^{nR_s}$, $\tilde{F} = 2^{n\tilde{R}_s}$, $\mathcal{B} = \mathcal{B}_n \triangleq \left\{\mathcal{B}_{1,n}(\mathbf{x}), \mathcal{B}_{2,n}(\mathbf{x})\right\}_{\mathbf{x}\in\mathcal{X}^n}$, and $\gamma = n(1-\epsilon_2)H(X)$, where $\epsilon_2 > 0$ and $\mathbf{X}^n$ is defined as in protocol A, i.e., is an i.i.d. sequence. Without loss of generality, we assume that for all $x \in \mathcal{X}$, we have $p_X(x) > 0$. Let $p_{\min} = \min_{x\in\mathcal{X}} \ p_X(x)$, where the minimum exists since the input alphabet $\mathcal{X}$ is finite[3]. Thus, the random variables $\log \frac{1}{p_X(X_i)}, i \in [1:n]$, are i.i.d. and each is bounded by the interval $[0, b_{\max}]$, where $b_{\max} = \log \frac{1}{p_{\min}}$.

We also have that $\bar{m} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{p_X}\left(\log \frac{1}{p_X(X_i)}\right) = H(X)$. Using Hoeffding's inequality in (2.6), we have, for any $\epsilon_2 > 0$, that

$$\mathbb{P}\left(\mathbf{X} \notin \mathcal{D}_\gamma\right) = \mathbb{P}_{p_{\mathbf{X}}}\left(\log \frac{1}{p_{\mathbf{X}}(\mathbf{X})} \le \gamma\right) \tag{5.20}$$

$$= \mathbb{P}_{p_X}\left(\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{p_X(X_i)} \le (1-\epsilon_2)H(X)\right) \tag{5.21}$$

$$\le \exp\left(\frac{-2\epsilon_2^2 H^2(X)}{b_{\max}^2} n\right) = \exp(-\beta_1 n), \tag{5.22}$$

where $\beta_1 = \frac{2\epsilon_2^2 H^2(X)}{b_{\max}^2} > 0$.

By substituting the choices for $\tilde{W}, \tilde{F}, \gamma$ and (5.22) in (5.17), we have, as long as $R_s + \tilde{R}_s < (1-\epsilon_2)H(X)$, that

$$\mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}(\tilde{P}_{WF}, p_W^U p_F^U)\right) \le 2\exp(-\beta n), \tag{5.23}$$

[3]If the input alphabet $\mathcal{X}$ is infinite, $\min_{x\in\mathcal{X}} \ p_X(x)$ might not exist. As a result, there might not be a finite upper bound on the random variables $\log \frac{1}{p_X(X_i)}$. In such a case, Hoeffding inequality can not be applied.

where $\beta_2 = \frac{\ln 2}{2}\left((1-\epsilon_2)H(X) - R_s - \tilde{R}_S\right)$ and $\beta = \min\{\beta_1, \beta_2\} > 0$. By applying (2.5) to (5.14) and (5.15), and using (5.23), we have

$$\mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}\right)\right) = \mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}\left(\tilde{P}_{WF}, p_W^U p_F^U\right)\right) \leq 2\exp(-\beta n). \quad (5.24)$$

Consider Slepian-Wolf decoder for protocol A. As long as $\tilde{R}_s \geq H(X|Y)$, we have [30, Theorem 10.1]

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}_n}\left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{X}} \neq \mathbf{X})\right) = 0. \quad (5.25)$$

Next, we observe

$$\begin{aligned}
&\mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\}\right)\right) \\
&\quad = \mathbb{E}_{\mathcal{B}_n} \sum_{\substack{w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}}: \\ \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z})\mathbb{1}\{\hat{\mathbf{x}}=\mathbf{x}\} \\ < \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}})}} \left[\tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{x}) - \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z})\mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\}\right] \quad (5.26) \\
&\quad = \mathbb{E}_{\mathcal{B}_n} \sum_{w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}}:\ \hat{\mathbf{x}}\neq\mathbf{x}} \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{x}) \quad (5.27) \\
&\quad = \mathbb{E}_{\mathcal{B}_n}\left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{X}} \neq \mathbf{X})\right). \quad (5.28)
\end{aligned}$$

Equation (5.27) follows because $\tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}}) > \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z})\mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\}$ holds if and only if $\mathbb{1}\{\hat{\mathbf{x}} = \mathbf{x}\} = 0$, where $\tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}})$ factorizes as $\tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z},\hat{\mathbf{x}}) = \tilde{P}(w,f,\mathbf{x},\mathbf{y},\mathbf{z})\tilde{P}(\hat{\mathbf{x}}|\mathbf{y},f)$ and $\tilde{P}(\hat{\mathbf{x}}|\mathbf{y},f) \leq 1$. Thus, using (5.25) and (5.28), we have that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}} = \mathbf{X}\}\right)\right) = 0, \quad (5.29)$$

as long as $\tilde{R}_s \geq H(X|Y)$.

Now, we apply Lemma 8 to protocol A. In Lemma 8, set $X = \mathbf{X}^n$, $\tilde{W} = 2^{nR_s}$, $\tilde{F} = 2^{n\tilde{R}_s}$, $\mathcal{B} = \mathcal{B}_n$, $Z_S = \mathbf{Z}_S^n$, for all $S \in \mathcal{S}$, and $\gamma = n(1-\tilde{\epsilon}_2)(1-\alpha)H(X|V)$, where $\tilde{\epsilon}_2 > 0$ and $\mathbf{X}^n, \mathbf{Z}_S^n, \mathcal{S}$ are defined as in protocol A. In order to calculate $\mathbb{P}_{p_{\mathbf{X}\mathbf{Z}_S}}\left((\mathbf{X}, \mathbf{Z}_S) \notin \mathcal{D}_\gamma^S\right)$, we only need to consider the pairs $(\mathbf{x}, \mathbf{z})$ such that $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) > 0$, since all the pairs $(\mathbf{x}, \mathbf{z})$ with $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) = 0$ belong to $\mathcal{D}_\gamma^S$, by the definition of $\mathcal{D}_\gamma^S$ in (5.18). Since the sequence $\mathbf{X}$ is i.i.d. and the channel $p_{V|X}$ is memoryless, we have, for all $(\mathbf{x}, \mathbf{z})$ with $p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) > 0$, that

$$p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{x}|\mathbf{z}) = p_{\mathbf{X}_S\mathbf{X}_{S^c}|\mathbf{X}_S\mathbf{V}_{S^c}}(\mathbf{x}_S, \mathbf{x}_{S^c}|\mathbf{x}_S, \mathbf{v}_{S^c}) \tag{5.30}$$

$$= p_{\mathbf{X}_{S^c}|\mathbf{V}_{S^c}}(\mathbf{x}_{S^c}|\mathbf{v}_{S^c}) = \prod_{i \in S^c} p_{X|V}(x_i|v_i). \tag{5.31}$$

Once again, using Hoeffding's inequality, we have, for all $S \in \mathcal{S}$,

$$\mathbb{P}_{p_{\mathbf{X}\mathbf{Z}_S}}\left((\mathbf{X}, \mathbf{Z}_S) \notin \mathcal{D}_\gamma^S\right) = \mathbb{P}_{p_{\mathbf{X}\mathbf{Z}_S}}\left(p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{X}|\mathbf{Z}_S) > 0,\ \log\frac{1}{p_{\mathbf{X}|\mathbf{Z}_S}(\mathbf{X}|\mathbf{Z}_S)} \leq \gamma\right) \tag{5.32}$$

$$= \mathbb{P}_{p_{X|V}}\left(\frac{1}{n-\mu}\sum_{i \in S^c} \log\frac{1}{p_{X|V}(X_i|V_i)} \leq (1-\tilde{\epsilon}_2)H(X|V)\right) \tag{5.33}$$

$$\leq \exp\left(-\tilde{\beta}(1-\alpha)n\right) = \delta^2, \tag{5.34}$$

where $\tilde{\beta} > 0$, and (5.33) follows from (5.31). From (5.34), $\lim_{n\to\infty} \delta^2 = 0$, and hence, for sufficiently large n, we have $\delta^2 \in \left(0, \frac{1}{4}\right)$. Thus, the conditions in Lemma 8 are satisfied.

Note that $\lim_{n\to\infty} n(\delta+\delta^2) = 0$, and $\lim_{n\to\infty} H_b(\delta^2) = H_b(\lim_{n\to\infty}\delta^2) = 0$ since H_b is a continuous function. Thus,

$$\lim_{n\to\infty}\tilde{\epsilon} = \epsilon_1 + (R_s+\tilde{R}_s)\lim_{n\to\infty} n(\delta+\delta^2) + \lim_{n\to\infty} H_b(\delta^2) = \epsilon_1. \tag{5.35}$$

By substituting the choices for $\tilde{W}, \tilde{F}, \gamma$, and $|\mathcal{S}||\mathcal{Z}^n| \leq \exp\left(n\left[\ln 2 + \ln\left(|\mathcal{X}|+|\mathcal{V}|\right)\right]\right)$ in (5.19), and using (5.35), we have that, for all $\epsilon_1, \epsilon_1' > 0$ and $\tilde{\epsilon} = \epsilon_1 + \epsilon_1'$, there exist $n^* \in \mathbb{N}$ and $\psi(\epsilon_1), \kappa > 0$ such that, for all $n \geq n^*$,

$$\mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right) \geq \tilde{\epsilon}\right) \leq \exp\left(-\psi(\epsilon_1)e^{\kappa n}\right), \tag{5.36}$$

as long as $R_s + \tilde{R}_s < (1-\tilde{\epsilon}_2)(1-\alpha)H(X|V)$.

Take $r > 0$ and let $D_n = \max_S \mathbb{D}(\tilde{P}_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S})$ and $\mathcal{K}_n \triangleq \{D_n \geq r\}$. Using (5.36), we have that $\sum_{n=1}^{\infty} \mathbb{P}_{\mathcal{B}_n}(\mathcal{K}_n) < \infty$. Thus, using the first Borel-Cantelli lemma yields

$$\mathbb{P}_{\mathcal{B}_n}\left(\mathcal{K}_n \text{ infinitely often (i.o.)}\right) = 0. \tag{5.37}$$

This implies that, for all $r > 0$, $\mathbb{P}_{\mathcal{B}_n}\left(\{D_n < r\} \text{ i.o.}\right) = 1$, i.e., the sequence D_n converges to zero almost surely. Thus, the sequence D_n converges to zero in probability as well. We conclude that, for $R_s + \tilde{R}_s < (1-\tilde{\epsilon}_2)(1-\alpha)H(X|V)$, we have

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right) > 0\right) = 0. \tag{5.38}$$

That is, protocol A is secure.

Next, we deduce that protocol B is also reliable and secure when $\tilde{R}_s \geq H(X|Y)$ and $R_s + \tilde{R}_s < (1-\tilde{\epsilon}_2)(1-\alpha)H(X|V)$. First, we show that the reliability in (5.29) holds for protocol B as well. We have

$$\begin{aligned}
&\mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) \\
&\leq \mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}\right) + \mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) \qquad (5.39) \\
&\leq \mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) + \mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) \\
&\qquad + \mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}\right) \qquad (5.40) \\
&= \mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) + 2\mathbb{V}\left(\tilde{P}_{WF}, p_W^U p_F^U\right), \qquad (5.41)
\end{aligned}$$

where (5.39) and (5.40) follow from the triangle inequality, and (5.41) follows since (5.14), (5.15) and (2.5) imply that

$$\begin{aligned}
\mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, \tilde{P}_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}\right) &= \mathbb{V}\left(\tilde{P}_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) \qquad (5.42) \\
&= \mathbb{V}\left(\tilde{P}_{WF}, p_W^U p_F^U\right). \qquad (5.43)
\end{aligned}$$

Substituting (5.23) and (5.29) in (5.41) yields

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right)\right) = 0. \qquad (5.44)$$

Second, we show that the secrecy property in (5.38) holds for protocol B. Using the union bound, we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right)>0\right)\\
&\quad\le \mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right)>0,\quad \text{and } \mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)>0\right)\\
&\qquad + \mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right)>0,\quad \text{and } \mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)=0\right) \qquad (5.45)\\
&\quad\le \mathbb{P}_{\mathcal{B}_n}\left(\mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)>0\right)+\mathbb{P}_{\mathcal{B}_n}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(\tilde{P}_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S}\right)>0\right). \qquad (5.46)
\end{aligned}$$

Equation (5.46) follows since $\mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)=0$ if and only if $\tilde{P}_{WF}(w,f)=p_W^U p_F^U$ for all w and f, and hence $P_{WF\mathbf{Z}_S}=p_W^U p_F^U P_{\mathbf{Z}_S|WF}=\tilde{P}_{WF}\tilde{P}_{\mathbf{Z}_S|WF}=\tilde{P}_{WF\mathbf{Z}_S}$, where

$$P_{\mathbf{Z}_S|WF}(\mathbf{z}|w,f)=\sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{Z}_S|\mathbf{X}}(\mathbf{z}|\mathbf{x})\tilde{P}_{\mathbf{X}|WF}(\mathbf{x}|w,f)=\tilde{P}_{\mathbf{Z}_S|WF}(\mathbf{z}|w,f). \qquad (5.47)$$

Using the exponential decay in (5.23) and Markov inequality, we have, for any $r>0$, that

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbb{P}_{\mathcal{B}_n}\left(\mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)>r\right) &\le \frac{1}{r}\sum_{n=1}^{\infty}\mathbb{E}_{\mathcal{B}_n}\left(\mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)\right) \qquad (5.48)\\
&\le \frac{2}{r}\sum_{n=1}^{\infty}\exp(-\beta n)<\infty, \qquad (5.49)
\end{aligned}$$

where $\beta>0$. Thus, using the Borel-Cantelli lemma, as in the derivation for (5.38), we have

$$\lim_{n\to\infty}\mathbb{P}_{\mathcal{B}_n}\left(\mathbb{V}(\tilde{P}_{WF},p_W^U p_F^U)>0\right)=0. \qquad (5.50)$$

By substituting (5.38) and (5.50) in (5.46), we get

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}_n} \left(\max_{S\in\mathcal{S}} \mathbb{D}\left(P_{WF\mathbf{Z}_S} || p_W^U p_F^U p_{\mathbf{Z}_S} \right) > 0 \right) = 0. \tag{5.51}$$

Now, we show the existence of a binning realization, and hence an encoder and decoder, such that the reliability and secrecy properties, in (5.44) and (5.51), hold for protocol B. By applying the selection lemma, Lemma 5, to the random sequence $\{\mathcal{B}_n\}_{n\geq 1}$ and the functions $\mathbb{V}\left(P_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, P_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right)$, $\mathbb{1}\left\{\max_{S\in\mathcal{S}}\mathbb{D}(P_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S})>0\right\}$, while using (5.44) and (5.51), there exists a sequence of binning realizations $\mathbf{b}_n^* = (b_{1,n}^*, b_{2,n}^*)$, with a corresponding joint distribution p^* for protocol B, such that

$$\lim_{n\to\infty} \mathbb{V}\left(p^*_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, p^*_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) = 0, \tag{5.52}$$

$$\lim_{n\to\infty} \mathbb{1}\left\{\max_{S\in\mathcal{S}}\mathbb{D}(p^*_{WF\mathbf{Z}_S}||p_W^U p_F^U p_{\mathbf{Z}_S})>0\right\} = 0, \tag{5.53}$$

where $W = b_{1,n}^*(\mathbf{X}^n)$ and $F = b_{2,n}^*(\mathbf{X}^n)$.

Next, we introduce the random variable $\hat{W}$ to the two joint distributions in (5.52), where $\hat{W}$ is a deterministic function of the random sequence $\hat{\mathbf{X}}^n$, i.e., $p^*_{\hat{W}|\hat{\mathbf{X}}}(\hat{w}|\hat{\mathbf{x}}) = \mathbb{1}\left\{\hat{w} = b_{1,n}^*(\hat{\mathbf{x}})\right\}$. Using (2.5) and (5.52), we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{V}\left(p^*_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}\hat{W}}, p^*_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{W}=W\}\right) \\
&= \lim_{n\to\infty} \mathbb{V}\left(p^*_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}\mathbb{1}\left\{\hat{W}=b_{1,n}^*(\hat{\mathbf{X}})\right\}, p^*_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\mathbb{1}\left\{\hat{W}=b_{1,n}^*(\hat{\mathbf{X}})\right\}\right) \qquad (5.54)\\
&= \lim_{n\to\infty} \mathbb{V}\left(p^*_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}}, p^*_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{\mathbf{X}}=\mathbf{X}\}\right) = 0, \qquad (5.55)
\end{aligned}$$

where (5.54) follows since $p^*_{\hat{W}|WF\mathbf{XYZ}_S\hat{\mathbf{X}}} = p^*_{\hat{W}|\hat{\mathbf{X}}} = \mathbb{1}\{\hat{W} = b^*_{1,n}(\hat{\mathbf{X}})\}$, and that $\hat{W} = W$ if and only if $\hat{\mathbf{X}} = \mathbf{X}$ and $\hat{W} = b^*_{1,n}(\hat{\mathbf{X}})$. We then have

$$\lim_{n\to\infty} \mathbb{E}_F\left(\mathbb{P}_{p^*}(\hat{W} \neq W|F)\right) = \lim_{n\to\infty} \sum_f p_F^U \sum_{w,\hat{w}:\ \hat{w}\neq w} p^*_{W\hat{W}|F}(w,\hat{w}|f) \tag{5.56}$$

$$= \lim_{n\to\infty} \sum_{w,\hat{w},f:\ \hat{w}\neq w} p^*_{W\hat{W}F}(w,\hat{w},f) \tag{5.57}$$

$$= \lim_{n\to\infty} \sum_{\substack{w,\hat{w},f:\\ p_W^U p_F^U \mathbb{1}\{\hat{w}=w\} \\ <p^*(w,\hat{w},f)}} \left[p^*_{W\hat{W}F}(w,\hat{w},f) - p_W^U p_F^U \mathbb{1}\{\hat{w}=w\}\right] \tag{5.58}$$

$$= \lim_{n\to\infty} \mathbb{V}\left(p^*_{W\hat{W}F}, p_W^U p_F^U \mathbb{1}\{\hat{W}=W\}\right) \tag{5.59}$$

$$= \lim_{n\to\infty} \mathbb{V}\left(p^*_{WF\mathbf{XYZ}_S\hat{\mathbf{X}}\hat{W}}, p^*_{WF\mathbf{XYZ}_S}\mathbb{1}\{\hat{W}=W\}\right) = 0. \tag{5.60}$$

Equation (5.58) follows because $p^*_{W\hat{W}F} > p_W^U p_F^U \mathbb{1}\{\hat{W}=W\}$ if and only if $\mathbb{1}\{\hat{W}=W\} = 0$ where $p^*_{W\hat{W}F}$ factorizes as $p_W^U p_F^U p^*_{\hat{W}|WF}$ and $p^*_{\hat{W}|WF} \leq 1$, while equation (5.60) follows from (2.5) and (5.55).

We also have that

$$\begin{aligned}
&\mathbb{P}_F\left(\max_S \mathbb{D}(p^*_{W\mathbf{Z}_S|F}||p_W^U p^*_{\mathbf{Z}_S|F}) > 0\right) \\
&\leq \mathbb{P}_F\left(\max_S \mathbb{D}\left(p^*_{W\mathbf{Z}_S|F}||p_W^U p^*_{\mathbf{Z}_S|F}\right) > 0, \quad \text{and } \max_S \mathbb{D}\left(p^*_{WF\mathbf{Z}_S^n}||p_W^U p_F^U p_{\mathbf{Z}_S}\right) = 0\right) \\
&\quad + \mathbb{P}_F\left(\max_S \mathbb{D}\left(p^*_{W\mathbf{Z}_S|F}||p_W^U p^*_{\mathbf{Z}_S|F}\right) > 0, \quad \text{and } \max_S \mathbb{D}\left(p^*_{WF\mathbf{Z}_S^n}||p_W^U p_F^U p_{\mathbf{Z}_S}\right) > 0\right)
\end{aligned} \tag{5.61}$$

$$\leq \mathbb{P}_F\left(\max_S \mathbb{D}\left(p^*_{W\mathbf{Z}_S|F}||p^U_W p^*_{\mathbf{Z}_S|F}\right) > 0, \text{ and } \forall S,\ p^*_{WF\mathbf{Z}_S}(w,f,\mathbf{z}) = p^U_W p^U_F p_{\mathbf{Z}_S}(\mathbf{z}), \forall w,f,\mathbf{z}\right)$$

$$+ \mathbb{P}_F\left(\max_S \mathbb{D}(p^*_{WF\mathbf{Z}^n_S}||p^U_W p^U_F p_{\mathbf{Z}_S}) > 0\right) \tag{5.62}$$

$$= \mathbb{1}\left\{\max_S \mathbb{D}(p^*_{WF\mathbf{Z}^n_S}||p^U_W p^U_F p_{\mathbf{Z}_S}) > 0\right\}, \tag{5.63}$$

where (5.62) follows since $\max_S \mathbb{D}\Big(p^*_{WF\mathbf{Z}^n_S}||p^U_W p^U_F p_{\mathbf{Z}_S}\Big) = 0$, if and only if, for all $S \in \mathcal{S}$, $p^*_{WF\mathbf{Z}_S}(w,f,\mathbf{z}) = p^U_W p^U_F p_{\mathbf{Z}_S}(\mathbf{z})$ for all w, f, and $\mathbf{z}$. (5.63) follows because the first probability term on the right hand side of (5.62) is equal to zero. Thus, using (5.53), we get

$$\lim_{n\to\infty} \mathbb{P}_F\left(\max_{S\in\mathcal{S}} \mathbb{D}(p^*_{W\mathbf{Z}_S|F}||p^U_W p^*_{\mathbf{Z}_S|F}) > 0\right) = 0. \tag{5.64}$$

Let us express the random variable F as an explicit function of n, i.e., $F = F_n = b^*_{2,n}(\mathbf{X}^n)$. In order to eliminate F_n from the channel model in protocol B, we apply the selection lemma, Lemma 5, to the random sequence $\{F_n\}_{n\geq 1}$ and the functions $\mathbb{P}_{p^*}\left(\hat{W} \neq W|F_n\right)$, $\mathbb{1}\left\{\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W\mathbf{Z}_S|F_n}||p^U_W p^*_{\mathbf{Z}_S|F_n}\right) > 0\right\}$, while using (5.60) and (5.64), which implies that there exists at least one realization $\{f^*_n\}$ such that

$$\lim_{n\to\infty} \mathbb{P}_{p^*}\left(\hat{W} \neq W|F_n = f^*_n\right) = 0, \text{ and} \tag{5.65}$$

$$\lim_{n\to\infty} \max_{S\in\mathcal{S}} I_{p^*}\left(W;\mathbf{Z}_S|F_n = f^*_n\right) = 0, \tag{5.66}$$

where I_{p^*} is the mutual information with respect to the distribution p^*. Equation (5.66) follows because $\lim_{n\to\infty} \mathbb{1}\{\max_{S\in\mathcal{S}} \mathbb{D}(p^*_{W\mathbf{Z}_S|F_n=f^*_n}||p^U_W p^*_{\mathbf{Z}_S|F_n=f^*_n}) > 0\} = 0$ implies that

there exists n' large enough such that, for all $n \geq n'$, we have

$$\max_S \mathbb{D}\left(p^*_{W\mathbf{Z}_S|F_n=f_n^*} || p^U_W p^*_{\mathbf{Z}_S|F_n=f_n^*}\right) = \max_S I_{p^*}\left(W; \mathbf{Z}_S | F_n = f_n^*\right) = 0. \tag{5.67}$$

Finally, let $\tilde{p}^*$ be the induced distribution for protocol A corresponding to $\mathbf{b}_n^*$. We use $\tilde{p}^*_{\mathbf{X}|W,F_n=f_n^*}$ as the encoder and $(\tilde{p}^*_{\hat{\mathbf{X}}|\mathbf{Y},F_n=f_n^*}, b^*_{1,n}(\hat{\mathbf{X}}))$ as the decoder for the original model. By combining the rate conditions $R_s + \tilde{R}_s < (1-\tilde{\epsilon}_2)(1-\alpha)H(X|V)$, $\tilde{R}_s \geq H(X|Y)$, and taking $\tilde{\epsilon}_2 \to 0$, the rate $R_s = \max_{p_X}[I(X;Y) - I(X;V) - \alpha H(X|V)]$ is achievable.

So far, we have considered the case $U = X$. Next, we prefix a discrete memoryless channel $p_{X|U}$ to the original channel model in Figure 5.1. The main channel for the generalized model is $p_{Y|U}$ and the wiretapper channel is described by $p_{X|U}$ and (5.2). The proof for this case follows similar steps to the proof above. In particular, for protocol A, we consider the i.i.d. input sequence $\mathbf{U}^n = [U_1, U_2, \cdots, U_n]$. When we apply Lemma 8 to protocol A, we set $\gamma = n(1-\tilde{\epsilon}_2)[\alpha H(U|X) + (1-\alpha)H(U|V)]$, and for $p_{\mathbf{U}|\mathbf{Z}_S}(\mathbf{u}|\mathbf{z}) > 0$, we have, for any $S \in \mathcal{S}$, that

$$p_{\mathbf{U}|\mathbf{Z}_S}(\mathbf{u}|\mathbf{z}) = p_{\mathbf{U}_S\mathbf{U}_{S^c}|\mathbf{X}_S\mathbf{V}_{S^c}}(\mathbf{u}_S, \mathbf{u}_{S^c}|\mathbf{x}_S, \mathbf{v}_{S^c}) \tag{5.68}$$

$$= p_{\mathbf{U}_S|\mathbf{X}_S\mathbf{V}_{S^c}}(\mathbf{u}_S|\mathbf{x}_S, \mathbf{v}_{S^c})\; p_{\mathbf{U}_{S^c}|\mathbf{U}_S\mathbf{X}_S\mathbf{V}_{S^c}}(\mathbf{u}_{S^c}|\mathbf{u}_S, \mathbf{x}_S, \mathbf{v}_{S^c}) \tag{5.69}$$

$$= p_{\mathbf{U}_S|\mathbf{X}_S}(\mathbf{u}_S|\mathbf{x}_S)\; p_{\mathbf{U}_{S^c}|\mathbf{V}_{S^c}}(\mathbf{u}_{S^c}|\mathbf{v}_{S^c}) \tag{5.70}$$

$$= \prod_{i\in S} p_{U|X}(u_i|x_i) \prod_{i\in S^c} p_{U|V}(u_i|v_i), \tag{5.71}$$

where (5.70) and (5.71) follow since the sequences $\mathbf{U}^n, \mathbf{X}^n$, and $\mathbf{V}^n$ are i.i.d. and the channels $p_{X|U}$ and $p_{V|X}$ are discrete memoryless channels. Using (5.71), the choice for γ, and Hoeffding's inequality, the conditions of Lemma 8 are satisfied, and we deduce the rate condition

$$R_s + \tilde{R}_s < (1 - \tilde{\epsilon}_2)[\alpha H(U|X) + (1 - \alpha)H(U|V)] \tag{5.72}$$

required for secrecy of protocol A. Note that $H(U|X) = H(U|X, V)$ because of the Markov chain $U - X - V$. By combining (5.72) with the rate condition $\tilde{R}_s \geq H(U|Y)$ required for the Slepian-Wolf decoder, we obtain the achievability of (5.5). The cardinality bound on $\mathcal{U}$, $|\mathcal{U}| \leq |\mathcal{X}|$, follows using [30, Appendix C]. This completes the achievability proof of Theorem 5.

5.5 Converse

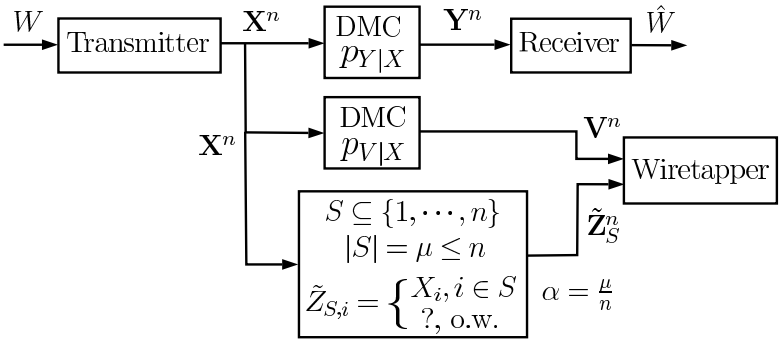


Fig. 5.3. A wiretap channel model whose secrecy capacity is equal to that of Figure 5.1.

Consider the channel model illustrated in Figure 5.3, where the wiretapper observes the outputs of two independent channels, with $\mathbf{X}^n$ being the input to both the channels. The first channel to the wiretapper is the discrete memroyless channel $p_{V|X}$ which outputs $\mathbf{V}^n$. The second channel is the wiretapper channel in the wiretap II channel model, i.e., the wiretapper chooses $S \subseteq [1:n]$ and observes $\tilde{\mathbf{Z}}_S^n = [\tilde{Z}_{S,1}, \cdots, \tilde{Z}_{S,n}]$, where $\tilde{Z}_{S,i} = X_i$ for $i \in S$, and $\tilde{Z}_{S,i} =?$, i.e., erasures, otherwise.

We show that, for $0 \leq \alpha \leq 1$, the strong secrecy capacity for this channel model, $C_s^{\mathrm{EQ}}(\alpha)$, is equal to the strong secrecy capacity of the original channel model, $C_s(\alpha)$, in (5.5). Since the main channels in the two models are the same, it suffices to show that $I(W;\mathbf{Z}_S^n) = I(W;\tilde{\mathbf{Z}}_S^n\mathbf{V}^n)$ for all $S \in \mathcal{S}$, where $\mathbf{Z}_S^n$ is defined as in (5.2). This follows because, for all $S \in \mathcal{S}$, we have

$$H(W|\tilde{\mathbf{Z}}_S\mathbf{V}) = H(W,\mathbf{X}|\tilde{\mathbf{Z}}_S,\mathbf{V}) - H(\mathbf{X}|W,\tilde{\mathbf{Z}}_S,\mathbf{V}) \tag{5.73}$$

$$= H(\mathbf{X}|\tilde{\mathbf{Z}}_S,\mathbf{V}) + H(W|\mathbf{X},\tilde{\mathbf{Z}}_S,\mathbf{V}) - H(\mathbf{X}|W,\tilde{\mathbf{Z}}_S,\mathbf{V}) \tag{5.74}$$

$$= H(\mathbf{X}|\tilde{\mathbf{Z}}_S,\mathbf{V}) - H(\mathbf{X}|W,\tilde{\mathbf{Z}}_S,\mathbf{V}) \tag{5.75}$$

$$= H(\mathbf{X}_S,\mathbf{X}_{S^c}|\mathbf{X}_S,\mathbf{V}_S,\mathbf{V}_{S^c}) - H(\mathbf{X}_S,\mathbf{X}_{S^c}|W,\mathbf{X}_S,\mathbf{V}_S,\mathbf{V}_{S^c}) \tag{5.76}$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S,\mathbf{V}_S,\mathbf{V}_{S^c}) - H(\mathbf{X}_{S^c}|W,\mathbf{X}_S,\mathbf{V}_S,\mathbf{V}_{S^c}) \tag{5.77}$$

$$= H(\mathbf{X}_{S^c}|\mathbf{X}_S,\mathbf{V}_{S^c}) - H(\mathbf{X}_{S^c}|W,\mathbf{X}_S,\mathbf{V}_{S^c}) \tag{5.78}$$

$$= H(\mathbf{X}|\mathbf{Z}_S) - H(\mathbf{X}|W,\mathbf{Z}_S) \tag{5.79}$$

$$= H(\mathbf{X},W|\mathbf{Z}_S) - H(\mathbf{X}|W,\mathbf{Z}_S) = H(W|\mathbf{Z}_S), \tag{5.80}$$

where (5.75) and (5.80) follow because $H(W|\mathbf{X}) = 0$, and (5.78) follows since the channel $p_{V|X}$ is memoryless which results in the Markov chains $\mathbf{X}_{S^c} - \mathbf{X}_S\mathbf{V}_{S^c} - \mathbf{V}_S$ and $\mathbf{X}_{S^c} - W\mathbf{X}_S\mathbf{V}_{S^c} - \mathbf{V}_S$.

Next, consider the channel model illustrated in Figure 5.4, which is the same as the channel model in Figure 5.3, except we replace the second channel to the wiretapper with a discrete memoryless erasure channel (DM-EC) with erasure probability $1 - \alpha$. The output of the second channel to the wiretapper is $\mathbf{Z}^n$. For this model, we have the Markov chain $\mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n$ since the two channels to the wiretapper are independent. Since the two channels to the wiretapper are discrete memoryless, we have

$$\begin{aligned} p_{\mathbf{VZ}|\mathbf{X}}(\mathbf{v}, \mathbf{z}|\mathbf{x}) &= p_{\mathbf{V}|\mathbf{X}}(\mathbf{v}|\mathbf{x})\ p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) \\ &= \prod_{i=1}^{n} p_{V|X}(v_i|x_i)\ p_{Z|X}(z_i|x_i) \\ &= \prod_{i=1}^{n} p_{VZ|X}(v_i, z_i|x_i). \end{aligned} \tag{5.81}$$

That is, the combined channel to the wiretapper is a discrete memoryless channel, making the channel model in Figure 5.4 a discrete memoryless wiretap channel. The strong secrecy capacity for this model $C_s^{\mathrm{EQ2}}(\alpha)$ is given by

$$C_s^{\mathrm{EQ2}}(\alpha) = \max_{U-X-YVZ} [I(U;Y) - I(U;VZ)]^+. \tag{5.82}$$

In order to compute $C_s^{\mathrm{EQ2}}(\alpha)$ in (5.82), we define the random variable $\Phi \sim \mathrm{Bern}(\alpha)$ whose n i.i.d. samples represent the erasure process in the DM-EC, where

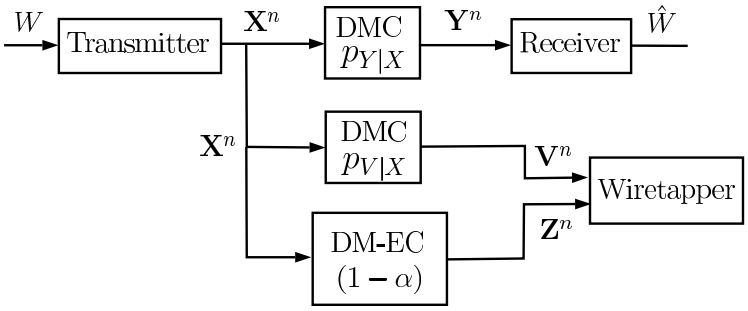


Fig. 5.4. A discrete memoryless equivalent wiretap channel model.

$\Phi = 0$ when $Z = X$ and $\Phi = 1$ when $Z =?$. Thus, Φ is determined by Z, and hence, the Markov chains $U - X - YVZ$ and $V - X - Z$ imply the Markov chains $U - X - YVZ\Phi$ and $V - X - Z\Phi$. Also, Φ is independent from X, since the erasure process is independent from the input to the channel. Thus, we have

$$\begin{aligned} p_{\Phi|UV}(\phi|u,v) &= \sum_{x\in\mathcal{X}} p_{\Phi X|UV}(\phi,x|u,v) \\ &= \sum_{x\in\mathcal{X}} p_{X|UV}(x|u,v)\; p_{\Phi|XUV}(\phi|x,u,v) \\ &= p_\Phi(\phi) \sum_{x\in\mathcal{X}} p_{X|UV}(x|u,v) = p_\Phi(\phi) \qquad (5.83) \\ p_{\Phi|V}(\phi|v) &= \sum_{x\in\mathcal{X}} p_{\Phi X|V}(\phi,x|v) \\ &= \sum_{x\in\mathcal{X}} p_{X|V}(x|v)\; p_{\Phi|XV}(\phi|x,v) \\ &= p_\Phi(\phi) \sum_{x\in\mathcal{X}} p_{X|V}(x|v) = p_\Phi(\phi). \qquad (5.84) \end{aligned}$$

where (5.83) and (5.84) follow since $p_{\Phi|XUV} = p_{\Phi|XV} = p_{\Phi|X} = p_\Phi$ due to the Markov chains $U - XV - \Phi$ and $V - X - \Phi$, and the independence of Φ and X. Since $p_{\Phi|UV} =$

$p_{\Phi|V} = p_\Phi$, then Φ and U are conditionally independent given V. Thus, we have

$$I(U;Z|V) = I(U;Z,\Phi|V) = I(U;Z|\Phi,V) \tag{5.85}$$

$$= \mathbb{P}(\Phi=0)I(U;Z|\Phi=0,V) + \mathbb{P}(\Phi=1)I(U;Z|\Phi=1,V) \tag{5.86}$$

$$= \alpha I(U;X|V) + (1-\alpha)I(U;?|V) \tag{5.87}$$

$$= \alpha I(U;X|V). \tag{5.88}$$

Substituting (5.88) in (5.82), we have

$$C_s^{\mathrm{EQ2}}(\alpha) = \max_{U-X-YV} [I(U;Y) - I(U;V) - \alpha I(U;X|V)]^+. \tag{5.89}$$

Next, we use similar arguments to [33, Section V-C] to show that $C_s^{\mathrm{EQ}}(\alpha) \leq C_s^{\mathrm{EQ2}}(\alpha)$ for any $0 \leq \alpha \leq 1$ and sufficiently large n. The idea is that when the number of erasures of the DM-EC in the latter channel model (Figure 5.4) is greater than or equal to $(1-\alpha)n$, the wiretapper's channel in the former (Figure 5.3) is better than its channel in the latter, since the wiretapper in the former is more capable and encounters a smaller number of erasures. Thus, $C_s^{\mathrm{EQ}}(\alpha) \leq C_s^{\mathrm{EQ2}}(\alpha)$ for this case. The result is established by using Sanov's theorem in method of types [21, Theorem 11.4.1] to show that the probability that the DM-EC causes erasures less than $(1-\alpha)n$ goes to 0 as $n \to \infty$.

In particular, we first show that, for $0 \leq \lambda < \alpha \leq 1$, we have $C_s^{\mathrm{EQ}}(\alpha) \leq C_s^{\mathrm{EQ2}}(\lambda)$. To do so, we show that every achievable strong secrecy rate for the channel model in Figure 5.3 is also achievable for the channel model in Figure 5.4. Fix λ such that

$0 \leq \lambda < \alpha \leq 1$, and let R_s be an achievable strong secrecy rate for the former channel model. Thus, there exists a sequence of $(n, 2^{nR_S})$ channel codes $\left\{\mathcal{C}_n^{\mathrm{EQ}}\right\}_{n\geq 1}$ such that

$$\lim_{n\to\infty} \mathbb{P}^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(\hat{W} \neq W\right) = 0, \tag{5.90}$$

$$\text{and } \lim_{n\to\infty} \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(W; \tilde{\mathbf{Z}}_S, \mathbf{V}\right) = 0, \tag{5.91}$$

where $\mathbb{P}^{(\mathcal{C}_n^{\mathrm{EQ}})}$ and $I^{(\mathcal{C}_n^{\mathrm{EQ}})}$ are the probability and the mutual information with respect to the joint distribution that corresponds to the code $\mathcal{C}_n^{\mathrm{EQ}}$.

We show that the rate R_s is also an achievable strong secrecy rate for the channel model in Figure 5.4 by showing that the sequence of $(n, 2^{nR_s})$ codes $\left\{\mathcal{C}_n^{\mathrm{EQ}}\right\}_{n\geq 1}$ satisfies the constraints $\lim_{n\to\infty} \mathbb{P}^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(\hat{W} \neq W\right) = 0$ and $\lim_{n\to\infty} \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W; \mathbf{Z}, \mathbf{V}) = 0$ for this channel model.

The main channel in the two models is the same, and hence, the sequence of $(n, 2^{nR_s})$ codes $\left\{\mathcal{C}_n^{\mathrm{EQ}}\right\}_{n\geq 1}$ achieves the reliability constraint for both channel models. Thus, it remains to show that $\left\{\mathcal{C}_n^{\mathrm{EQ}}\right\}_{n\geq 1}$ achieves $\lim_{n\to\infty} \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W; \mathbf{Z}, \mathbf{V}) = 0$.

Since $\lim_{n\to\infty} \max_{S\in\mathcal{S}} I(W; \tilde{\mathbf{Z}}_S, \mathbf{V}) = 0$, then for any $\epsilon_0 > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have

$$\max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(W; \tilde{\mathbf{Z}}_S, \mathbf{V}\right) = I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W; \mathbf{V}) + \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(W; \tilde{\mathbf{Z}}_S | \mathbf{V}\right) \leq \frac{\epsilon_0}{2}. \tag{5.92}$$

Let us define $\tilde{\mathcal{Z}} \triangleq \mathcal{X} \cup \{?\}$. For every $\mathbf{z}^n \in \tilde{\mathcal{Z}}^n$, define

$$\mathcal{N}(\mathbf{z}^n) \triangleq \{k \in [1:n] : z_k = ?\} \tag{5.93}$$

$$\Theta(\mathbf{z}^n) \triangleq \mathbb{1}\Big\{|\mathcal{N}(\mathbf{z}^n)| \leq \lceil (1-\alpha)n \rceil\Big\}. \tag{5.94}$$

That is, $\mathcal{N}(\mathbf{z}^n)$ represents the number of erasures in the sequence $\mathbf{z}^n$, while $\Theta(\mathbf{z}^n)$ indicates whether the sequence $\mathbf{z}^n$ has erasures less than or equal to $\lceil (1-\alpha)n \rceil$.

For simplicity of notation, we drop the superscripts $\mathcal{C}_n^{\mathrm{EQ}}$ in (5.92); it is understood implicitly that the mutual information is calculated with respect to the code $\mathcal{C}_n^{\mathrm{EQ}}$. Since $\Theta(\mathbf{Z}^n)$ is a deterministic function of $\mathbf{Z}^n$, the Markov chains $W - \mathbf{X}^n - \mathbf{V}^n\mathbf{Z}^n$ and $W\mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n$ imply the Markov chains $W - \mathbf{X}^n - \mathbf{V}^n\mathbf{Z}^n\Theta(\mathbf{Z}^n)$ and $W\mathbf{V}^n - \mathbf{X}^n - \mathbf{Z}^n\Theta(\mathbf{Z}^n)$. Also, $\Theta(\mathbf{Z}^n)$ is independent from $\mathbf{X}^n$. Thus, we have

$$\begin{aligned}
p_{\Theta(\mathbf{Z})|W\mathbf{V}}(\theta|w,\mathbf{v}) &= \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\Theta(\mathbf{Z})\mathbf{X}|W\mathbf{V}}(\theta,\mathbf{x}|w,\mathbf{v}) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{X}|W\mathbf{V}}(\mathbf{x}|w,\mathbf{v})\; p_{\Theta(\mathbf{Z})|\mathbf{X}W\mathbf{V}}(\theta|\mathbf{x},w,\mathbf{v}) \\
&= p_{\Theta(\mathbf{Z})}(\theta) \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{X}|W\mathbf{V}}(\mathbf{x}|w,\mathbf{v}) = p_{\Theta(\mathbf{Z})}(\theta)
\end{aligned} \tag{5.95}$$

$$\begin{aligned}
p_{\Theta(\mathbf{Z})|\mathbf{V}}(\theta|\mathbf{v}) &= \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\Theta(\mathbf{Z})\mathbf{X}|\mathbf{V}}(\theta,\mathbf{x}|\mathbf{v}) \\
&= \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{X}|\mathbf{V}}(\mathbf{x}|\mathbf{v})\; p_{\Theta(\mathbf{Z})|\mathbf{X}\mathbf{V}}(\theta|\mathbf{x},\mathbf{v}) \\
&= p_{\Theta(\mathbf{Z})}(\theta) \sum_{\mathbf{x}\in\mathcal{X}^n} p_{\mathbf{X}|\mathbf{V}}(\mathbf{x}|\mathbf{v}) = p_{\Theta(\mathbf{Z})}(\theta).
\end{aligned} \tag{5.96}$$

From (5.95) and (5.96), W and $\Theta(\mathbf{Z})$ are conditionally independent given $\mathbf{V}$, and hence,

$$I(W;\mathbf{Z}|\mathbf{V}) = I(W;\mathbf{Z},\Theta(\mathbf{Z})|\mathbf{V}) \tag{5.97}$$

$$= I(W;\Theta(\mathbf{Z})|\mathbf{V}) + I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})) \tag{5.98}$$

$$= I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})) \tag{5.99}$$

$$= \mathbb{P}(\Theta(\mathbf{Z})=0)I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})=0) + \mathbb{P}(\Theta(\mathbf{Z})=1)I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})=1). \tag{5.100}$$

We upper bound each term in the right hand side of (5.100). The first term is upper bounded by

$$I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})=0) = I\left(W;\mathbf{Z}|\mathbf{V},\{|\mathcal{N}(\mathbf{Z})| > \lceil(1-\alpha)n\rceil\}\right) \tag{5.101}$$

$$\leq I\big(W;\mathbf{Z}|\mathbf{V},\big\{|\mathcal{N}(\mathbf{Z})| = \lceil(1-\alpha)n\rceil\big\}\big) \tag{5.102}$$

$$\leq \max_{S\in\mathcal{S}} I(W;\tilde{\mathbf{Z}}_S|\mathbf{V}). \tag{5.103}$$

We also have that

$$I(W;\mathbf{Z}|\mathbf{V},\Theta(\mathbf{Z})=1) \leq H(\mathbf{Z}) \leq n\log(|\mathcal{X}|+1). \tag{5.104}$$

Next we upper bound $\mathbb{P}(\Theta(\mathbf{Z})=1)$. Take ν such that $\lambda<\nu<\alpha$, and hence, we have $\lceil(1-\alpha)n\rceil \leq (1-\nu)n < (1-\lambda)n$. Let $\Phi_1,\Phi_2,\cdots,\Phi_n$ be a sequence of i.i.d. binary random variables which represents the erasure process of the DM-EC in the model in Figure 5.4 ($\Phi_i=1$ if $Z_i=X_i$, and $\Phi_i=0$ if $Z_i=?$), where Φ_i is distributed according to $Q_\Phi = \text{Bern}(\lambda)$. Let Q_Φ^n be the n-letter distribution of the sequence $\{\Phi_i\}_{i=1}^n$. For each $\xi=\frac{l}{n}$, with $l\in[\lceil\nu n\rceil : n]$, i.e., $\nu\leq\xi<1$, define the distribution $P_\Phi^{(\xi)} = \text{Bern}(\xi)$, and let $\mathcal{P}$ be the set of all of these distributions. Let $T(P)$ denote the type class of the

distribution P, i.e., all possible n-length sequences with the type (empirical distribution) P [21, Section 11.1]. Define the set $\mathcal{T} \triangleq \left\{T(P_\Phi^{(\xi)}) : (1-\xi) \leq (1-\nu)\right\}$. Using Sanov's theorem [21, Theorem 11.4.1], we have

$$\mathbb{P}(\Theta(\mathbf{Z})=1) = \mathbb{P}_{Q_\Phi^n}\left(|\mathcal{N}(\mathbf{Z})| \leq \lceil (1-\alpha)n \rceil\right) \tag{5.105}$$

$$\leq \mathbb{P}_{Q_\Phi^n}\left(|\mathcal{N}(\mathbf{Z})| \leq (1-\nu)n\right) \tag{5.106}$$

$$= \mathbb{P}_{Q_\Phi^n}\left(\left|\left\{k \in [1:n] : \Phi_k = 1\right\}\right| \leq (1-\nu)n\right) \tag{5.107}$$

$$= \mathbb{P}_{Q_\Phi^n}(\mathcal{T}) = \mathbb{P}_{Q_\Phi^n}(\mathcal{P}) \leq (n+1)^2\, 2^{-n\mathbb{D}(P_\Phi^*||Q_\Phi)}, \tag{5.108}$$

where

$$P_\Phi^* = \underset{P_\Phi^{(\xi)} \in \mathcal{P}}{\operatorname{argmin}}\, \mathbb{D}(P_\Phi^{(\xi)}||Q_\Phi) = \underset{\xi:\, \xi \geq \nu}{\operatorname{argmin}} \left(\xi \log \frac{\xi}{\lambda} + (1-\xi)\frac{1-\xi}{1-\lambda}\right) = \text{Bern}(\nu). \tag{5.109}$$

Note that $\mathbb{D}(P_\Phi^*||Q_\Phi) > 0$ since $\nu \neq \lambda$.

Using (5.104) and (5.108), the second term in the right hand side of (5.100) is upper bounded by

$$\log(|\mathcal{X}|+1)n(n+1)^2\, 2^{-n\mathbb{D}(P_\Phi^*||Q_\Phi)} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.110}$$

Thus, for $\epsilon_0 > 0$, there exists $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$,

$$\mathbb{P}(\Theta(\mathbf{Z})=1) I(W; \mathbf{Z}|\mathbf{V}, \Theta(\mathbf{Z})=1) \leq \frac{\epsilon_0}{2}. \tag{5.111}$$

Using (5.92), (5.100), (5.103), and (5.111), we have, for sufficiently large n, that

$$I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W;\mathbf{Z},\mathbf{V}) = I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W;\mathbf{V}) + I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W;\mathbf{Z}|\mathbf{V}) \tag{5.112}$$

$$\leq I^{(\mathcal{C}_n^{\mathrm{EQ}})}(W;\mathbf{V}) + \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n^{\mathrm{EQ}})}\left(W;\tilde{\mathbf{Z}}_S|\mathbf{V}\right) + \frac{\epsilon_0}{2} \leq \epsilon_0. \tag{5.113}$$

Thus, for $0 \leq \lambda < \alpha \leq 1$, we have $C_s^{\mathrm{EQ}}(\alpha) \leq C_s^{\mathrm{EQ2}}(\lambda)$. The right hand side of (5.89) is a continuous function of α, for $0 < \alpha < 1$ [33, Lemma 6]. Thus, by taking $\lambda \to \alpha$, we have $C_s^{\mathrm{EQ}}(\alpha) \leq C_s^{\mathrm{EQ2}}(\alpha)$. Note that for $\alpha = 0, 1$, we have $C_s^{\mathrm{EQ}}(\alpha) = C_s^{\mathrm{EQ2}}(\alpha)$. Thus, the secrecy capacity of the original model in Figure 5.1 is upper bounded by (5.89). This completes the proof for Theorem 5.

5.6 Discussion

In the converse proof for Theorem 5, we have shown that the strong secrecy capacity $C_s(\alpha)$ for the generalized wiretap channel model is equal to the strong secrecy capacity when the wiretapper, in addition to observing μ transmitted symbols of its choice noiselessly, observes the whole sequence $\mathbf{V}^n$. This is not surprising since observing noisy versions of the transmitted symbols through the discrete memroyless channel $p_{V|X}$ in the same positions where noiseless versions are available does not increase the wiretapper's information about the message. The expression for the strong secrecy capacity in (5.5) is thus intuitive where $I(U,V)$ represents the secrecy cost due to observing the whole sequence $\mathbf{V}^n$, and $\alpha I(U;X|V)$ represents the secrecy cost due to observing a fraction α of the transmitted symbols noiselessly, given the wiretapper's knowledge of the V outputs in these positions. Furthermore, the alternative characterization for the

strong secrecy capacity in (5.6) is again intuitively pleasing, where the overall secrecy cost is represented by a weighted sum of the secrecy costs due to the noiseless and the noisy observations at the wiretapper, i.e., $\alpha I(U;X)$ and $(1-\alpha)I(U;V)$.

It is worth noting that a problem similar to the model considered in this chapter appears in the context of Quantum Cryptography when a transmitter and a receiver wish to agree on a secret key over a quantum channel in the presence of an external adversary [10, 100]. The adversary can apply any arbitrary sequence of operations, allowed by the laws of quantum physics, on the quantum states exchanged between the transmitter and receiver. The security of the key follows from the impossibility of applying such operations on a quantum mechanical system without changing its overall state. The legitimate terminals, by communicating over an additional classical error-free channel, can estimate the number of errors in the system, caused by the adversary, and abort the key agreement protocol if the number of errors exceeds a certain threshold. To sum up, like the models considered in this chapter, the adversary in the problem described above is limited only in the fraction of time of being active. We refer the reader to [116, 119], and references therein, for a comprehensive treatment of the problems and utilized tools in quantum information theory.

Finally, we note that extending the proposed achievability approach in this chapter to the case of a non-uniform message at the transmitter, i.e., semantic secrecy, does not appear straightforward. In particular, in order to handle the case of a non-uniform message, we would need to characterize the distribution of the source $\mathbf{X}^n$ given the wiretapper's observation $\mathbf{Z}_S^n$, when conditioned over each key realization w, i.e., $p_{\mathbf{X}^n|\mathbf{Z}_S^n,W}(\mathbf{x}|\mathbf{z},w)$, for all $\mathbf{x}\in\mathcal{X}^n, \mathbf{z}\in\mathcal{Z}^n$, and $w\in[1:2^{nR_s}]$, which is not easy.

5.7 Conclusion

In this chapter, we have introduced a wiretap channel model that generalizes the classical wiretap channel models, and derived its strong secrecy capacity. The model generalizes the classical wiretap channel [23, 121] to one with a wiretapper which chooses a fixed-length subset of the transmitted codeword symbols to perfectly access, and generalizes the wiretap channel II with a discrete memoryless main channel in [85] to one with a wiretapper which observes the output of a noisy channel instead of the erasures. The wiretapper in this model is still passive, yet it is more capable than a classical wiretapper since it can tap a subset of the symbols of its choosing noiselessly, while still receiving the remaining symbols through a channel. Our secrecy capacity result quantifies the secrecy cost of this additional capability of the wiretapper, with respect to the previous wiretap channel models. Exploring the multi-terminal extensions of this generalized model is the natural next step, similar to multi-terminal extensions for Wyner's original model, e.g., [29, 41, 42, 60, 64, 112] , which is the focus of the following two chapters.

Chapter 6

Generalizing Multiple Access Wiretap and Wiretap II Channel Models

6.1 Introduction

In Chapter 4, we have introduced a discrete memoryless main channel to the wiretap channel II model, and derived inner and outer bounds for its capacity-equivocation region. In Chapter 5, we have introduced and derived the strong secrecy capacity of the generalized wiretap channel in which the main channel is noisy and the wiretapper, besides noiselessly observing a subset of its choice of the transmitted symbols, observes the remainder through a noisy channel. Investigating the multi-terminal extensions of this generalized wiretap model is the natural next step, much like what happened with Wyner's wiretap channel.

In this chapter, we extend our generalized wiretap channel model to the multiple access scenario [113]. In particular, we first consider the special case of the multiple access wiretap channel II with a discrete memoryless main channel, and propose three different attack models for the wiretapper. In each of these models, the wiretapper chooses a fixed-length subset of the channel uses and observes erasures outside this subset. In the first wiretapping model, the wiretapper, in each position of the subset, decides to observe either the first or the second user's symbol. In the second model, the wiretapper observes a noiseless superposition of the two transmitted symbols in the positions of the

subset, while in the third model, the wiretapper observes the transmitted symbols of both users.

The first attack model is a setting in which the wiretapper is able to tap one of the two transmissions but not both. For instance, if two transmitters are distant from each other, the wiretapper may need get close to one in order to obtain noise-free observations, and thus is able to tap one at a time. The second attack model mimics a medium that superposes both transmissions (e.g., wireless), where the attacker is close enough to both transmitters. In the third attack model, the wiretapper is able to tap both codewords individually, which can be interpreted as the wiretapper being able to obtain noiseless (partial) side information about both transmitted codewords.

For each of these models, we derive an achievable strong secrecy rate region. Even though the third attack model, in which the wiretapper sees the transmitted symbols of *both* users, is stronger than the first, the ability of the wiretapper in the first model to choose which user's symbol to tap into results in identical achievable strong secrecy rate regions for the two models. That is, each transmitter designs their encoding according to the worst case scenario in which the wiretapper chooses to see its symbols in all positions of the subset. The achievable secrecy rate region for the second attack model is shown to be larger than the achievable secrecy rate region for the other two models, demonstrating the intrinsic cooperation introduced by superposition.

After obtaining these insights, we generalize these models by replacing the wiretapper's erasures with noisy channel outputs as we have done in Chapter 5 for the single user channel. In particular, we generalize the multiple access wiretap channel II with a discrete memoryless main channel, under the third wiretapping scenario, i.e., the

strongest attack model, to the case when the wiretapper observes the remainder of the codewords of both users separately through a discrete memoryless channel. This model also generalizes the multiple access wiretap channel in [112] to the case when the wiretapper is provided with a subset of noiseless observations of its choice for the transmitted symbols of both users. An achievable strong secrecy rate region, which quantifies the secrecy cost of the additional capability of the wiretapper in this model with respect to the multiple access wiretap channel in [112, 125], is derived.

Achievability of the strong secrecy rate regions for all the proposed models is established by muti-terminal extensions of methods proposed in Chapter 5. In particular, for each of the proposed models, a corresponding dual multi-terminal secret key agreement problem in the source model is introduced. In this dual model, two independent sources wish to agree on two indepedent keys with a common decoder in the presence of a *compound* wiretapping source. We solve the problem in the dual source model, and convert the solution to the original channel model by means of deriving the joint distributions of the two problems to become almost identical, in the total variation distance sense. The technical challenge in this chapter lies in generalizing the tool utilized for establishing secrecy of the key in the dual source model from the single source case, Lemma 8, to the case of two *independent* sources. This is done by adapting the lemma in order to establish all the *corner* (extreme) points of the rate region for the two keys, generated at the independent sources, such that the convergence rate for the probability of the two keys being independent from the wiretapper's observation is *doubly-exponential*. Time sharing between the resulting corner points produces the desired rate region. As

in Chapter 5, this doubly-exponential convergence rate is needed in order to *exhaust* the exponentially many possible strategies for the wiretapper.

The remainder of the chapter is organized as follows. Section 6.2 describes the channel models considered in this chapter. Section 6.3 presents the main results. The proofs of the results are presented in Sections 6.4 and 6.5. Section 6.6 concludes the chapter.

6.2 Channel Models

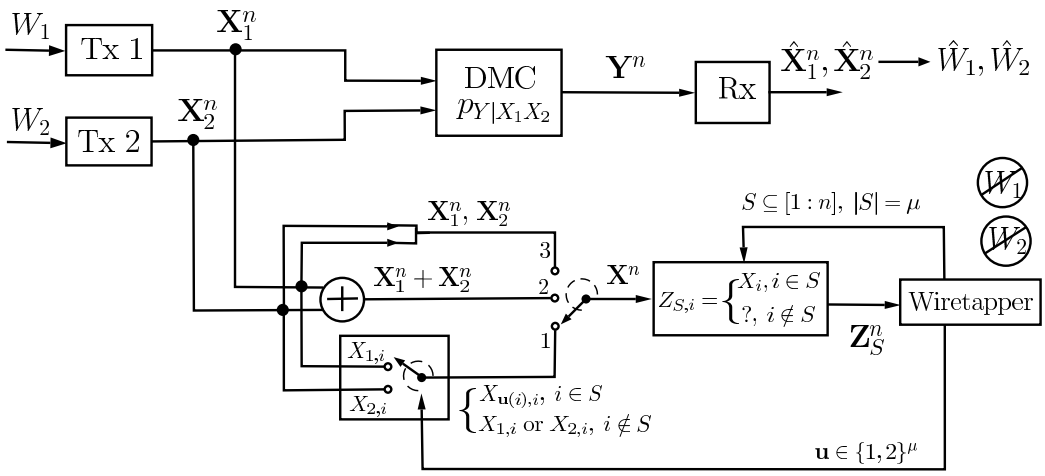


Fig. 6.1. The two-user multiple access wiretap channel II with a noisy main channel.

We describe the channel models we consider in this chapter. In Section 6.2.1, we present the multiple access wiretap channel II with a noisy main channel under the three aforementioned attack models for the wiretapper. Section 6.2.2 describes a generalized multiple access wiretap channel model which generalizes the strongest attack model in Section 6.2.1.

6.2.1 The Multiple Access Wiretap Channel II with a Noisy Main Channel

Consider the channel model in Fig. 6.1. The main channel $\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, p_{Y|X_1X_2}\}$ is a discrete memoryless channel consisting of two finite input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, a finite output alphabet $\mathcal{Y}$, and a transition probability distribution $p_{Y|X_1X_2}$. Each transmitter wishes to reliably communicate an independent message to a common receiver and to keep it secret from the wiretapper. To do so, transmitter j maps its message, W_j, uniformly distributed over $[1 : 2^{nR_j}]$, into the transmitted codeword $\mathbf{X}_j^n = [X_{j,1}, X_{j,2}, \cdots, X_{j,n}] \in \mathcal{X}_j^n$ using a stochastic encoder, $j = 1, 2$. The receiver observes the sequence $\mathbf{Y}^n = [Y_1, Y_2, \cdots, Y_n] \in \mathcal{Y}^n$ and outputs the estimates $\hat{W}_j, j = 1, 2$, of the transmitted messages. As shown in Fig. 6.1, we consider the following three models for the wiretapper channel.

6.2.1.1 Model 1

This model is described in Fig. 6.1, when the switch is on position 1. The wiretapper chooses the subset $S_p \in \mathcal{S}_p$ and the sequence $\mathbf{u} = [u_1, u_2, \cdots, u_\mu] \in \{1, 2\}^\mu$, where $\mathcal{S}_p \triangleq \{S_p \subseteq [1 : n] : \ |S_p| = \mu \leq n\}$. That is, S_p represents the set of positions noiselessly tapped by the wiretapper and $\mathbf{u}$ represents its sequence of decisions to observe *either the first or the second user* codeword symbols. We define the fraction of the tapped symbols by the wiretapper as

$$\alpha = \frac{\mu}{n}, \qquad 0 \leq \alpha \leq 1. \tag{6.1}$$

Let $S_p(k)$ and $\mathbf{u}(k)$ denote the kth elements of the subset S_p and the sequence $\mathbf{u}$, where $k = 1, 2, \cdots, \mu$. Let $\mathcal{S}$ be the set of all possible strategies for the wiretapper, where $\mathcal{S}$ is defined as

$$\mathcal{S} \triangleq \left\{ (S_p(k), \mathbf{u}(k)) : S_p \in \mathcal{S}_p,\ \mathbf{u} \in \{1,2\}^{\mu},\ k = 1, 2, \cdots, \mu \right\}. \tag{6.2}$$

For $S \in \mathcal{S}$, the wiretapper observes $\mathbf{Z}_S^n = [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$, where

$$Z_{S,i} = \begin{cases} X_{j,i}, & (i,j) \in S \\ ?, & (i,j) \notin S, \end{cases} \tag{6.3}$$

and the alphabet $\mathcal{Z} \triangleq \{\mathcal{X}_1 \cup \mathcal{X}_2\} \cup \{?\}$.

6.2.1.2 Model 2

The model is described in Fig. 6.1, when the switch is on position 2. The wiretapper chooses the subset $S \in \mathcal{S}$, where we redfine the set $\mathcal{S}$ as

$$\mathcal{S} \triangleq \{S \subseteq [1:n] :\ |S| = \mu \leq n\}. \tag{6.4}$$

The wiretapper then observes $\mathbf{Z}_S^n = [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$, where

$$Z_{S,i} = \begin{cases} X_{1,i} + X_{2,i}, & i \in S \\ ?, , & i \notin S, \end{cases} \tag{6.5}$$

and $\mathcal{Z} \triangleq \{\mathcal{X}_1 + \mathcal{X}_2\} \cup \{?\}$. That is, the wiretapper observes noiseless *superposition of the two users codeword symbols* in the positions of the subset S, and erasures otherwise. The ratio α is defined as in (6.1). Note that in the definition of the set $\mathcal{Z}$, we consider natural addition over the alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, i.e., $\mathcal{X}_1 + \mathcal{X}_2 \triangleq \{x_1 + x_2 : \ x_1 \in \mathcal{X}_1, \ x_2 \in \mathcal{X}_2\}$.

6.2.1.3 Model 3

The model is described in Fig. 6.1, when the switch is on position 3. The wiretapper chooses the subset $S \in \mathcal{S}$, with $\mathcal{S}$ defined as in (6.4), and observes $\mathbf{Z}_S^n = [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$, where

$$Z_{S,i} = \begin{cases} \{X_{1,i}, X_{2,i}\}, & i \in S \\ ?, , & i \notin S, \end{cases} \tag{6.6}$$

and $\mathcal{Z} \triangleq \{\mathcal{X}_1 \times \mathcal{X}_2\} \cup \{?\}$. That is, the wiretapper observes *the transmitted codeword symbols of both users* in the positions of the subset S, and erasures otherwise.

Next, we present a generalized multiple access wiretap channel model which extends the strongest attack model in Section 6.2.1.3 to the case when the wiretapper sees noisy observations, instead of erasures, outside the subset it chooses.

6.2.2 The Generalized Multiple Access Wiretap Channel

Consider the channel model in Fig. 6.2. The main channel in this model is identical to the main channel in Section 6.2.1. The wiretapper however chooses the subset $S \in \mathcal{S}$, with $\mathcal{S}$ defined as in (6.4), and observes $\mathbf{Z}_S^n = [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$,

where

$$Z_{S,i} = \begin{cases} \{X_{1,i}, X_{2,i}\}, & i \in S \\ V_i, & i \notin S. \end{cases} \tag{6.7}$$

$\mathbf{V}^n = [V_1, V_2, \cdots, V_n] \in \mathcal{V}^n$ is the n-letter output of the discrete memoryless multiple access channel $p_{V|X_1X_2}$, $\mathcal{V}$ is a finite alphabet, and $\mathcal{Z} \triangleq \{\mathcal{X}_1 \times \mathcal{X}_2\} \cup \mathcal{V}$.

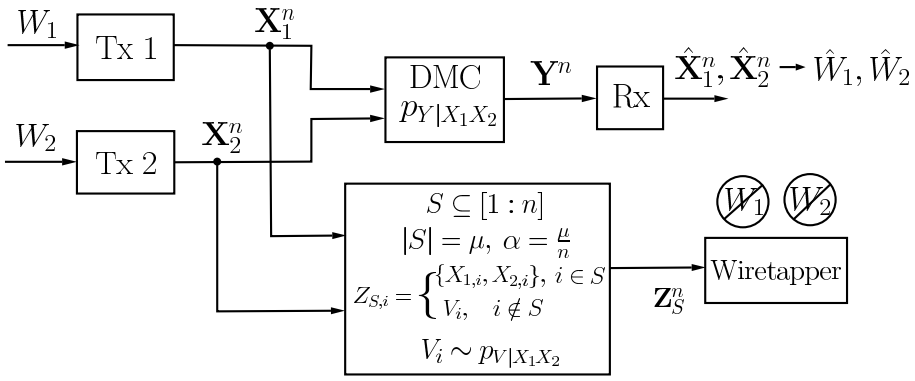


Fig. 6.2. The generalized two-user multiple access wiretap channel.

For the channel models described in Sections 6.2.1 and 6.2.2, an $(n, 2^{nR_1}, 2^{nR_2})$ channel code $\mathcal{C}_n \triangleq \{\mathcal{C}_{1,n}, \mathcal{C}_{2,n}\}$ consists of two message sets $\mathcal{W}_1 = [1 : 2^{nR_1}]$, $\mathcal{W}_2 = [1 : 2^{nR_2}]$; two stochastic encoders $P^{(\mathcal{C}_{1,n})}_{\mathbf{X}_1^n|W_1}$, $P^{(\mathcal{C}_{2,n})}_{\mathbf{X}_2^n|W_2}$, and a decoder at the receiver. (R_1, R_2) is an achievable strong secrecy rate pair if there exists a sequence of $(n, 2^{nR_1}, 2^{nR_2})$ codes, $\{\mathcal{C}_n\}_{n\geq 1}$, such that

$$\lim_{n\to\infty} \mathbb{P}^{(\mathcal{C}_n)} \left(\bigcup_{j=1,2} (\hat{W}_j \neq W_j) \right) = 0, \tag{6.8}$$

$$\text{and} \quad \lim_{n\to\infty} \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n)} \left(W_1, W_2; \mathbf{Z}_S^n \right) = 0, \tag{6.9}$$

where $\mathbb{P}^{(\mathcal{C}_n)}$ and $I^{(\mathcal{C}_n)}$ are the probability and the mutual information with respect to the joint distribution that corresponds to the code $\mathcal{C}_n$. Strong secrecy capacity region for the channel is the supremum of all achievable strong secrecy rate pairs (R_1, R_2).

6.3 Main Results

We first present achievable strong secrecy rate regions for the two-user multiple access wiretap channel II with a discrete memoryless main channel, under the attack models for the wiretapper described in Sections 6.2.1.1 and 6.2.1.2.

Theorem 6. For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the multiple access wiretap channel II in Fig. 6.1 under the wiretapper model 1, $\mathcal{R}^{(1)}(\alpha)$, is given by the convex hull of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y|U_2) - \alpha I(U_1; X_1), \tag{6.10}$$

$$R_2 \leq I(U_2; Y|U_1) - \alpha I(U_2; X_2), \tag{6.11}$$

$$R_1 + R_2 \leq I(U_1, U_2; Y) - \alpha I(U_1, U_2; X_1, X_2), \tag{6.12}$$

for some distribution $p_{U_1X_1}p_{U_2X_2}$ which satisfies the Markov chains $U_1 - X_1 - Y$ and $U_2 - X_2 - Y$.

Remark 9. The achievable strong secrecy rate region for the wiretapper model 1 in Theorem 6 is identical to the achievable region for the more capable wiretapper in model 3, see Corollary 5. When the wiretapper has the ability of choosing to observe either symbol in every tapped position, each user ought to design their transmission according

to the worst case scenario in which the wiretapper decides to observe only its symbols in all the positions it taps. This results in an achievable rate region for the wiretapper model 1 as when the wiretapper observes both users symbols in each position it taps.

Theorem 7. For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the multiple access wiretap channel II in Fig. 6.1 under the wiretapper model 2, $\mathcal{R}^{(2)}(\alpha)$, is given by the convex hull of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y|U_2) - \alpha I(U_1; X_1 + X_2), \tag{6.13}$$

$$R_2 \leq I(U_2; Y|U_1) - \alpha I(U_2; X_1 + X_2), \tag{6.14}$$

$$R_1 + R_2 \leq I(U_1, U_2; Y) - \alpha I(U_1, U_2; X_1 + X_2), \tag{6.15}$$

for some distribution $p_{U_1X_1}p_{U_2X_2}$ which satisfies the Markov chains $U_1 - X_1 - Y$ and $U_2 - X_2 - Y$.

Remark 10. The achievable strong secrecy rate region for the wiretapper models 1 and 3 is included in the achievable region for the wiretapper model 2, i.e., $\mathcal{R}^{(1)}(\alpha) \subseteq \mathcal{R}^{(2)}(\alpha)$. This follows due to the Markov chains $U_1 - X_1 - (X_1 + X_2)$; $U_2 - X_2 - (X_1 + X_2)$, and $(U_1, U_2) - (X_1, X_2) - (X_1 + X_2)$. By data processing inequality, we have

$$I(U_j; X_j) \geq I(U_j; X_1 + X_2), \qquad j = 1, 2, \tag{6.16}$$

$$I(U_1, U_2; X_1, X_2) \geq I(U_1, U_2; X_1 + X_2). \tag{6.17}$$

Next, we present achievable strong secrecy rate regions for the generalized multiple access wiretap channel in Fig. 6.2.

Theorem 8. For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the generalized multiple access wiretap channel in Fig. 6.2, $\mathcal{R}(\alpha)$, is given by the convex hull of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y|U_2) - I(U_1; V) - \alpha I(U_1; X_1|V), \tag{6.18}$$

$$R_2 \leq I(U_2; Y|U_1) - I(U_2; V) - \alpha I(U_2; X_2|V), \tag{6.19}$$

$$R_1 + R_2 \leq I(U_1, U_2; Y) - I(U_1, U_2; V) - \alpha I(U_1, U_2; X_1, X_2|V), \tag{6.20}$$

for some distribution $p_{U_1X_1}p_{U_2X_2}$ which satisfies the Markov chains $U_1 - X_1 - (Y, V)$ and $U_2 - X_2 - (Y, V)$.

Corollary 5. For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the multiple access wiretap channel II in Section 6.2.1.3, i.e., in Fig. 6.1 under the wiretapper model 3, $\mathcal{R}^{(3)}(\alpha)$, is given by the convex hull of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y|U_2) - \alpha I(U_1; X_1), \tag{6.21}$$

$$R_2 \leq I(U_2; Y|U_1) - \alpha I(U_2; X_2), \tag{6.22}$$

$$R_1 + R_2 \leq I(U_1, U_2; Y) - \alpha I(U_1, U_2; X_1, X_2), \tag{6.23}$$

for some distribution $p_{U_1X_1}p_{U_2X_2}$ which satisfies the Markov chains $U_1 - X_1 - Y$ and $U_2 - X_2 - Y$.

Corollary 5 follows directly from Theorem 8 by setting $V = \text{const.}$, i.e., the channel $p_{V|X_1X_2}$ is an erasure channel with erasure probability one. The proofs for Theorems 6, 7, and 8, are provided in Sections 6.4 and 6.5.

Remark 11. By setting the size of the subset S to zero, i.e., $\alpha = 0$, in Theorem 8, we obtain the achievable strong secrecy rate region in [125, Theorem 1] for the two user multiple access wiretap channel. The same region was derived under a weak secrecy criterion in [107, 113].

6.4 Proof of Theorem 6

The achievability proof for Theorem 6 follows the same key steps in Section 5.4, with the need of extending the technique to address the multi-terminal setting as will be explained shortly. In particular, we first assume the availability of common randomness at all terminals of the original channel model. We then define a dual *multi-terminal* secret key agreement problem in the source model, which introduces a set of random variables similar to those introduced by the original problem with the assumed common randomness. We then solve for rate conditions which result in the induced joint distributions from the two models to be almost identical in the total variation distance sense. We also provide rate conditions which satisfy certain reliability and secrecy (independence) conditions in the source model. Next, we use the closeness of the induced joint distributions to show that, under the same rate conditions, the desired reliability and secrecy properties in the original channel model are satisfied. Finally, we eliminate the common randomness from the channel model by conditioning on a certain instance of that randomness.

The outline of achievability is hence threefold: (i) Reliability of the keys in the dual source model, (ii) Security of the keys in the dual source model, and (iii) Closeness of the induced joint distributions. Reliability of the keys follows from Slepian-Wolf source coding theorem for multiple sources [30, Theorem 10.3]. Closeness of joint distributions, and converting the reliability and security conditions from the dual model to the original problem, are ensured by deriving an *exponential* convergence rate for the average total variation distance between the two distributions. This is done using a rather straightforward generalization of Lemma 7 in Chapter 5.

The main challenge in the proof lies in ensuring security for the keys in the dual source model, which requires *doubly-exponential* convergence rate for the probability of the *two keys* being uniform and independent from the wiretapper's observation, in the Kullback-Leibler divergence sense. This is established by adapting the lemma derived for the single source case, Lemma 8 in Chapter 5, so that we derive the corner points of the rate region, for the two keys, that satisfies the doubly-exponential convergence. Time sharing between these corner points hence results in the desired rate region.

Let us first fix the distribution $p_{U_1X_1}p_{U_2X_2} = p_{U_1}p_{U_2}p_{X_1|U_1}p_{X_2|U_2}$. Let $p_{Y|U_1U_2}$ be the distribution resulting from concatenating the discrete memoryless channels $p_{Y|X_1X_2}$ and $p_{X_1X_2|U_1U_2} = p_{X_1|U_1}p_{X_2|U_2}$, where $p_{Y|X_1X_2}$ is the main channel transition probability distribution for the model in Section 6.2.1. That is,

$$p_{Y|U_1U_2}(y|u_1,u_2) = \sum_{x_1,x_2 \in \mathcal{X}_1 \times \mathcal{X}_2} p_{X_1|U_1}(x_1|u_1)\, p_{X_2|U_2}(x_2|u_2)\, p_{Y|X_1X_2}(y|x_1,x_2). \quad (6.24)$$

As in Section 5.4, we describe the following two protocols.

Protocol A: This protocol describes a multi-terminal secret key agreement problem in the source model as shown in Fig. 6.3. Let $\mathbf{U}_1^n, \mathbf{U}_2^n, \mathbf{Y}^n$ be i.i.d. sequences according to $p_{U_1}p_{U_2}p_{Y|U_1U_2}$. Source encoder j observes $\mathbf{U}_j$, $j = 1, 2$. The sequence $\mathbf{U}_j$ is randomly and independently binned into the two indices $W_j = \mathcal{B}_1^{(j)}(\mathbf{X}_j)$, $F_j = \mathcal{B}_2^{(j)}(\mathbf{X}_j)$; $\mathcal{B}_1^{(j)}$, $\mathcal{B}_2^{(j)}$ are independent and uniform over $[1 : 2^{nR_j}]$, $[1 : 2^{n\tilde{R}_j}]$, respectively. $F_j, j = 1, 2$, represent the public messages transmitted noiselessly to the common decoder and perfectly accessed by the wiretapper and $W_j, j = 1, 2$, represent the independent confidential keys generated at the two encoders. The decoder observes the i.i.d. sequence $\mathbf{Y}$ and the public messages F_1, F_2, and outputs the estimates $\hat{\mathbf{U}}_1$, $\hat{\mathbf{U}}_2$, $\hat{W}_1$, $\hat{W}_2$.

Let $\mathcal{S}$ and $\mathbf{Z}_S$, for all $S \in \mathcal{S}$, be defined as in (6.2) and (6.3). The wiretapper chooses the strategy $S \in \mathcal{S}$ whose realization is unknown to the legitimate terminals. The wiretapper can thus be represented by the source $\mathbf{Z}_S \triangleq \{\mathcal{Z}, p_{\mathbf{Z}_S}, S \in \mathcal{S}\}$ whose distribution is only known to belong to the finite class $\{p_{\mathbf{Z}_S}\}_{S\in\mathcal{S}}$; the cardinality of the set $\mathcal{S}$ of all possible wiretapper's strategies for the attack model 1 is upper bounded as

$$|\mathcal{S}| = \binom{n}{\mu} \times 2^{\mu} = \binom{n}{\alpha n} \times 2^{\alpha n} < 2^n \times 2^{\alpha n} = 2^{(1+\alpha)n}. \tag{6.25}$$

Protocol A hence introduces the random variables $W_{[1:2]}$, $F_{[1:2]}$, $\mathbf{U}_{[1:2]}$, $\mathbf{Y}$, $\mathbf{Z}_S$, $\hat{\mathbf{U}}_{[1:2]}$, $\hat{W}_{[1:2]}$. The induced distribution over these variables is given by

$$\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}} = p_{\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\tilde{P}_{W_{[1:2]}F_{[1:2]}|\mathbf{U}_{[1:2]}}\tilde{P}_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}} \tag{6.26}$$

$$= p_{\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\tilde{P}_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}}\mathbb{1}\left\{\mathcal{B}_1^{(j)}(\mathbf{U}_j) = W_j, \mathcal{B}_2^{(j)}(\mathbf{U}_j) = F_j, \forall j = 1, 2\right\} \tag{6.27}$$

$$= \tilde{P}_{W_{[1:2]}F_{[1:2]}}\tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}}\, p_{\mathbf{Y}\mathbf{Z}_S|\mathbf{U}_{[1:2]}}\tilde{P}_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}}. \tag{6.28}$$

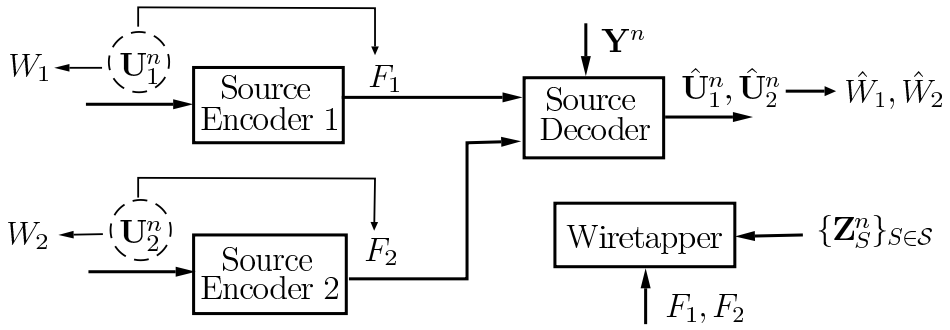


Fig. 6.3. Protocol A: Multi-terminal secret key agreement problem in the source model.

Protocol B: This protocol is described as the original channel model in Section 6.2.1.1, with assuming the availability of common randomness F_1, F_2, at all terminals. F_1, F_2, are independent, uniform over $[1 : 2^{n\tilde{R}_1}]$, $[1 : 2^{n\tilde{R}_2}]$, and independent from all other variables. We utilize here the encoders and decoder in (6.28):

$$P_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}} = \tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}}, \quad \text{and} \quad P_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}} = \tilde{P}_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}}. \tag{6.29}$$

The induced joint distribution for protocol B is thus given by

$$P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}} = p^U_{W_{[1:2]}} p^U_{F_{[1:2]}} \tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}} p_{\mathbf{Y}\mathbf{Z}_S|\mathbf{U}_{[1:2]}} \tilde{P}_{\hat{\mathbf{U}}_{[1:2]}|\mathbf{Y}F_{[1:2]}}. \tag{6.30}$$

Remark 12. As in Section 5.4, we have ignored the $\hat{W}$ variables from the joint distributions in (6.28), (6.30), as we will introduce them later as deterministic functions of the $\hat{\mathbf{U}}$ random vectors, after fixing the binning functions.

Remark 13. Notice that $\tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}}$ factorizes as $\tilde{P}_{\mathbf{U}_1|W_1F_1}\tilde{P}_{\mathbf{U}_2|W_2F_2}$. That is, the common randomness F_i available at the jth transmitter, $i, j = 1, 2, i \neq j$, is not utilized to generate $\mathbf{U}_j$. The common randomness $F_i, i = 1, 2$, represents the realization of

transmitter i's codebook, which is known at all terminals. However, the transmitted codeword at one transmitter does not depend on the codebook of the other transmitter.

Before continuing with the proof, we state the following lemmas.

6.4.1 Useful Lemmas

By comparing the joint distributions for protocols A and B in (6.28) and (6.30), we find that they only differ in the distribution for $W_{[1:2]}$, $F_{[1:2]}$. In particular, $W_{[1:2]}$ and $F_{[1:2]}$ are independent and uniform in protocol B, while their distribution in protocol A is determined by the random binning of $\mathbf{U}_1$, $\mathbf{U}_2$. The following lemma generalizes Lemma 7 in Chapter 5, in order to provide conditions on the binning rates such that the random binning of $\mathbf{U}_1$, $\mathbf{U}_2$, results in a distribution for the bins that is close, in the total variation distance, to independent uniform distributions. Once again, the *exponential* convergence rate provided by the lemma is needed for converting the secrecy (independence) condition, established for the source model in protocol A, to the original channel model in protocol B.

Lemma 9. Let $X_1 \triangleq \{\mathcal{X}_1, p_{X_1}\}$ and $X_2 \triangleq \{\mathcal{X}_2, p_{X_2}\}$ be two independent sources. The source $X_j, j = 1, 2$, is randomly binned into the two indices $W_j = \mathcal{B}_1^{(j)}(X_j)$ and $F_j = \mathcal{B}_2^{(j)}(X_j)$, where $\mathcal{B}_1^{(j)}$ and $\mathcal{B}_2^{(j)}$ are independent and uniformly distributed over $[1 : \tilde{W}_j]$ and $[1 : \tilde{F}_j]$. Let $\mathcal{B} \triangleq \left\{\mathcal{B}_1^{(j)}(x_j), \mathcal{B}_2^{(j)}(x_j) : x_j \in \mathcal{X}_j, j = 1, 2\right\}$, and for $\gamma_j > 0, j = 1, 2$, define

$$\mathcal{D}_{\gamma_j} \triangleq \left\{ x_j \in \mathcal{X}_j : \log \frac{1}{p_{X_j}(x_j)} > \gamma_j \right\}. \tag{6.31}$$

Then, we have

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(P_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right)\right) \leq \sum_{j=1}^{2}\left(\mathbb{P}_{P_{X_j}}\left(X_j \notin \mathcal{D}_{\gamma_j}\right) + \frac{1}{2}\sqrt{\tilde{W}_j\tilde{F}_j 2^{-\gamma_j}}\right), \quad (6.32)$$

where P is the induced distribution over $W_{[1:2]}$ and $F_{[1:2]}$.

Proof: Lemma 9 is a generalization of Lemma 7. In particular, using the triangle inequality,

$$\mathbb{V}\left(P_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right) = \mathbb{V}\left(P_{W_1F_1}P_{W_2F_2}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right) \quad (6.33)$$

$$\leq \mathbb{V}\left(P_{W_1F_1}P_{W_2F_2}, p^U_{W_1}p^U_{F_1}P_{W_2F_2}\right) + \mathbb{V}\left(p^U_{W_1}p^U_{F_1}P_{W_2F_2}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right) \quad (6.34)$$

$$= \sum_{j=1,2}\mathbb{V}\left(P_{W_jF_j}, p^U_{W_j}p^U_{F_j}\right), \quad (6.35)$$

where (6.33) follows since X_1 and X_2 are independent, and hence $\{W_1, F_1\}$ and $\{W_2, F_2\}$ are independent as well. Using Lemma 7, we have, for $j = 1, 2$,

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(P_{W_jF_j}, p^U_{W_j}p^U_{F_j}\right)\right) \leq \mathbb{P}_{P_{X_j}}\left(X_j \notin \mathcal{D}_{\gamma_j}\right) + \frac{1}{2}\sqrt{\tilde{W}_j\tilde{F}_j 2^{-\gamma_j}}, \quad (6.36)$$

which completes the proof for Lemma 9. ■

Lemma 10 below generalizes Lemma 8 in Chapter 5. In particular, the lemma provides conditions on the binning rates required for a *doubly-exponential* convergence rate for the probability of $W_{[1:2]}$ and $F_{[1:2]}$ being independent from one another, uniform, and independent from $\mathbf{Z}_S$.

Lemma 10. Let $X_1 \triangleq \{\mathcal{X}_1, p_{X_1}\}$ and $X_2 \triangleq \{\mathcal{X}_2, p_{X_2}\}$ be two sources, both are correlated with the source $\{Z_S\} \triangleq \left\{\mathcal{Z}, p_{Z_S}\right\}, S \in \mathcal{S}$. The alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Z}$, and $\mathcal{S}$, are finite. For $j = 1, 2$, the source X_j is randomly binned into the two indices W_j and F_j as in Lemma 9. For $\gamma_j, \gamma_{ij} > 0, i, j = 1, 2, i \neq j$, and for any $S \in \mathcal{S}$, define

$$\mathcal{D}_j^S \triangleq \left\{(x_{[1:2]}, z) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Z} : (x_j, z) \in \mathcal{D}_{\gamma_j}^S, (x_{[1:2]}, z) \in \mathcal{D}_{\gamma_{ij}}^S\right\}, \tag{6.37}$$

$$\text{where} \quad \mathcal{D}_{\gamma_j}^S \triangleq \left\{(x_j, z) \in \mathcal{X}_j \times \mathcal{Z} : \log \frac{1}{p_{X_j|Z_S}(x_j|z)} > \gamma_j\right\}, \tag{6.38}$$

$$\text{and} \quad \mathcal{D}_{\gamma_{ij}}^S \triangleq \left\{(x_{[1:2]}, z) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Z} : \log \frac{1}{p_{X_i|X_j Z_S}(x_i|x_j, z)} > \gamma_{ij}\right\}. \tag{6.39}$$

If there exists a $\delta \in (0, \frac{1}{2})$ such that for $j = 1, 2$, and for all $S \in \mathcal{S}$, we have

$$\mathbb{P}_{p_{X_{[1:2]}Z_S}}\left((X_{[1:2]}, Z_S) \in \mathcal{D}_j^S\right) \geq 1 - \delta^2, \tag{6.40}$$

then, we have, for every $\epsilon \in [0, 1]$, that

$$\begin{aligned}\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}(P_{W_{[1:2]}F_{[1:2]}Z_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{Z_S}) \geq 2\tilde{\epsilon}\right) \\ \leq |\mathcal{S}||\mathcal{Z}|\min_{i,j=1,2,i\neq j}\left\{\exp\left(\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_j}}{3\tilde{W}_j\tilde{F}_j}\right)\right) + \exp\left(\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_{ij}}}{3\tilde{W}_i\tilde{F}_i}\right)\right)\right\},\end{aligned} \tag{6.41}$$

where P is the induced distribution over $W_{[1:2]}$ and $F_{[1:2]}$,

$$\tilde{\epsilon} = \max_{j=1,2}\left\{\epsilon + (\delta + \delta^2)\log(\tilde{W}_j\tilde{F}_j) + H_b(\delta^2)\right\}, \tag{6.42}$$

and H_b is the binary entropy function.

Proof: The proof is provided in Appendix H. ■

6.4.2 Proof

We first apply Lemma 9 to the source model in protocol A. In Lemma 9, set $X_j = \mathbf{U}_j$, $\tilde{W}_j = 2^{nR_j}$, and $\tilde{F}_j = 2^{n\tilde{R}_j}$, for $j = 1, 2$; $\mathbf{U}_j, \tilde{W}_j, \tilde{F}_j$ are defined as in protocol A. Let $\mathcal{D}_{\gamma_j}$ be defined as in (6.31) with $X_j = \mathbf{U}_j$ for $j = 1, 2$. For $\epsilon_j > 0, j = 1, 2$, choose $\gamma_j = n(1 - \epsilon_j)H(U_j)$. Without loss of generality, assume that for all $\mathbf{u}_j, j = 1, 2$, $p_{\mathbf{U}_j}(\mathbf{u}_j) > 0$. Using Hoeffding's inequality,

$$\mathbb{P}_{p_{\mathbf{U}_j}}\left(\mathbf{U}_j \notin \mathcal{D}_{\gamma_j}\right) = \mathbb{P}\left(\log \frac{1}{p_{\mathbf{U}_j}(\mathbf{U}_j)} \leq \gamma_j\right) \leq \exp(-\beta_j n), \tag{6.43}$$

where $\beta_j > 0$. By substituting the choices for $\tilde{W}_j, \tilde{F}_j, \gamma_j$, and (6.43) in (6.32), as long as

$$R_1 + \tilde{R}_1 < (1 - \epsilon_1)H(U_1) \tag{6.44}$$

$$R_2 + \tilde{R}_2 < (1 - \epsilon_2)H(U_2), \tag{6.45}$$

there exists a $\beta > 0$ such that

$$\begin{aligned}&\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}\right)\right)\\ &\quad = \mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right)\right) \leq 4\exp(-\beta n).\end{aligned} \tag{6.46}$$

Next, we establish a reliability condition for the source model in protocol A. We utilize a Slepian-Wolf decoder [108], which implies that [30, Theorem 10.3]

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{U}}_{[1:2]} \neq \mathbf{U}_{[1:2]}) \right) = 0, \tag{6.47}$$

as long as

$$\tilde{R}_1 \geq H(U_1|U_2, Y), \tag{6.48}$$

$$\tilde{R}_2 \geq H(U_2|U_1, Y), \tag{6.49}$$

$$\tilde{R}_1 + \tilde{R}_2 \geq H(U_1, U_2|Y). \tag{6.50}$$

Using (6.47) and [126, Lemma 1], which is a variation on the Slepian-Wolf source coding theorem, we have, for all $S \in \mathcal{S}$,

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(\tilde{P}_{W_{[1:2]} F_{[1:2]} \mathbf{U}_{[1:2]} \mathbf{Y} \mathbf{Z}_S \hat{\mathbf{U}}_{[1:2]}}, \tilde{P}_{W_{[1:2]} F_{[1:2]} \mathbf{U}_{[1:2]} \mathbf{Y} \mathbf{Z}_S} \mathbb{1}\{\hat{\mathbf{U}}_{[1:2]} = \mathbf{U}_{[1:2]}\} \right) \right) \\ &\quad = \lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{P}_{\tilde{P}}(\hat{\mathbf{U}}_{[1:2]} \neq \mathbf{U}_{[1:2]}) \right) = 0. \end{aligned} \tag{6.51}$$

Next, we use Lemma 10 to establish the secrecy condition for the source model in protocol A. In Lemma 10, for $j = 1, 2$, set $X_j = \mathbf{U}_j$, $\tilde{W}_j = 2^{nR_j}$, $\tilde{F}_j = 2^{n\tilde{R}_j}$, $Z_S = \mathbf{Z}_S$, for all $S \in \mathcal{S}$, where $\mathbf{U}_j, \mathcal{S}, \mathbf{Z}_S$ are defined as in protocol A. In addition, let $\mathcal{D}_j^S, \mathcal{D}_{\gamma_j}^S$, and $\mathcal{D}_{\gamma_{ij}}^S$ be defined as in (6.37)-(6.39), with $X_j = \mathbf{U}_j$ and $Z_S = \mathbf{Z}_S$.

For $S \in \mathcal{S}$, define $\overline{S}_j \triangleq \{k : (k, j) \in S\}$. That is, $\overline{S}_j$ is the set of positions in which the wiretapper observes the jth transmitter's symbols. For $j = 1, 2$, let $|\overline{S}_j| = \mu_j$,

and hence $\mu_1 + \mu_2 = \mu$. Thus, we have

$$H(\mathbf{U}_1|\mathbf{Z}_S) = H(\mathbf{U}_1|\mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) = H(\mathbf{U}_{1,\overline{S}_1}, \mathbf{U}_{1,\overline{S}_1^c}|\mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) \tag{6.52}$$

$$= H(\mathbf{U}_{1,\overline{S}_1}|\mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) + H(\mathbf{U}_{1,\overline{S}_1^c}|\mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}, \mathbf{U}_{1,\overline{S}_1}) \tag{6.53}$$

$$= H(\mathbf{U}_{1,\overline{S}_1}|\mathbf{X}_{1,\overline{S}_1}) + H(\mathbf{U}_{1,\overline{S}_1^c}) = \mu_1 H(U_1|X_1) + (n-\mu_1)H(U_1) \tag{6.54}$$

$$H(\mathbf{U}_2|\mathbf{Z}_S) = \mu_2 H(U_2|X_2) + (n-\mu_2)H(U_2) \tag{6.55}$$

$$H(\mathbf{U}_1|\mathbf{U}_2,\mathbf{Z}_S) = H(\mathbf{U}_1|\mathbf{U}_2, \mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) = H(\mathbf{U}_{1,\overline{S}_1}, \mathbf{U}_{1,\overline{S}_1^c}|\mathbf{U}_2, \mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) \tag{6.56}$$

$$= H(\mathbf{U}_{1,\overline{S}_1}|\mathbf{U}_2, \mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}) + H(\mathbf{U}_{1,\overline{S}_1^c}|\mathbf{U}_2, \mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}, \mathbf{U}_{1,\overline{S}_1}) \tag{6.57}$$

$$= H(\mathbf{U}_{1,\overline{S}_1}|\mathbf{X}_{1,\overline{S}_1}) + H(\mathbf{U}_{1,\overline{S}_1^c}) = \mu_1 H(U_1|X_1) + (n-\mu_1)H(U_1) \tag{6.58}$$

$$H(\mathbf{U}_2|\mathbf{U}_1,\mathbf{Z}_S) = \mu_2 H(U_2|X_2) + (n-\mu_2)H(U_2), \tag{6.59}$$

where (6.54) follows since $\{\mathbf{U}_{1,\overline{S}_1}, \mathbf{X}_{1,\overline{S}_1}\}$ are independent from $\mathbf{X}_{2,\overline{S}_2}$, and $\mathbf{U}_{1,\overline{S}_1^c}$ is independent from $\{\mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}, \mathbf{U}_{1,\overline{S}_1}\}$, since $\mathbf{U}_1$ is an i.i.d. sequence and $p_{X_1|U_1}$ is a discrete memoryless channel. Similarly, (6.58) follows since $\{\mathbf{U}_{1,\overline{S}_1}, \mathbf{X}_{1,\overline{S}_1}\}$ are independent from $\{\mathbf{U}_2, \mathbf{X}_{2,\overline{S}_2}\}$, and $\mathbf{U}_{1,\overline{S}_1^c}$ is independent from $\{\mathbf{U}_2, \mathbf{X}_{1,\overline{S}_1}, \mathbf{X}_{2,\overline{S}_2}, \mathbf{U}_{1,\overline{S}_1}\}$.

In addition, for the tuples $(\mathbf{x}_{[1,2]}, \mathbf{z})$ with $p_{\mathbf{X}_j|\mathbf{Z}_S}(\mathbf{x}_j|\mathbf{z}) > 0$ and $p_{\mathbf{X}_i|\mathbf{X}_j\mathbf{Z}_S}(\mathbf{x}_i|\mathbf{x}_j, \mathbf{z}) > 0$, where $i, j = 1, 2, i \neq j$, we have, for all $S \in \mathcal{S}$, that

$$p_{\mathbf{U}_j|\mathbf{Z}_S}(\mathbf{u}_j|\mathbf{z}) = p(\mathbf{u}_{j,\overline{S}_j}, \mathbf{u}_{j,\overline{S}_j^c}|\mathbf{x}_{j,\overline{S}_j}, \mathbf{x}_{i,\overline{S}_i}) = p(\mathbf{u}_{j,\overline{S}_j}|\mathbf{x}_{j,\overline{S}_j}, \mathbf{x}_{i,\overline{S}_i})\, p(\mathbf{u}_{j,\overline{S}_j^c}|\mathbf{u}_{j,\overline{S}_j}, \mathbf{x}_{j,\overline{S}_j}, \mathbf{x}_{i,\overline{S}_i})$$

$$= p(\mathbf{u}_{j,\overline{S}_j}|\mathbf{x}_{j,\overline{S}_j})\, p(\mathbf{u}_{j,\overline{S}_j^c}) = \prod_{k\in\overline{S}_j} p(u_{j,k}|x_{j,k}) \prod_{k\in\overline{S}_j^c} p(u_{j,k}), \tag{6.60}$$

$$p_{\mathbf{U}_i|\mathbf{U}_j\mathbf{Z}_S}(\mathbf{u}_i|\mathbf{u}_j,\mathbf{z}) = p(\mathbf{u}_{i,\overline{S}_i}|\mathbf{x}_{i,\overline{S}_i})\; p(\mathbf{u}_{i,\overline{S}_i^c}) = \prod_{k\in\overline{S}_i} p(u_{i,k}|x_{i,k}) \prod_{k\in\overline{S}_i^c} p(u_{i,k}). \tag{6.61}$$

For $i,j=1,2, i\neq j$, and $\tilde{\epsilon}_j>0$, let

$$\gamma_j = (1-\tilde{\epsilon}_j)\min_{S\in\mathcal{S}} H(\mathbf{U}_j|\mathbf{Z}_S) = (1-\tilde{\epsilon}_j)[\mu H(U_j|X_j) + (n-\mu)H(U_j)], \tag{6.62}$$

$$\gamma_{ij} = (1-\tilde{\epsilon}_j)\min_{S\in\mathcal{S}} H(\mathbf{U}_i|\mathbf{U}_j,\mathbf{Z}_S) = (1-\tilde{\epsilon}_j)[\mu H(U_i|X_i) + (n-\mu)H(U_i)], \tag{6.63}$$

where (6.62) and (6.63) follow from (6.54), (6.55), (6.58), (6.59), and the fact that $\mu_j H(U_j|X_j) + (n-\mu_j)H(U_j)$ is minimized by $\mu_j=\mu$, which occurs when $S=\{(k,j): k\in\mathcal{S}_p\}$, i.e., when the wiretapper observes the symbols of the jth transmitter in all the positions it chooses.

Using Hoeffding inequality and the definition of $\mathcal{D}_{\gamma_j}^S$ in (6.38), we have, for all $S\in\mathcal{S}$,

$$\mathbb{P}_{p_{\mathbf{U}_j\mathbf{Z}_S}}\left((\mathbf{U}_j,\mathbf{Z}_S)\notin\mathcal{D}_{\gamma_j}^S\right) = \mathbb{P}_{p_{\mathbf{U}_j\mathbf{Z}_S}}\left(\log\frac{1}{p_{\mathbf{U}_j|\mathbf{Z}_S}(\mathbf{U}_j|\mathbf{Z}_S)}\le\gamma_j\right) \tag{6.64}$$

$$= \mathbb{P}_{p_{\mathbf{U}_j\mathbf{Z}_S}}\left(\sum_{k\in\overline{S}_j}\log\frac{1}{p(U_{j,k}|X_{j,k})} + \sum_{k\in\overline{S}_j^c}\log\frac{1}{p(U_{j,k})} \le (1-\tilde{\epsilon}_j)[\mu H(U_j|X_j)+(n-\mu)H(U_j)]\right) \tag{6.65}$$

$$\le \mathbb{P}_{p_{\mathbf{U}_j\mathbf{Z}_S}}\left(\sum_{k\in\overline{S}_j}\log\frac{1}{p(U_{j,k}|X_{j,k})} + \sum_{k\in\overline{S}_j^c}\log\frac{1}{p(U_{j,k})} \le (1-\tilde{\epsilon}_j)[\mu_j H(U_j|X_j)+(n-\mu_j)H(U_j)]\right) \tag{6.66}$$

$$\le \exp(-\tilde{\beta}_j n), \tag{6.67}$$

where $\tilde{\beta}_j > 0$ for $j = 1, 2$, and (6.66) follows because, for all $S \in \mathcal{S}$,

$$\mu H(U_j|X_j) + (n-\mu)H(U_j) \leq \mu_j H(U_j|X_j) + (n-\mu_j)H(U_j). \tag{6.68}$$

Note that, for any finite γ_j, in order to compute the probability on the left hand side of (6.64), we only need to consider the tuples $(\mathbf{u}_j, \mathbf{z})$ with $p_{\mathbf{U}_j|\mathbf{Z}_S}(\mathbf{u}_j|\mathbf{z}) > 0$.

Similarly, for $i, j = 1, 2, i \neq j$ and all $S \in \mathcal{S}$, using Hoeffding's inequality, (6.61), (6.63), and the definition for $\mathcal{D}^S_{\gamma_{ij}}$ in (6.39), we have

$$\mathbb{P}_{p_{\mathbf{U}_{[1:2]}\mathbf{Z}_S}}\left((\mathbf{U}_{[1:2]}, \mathbf{Z}_S) \notin \mathcal{D}^S_{\gamma_{ij}}\right) = \mathbb{P}_{p_{\mathbf{U}_{[1:2]}\mathbf{Z}_S}}\left(\log \frac{1}{p_{\mathbf{U}_i|\mathbf{U}_j\mathbf{Z}_S}(\mathbf{U}_i|\mathbf{U}_j, \mathbf{Z}_S)} \leq \gamma_{ij}\right) \leq \exp(-\tilde{\beta}_i n). \tag{6.69}$$

Taking $\delta^2 = 2\exp(-\tilde{\beta}n)$, where $\tilde{\beta} = \min\{\tilde{\beta}_1, \tilde{\beta}_2\}$, yields

$$\mathbb{P}_{p_{\mathbf{U}_{[1:2]}\mathbf{Z}_S}}\left((\mathbf{U}_{[1:2]}, \mathbf{Z}_S) \notin \mathcal{D}^S_j\right) \leq \delta^2, \tag{6.70}$$

for $j = 1, 2$ and all $S \in \mathcal{S}$. Note that $\lim_{n\to\infty} \delta^2 = 0$, and hence, for n sufficiently large, $\delta^2 \in (0, \frac{1}{4})$. Thus, the conditions for Lemma 10 are satisfied. As in (5.35), we have

$$\lim_{n\to\infty} \tilde{\epsilon} = \epsilon + \lim_{n\to\infty} (\delta + \delta^2) \log(\tilde{W}_j \tilde{F}_j) + \lim_{n\to\infty} H_b(\delta^2) = \epsilon. \tag{6.71}$$

By substituting the choices for $\tilde{W}_j, \tilde{F}_j, \gamma_j, \gamma_{ij}$, where $i, j = 1, 2,\ i \neq j$, and

$$|\mathcal{S}||\mathcal{Z}^n| \leq \exp(n[(1+\alpha)\ln 2 + \ln(|\mathcal{X}_1| + |\mathcal{X}_2| + 1)]), \tag{6.72}$$

in (6.41), and using (6.71), we have, for every $\epsilon, \epsilon' > 0$, $\tilde{\epsilon} = \epsilon + \epsilon'$, there exist $n^* \in \mathbb{N}$ and $\kappa_\epsilon, \tilde{\kappa} > 0$ such that for all $n \geq n^*$,

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S} || p^U_{W_{[1:2]}} p^U_{F_{[1:2]}} p_{\mathbf{Z}_S}\right) \geq 2\tilde{\epsilon}\right) \leq \exp\left(-\kappa_\epsilon e^{\tilde{\kappa} n}\right), \tag{6.73}$$

as long as

$$R_1 + \tilde{R}_1 \leq (1-\tilde{\epsilon}_1)\,[\alpha H(U_1|X_1) + (1-\alpha)H(U_1)], \tag{6.74}$$

$$R_2 + \tilde{R}_2 \leq (1-\tilde{\epsilon}_2)\,[\alpha H(U_2|X_2) + (1-\alpha)H(U_2)]. \tag{6.75}$$

By applying the first Borel-Cantelli Lemma to (6.73), we get

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S} || p^U_{W_{[1:2]}} p^U_{F_{[1:2]}} p_{\mathbf{Z}_S}\right) > 0\right) = 0. \tag{6.76}$$

In addition, using Markov's inequality and (6.46), we have, for any $r > 0$, that

$$\sum_{n=1}^{\infty} \mathbb{P}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}} p^U_{F_{[1:2]}}\right) > r\right) \leq \frac{4}{r}\sum_{n=1}^{\infty} \exp(-\beta n) < \infty. \tag{6.77}$$

Using the first Borel-Cantelli lemma, it follows from (6.77) that

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}} p^U_{F_{[1:2]}}\right) > 0\right) = 0. \tag{6.78}$$

Now, we show that the reliability and secrecy conditions in (6.51) and (6.76) hold as well for the channel model in protocol B. First, as in (5.39)-(5.41), we have

$$\begin{aligned}
&\mathbb{V}\left(P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\mathbb{1}\{\hat{\mathbf{U}}_{[1:2]}=\mathbf{U}_{[1:2]}\}\right)\\
&\leq \mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, \tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\mathbb{1}\{\hat{\mathbf{U}}_{[1:2]}=\mathbf{U}_{[1:2]}\}\right)\\
&\qquad + 2\mathbb{V}\left(P_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right). \qquad (6.79)
\end{aligned}$$

Thus, using (6.46), (6.51), and (6.79), we have

$$\lim_{n\to\infty}\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\mathbb{1}\{\hat{\mathbf{U}}_{[1:2]}=\mathbf{U}_{[1:2]}\}\right)\right)=0. \qquad (6.80)$$

Second, for the secrecy condition, using the union bound, we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right)>0\right)\\
&\leq \mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right)>0\right)+\mathbb{P}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[1:2]}F_{[1:2]}}, p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}\right)>0\right). \qquad (6.81)
\end{aligned}$$

Thus, using (6.76), (6.78), and (6.81), we have

$$\lim_{n\to\infty}\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right)>0\right)=0. \qquad (6.82)$$

By applying the selection lemma to (6.80) and (6.82), there is at least one binning realization $\mathbf{b}^*=\{b_1^{*(j)}, b_2^{*(j)} : j=1,2\}$, with a corresponding joint distribution p^* for

protocol B such that

$$\lim_{n\to\infty} \mathbb{V}\left(p^*_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, p^*_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\mathbb{1}\{\hat{\mathbf{U}}_{[1:2]} = \mathbf{U}_{[1:2]}\}\right) = 0, \tag{6.83}$$

$$\text{and } \lim_{n\to\infty} \mathbb{1}\left\{\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right) > 0\right\} = 0, \tag{6.84}$$

where $W_j = b_1^{*(j)}(\mathbf{U}_j)$ and $F_j = b_2^{*(j)}(\mathbf{U}_j)$, $j = 1, 2$.

Next, we introduce the $\hat{W}$ variables to the joint distributions in (6.83). For $j = 1, 2$, $\hat{W}_j$ is a deterministic function of the random sequence $\hat{\mathbf{U}}_j$. In particular, $p^*_{\hat{W}_j|\hat{\mathbf{U}}_j}(\hat{w}_j|\hat{\mathbf{u}}_j) = \mathbb{1}\left\{\hat{w}_j = b_1^{*(j)}(\hat{\mathbf{u}})\right\}$. Using (6.83) and a similar analysis as in (5.54)-(5.60) in Chapter 5, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}_{F_{[1:2]}}\left(\mathbb{P}_{p^*}\left(\hat{W}_{[1:2]} \neq W_{[1:2]}|F_{[1:2]}\right)\right) \\
&\quad = \lim_{n\to\infty} \mathbb{V}\left(p^*_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S\hat{\mathbf{U}}_{[1:2]}}, p^*_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}\mathbf{Z}_S}\mathbb{1}\{\hat{\mathbf{U}}_{[1:2]} = \mathbf{U}_{[1:2]}\}\right) = 0.
\end{aligned} \tag{6.85}$$

Using the union bound, we also have

$$\begin{aligned}
&\mathbb{P}_{F_{[1:2]}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[1:2]}\mathbf{Z}_S|F_{[1:2]}}||p^U_{W_{[1:2]}}p^*_{\mathbf{Z}_S|F_{[1:2]}}\right) > 0\right) \\
&\quad = \mathbb{P}\left(\max_{S\in\mathcal{S}} \mathbb{D}(p^*_{W_{[1:2]}\mathbf{Z}_S|F_{[1:2]}}||p^U_{W_{[1:2]}}p^*_{\mathbf{Z}_S|F_{[1:2]}}) > 0, \text{ and } \forall S,\ p^*_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S} = p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right) \\
&\qquad\qquad + \mathbb{1}\left\{\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right) > 0\right\} \qquad (6.86) \\
&\quad = \mathbb{1}\left\{\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[1:2]}F_{[1:2]}\mathbf{Z}_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{\mathbf{Z}_S}\right) > 0\right\}, \qquad (6.87)
\end{aligned}$$

where (6.87) follows since the first term on the right hand side of (6.86) is equal to zero. Thus, using (6.84) and (6.87), we have

$$\lim_{n\to\infty} \mathbb{P}_{F_{[1:2]}} \left(\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[1:2]}\mathbf{Z}_S|F_{[1:2]}} || p^U_{W_{[1:2]}} p^*_{\mathbf{Z}_S|F_{[1:2]}} \right) > 0 \right) = 0 \tag{6.88}$$

Once again, applying the selection lemma to (6.85) and (6.88), implies that there is at least one realization $f^*_{[1:2]}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{W}_{[1:2]} \neq W_{[1:2]} | F_{[1:2]} = f^*_{[1:2]} \right) = 0, \tag{6.89}$$

$$\lim_{n\to\infty} \max_{S\in\mathcal{S}} I\left(W_{[1:2]}; \mathbf{Z}_S | F_{[1:2]} = f^*_{[1:2]} \right) = 0. \tag{6.90}$$

Let $\tilde{p}^*$ be the induced joint distribution for protocol A which corresponds to the binning realization $\mathbf{b}^*$. We identify $\{\tilde{p}^*(\mathbf{u}_j|w_j, f^*_j), p(\mathbf{x}_j|\mathbf{u}_j), j = 1, 2\}$, $\{\tilde{p}^*(\hat{\mathbf{u}}_{[1:2]}|\mathbf{y}, f^*_{[1:2]}), \{b_1^{*(j)}(\hat{\mathbf{u}}_j)\}, j = 1, 2\}$, as the encoders and the decoder for the original channel model.

By combining the rate conditions in (6.44), (6.45), (6.48)-(6.50), (6.74), and (6.75), and taking $\tilde{\epsilon}_1, \tilde{\epsilon}_2 \to 0$, we obtain the achievable strong secrecy rate region in (6.10)-(6.12). The convex hull follows by time sharing independent codes and the fact that maximizing the secrecy constraint over S in the whole block-length is upper bounded by its maximization over the individual segments of the time sharing.

6.5 Proofs of Theorems 7 and 8

The proof for Theorem 7 follows similar steps as in the proof for Theorem 6. The difference is that $\mathcal{S}$ and $\mathbf{Z}_S$, for all $S \in \mathcal{S}$, in protocol A are defined as in (6.4), (6.5).

We thus have, for $i, j = 1, 2,\ i \neq j$, and all $S \in \mathcal{S}$, that

$$H(\mathbf{U}_j|\mathbf{Z}_S) = H(\mathbf{U}_{j,S}, \mathbf{U}_{j,S^c}|\mathbf{X}_{1,S} + \mathbf{X}_{2,S}) \tag{6.91}$$

$$= H(\mathbf{U}_{j,S}|\mathbf{X}_{1,S} + \mathbf{X}_{2,S}) + H(\mathbf{U}_{j,S^c}|\mathbf{U}_{j,S}, \mathbf{X}_{1,S} + \mathbf{X}_{2,S}) \tag{6.92}$$

$$= H(\mathbf{U}_{j,S}|\mathbf{X}_{1,S} + \mathbf{X}_{2,S}) + H(\mathbf{U}_{j,S^c}) \tag{6.93}$$

$$= \mu H(U_j|X_1 + X_2) + (n - \mu)H(U_j) \tag{6.94}$$

$$H(\mathbf{U}_i|\mathbf{U}_j, \mathbf{Z}_S) = H(\mathbf{U}_{i,S}, \mathbf{U}_{i,S^c}|\mathbf{U}_j, \mathbf{X}_{1,S} + \mathbf{X}_{2,S}) \tag{6.95}$$

$$= H(\mathbf{U}_{i,S}|\mathbf{U}_j, \mathbf{X}_{1,S} + \mathbf{X}_{2,S}) + H(\mathbf{U}_{i,S^c}|\mathbf{U}_{i,S}, \mathbf{U}_j, \mathbf{X}_{1,S} + \mathbf{X}_{2,S}) \tag{6.96}$$

$$= H(\mathbf{U}_{i,S}|\mathbf{U}_{j,S}, \mathbf{X}_{1,S} + \mathbf{X}_{2,S}) + H(\mathbf{U}_{i,S^c}) \tag{6.97}$$

$$= \mu H(U_i|U_j, X_1 + X_2) + (n - \mu)H(U_i). \tag{6.98}$$

Thus, in applying Lemma 10 to the source model in protocol A, for $i, j = 1, 2,\ i \neq j$, and $\tilde{\epsilon}_j > 0$, we choose

$$\gamma_j = (1 - \tilde{\epsilon}_j) \min_{S \in \mathcal{S}} H(\mathbf{U}_j|\mathbf{Z}_S) = (1 - \tilde{\epsilon}_j)[\mu H(U_j|X_1 + X_2) + (n - \mu)H(U_j)] \tag{6.99}$$

$$\gamma_{ij} = (1 - \tilde{\epsilon}_j) \min_{S \in \mathcal{S}} H(\mathbf{U}_i|\mathbf{U}_j, \mathbf{Z}_S) = (1 - \tilde{\epsilon}_j)[\mu H(U_i|U_j, X_1 + X_2) + (n - \mu)H(U_i)]. \tag{6.100}$$

Using Hoeffding inequality, the conditions of the lemma are satisfied, and the rate conditions required for the secrecy property in (6.76) are

$$R_1 + \tilde{R}_1 \leq \alpha H(U_1|X_1 + X_2) + (1 - \alpha)H(U_1) \tag{6.101}$$

$$R_2 + \tilde{R}_2 \leq \alpha H(U_2|X_1 + X_2) + (1-\alpha)H(U_2) \tag{6.102}$$

$$R_1 + R_2 + \tilde{R}_1 + \tilde{R}_2 \leq \alpha H(U_{[1:2]}|X_1 + X_2) + (1-\alpha)H(U_{[1:2]}). \tag{6.103}$$

These conditions, combined with the rate conditions for the Slepian-Wolf decoder, which are

$$\tilde{R}_1 \geq H(U_1|U_2, Y), \qquad \tilde{R}_2 \geq H(U_2|U_1, Y), \tag{6.104}$$

$$\tilde{R}_1 + \tilde{R}_2 \geq H(U_{[1:2]}|Y), \tag{6.105}$$

and using time sharing, establish the achievability for the strong secrecy rate region in Theorem 7.

Remark 14. By setting $j = 1, i = 2$, instead of the minimum in the right hand side of (6.41), Lemma 2 results in the maximum binning rate $R_1 + \tilde{R}_1$ of the source $\mathbf{U}_1$, and the corresponding maximum conditional binning rate $R_2 + \tilde{R}_2$ for the source $\mathbf{U}_2$ given $R_1 + \tilde{R}_1$, such that the probability in the left hand side of (6.41) is vanishing. In other words, Lemma 10 provides the corner points of the binning rate region such that the probability, over the random binning of the sources, that the bins are independent, uniform, and independent from the wiretapper's observation, is vanishing.

Similarly, the proof for Theorem 8 follows similar steps as in the proof for Theorem 6. In protocol A, $\mathcal{S}$ and $\mathbf{Z}_S$ for all $S \in \mathcal{S}$ are defined as in (6.7) in Section 6.2.2. The sequences $\mathbf{U}_1, \mathbf{U}_2$ are i.i.d. and the channel $p_{V|U_{[1:2]}}$ is a discrete memoryless channel, since it results from concatenating the two discrete memoryless channels $p_{V|X_{[1:2]}}$ and

$p_{X_{[1:2]}|U_{[1:2]}}$. Thus, we have, for $i,j=1,2,\ i\neq j$, and all $S\in\mathcal{S}$,

$$H(\mathbf{U}_j|\mathbf{Z}_S) = H(\mathbf{U}_{j,S},\mathbf{U}_{j,S^c}|\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) \tag{6.106}$$

$$= H(\mathbf{U}_{j,S}|\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) + H(\mathbf{U}_{j,S^c}|\mathbf{U}_{j,S},\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) \tag{6.107}$$

$$= H(\mathbf{U}_{j,S}|\mathbf{X}_{j,S}) + H(\mathbf{U}_{j,S^c}|\mathbf{V}_{S^c}) \tag{6.108}$$

$$= \mu H(U_j|X_j) + (n-\mu)H(U_j|V) \tag{6.109}$$

$$H(\mathbf{U}_i|\mathbf{U}_j,\mathbf{Z}_S) = H(\mathbf{U}_{i,S},\mathbf{U}_{i,S^c}|\mathbf{U}_j,\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) \tag{6.110}$$

$$= H(\mathbf{U}_{i,S}|\mathbf{U}_j,\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) + H(\mathbf{U}_{i,S^c}|\mathbf{U}_{i,S},\mathbf{U}_{j,S},\mathbf{U}_{j,S^c},\mathbf{X}_{1,S},\mathbf{X}_{2,S},\mathbf{V}_{S^c}) \tag{6.111}$$

$$= H(\mathbf{U}_{i,S}|\mathbf{X}_{i,S}) + H(\mathbf{U}_{i,S^c}|\mathbf{U}_{j,S^c},\mathbf{V}_{S^c}) \tag{6.112}$$

$$= \mu H(U_i|X_i) + (n-\mu)H(U_i|U_j,V), \tag{6.113}$$

where (6.108) follows due to the Markov chains $\mathbf{U}_{j,S}-\mathbf{X}_{j,S}-(\mathbf{X}_{i,S},\mathbf{V}_{S^c})$ and $(\mathbf{U}_{j,S},\mathbf{X}_{1,S},\mathbf{X}_{2,S})-\mathbf{V}_{S^c}-\mathbf{U}_{j,S^c}$. Equation (6.112) follows from the Markov chains $\mathbf{U}_{i,S}-\mathbf{X}_{i,S}-(\mathbf{U}_j,\mathbf{X}_{j,S},\mathbf{V}_{S^c})$ and $(\mathbf{U}_{i,S},\mathbf{U}_{j,S},\mathbf{X}_{1,S},\mathbf{X}_{2,S})-(\mathbf{U}_{j,S^c},\mathbf{V}_{S^c})-\mathbf{U}_{i,S^c}$. These Markov chains follow since the sequences $\mathbf{U}_1,\mathbf{U}_2$ are i.i.d. and the channels $p_{X_1|U_1}, p_{X_2|U_2}, p_{V|U_{[1:2]}}$ are discrete memoryless.

Thus, for $i,j=1,2,\ i\neq j$, and $\tilde{\epsilon}_j>0$, by choosing

$$\gamma_j = (1-\tilde{\epsilon}_j)\min_{S\in\mathcal{S}}\ H(\mathbf{U}_j|\mathbf{Z}_S) = (1-\tilde{\epsilon}_j)[\mu H(U_j|X_j) + (n-\mu)H(U_j|V)] \tag{6.114}$$

$$\gamma_{ij} = (1-\tilde{\epsilon}_j)\min_{S\in\mathcal{S}}\ H(\mathbf{U}_i|\mathbf{U}_j,\mathbf{Z}_S) = (1-\tilde{\epsilon}_j)[\mu H(U_i|X_i) + (n-\mu)H(U_i|U_j,V)], \tag{6.115}$$

and using Hoeffding inequality, the conditions of Lemma 10 are satisfied. The rate conditions needed for the secrecy property in (6.76) are

$$R_1 + \tilde{R}_1 \leq \alpha H(U_1|X_1) + (1-\alpha)H(U_1|V) \tag{6.116}$$

$$R_2 + \tilde{R}_2 \leq \alpha H(U_2|X_2) + (1-\alpha)H(U_2|V) \tag{6.117}$$

$$R_1 + R_2 + \tilde{R}_1 + \tilde{R}_2 \leq \alpha H(U_{[1:2]}|X_{[1:2]}) + (1-\alpha)H(U_{[1:2]}|V). \tag{6.118}$$

Combining (6.116)-(6.118) with the rate conditions required for the Slepian-Wolf decoder in (6.104) and (6.105), and using time sharing, establish the achievability for the strong secrecy rate region in Theorem 8.

6.6 Conclusion

In this chapter, we have studied the extension of the wiretap channel II with a noisy main channel in Chapter 4 and the generalized wiretap channel model in Chapter 5 to the multiple access setting. For the multiple access wiretap channel II with a noisy main channel, we have proposed three attack models for the wiretapper and derived an achievable strong secrecy rate region for each. We have generalized the strongest attack model, in which the wiretapper observes the transmitted symbols of both users in the positions of the subset it chooses, to the case when the wiretapper observes the outputs of a noisy multiple access channel instead of erasures outside this subset, proposing a *generalized* multiple access wiretap model. We have derived an achievable strong secrecy rate region for this generalized model. This model generalizes the multiple access wiretap channel in [112, 113] as well to the case when the wiretapper is provided with noiseless

observations for a subset, of its choice, of the transmitted codeword symbols of both uses. The tools we have utilized for achievability extend the set of tools we have developed for the single-user scenario in Chapter 5 to a multi-user setting.

Chapter 7

Generalized Multi-receiver Wiretap Channel Models

7.1 Introduction

Among the multi-terminal extensions for Wyner's wiretap channel, various multi-receiver models have been investigated. The broadcast wiretap channel with an external wiretapper has been studied in [9,26,34]. References [9,26] have characterized the secrecy capacity for several special classes of the model such the physically degraded and semi-deterministic broadcast channels with more noisy wiretappers. The two user broadcast and interference channels with confidential messages have been introduced and studied in [67]. The three-receiver broadcast channel with common and confidential message sets has been studied in [19].

In this chapter, we study the extension of our generalized wiretap channel model, introduced in Chapter 5, into three different muli-receiver settings. We first consider the two-user broadcast wiretap channel model with a common message and an external wiretapper. In this model, the wiretapper, besides choosing a subset of the transmitted codeword symbols to noiselessly tap into, observes the remainder through a discrete memeoryless channel. Next, we introduce generalized models for the two-user broadcast channel with confidential messages, and the two-user interference channel with confidential messages. In the generalized broadcast channel with confidential messages model, each receiver, besides its noisy observations, is provided with noiseless observations for

a subset, of its choice, of the transmitted codeword. In the same spirit, for the generalized interference channel with confidential messages model, both receivers are provided with subsets of their choice of noiseless observations for the transmitted symbols of both codewords.

We derive an achievable strong secrecy rate region for each of the three proposed models. Similar to the multiple-access generalization in Chapter 6, achievability is established by solving dual *multi-terminal* secret key agreement problems in the source model, and converting the solution to the original channel models. The achievability proofs in this chapter however require extending Lemmas 7, 8, 9, and 10, derived for the cases of a single source and two independent sources, into the case of *more than two* and *correlated* sources.

For the generalized broadcast wiretap channel, the derived achievable strong secrecy rate region extends Marton's inner bound for the broadcast channel with a common message [30, Theorem 8.4] to the proposed setting. Additionally, the derived rate region quantifies the secrecy cost due to the additional capabilities at the wiretapper. We characterize the strong secrecy capacity regions for two special classes of the generalized broadcast wiretap channel model. We first consider the class with deterministic channels to the legitimate receivers. Second, we consider the class with degraded receivers and a certain range for the noiselessly tapped ratio by the wiretapper which results in the wiretapper being more noisy than both receivers. These results establish the optimality of the proposed achievability scheme for the two aforementioned classes of the generalized broadcast wiretap channel.

For the generalized broadcast and interference channels with confidential messages, we observe that the derived achievable rate regions highlight the role of the size of the subset at each receiver which induces a trade-off between the secrecy rates for the two receivers. We further focus on the case of the generalized broadcast channel with confidential messages when one receiver's noisy observations are degraded with respect to the other receiver's noisy observations, and only the degraded receiver is provided with the subset of noiseless observations. In the achievable rate region for this case, the receiver with the degraded noisy observations achieves a positive secrecy rate after a certain threshold on its noiseless observations. That is, the weaker receiver is aided to the point of achieving a positive rate by the symbols he chooses to tap.

The remainder of the chapter is organized as follows. Section 7.2 describes the channel models considered in this chapter. The main results are provided in Section 7.3. The proofs for these results are provided in Sections 7.4 and 7.5. Section 7.6 concludes the chapter.

7.2 Channel Models

We present the following three generalized multi-receiver wiretap channel models. We begin with the generalized model for the two-user broadcast wiretap channel.

7.2.1 Generalized Broadcast Wiretap Channel

Consider the channel model in Figure 7.1. The legitimate channel is a discrete memoryless channel which consists of a finite input alphabet $\mathcal{X}$, two finite output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, and a transition probability distribution $p_{Y_1Y_2|X}$. The transmitter sends a

common message W_0 to both receivers, and a private message W_j to receiver j, where $j = 1, 2$, while keeping these three messages secret from the external wiretapper. The messages W_0, W_1, and W_2, are independent and uniformly distributed over $[1 : 2^{nR_0}]$, $[1 : 2^{nR_1}]$, and $[1 : 2^{nR_2}]$, respectively. The wiretapper chooses the subset $S \in \mathcal{S}$, where

$$\mathcal{S} \triangleq \{S \subseteq [1 : n] : |S| = \mu \leq n\}, \tag{7.1}$$

and observes the sequence $\mathbf{Z}_S^n = [Z_{S,1}, Z_{S,2}, \cdots, Z_{S,n}] \in \mathcal{Z}^n$, with

$$Z_{S,i} = \begin{cases} X_i, & i \in S \\ V_i, & \text{otherwise,} \end{cases} \tag{7.2}$$

$\mathbf{V}^n = [V_1, V_2, \cdots, V_n] \in \mathcal{V}^n$ is the n-letter output of the discrete memoryless channel $p_{V|X}$ when $\mathbf{X}^n$ is the input. The alphabet $\mathcal{Z}$ is given by $\mathcal{Z} = \mathcal{X} \cup \mathcal{V}$. Let $\alpha = \frac{\mu}{n}$, $0 \leq \alpha \leq 1$, denotes the ratio of the tapped symbols by the wiretapper.

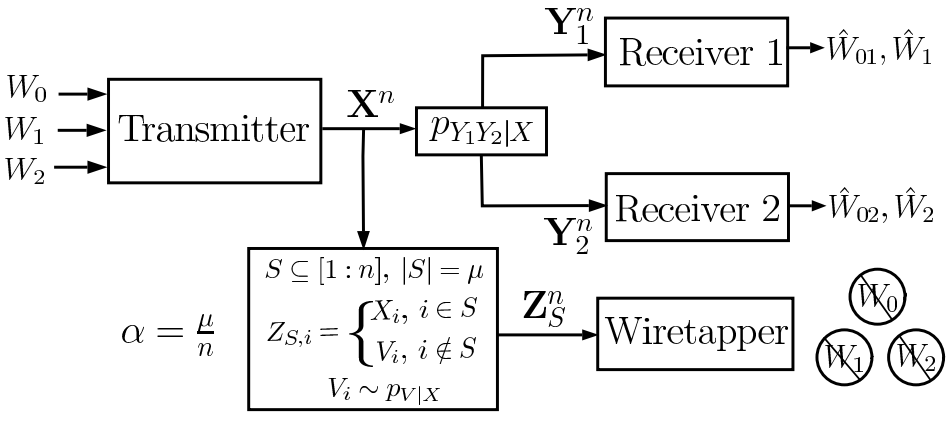


Fig. 7.1. The generalized two-user broadcast wiretap channel model.

The transmitter uses a stochastic encoder to encode $W_{[0:2]} = \{W_0, W_1, W_2\}$ into the codeword $\mathbf{X}^n = [X_1, X_2, \cdots, X_n] \in \mathcal{X}^n$. Receiver j, $j = 1, 2$, observes the sequence $\mathbf{Y}_j^n = [Y_{j,1}, Y_{j,2}, \cdots, Y_{j,n}] \in \mathcal{Y}_j^n$ and outputs the estimates $\hat{W}_{0,j}$ and $\hat{W}_j$ of its desired messages. An $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2})$ channel code $\mathcal{C}_n$ consists of three message sets $[1 : 2^{nR_0}]$, $[1 : 2^{nR_1}]$, and $[1 : 2^{nR_2}]$, a stochastic encoder $P^{(\mathcal{C}_n)}_{\mathbf{X}^n|W_{[0:2]}}$, and two decoders. A rate tuple (R_0, R_1, R_2) is achievable, with strong secrecy, if there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2})$ channel codes $\{\mathcal{C}_n\}_{n\geq 1}$ such that

$$\lim_{n\to\infty} \mathbb{P}^{(\mathcal{C}_n)} \left(\bigcup_{j=1,2} (\hat{W}_{0,j}, \hat{W}_j) \neq (W_0, W_j) \right) = 0, \quad \textbf{(Reliability),} \tag{7.3}$$

$$\lim_{n\to\infty} \max_{S\in\mathcal{S}} I^{(\mathcal{C}_n)} \left(W_0, W_1, W_2; \mathbf{Z}_S^n \right) = 0, \quad \textbf{(Strong Secrecy)}. \tag{7.4}$$

The strong secrecy capacity region for the model is the closure of all achievable (R_0, R_1, R_2).

Next, we propose generalized models for the two-user broadcast and interference channels with confidential messages.

7.2.2 Generalized Broadcast Channel with Confidential Messages

Consider the channel model described in Figure 7.2. The channel $\{\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, p_{Y_1Y_2|X}\}$ is discrete memoryless with a finite input alphabet $\mathcal{X}$, two finite output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a transition probability distribution $p_{Y_1Y_2|X}$. The transmitter sends a message W_j to receiver j, $j = 1, 2$, while keeping W_j secret from the other receiver, i.e., receiver i, where $i = 1, 2$, and $i \neq j$. The messages W_1 and W_2 are independent and uniformly distributed over $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$, respectively. The transmitter encodes the messages W_1 and W_2 into the codeword $\mathbf{X}^n \in \mathcal{X}^n$ using a stochastic encoder. Receiver

j, $j = 1, 2$, besides observing $\mathbf{Y}_j^n \in \mathcal{Y}_j^n$, chooses the subset $S_j \in \mathcal{S}_j$, where

$$\mathcal{S}_j \triangleq \left\{ S_j \subseteq [1:n] : |S_j| = \mu_j, \alpha_j = \frac{\mu_j}{n} \right\}, \tag{7.5}$$

and observes $\mathbf{Z}_{S_j}^n = [Z_{S_j,1}, Z_{S_j,2}, \cdots, Z_{S_j,n}] \in \mathcal{Z}^n$, with $\mathcal{Z} \triangleq \{\mathcal{X}_1 \times \mathcal{X}_2\} \cup \{?\}$, and

$$Z_{S_j,i} = \begin{cases} X_i, & i \in S_j \\ \text{`?'}, & \text{otherwise.} \end{cases} \tag{7.6}$$

Receiver j, upon observing $\mathbf{Y}_j^n, \mathbf{Z}_{S_j}^n$, outputs the estimate $\hat{W}_j$ of its desired message.

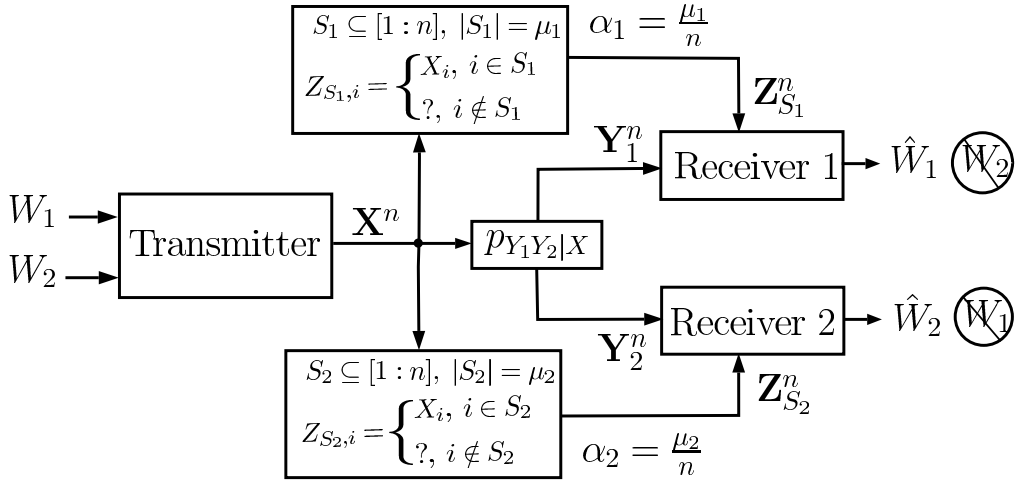


Fig. 7.2. The generalized two-user broadcast channel with confidential messages.

An $(n, 2^{nR_1}, 2^{nR_2})$ channel code $\mathcal{C}_n$ consists of two message sets, one stochastic encoder $P_{\mathbf{X}|W_1W_2}^{(\mathcal{C}_n)}$, and two decoders. The strong secrecy rate pair (R_1, R_2) is achievable

if there exists a sequence of channel codes $\{\mathcal{C}_n\}_{n\geq 1}$ such that

$$\lim_{n\to\infty} \mathbb{P}^{(\mathcal{C}_n)}\left((\hat{W}_1,\hat{W}_2)\neq(W_1,W_2)\right)=0, \qquad \textbf{(Reliability),} \tag{7.7}$$

$$\lim_{n\to\infty}\max_{S_2\in\mathcal{S}_2} I^{(\mathcal{C}_n)}(W_1;\mathbf{Y}_2^n,\mathbf{Z}_{S_2}^n)=0, \qquad \textbf{(Strong Secrecy against Receiver 2),} \tag{7.8}$$

$$\lim_{n\to\infty}\max_{S_1\in\mathcal{S}_1} I^{(\mathcal{C}_n)}(W_2;\mathbf{Y}_1^n,\mathbf{Z}_{S_1}^n)=0, \qquad \textbf{(Strong Secrecy against Receiver 1).} \tag{7.9}$$

The strong secrecy capacity region is the closure of all the achievable strong secrecy rate pairs (R_1, R_2).

7.2.3 Generalized Interference Channel with Confidential Messages

Consider the channel model in Figure 7.3. The channel $p_{Y_1Y_2|X_1X_2}$ is a discrete memoryless channel with two finite input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, and two finite output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$. Transmitter $j, j = 1, 2$, wishes to send a message W_j reliably to receiver j, while keeping W_j secret from the other user's receiver. W_1 and W_2 are independent and uniformly distributed over $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$, respectively. Transmitter j maps W_j into the codeword $\mathbf{X}_j^n \triangleq [X_{j,1}, X_{j,2}, \cdots, X_{j,n}] \in \mathcal{X}_j^n$ using a stochastic encoder. As in Section 7.2.2, receiver j, $j = 1, 2$, (i) chooses the subset $S_j \in \mathcal{S}_j$ where $\mathcal{S}_j$ is defined as in (7.5), (ii) observes $\mathbf{Y}_j^n \in \mathcal{Y}_j^n$ and $\mathbf{Z}_{S_j}^n = [Z_{S_j,1}, \cdots, ZS_j, n] \in \mathcal{Z}^n$, where

$\mathcal{Z} \triangleq \{\mathcal{X}_1 \times \mathcal{X}_2\} \cup \{?\}$, and

$$Z_{S_j,i} = \begin{cases} \{X_{1,i}, X_{2,i}\} & i \in S_j \\ \text{`?'}, & \text{otherwise.} \end{cases} \tag{7.10}$$

$\mathcal{Z} \triangleq \mathcal{X} \cup \{?\}$, and (iii) outputs the estimate $\hat{W}_j$ of its desired message.

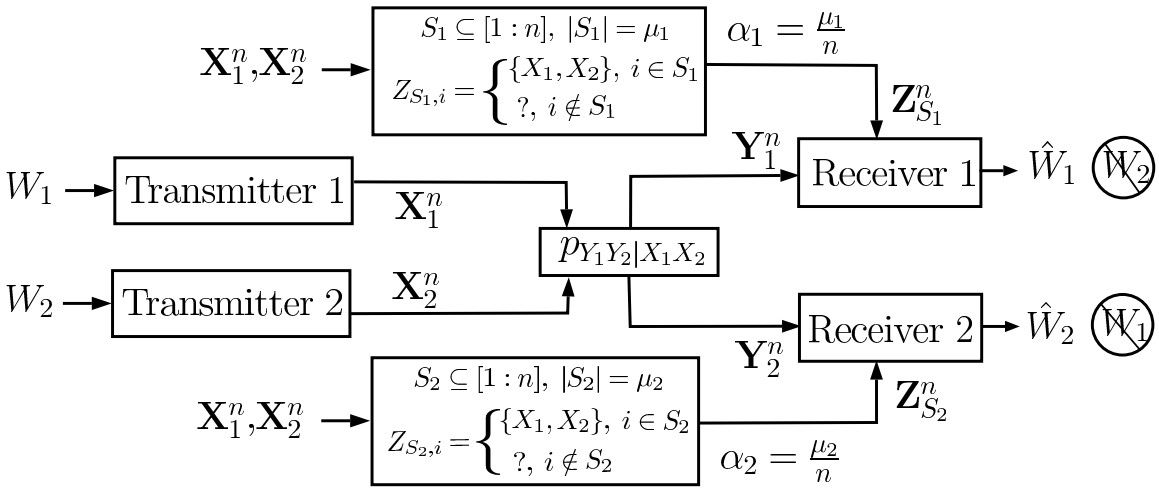


Fig. 7.3. The generalized two-user interference channel with confidential messages.

An $(n, 2^{nR_1}, 2^{nR_2})$ channel code $\mathcal{C}_n \triangleq \{\mathcal{C}_{1,n}, \mathcal{C}_{2,n}\}$ consists of two message sets, two stochastic encoders $P_{\mathbf{X}_j|W_j}^{(\mathcal{C}_{j,n})}$, $j = 1, 2$, and two decoders. (R_1, R_2) is an achievable strong secrecy rate pair if there exists a sequence of channel codes $\{\mathcal{C}_n\}_{n\geq 1}$ such that (7.7)-(7.9) hold. The strong secrecy capacity region is the closure of all achievable strong secrecy rate pairs (R_1, R_2).

7.3 Main Results

In this section, we present the main results of this chapter. Section 7.3.1 provides an achievable strong secrecy rate region for the generalized two-user broadcast wiretap channel model. Sections 7.3.1.1 and 7.3.1.2 identify the strong secrecy capacity regions for two special classes of the model.

7.3.1 Broadcast Wiretap Channel

The following theorem is an achievable strong secrecy rate region for the generalized two-user broadcast wiretap channel model:

Theorem 9. *For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the generalized broadcast wiretap channel model in Section 7.2.1 is given by the convex hull of all the rate tuples (R_0, R_1, R_2) which satisfy*

$$R_0 + R_j \leq \left[I(U_0, U_j; Y_j) - I(U_0, U_j; V) - \alpha I(U_0, U_j; X|V)\right]^+, \qquad j = 1, 2, \tag{7.11}$$

$$R_0 + R_1 + R_2 \leq \Big[\min\{I(U_0; Y_1), I(U_0; Y_2)\} + I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) - I(U_1; U_2|U_0) - I(U_0, U_1, U_2; V) - \alpha I(U_0, U_1, U_2; X|V)\Big]^+, \tag{7.12}$$

$$2R_0 + R_1 + R_2 \leq \Big[I(U_0, U_1; Y_1) + I(U_0, U_2; Y_2) - I(U_1; U_2|U_0) - I(U_0; V) - I(U_0, U_1, U_2; V) - \alpha[I(U_0; X|V) + I(U_0, U_1, U_2; X|V)\Big]^+, \tag{7.13}$$

for some probability distribution $p_{U_0U_1U_2XY_1Y_2V}$, over the random variables of the model, which factorizes as $p_{U_0U_1U_2}\ p_{X|U_0U_1U_2}\ p_{Y_1Y_2V|X}$. That is, $(U_0, U_1, U_2) - X - Y_1Y_2V$ forms a Markov chain.

Proof. The proof is provided in Section 7.4. □

Next, we provide the following remarks about Theorem 9.

Remark 15. *By setting $\alpha = 0$ in Theorem 9, we obtain the achievable strong secrecy rate region for the two-user broadcast wiretap channel in [126, Theorem 3]. That is, the terms in (7.11)-(7.13) multiplied by α determine the secrecy cost, with respect to the broadcast wiretap channel, of the additional capability of the wiretapper to choose αn noiseless codeword symbols.*

Remark 16. *By setting $\alpha = 0$ and $V =$ constant in (7.11)-(7.13), we obtain Marton's inner bound for the two-user broadcast channel with a common message [30, Theorem 8.4].*

Next, we characterize the strong secrecy capacity regions for two classes of the generalized broadcast wiretap channel. We consider the case of no common message, i.e., $W_0 = 0$. In particular, we show that the achievable strong secrecy rate region in Theorem 9 is tight for these two special classes of the generalized broadcast wiretap channel model.

7.3.1.1 Broadcast Wiretap Channel with Deterministic Receivers

We first consider the class of the generalized broadcast wiretap channel in Figure 7.1 when both Y_1 and Y_2 are deterministic functions of the input X, i.e., there exit deterministic functions f_1 and f_2 such that $Y_1 = f_1(X)$ and $Y_2 = f_2(X)$.

Theorem 10. *For $0 \leq \alpha \leq 1$, the strong secrecy capacity region of the generalized broadcast wiretap channel with deterministic receivers is the set of all rate pairs (R_1, R_2)*

satisfying

$$R_j \leq (1-\alpha)H(Y_j|V), \quad j=1,2, \tag{7.14}$$

$$R_1 + R_2 \leq (1-\alpha)H(Y_1, Y_2|V). \tag{7.15}$$

Proof: The achievability of the rate region in Theorem 10 follows from the achievable strong secrecy rate region in Theorem 9 by setting $U_0 =$ constant, $U_1 = Y_1$, and $U_2 = Y_2$, in (7.11)-(7.13).

The converse is established as follows. We first use similar steps as in Section 5.5 to show that the secrecy capacity of the generalized broadcast wiretap channel is upper bounded by the secrecy capacity when the wiretapper observes the outputs of two discrete memoryless channel, where the first discrete memoryless channel is an erasure channel with erasure probability $(1-\alpha)$ and the second discrete memoryless channel is $p_{V|X}$. Next, for the resulting discrete memoryless setting, we evaluate the upper bound for the discrete memoryless *deterministic* broadcast wiretap channel in [9, Theorem 4], which results in the region in (7.11)-(7.13). ■

7.3.1.2 Broadcast Wiretap Channel with Degraded Receivers and More Noisy Wiretapper

We next consider the class of the generalized broadcast wiretap channel when (i) Y_2 is a degraded version of Y_1, i.e., $X - Y_1 - Y_2$ forms a Markov chain, and (ii) the wiretapper, in the corresponding discrete memoryless setting, is *more noisy* than both receivers. That is, the wiretapper which observes the outputs two discrete memoryless

channel, where the first channel is an erasure channel with erasure probability $(1-\alpha)$ and the second channel is $p_{V|X}$, is more noisy than both receivers. The condition of a more noisy wiretapper can hence be described as follows: For all random variables U such $U - X - (Y_2, V)$ forms a Markov chain, we

$$\alpha I(U;X|V) \leq I(U;Y_2) - I(U;V). \tag{7.16}$$

Theorem 11. *For $0 \leq \alpha \leq 1$ such that (7.16) holds, the strong secrecy capacity of the generalized broadcast wiretap channel with degraded receivers is the set of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq I(X;Y_1|U,Q) - I(X;V|U,Q) - \alpha H(X|V,U,Q), \tag{7.17}$$

$$R_2 \leq I(U;Y_2|Q) - I(U;V|Q) - \alpha I(U;X|V,Q), \tag{7.18}$$

so that $Q - U - X - Y_1Y_2V$ forms a Markov chain, where Q represents a time sharing random variable.

Proof: For achievability, we set $U_0 = U_2 = U$ and $U_1 = X$ in (3.6)-(5.5). The converse is established as in the previous theorem, while utilizing the upper bound in [9, Theorem 4]. ■

7.3.2 Generalized Broadcast Channel with Confidential Messages

Next, we provide an achievable strong secrecy rate region for generalized broadcast channel with confidential messages model described in Section 7.2.2.

Theorem 12. *For $0 \leq \alpha_1, \alpha_2 \leq 1$, an achievable strong secrecy rate region for the generalized broadcast channel with confidential messages in Section 7.2.2 is given by the convex hull of all rate pairs (R_1, R_2) satisfying:*

$$R_1 \leq \left[I(U_1;Y_1) + \alpha_1 I(U_1;X|Y_1) - I(U_1;U_2) - I(U_1;Y_2|U_2) - \alpha_2 I(U_1;X|U_2,Y_2)\right]^+, \tag{7.19}$$

$$R_2 \leq \left[I(U_2;Y_2) + \alpha_2 I(U_2;X|Y_2) - I(U_1;U_2) - I(U_2;Y_1|U_1) - \alpha_1 I(U_2;X|U_1,Y_1)\right]^+, \tag{7.20}$$

for some probability distribution $p_{U_1U_2XY_1Y_2}$ wghich factorizes as $p_{U_1U_2}p_{X|U_1U_2}p_{Y_1Y_2|X}$. That is, $(U_1,U_2) - X - (Y_1,Y_2)$ forms a Markov chain.

Proof: The proof is provided in Section 7.5. ■

Remark 17. *In Theorem 12, by setting $\alpha_1 = \alpha_2 = 0$ in (7.19) and (7.20), we obtain the achievable secrecy rate region for the broadcast channel with confidential messages in [67, Theorem 4].*

7.3.2.1 Generalized Broadcast Channel with Confidential Messages and a Degraded Receiver

We now highlight the special instance of the generalized broadcast channel with confidential messages in Section 7.2.2 when one receiver is degraded with respect to the other, and only the degraded receiver is provided with a subset of noiseless observations of the transmitted codeword symbols. For the generalized broadcast channel with confidential messages model, when $\alpha_1 = 0$, $\alpha_2 = \alpha$, and the channel $p_{Y_1Y_2|X}$ is degraded, i.e.,

$X - Y_1 - Y_2$ forms a Markov chain, we have the following achievable strong secrecy rate region for the model.

Corollary 6. *For $0 \leq \alpha \leq 1$, an achievable strong secrecy rate region for the degraded broadcast channel with confidential messages, with the degraded receiver is provided by αn noiseless transmitted codeword symbols of its choice, is the convex hull of all rate pairs (R_1, R_2) satisfying:*

$$R_1 \leq [I(U_1; Y_1|Y_2) - I(U_1; U_2|Y_2) - \alpha I(U_1; X|U_2, Y_2)]^+ \tag{7.21}$$

$$R_2 \leq [\alpha I(U_2; X|Y_2) - I(U_1; U_2|Y_1) - I(U_2; Y_1|Y_2)]^+ , \tag{7.22}$$

for some probability distribution $p_{U_1U_2XY_1Y_2}$ which factorizes as $p_{U_1U_2}p_{X|U_1U_2}p_{Y_1|X}p_{Y_2|Y_1}$. That is, $(U_1, U_2) - X - Y_1 - Y_2$ forms a Markov chain.

Proof: Corollary 6 follows from Theorem 12 by setting $\alpha_1 = 0$, $\alpha_2 = \alpha$, and $p_{Y_1Y_2|X} = p_{Y_1|X}p_{Y_2|Y_1}$. ■

Remark 18. *For the broadcast channel with confidential messages in [67, Theorem 4], receiver 2 has zero secrecy rate, $R_2 = 0$, when its observation is a degraded version from receiver 1's observation, i.e., Y_2 is degraded with respect to Y_1. By contrast, for the generalized broadcast channel with confidential messages in Section 7.2.2, when $\alpha_1 = 0$, $\alpha_2 = \alpha$, and Y_2 is degraded with respect to Y_1, Corollary 6 implies that receiver 2 has a positive strong secrecy rate after a certain threshold on α. For example, by setting $U_1 =$ constant and $U_2 = X$, and for $H(X|Y_2) \neq 0$, we have that $R_2 > 0$ if the following*

condition is satisfied:

$$\frac{H(X|Y_2) - H(X|Y_1)}{H(X|Y_2)} < \alpha \leq 1. \tag{7.23}$$

In general, for the generalized broadcast wiretap channel model with $\alpha_1 = 0$, $\alpha_2 = \alpha$, *and* Y_2 *is degraded with respect to* Y_1, *the strong secrecy rate for receiver* 2 *is positive, i.e.,* $R_2 > 0$, *if there exist* U_1 *and* U_2 *such that* $(U_1, U_2) - X - Y_1 - Y_2$ *forms a Markov chain, and*

$$\alpha I(U_2; X|Y_2) > I(U_2; Y_1|Y_2) + I(U_1; U_2|Y_1). \tag{7.24}$$

7.3.3 Generalized Interference Channel with Confidential Messages

Finally, we provide an achievable strong secrecy rate region for generalized interference channel with confidential messages model described in Section 7.2.3.

Theorem 13. *For* $0 \leq \alpha_1, \alpha_2 \leq 1$, *an achievable strong secrecy rate region for generalized interference channel with confidential messages in Section 7.2.3 is given by the convex hull of all rate pairs* (R_1, R_2) *satisfying:*

$$R_1 \leq \left[I(U_1; Y_1) + \alpha_1 I(U_1; X_1, X_2|Y_1) - I(U_1; Y_2|U_2) - \alpha_2 I(U_1; X_1, X_2|U_2, Y_2)\right]^+, \tag{7.25}$$

$$R_2 \leq \left[I(U_2; Y_2) + \alpha_2 I(U_2; X_1, X_2|Y_2) - I(U_2; Y_1|U_1) - \alpha_1 I(U_2; X_1, X_2|U_1, Y_1)\right]^+, \tag{7.26}$$

for some probability distribution $p_{U_1U_2X_1X_2Y_1Y_2}$ *which factorizes as* $p_{U_1}p_{U_2}p_{X_1|U_1}p_{X_2|U_2}p_{Y_1Y_2|X_1X_2}$. *That is,* $(U_1, U_2) - (X_1, X_2) - (Y_1, Y_2)$ *forms a Markov chain.*

Proof. The Proof is provided in Section 7.5. □

Remark 19. *In Theorem 13, setting* $\alpha_1 = \alpha_2 = 0$ *in (7.25) and (7.26) yields the achievable secrecy rate region for the interference channel with confidential messages in [67, Theorem 2].*

Remark 20. *By comparing (7.25) and (7.26) to the region in [67, Theorem 2], we observe that the term* $\alpha_j I(U_j; X_1, X_2|Y_j)$, *for* $j = 1, 2$, *represents the secrecy rate gain for user* j *due to its noiseless observations, and the term* $\alpha_i I(U_j; X_1, X_2|U_i, Y_i)$ *represents the secrecy penalty at user* j *due to the noiseless observations of user* i, *where* $i, j = 1, 2$ *and* $i \neq j$.

7.4 Proof of Theorem 9

The proof for Theorem 9 follows the same key steps in Sections 5.4 and 6.4, with the need of extending the technique and the utilized tools to address the setting of *multiple correlated sources*. In the original channel model, along with using stochastic encoding for secrecy, we utilize a combination of superposition and Marton coding as in [30, Chapter 8]. We hence define the correlated auxiliary random variables U_0, U_1, and U_2, according to the distribution $p_{U_{[0:2]}}p_{X|U_{[0:2]}}$. The message W_0 is represented by the codeword $\mathbf{U}_0$, and the messages W_1 and W_2 are superposed over W_0 through the codewords $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively. Decoder j, $j = 1, 2,$, thus decodes W_0 from its estimate for the codeword $\mathbf{U}_0$, denoted as $\hat{\mathbf{U}}_{0,j}$, and decodes W_j from its estimates for

both the codewords $\mathbf{U}_0$ and $\mathbf{U}_j$, i.e., $\hat{\mathbf{U}}_{0,j}$ and $\hat{\mathbf{U}}_j$. In the dual secret key agreement problem in the source model, we define the sources' noisy observations according to the combined superposition and Marton coding. That is, we consider *three correlated sources*, where one source observes the sequence $\mathbf{U}_0$, and each of other two sources observes the sequence $\mathbf{U}_0\mathbf{U}_j$, where $j = 1, 2$..

We now describe the achievability proof in detail. Let us fix the probability distribution $p_{U_{[0:2]}X} = p_{U_{[0:2]}} p_{X|U_{[0:2]}}$, and let $p_{Y_{[1:2]}|U_{[0:2]}}$ be the distribution resulting from concatenating the discrete memoryless channels $p_{X|U_{[0:2]}}$ and $p_{Y_{[1:2]}|X}$, where $p_{Y_{[1:2]}|X}$ is the transition probability distribution for the legitimate channel in Figure 7.1. That is,

$$p_{Y_{[1:2]}|U_{[0:2]}}(y_{[1:2]}|u_{[0:2]}) = \sum_{x\in\mathcal{X}} p_{Y_{[1:2]}|X}(y_{[1:2]}|x)\; p_{X|U_{[0:2]}}(x|u_{[0:2]}). \tag{7.27}$$

As in Sections 5.4 and 6.4, we describe the following two protocols:

Protocol A: The protocol is described in Figure 7.4. Let $\mathbf{U}_0^n$, $\mathbf{U}_1^n$, $\mathbf{U}_2^n$, $\mathbf{Y}_1^n$, and $\mathbf{Y}_2^n$ be i.i.d. sequences according to the distribution $p_{U_{[0:2]}} p_{Y_{[1:2]}|U_{[0:2]}}$. Source $\mathbf{U}_0$ is randomly and independently binned into the two indices $W_0 = \mathcal{B}_{10}(\mathbf{U}_0)$, $F_0 = \mathcal{B}_{20}(\mathbf{U}_0)$, and source $\mathbf{U}_0\mathbf{U}_j$, $j = 1, 2$, is randomly and independently binned into the two indices $W_j = \mathcal{B}_{1j}(\mathbf{U}_0\mathbf{U}_j)$ and $F_j = \mathcal{B}_{2j}(\mathbf{U}_0\mathbf{U}_j)$. $\mathcal{B}_{1t}$ and $\mathcal{B}_{2t}$, where $t = 0, 1, 2$, are independent and uniformly distributed over $[1 : 2^{nR_t}]$ and $[1 : 2^{n\tilde{R}_t}]$, respectively. Decoder j observes the public messages F_0 and F_j and the sequence $\mathbf{Y}_j$, and outputs the estimates $\hat{\mathbf{U}}_{0,j}$ and $\hat{\mathbf{U}}_j$ of the codewords $\mathbf{U}_0$ and $\mathbf{U}_j$, and the estimates $\hat{W}_{0,j}$ and $\hat{W}_j$ of its desired messages. The wiretapper chooses the subset $S \in \mathcal{S}$ and observes $\mathbf{Z}_S$ as in (7.2). The distribution of $\mathbf{Z}_S$ is only known to belong to the finite class $\{p_{\mathbf{Z}_S}\}_{S\in\mathcal{S}}$, with $|\mathcal{S}| < 2^{\alpha n}$. The induced

joint distribution of protocol A is thus given by

$$\tilde{P}_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_{[1:2]}} = p_{\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S}\tilde{P}_{\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_1|\mathbf{Y}_1F_0F_1}\tilde{P}_{\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_2|\mathbf{Y}_2F_0F_2}$$

$$\times \mathbb{1}\left\{\mathcal{B}_{1j}(\mathbf{U}_0\mathbf{U}_j) = W_j, \mathcal{B}_{2j}(\mathbf{U}_0\mathbf{U}_j) = F_j, j = 1,2\right\}\ \mathbb{1}\left\{\mathcal{B}_{10}(\mathbf{U}_0) = W_0, \mathcal{B}_{20}(\mathbf{U}_0) = F_0\right\} \tag{7.28}$$

$$= \tilde{P}_{W_{[0:2]}F_{[0:2]}}\tilde{P}_{\mathbf{U}_{[0:2]}|W_{[0:2]}F_{[0:2]}}p_{\mathbf{Y}_{[1:2]}\mathbf{Z}_S|\mathbf{U}_{[0:2]}}\tilde{P}_{\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_1|\mathbf{Y}_1C_0C_1}\tilde{P}_{\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_2|\mathbf{Y}_2F_0F_2}. \tag{7.29}$$

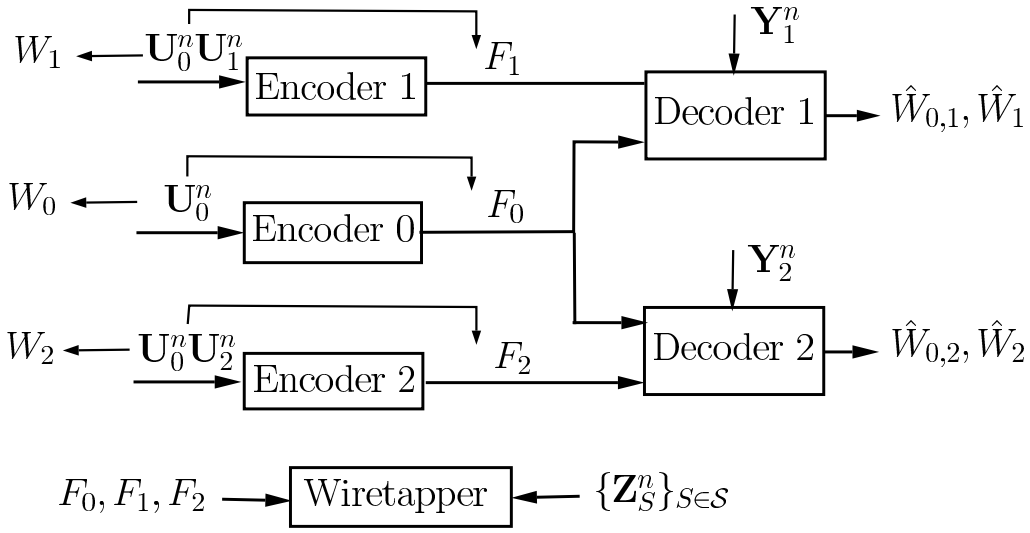


Fig. 7.4. Multi-terminal secret key agreement problem in the source model.

Protocol B: This protocol is the original channel model in Figure 7.1 with added common randomness F_t, $t = 0, 1, 2$, available to all terminals and uniformly distributed over $[1 : 2^{n\tilde{R}_t}]$. We utilize here the encoder and decoders in (7.29). The induced joint distribution is given by

$$P_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_{[1:2]}}$$

$$= p^U_{W_{[0:2]}} p^U_{F_{[0:2]}} \tilde{P}_{\mathbf{U}_{[0:2]}|W_{[0:2]}F_{[0:2]}} p_{\mathbf{Y}_{[1:2]}\mathbf{Z}_S|\mathbf{U}_{[0:2]}} \tilde{P}_{\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_1|\mathbf{Y}_1 C_0 C_1} \tilde{P}_{\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_2|\mathbf{Y}_2 F_0 F_2}. \quad (7.30)$$

Notice that, although F_i is available at receiver j, where $i, j = 1, 2$, and $i \neq j$, it is not used to decode W_0 and W_j. We next state the following two lemmas which extend Lemmas 9 and 10 in the previous chapter to the case of *multiple correlated* sources.

Lemma 11. *Let $X_{[1:T]}$ be T correlated sources according to the distribution $p_{X_{[1:T]}}$. Each source $X_t \in \mathcal{X}_t$, where $t \in [1:T]$, is randomly binned into the two indices $W_t = \mathcal{B}_{1t}(X_t)$ and $F_t = \mathcal{B}_{2t}(X_t)$. $\mathcal{B}_{1t}$ and $\mathcal{B}_{2t}$ are independent and uniformly distributed over $[1:\tilde{W}_t]$ and $[1:\tilde{F}_t]$, respectively. Define*

$$\mathcal{B} \triangleq \{\mathcal{B}_{1t}(x_t), \mathcal{B}_{2t}(x_t) : t \in [1:T], x_t \in \mathcal{X}_t\} \quad (7.31)$$

$$\mathcal{J} \triangleq \{J : J \subseteq [1:T], J \neq \emptyset\}. \quad (7.32)$$

For $J \in \mathcal{J}$ and $\gamma^{(J)} > 0$, define

$$\mathcal{D} \triangleq \{x_{[1:T]} \in \mathcal{X}_{[1:T]} : x_J \in \mathcal{D}_{\gamma^{(J)}}, \forall J \in \mathcal{J}\},$$

$$\mathcal{D}_{\gamma^{(J)}} \triangleq \{x_J \in \mathcal{X}_J : -\log p_{X_J}(x_J) > \gamma^{(J)}\}, \quad (7.33)$$

where, for $J \in \mathcal{J}$, $\mathcal{X}_J$ denotes the Cartesian product $\prod_{t\in J} \mathcal{X}_t$. Let $\tilde{W}_J = \prod_{t\in J} \tilde{W}_t$ and $\tilde{F}_J = \prod_{t\in J} \tilde{F}_t$. Then, we have

$$\mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{W_{[1:T]}F_{[1:T]}}, p^U_{W_{[1:T]}} p^U_{F_{[1:T]}} \right) \right) \leq \mathbb{P}_{p_{X_{[1:T]}}} \left(X_{[1:T]} \notin \mathcal{D} \right) + \frac{1}{2} \sum_{J\in\mathcal{J}} \sqrt{\tilde{W}_J \tilde{F}_J 2^{-\gamma^{(J)}}}, \quad (7.34)$$

where P is the induced distribution over $W_{[1:T]}$ and $F_{[1:T]}$.

Proof: The proof is provided in Appendix I. ■

Lemma 12. *Let $X_{[1:T]}$ be T correlated sources, which are correlated with the source $\{Z_S\} \triangleq \{\mathcal{Z}, p_{Z_S}\}$, where $S \in \mathcal{S}$, according to the distribution $p_{X_{[1:T]}Z_S}$. All the alphabets of $\{\mathcal{X}_t\}_{t=1}^T$, $\mathcal{Z}$, and $\mathcal{S}$, are finite. Each source X_t is randomly binned into the two indices W_t and F_t as in Lemma 11. Let $\mathcal{P}$ be the set of all possible permutations of $[1:T]$. For all $\mathbf{p} \in \mathcal{P}$ and $t \in [1:T]$, let $\gamma_t^{\mathbf{p}} > 0$, and define*

$$\mathcal{D}_{\mathbf{p}}^S \triangleq \left\{(x_{[1:T]}, z) \in \mathcal{X}_{[1:T]} \times \mathcal{Z} : (x_{p_{1:t}}, z) \in \mathcal{D}_{\gamma_t^{\mathbf{p}}}^S, \forall t \in [1:T]\right\} \tag{7.35}$$

where $\mathbf{p} \triangleq [p_1 \cdots p_T]$, $x_{p_{1:t}} \triangleq \{x_{p_1}, \cdots, x_{p_t}\}$, $x_{p_{1:0}} = \emptyset$, and

$$\mathcal{D}_{\gamma_t^{\mathbf{p}}}^S \triangleq \left\{(x_{p_{1:t}}, z) : \log \frac{1}{p_{X_{p_t}|X_{p_{1:t-1}}Z_S}(x_{p_t}|x_{p_{1:t-1}}, z)} > \gamma_t^{\mathbf{p}}\right\}. \tag{7.36}$$

If there exists $\delta \in (0, \frac{1}{2})$ such that for all $S \in \mathcal{S}$ and $\mathbf{p} \in \mathcal{P}$,

$$\mathbb{P}_{p_{X_{[1:T]}Z_S}}\left(\left(X_{[1:T]}, Z_S\right) \in \mathcal{D}_{\mathbf{p}}^S\right) \geq 1 - \delta^2 \tag{7.37}$$

then, for $\epsilon \in [0, 1]$, we have

$$\begin{aligned}\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}(P_{W_{[1:T]}F_{[1:T]}Z_S}||p^U_{W_{[1:T]}}p^U_{F_{[1:T]}}p_{Z_S}) \geq T\tilde{\epsilon}\right)& \\ \leq |\mathcal{S}||\mathcal{Z}| \min_{\mathbf{p}\in\mathcal{P}} \sum_{t=1}^T \exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_t^{\mathbf{p}}}}{3\tilde{W}_{p_t}\tilde{F}_{p_t}}\right),&\end{aligned} \tag{7.38}$$

where $\tilde{\epsilon} = \max_t\{\epsilon + (\delta+\delta^2)\log(\tilde{W}_t\tilde{F}_t) + H_b(\delta^2)\}$*,* H_b *is the binary entropy function, and* P *is the induced distribution.*

Proof: The proof is provided in Appendix J. ■

We now apply Lemma 11 to the source model in Figure7.4. Set $X_1 = \mathbf{U}_0$, $X_2 = \mathbf{U}_0\mathbf{U}_1$, $X_3 = \mathbf{U}_0\mathbf{U}_2$, $\tilde{W}_t = 2^{nR_t}$, and $\tilde{F}_t = 2^{n\tilde{R}_t}$, $t = 0, 1, 2$, where $\mathbf{U}_{[0:2]}$, $W_{[0:2]}$, and $F_{[0:2]}$, are as in protocol A. For $\epsilon' > 0$ and $J \subseteq [1:3]$, $J \neq \emptyset$, let $\gamma^{(J)} = (1-\epsilon')H(X_J)$. For $J = \{1\}$, using Hoeffding inequality, we have

$$\mathbb{P}\left(X_1 \notin \mathcal{D}_{\gamma^{(\{1\})}}\right) = \mathbb{P}_{p_{\mathbf{U}_0}}\left(-\log p(\mathbf{U}_0) \leq \gamma^{(\{1\})}\right) \tag{7.39}$$

$$= \mathbb{P}\left(\sum_{k=1}^{n}(-\log p(U_{0,k})) \leq n(1-\epsilon')H(U_0)\right) \leq \exp(-\beta^{(\{1\})}n), \tag{7.40}$$

where $\beta^{(\{1\})} > 0$. Similarly, for $J \subseteq [1:3]$, $J \neq \emptyset$, there exists $\beta^{(J)} > 0$ such that

$$\mathbb{P}\left(X_J \notin \mathcal{D}_{\gamma^{(J)}}\right) \leq \exp(-\beta^{(J)}n). \tag{7.41}$$

Using (7.33), there exists $\bar{\beta} > 0$ such that

$$\mathbb{P}\left(X_{[1:3]} \notin \mathcal{D}\right) \leq \sum_{J\subseteq[1:3],J\neq\emptyset} \mathbb{P}\left(X_J \notin \mathcal{D}_{\gamma^{(J)}}\right) \leq \exp(-\bar{\beta}n). \tag{7.42}$$

Substituting the choices for $\tilde{W}_t$, $\tilde{F}_t$, $\gamma^{(J)}$, and (7.42) in (7.34), there exists $\beta > 0$ such that, for any $S \in \mathcal{S}$,

$$\mathbb{E}_{\mathcal{B}}\mathbb{V}\left(\tilde{P}_{W_{[0:2]}F_{[0:2]}}, p^U_{W_{[0:2]}}p^U_{F_{[0:2]}}\right) \leq \exp(-\beta n), \tag{7.43}$$

as long as

$$R_0 + \tilde{R}_0 < (1-\epsilon')H(U_0) \tag{7.44}$$

$$R_0 + \tilde{R}_0 + R_j + \tilde{R}_j < (1-\epsilon')H(U_0U_j),\ j=1,2, \tag{7.45}$$

$$R_0 + \tilde{R}_0 + R_1 + \tilde{R}_1 + R_2 + \tilde{R}_2 < (1-\epsilon')H(U_{[0:2]}), \tag{7.46}$$

Now, for reliability of protocol A, we use Slepian-Wolf decoders at both users. Using [126, Lemma 1], for any $S \in \mathcal{S}$,

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(\begin{array}{c} \tilde{P}_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S} \mathbb{1}\left\{\hat{\mathbf{U}}_{0,1} = \hat{\mathbf{U}}_{0,2} = \mathbf{U}_0, \hat{\mathbf{U}}_j = \mathbf{U}_j, j=1,2\right\}, \\ \tilde{P}_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_{[1:2]}} \end{array} \right) \right) = 0, \tag{7.47}$$

as long as, for $j=1,2$,

$$\tilde{R}_0 + \tilde{R}_j > H(U_0, U_j|Y_j) \tag{7.48}$$

$$\tilde{R}_j > H(U_j|U_0Y_j). \tag{7.49}$$

Next, in Lemma 12, set $X_1 = \mathbf{U}_0$, $X_2 = \mathbf{U}_0\mathbf{U}_1$, $X_3 = \mathbf{U}_0\mathbf{U}_2$, $\tilde{W}_t = 2^{nR_t}$, $\tilde{F}_t = 2^{n\tilde{R}_t}$, $t=0,1,2$, and $Z_S = \mathbf{Z}_S, \forall S \in \mathcal{S}$. $\mathbf{U}_{[0:2]}$, $\mathbf{Z}_S$,and $\mathcal{S}$ are defined as in protocol A. Let us first consider $\mathbf{p} = \bar{\mathbf{p}} = [1:3]$. Since $p_{V|U_{[0:2]}}$ is a discrete memroyless channel, which results from concatenating the two discrete memoryless channels $p_{V|X}$ and $p_{X|U_{[0:2]}}$,

and $\mathbf{U}_{[0:2]}$ are i.i.d. sequences, then for all $S \in \mathcal{S}$, we have

$$H(X_1|Z_S) = H(\mathbf{U}_0|\mathbf{X}_S\mathbf{V}_{S^c}) = H(\mathbf{U}_{0,S}|\mathbf{X}_S) + H(\mathbf{U}_{0,S^c}|\mathbf{V}_{S^c}) \tag{7.50}$$

$$= \mu H(U_0|X) + (n-\mu)H(U_0|V) \tag{7.51}$$

$$H(X_2|X_1, Z_S) = H(\mathbf{U}_0\mathbf{U}_1|\mathbf{U}_0\mathbf{X}_S\mathbf{V}_{S^c}) \tag{7.52}$$

$$= \mu H(U_1|U_0, X) + (n-\mu)H(U_1|U_0V) \tag{7.53}$$

$$H(X_3|X_{[1:2]}, Z_S) = \mu H(U_2|U_{0:1}X) + (n-\mu)H(U_2|U_{0:1}V). \tag{7.54}$$

By Hoeffding inequality and the definition of $\mathcal{D}^S_{\bar{\mathbf{p}}}$ in (7.35) and (7.36), with $\bar{\epsilon} > 0$, and

$$\gamma_1^{\bar{\mathbf{p}}} = (1-\bar{\epsilon})[\mu H(U_0|X) + (n-\mu)H(U_0|V)]$$

$$\gamma_2^{\bar{\mathbf{p}}} = (1-\bar{\epsilon})[\mu H(U_1|U_0X) + (n-\mu)H(U_1|U_0V)]$$

$$\gamma_3^{\bar{\mathbf{p}}} = (1-\bar{\epsilon})[\mu H(U_2|U_0U_1X) + (n-\mu)H(U_2|U_0U_1V)],$$

there exists $\beta_{\bar{\mathbf{p}}} > 0$ such that

$$\mathbb{P}\left(\left(X_{[1:3]}, Z_S\right) \notin \mathcal{D}^S_{\bar{\mathbf{p}}}\right) \leq \exp\left(-\beta_{\bar{\mathbf{p}}} n\right). \tag{7.55}$$

Similarly, for any $\mathbf{p}$ which is a permutation of $[1:3]$, letting

$$\gamma_t^{\mathbf{p}} = (1-\bar{\epsilon}) \min_{S\in\mathcal{S}} H\left(X_{p_t}|X_{p_{1:t-1}}, Z_S\right), \tag{7.56}$$

with $X_{p_{[1:0]}} = \emptyset$, there exists $\beta_{\mathbf{p}} > 0$ such that

$$\mathbb{P}\left(\left(X_{[1:3]}, Z_S\right) \notin \mathcal{D}_{\mathbf{p}}^S\right) \leq \exp(-\beta_{\mathbf{p}} n). \tag{7.57}$$

Taking $\delta^2 = \exp(-\tilde{\beta} n)$ and $\tilde{\beta} = \min_{\mathbf{p}} \beta_{\mathbf{p}}$, we have, for all $\mathbf{p}$

$$\mathbb{P}\left(\left(X_{[1:3]}, Z_S\right) \notin \mathcal{D}_{\mathbf{p}}^S\right) \leq \delta^2. \tag{7.58}$$

Note that $\lim_{n\to\infty} \delta^2 = 0$, and hence, for n large enough, $\delta^2 \in (0, \frac{1}{4})$. Thus, the conditions of Lemma 12 are satisfied.

Substituting the choices for $\tilde{W}_t$, $\tilde{F}_t$, $\gamma_t^{\mathbf{P}}$, for $t = 1, 2, 3$, and all $\mathbf{p}$, and

$$|\mathcal{S}||\mathcal{Z}^n| \leq \exp\left(n\left[\ln 2 + \ln(|\mathcal{X}| + |\mathcal{V}|)\right]\right), \tag{7.59}$$

in (7.38), we have, for all $\epsilon, \epsilon_1 > 0$ and $\tilde{\epsilon} = \epsilon + \epsilon_1$, there exists $n^* \in \mathbb{N}$ and $\kappa_\epsilon, \tilde{\kappa} > 0$ such that for all $n \geq n^*$,

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{W_{[0:2]}F_{[0:2]}\mathbf{Z}_S} || p^U_{W_{[0:2]}} p^U_{F_{[0:2]}} p_{\mathbf{Z}_S}\right) \geq 3\tilde{\epsilon}\right) \leq \exp\left(-\kappa_\epsilon e^{\tilde{\kappa} n}\right), \tag{7.60}$$

as long as

$$R_0 + \tilde{R}_0 < (1-\bar{\epsilon})\left[\alpha H(U_0|X) + (1-\alpha) H(U_0|V)\right] \tag{7.61}$$

$$R_0 + \tilde{R}_0 + R_j + \tilde{R}_j < (1-\bar{\epsilon})\left[\alpha H(U_0, U_j|X) + (1-\alpha) H(U_0, U_j|V)\right], \quad j = 1, 2, \tag{7.62}$$

$$R_0 + \tilde{R}_0 + R_1 + \tilde{R}_1 + R_2 + \tilde{R}_2 < (1-\bar{\epsilon})\left[\alpha H(U_0, U_1, U_2|X) + (1-\alpha)H(U_0, U_1, U_2|V)\right]. \tag{7.63}$$

Remark 21. *Note that for each* $\mathbf{p} \in \mathcal{P}$*, Lemma 12 results in the maximum binning rate* $R_{p_1} + \tilde{R}_{p_1}$ *for the source* X_{p_1}*, and then the maximum conditional binning rate for the source* X_{p_2} *given* $R_{p_1} + \tilde{R}_{p_1}$*, and so on and so forth, so that the probability in the left hand side of (7.38) is vanishing. That is, for each* $\mathbf{p}$*, Lemma 12 results in one corner point in the binning rate region for the sources* $X_{[1:T]}$ *such that* $W_{[1:T]}$ *and* $F_{[1:T]}$ *are independent, uniformly distributed, and all are independent from the wiretapper's observation.*

By the Borel-Cantelli lemma, it follows from (7.43) and (7.60) that

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_{[0:2]}F_{[0:2]}}, p^U_{W_{[0:2]}} p^U_{F_{[0:2]}}\right) > 0\right) = 0, \tag{7.64}$$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(\tilde{P}_{W_{[0:2]}F_{[0:2]}\mathbf{Z}_S} || p^U_{W_{[0:2]}} p^U_{F_{[0:2]}} p_{\mathbf{Z}_S}\right) > 0\right) = 0. \tag{7.65}$$

Using similar steps as in Section 5.4, we first use (7.43) and (7.64) to show that (7.47) and (7.65) hold as well for protocol B. That is,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(P_{W_{[0:2]}F_{[0:2]}\mathbf{Z}_S} || p^U_{W_{[0:2]}} p^U_{F_{[0:2]}} p_{\mathbf{Z}_S}\right) > 0\right) = 0, \tag{7.66}$$

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\begin{array}{c} P_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S} \mathbb{1}\left\{\hat{\mathbf{U}}_{0,1} = \hat{\mathbf{U}}_{0,2} = \mathbf{U}_0, \hat{\mathbf{U}}_j = \mathbf{U}_j, j = 1,2\right\}, \\ P_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_{[1:2]}} \end{array}\right)\right) = 0. \tag{7.67}$$

Next, we apply the selection lemma, Lemma 5, to (7.66) and (7.67) to show the existence of a binning realization $\mathbf{b}^*$, with a corresponding joint distribution p^* for protocol B, such that

$$\lim_{n\to\infty} \mathbb{V}\left(\begin{array}{c} p^*_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S} \mathbb{1}\left\{\hat{\mathbf{U}}_{0,1}=\hat{\mathbf{U}}_{0,2}=\mathbf{U}_0, \hat{\mathbf{U}}_j=\mathbf{U}_j, j=1,2\right\}, \\ p^*_{W_{[0:2]}F_{[0:2]}\mathbf{U}_{[0:2]}\mathbf{Y}_{[1:2]}\mathbf{Z}_S\hat{\mathbf{U}}_{0,1}\hat{\mathbf{U}}_{0,2}\hat{\mathbf{U}}_{[1:2]}} \end{array} \right) = 0, \tag{7.68}$$

$$\lim_{n\to\infty} \mathbb{1}\left\{ \max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[0:2]}F_{[0:2]}\mathbf{Z}_S} || p^U_{W_{[0:2]}} p^U_{F_{[0:2]}} p_{\mathbf{Z}_S} \right) > 0 \right\} = 0, \tag{7.69}$$

where $W_0 = b^*_{10}(\mathbf{U}_0)$, $F_0 = b^*_{20}(\mathbf{U}_0)$, $W_j = b^*_{1j}(\mathbf{U}_0\mathbf{U}_j)$, and $F_j = b^*_{2j}(\mathbf{U}_0\mathbf{U}_j)$, $j=1,2$. We finally introduce the $\hat{W}$ variables to (7.68), and use the union bound with (7.69), to show that

$$\lim_{n\to\infty} \mathbb{E}_{F_{[0:2]}}\left(\mathbb{P}_{p^*}\left(\bigcup_{j=1,2} \left(\hat{W}_{0,j}, \hat{W}_j\right) \neq (W_0, W_j) \,|F_{[0:2]} \right) \right) = 0$$

$$\lim_{n\to\infty} \mathbb{P}_{F_{[0:2]}}\left(\max_{S\in\mathcal{S}} \mathbb{D}\left(p^*_{W_{[0:2]}\mathbf{Z}_S|F_{[0:2]}} || p^U_{W_{[0:2]}} p^*_{\mathbf{Z}_S|F_{[0:2]}} \right) > 0 \right) = 0,$$

which are used to show the existence of $c^*_{[0:2]}$ such that both the reliability and secrecy constraints in (7.3), (7.4) hold.

Let $\tilde{p}^*$ be the distribution in protocol A that corresponds to the binning realization $\mathbf{b}^*$. We identify $\tilde{p}^*(\mathbf{u}_{[0:2]}|w_{[0:2]}, {f^*}_{[0:2]})$ and $(\tilde{p}^*(\hat{\mathbf{u}}_{0,j}, \hat{\mathbf{u}}_j|\mathbf{y}_j, c^*_0, c^*_j), b^*_{10}(\hat{\mathbf{u}}_{0,j}), b^*_{1j}(\hat{\mathbf{u}}_{0,j}, \hat{\mathbf{u}}_j), j =$

$1, 2$) as the encoder and decoders for the original model. Finally, applying Fourier-Motzkin elimination to the rate conditions in (7.44)-(7.46), (7.48)-(7.49), and (7.61)-(7.63), results in the rate region in (3.6)-(5.5). The convex hull follows by time sharing independent codes.

7.5 Proofs of Theorems 12 and 13

We first prove Theorem 12. In this proof, we utilize Lemmas 11 and 12 in Section 7.4. The key steps of the proof are similar to the proof in Section 7.4. The difference however lies in the need for careful definition and treatment of the reliability and secrecy (independence) conditions in the dual secret key agreement problem, in order to address the multiple security conditions in the original channel model, i.e., (7.8) and (7.9).

We first define the correlated auxiliary random variables U_1 and U_2. Let us fix the distribution $p_{U_{[1:2]}X} = p_{U_{[1:2]}} p_{X|U_{[1:2]}}$. Let $p_{Y_{[1:2]}|U_{[1:2]}}$ be the distribution resulting from concatenating the two discrete memoryless channels $p_{X|U_{[1:2]}}$ and $p_{Y_{[1:2]}|X}$, where $p_{Y_{[1:2]}|X}$ is the transition probability distribution in Figure 7.2. Next, we define the following two protocols and describe the joint distribution induced by each of them.

Protocol A: This protocol describes the dual secret key agreement problem in Figure 7.5, where $\mathbf{U}_1^n$, $\mathbf{U}_2^n$,$\mathbf{Y}_1^n$, and $\mathbf{Y}_2^n$ are i.i.d. sequences according to the distribution $p_{U_{[1:2]}} p_{Y_{[1:2]}|U_{[1:2]}}$. Notice that the noisy observations at the source encoders, $\mathbf{U}_1$ and $\mathbf{U}_2$ correspond to the correlated auxiliary variables utilized in Marton's coding to separately encode the messages W_1 and W_2 [30, Chapter 8]. The source $\mathbf{U}_j$ is randomly and independently binned into the indices $W_j = \mathcal{B}_{1j}(\mathbf{U}_j)$, $F_j = \mathcal{B}_{2j}(\mathbf{U}_j)$, where $\mathcal{B}_{1j}$ and $\mathcal{B}_{2j}$ are independent and uniformly distributed over $[1 : 2^{nR_j}]$ and $[1 : 2^{n\tilde{R}_j}], j = 1, 2$.

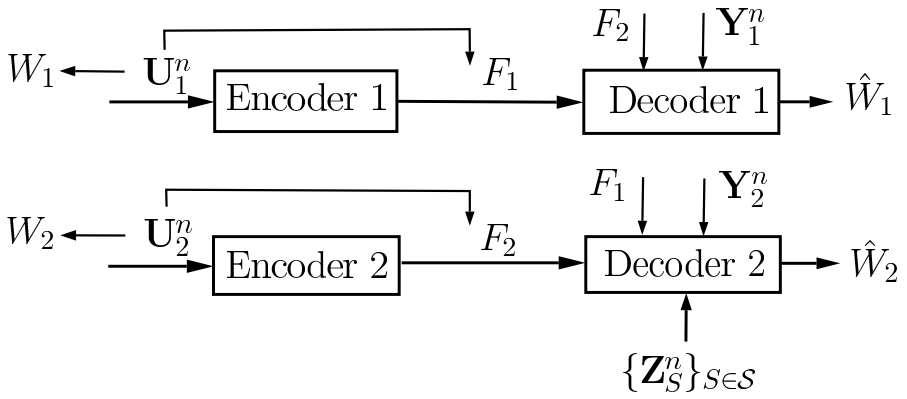


Fig. 7.5. Dual source model for the channel model in Fig. 7.2.

Decoder j (i) observes F_1, F_2, and the sequence $\mathbf{Y}_j$, (ii) chooses $S_j \in \mathcal{S}_j$ and observes $\mathbf{Z}_{S_j}$ as in (7.5) and (7.6), and (iii) outputs the estimates $\hat{\mathbf{U}}_j$ and $\hat{W}_j$. The message F_j is public to decoder i, while the key W_j should be kept secret from decoder i, $i, j = 1, 2$, and $i \neq j$. The realization of S_j, $j = 1, 2$, is unknown to the other decoder. The induced joint distribution for protocol A is

$$\tilde{P}_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}_1\mathbf{Z}_{S_1}\mathbf{Y}_2\mathbf{Z}_{S_2}\hat{\mathbf{U}}_{[1:2]}}$$

$$= p_{\mathbf{U}_{[1:2]}\mathbf{Y}_1\mathbf{Z}_{S_1}\mathbf{Y}_2\mathbf{Z}_{S_2}}\ \tilde{P}_{\hat{\mathbf{U}}_1|\mathbf{Y}_1\mathbf{Z}_{S_1}F_1}\tilde{P}_{\hat{\mathbf{U}}_2|\mathbf{Y}_2\mathbf{Z}_{S_2}F_2}\ \mathbb{1}\left\{\mathcal{B}_{1j}(\mathbf{U}_j) = W_j, \mathcal{B}_{2j}(\mathbf{U}_j) = F_j, j = 1, 2, \right\} \tag{7.70}$$

$$= \tilde{P}_{W_{[1:2]}F_{[1:2]}}\tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}}\ p_{\mathbf{Y}_1\mathbf{Z}_{S_1}\mathbf{Y}_2\mathbf{Z}_{S_2}|\mathbf{U}_{[1:2]}}\ \tilde{P}_{\hat{\mathbf{U}}_1|\mathbf{Y}_1\mathbf{Z}_{S_1}F_1}\tilde{P}_{\hat{\mathbf{U}}_2|\mathbf{Y}_2\mathbf{Z}_S F_2}. \tag{7.71}$$

Protocol B: This protocol describes the channel model in Figure 7.2 with assumed common randomness F_j, $j = 1, 2$, uniformly distributed over $[1 : 2^{n\tilde{R}_j}]$, independent from all other variables, and available to all terminals. We utilize $\tilde{P}_{\mathbf{U}_{[1:2]}|W_{1:2}F_{[1:2]}}$ and $\tilde{P}_{\hat{\mathbf{U}}_1|\mathbf{Y}_1\mathbf{Z}_{S_1}F_1}$, $\tilde{P}_{\hat{\mathbf{U}}_2|\mathbf{Y}_2\mathbf{Z}_{S_2}F_2}$ in (7.71) as the encoder and decoders for this protocol. The

induced joint distribution for protocol B is

$$P_{W_{[1:2]}F_{[1:2]}\mathbf{U}_{[1:2]}\mathbf{Y}_1\mathbf{Z}_{S_1}\mathbf{Y}_2\mathbf{Z}_{S_2}\hat{\mathbf{U}}_{[1:2]}}$$

$$= p^U_{W_{[1:2]}} p^U_{F_{[1:2]}} \tilde{P}_{\mathbf{U}_{[1:2]}|W_{[1:2]}F_{[1:2]}}\ p_{\mathbf{Y}_1\mathbf{Z}_{S_1}\mathbf{Y}_2\mathbf{Z}_{S_2}|\mathbf{U}_{[1:2]}} \tilde{P}_{\hat{\mathbf{U}}_1|\mathbf{Y}_1\mathbf{Z}_{S_1}F_1} \tilde{P}_{\hat{\mathbf{U}}_2|\mathbf{Y}_2\mathbf{Z}_{S_2}F_2}. \tag{7.72}$$

In the channel model in protocol B, although the common randomness F_i is available at receiver j, $i, j = 1, 2$, $i \neq j$, it is not utilized for decoding W_j. The encoders in the source model are chosen accordingly, c.f. (7.71). In the dual source model, since $\mathbf{U}_1$ and $\mathbf{U}_2$, are *correlated*, we utilize Lemmas 11 and 12 in the previous section. We divide the remainder of the proof into the following steps:

7.5.1 Closeness of joint induced distributions

In Lemma 11, set $X_1 = \mathbf{U}_1$, $X_2 = \mathbf{U}_2$, $\tilde{W}_1 = 2^{nR_1}$, $\tilde{F}_1 = 2^{n\tilde{R}_1}$, $\tilde{W}_2 = 1$, and $\tilde{F}_2 = 2^{n\tilde{R}_2}$, where $\mathbf{U}_{[1:2]}$ are defined as in protocol A. For $\epsilon' > 0$, by setting $\gamma_j = n(1-\epsilon')H(U_j), j = 1, 2$, $\gamma_{1,2} = n(1-\epsilon')H(U_{[1:2]})$, and using Hoeffding's inequality, we have

$$\mathbb{P}_{p_{\mathbf{U}_{[1:2]}}}\left(\mathbf{U}_{[1:2]} \notin \mathcal{D}\right) \leq \exp(-\beta_1' n), \tag{7.73}$$

where $\beta_1' > 0$. By substituting the choices for $\tilde{W}_j$, $\tilde{F}_j$, γ_j, $j = 1, 2$, and $\gamma_{1,2}$, in (7.34), there exists $\beta_1 > 0$ such that

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_1F_{[1:2]}}, p^U_{W_1} p^U_{F_{[1:2]}}\right)\right) \leq \exp(-\beta_1 n), \tag{7.74}$$

as long as,

$$R_1 + \tilde{R}_1 \leq (1-\epsilon')H(U_1), \quad \tilde{R}_2 \leq (1-\epsilon')H(U_2) \tag{7.75}$$

$$R_1 + \tilde{R}_1 + \tilde{R}_2 \leq (1-\epsilon')H(U_1, U_2). \tag{7.76}$$

Similarly, by setting $X_1 = \mathbf{U}_1$, $X_2 = \mathbf{U}_2$, $\tilde{W}_1 = 1$, $\tilde{F}_1 = 2^{n\tilde{R}_1}$, $\tilde{W}_2 = 2^{nR_2}$, and $\tilde{F}_2 = 2^{n\tilde{R}_2}$, in Lemma 11, there exists $\beta_2 > 0$ such that

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_2 F_{[1:2]}}, p^U_{W_2} p^U_{F_{[1:2]}}\right)\right) \leq \exp(-\beta_2 n), \tag{7.77}$$

as long as,

$$\tilde{R}_1 \leq (1-\epsilon')H(U_1), \quad R_2 + \tilde{R}_2 \leq (1-\epsilon')H(U_2) \tag{7.78}$$

$$\tilde{R}_1 + R_2 + \tilde{R}_2 \leq (1-\epsilon')H(U_1, U_2). \tag{7.79}$$

Using (7.71), (7.72), (7.74) and (7.77), we have, for $j = 1, 2$, $S_j \in \mathcal{S}_j$,

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j} \hat{\mathbf{U}}_j}, P_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j} \hat{\mathbf{U}}_j}\right)\right) \leq \exp(-\beta_j n). \tag{7.80}$$

Also, by the Borel-Cantelli lemma and Markov inequality, it follows from (7.74) and (7.77) that, for $j = 1, 2$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}}\left(\mathbb{V}\left(\tilde{P}_{W_j F_{[1:2]}}, p^U_{W_j} p^U_{F_{[1:2]}}\right) > 0\right) = 0. \tag{7.81}$$

7.5.2 Reliable decoding at source decoder j

In protocol A, for reliable communication of the source $\mathbf{U}_j$, decoder j employs Slepian-Wolf source decoder. Since $\mathbf{U}_j$ is an i.i.d. sequence and $p_{Y_j|U_{[1:2]}}$ is a discrete memoryless channel, then, for any $S_j \in \mathcal{S}_j$, $j = 1, 2$,

$$H(\mathbf{U}_j|\mathbf{Y}_j, \mathbf{Z}_{S_j}) = H(\mathbf{U}_{j,S_j}, \mathbf{U}_{j,S_j^c}|\mathbf{Y}_{j,S_j}, \mathbf{Y}_{j,S_j^c}, \mathbf{X}_{S_j}) \tag{7.82}$$

$$= H(\mathbf{U}_{j,S_j}|\mathbf{X}_{S_j}, \mathbf{Y}_{j,S_j}) + H(\mathbf{U}_{j,S_j^c}|\mathbf{Y}_{j,S_j^c}) \tag{7.83}$$

$$= \mu_j H(U_j|X) + (n - \mu_j)H(U_j|Y_j), \tag{7.84}$$

where (7.84) follows since $U_j - X - Y_j$ forms a Markov chain. Using [126, Lemma 1], for $j = 1, 2$, and any $S_j \in \mathcal{S}_j$,

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(\tilde{P}_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j} \hat{\mathbf{U}}_j}, \tilde{P}_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j}} \mathbb{1}\{\hat{\mathbf{U}}_j = \mathbf{U}_j\} \right) \right) = 0, \tag{7.85}$$

as long as,

$$\tilde{R}_j \geq \alpha_j H(U_j|X) + (1 - \alpha_j)H(U_j|Y_j). \tag{7.86}$$

7.5.3 Secrecy against source decoder j

Set $X_1 = \mathbf{U}_1$, $X_2 = \mathbf{U}_2$, $\tilde{W}_1 = 2^{nR_1}$, $\tilde{F}_1 = 2^{n\tilde{R}_1}$, $\tilde{W}_2 = 1$, $\tilde{F}_2 = 2^{n\tilde{R}_2}$, $\mathcal{S} = \mathcal{S}_2$, and $Z_S = \mathbf{Y}_2\mathbf{Z}_{S_2}$ in Lemma 12, where $\mathbf{U}_{[1:2]}$, $\mathbf{Y}_2$, $\mathcal{S}_2$, $\mathbf{Z}_{S_2}$ are as in protocol A. For

$\epsilon'' > 0, j = 1, 2$, by choosing

$$\gamma_{1,2} = (1 - \epsilon'')[\mu_2 H(U_1|U_2, X) + (n - \mu_2)H(U_1|U_2, Y_2)] \tag{7.87}$$

$$\gamma_{2,1} = (1 - \epsilon'')[\mu_2 H(U_2|U_1, X) + (n - \mu_2)H(U_2|U_1 Y_2)], \tag{7.88}$$

$$\gamma_j = (1 - \epsilon'')[\mu_2 H(U_j|X) + (n - \mu_2)H(U_j|Y_2)], \tag{7.89}$$

using Hoeffding's inequality, there exists $\tilde{\beta} > 0$ such that for all $S_2 \in \mathcal{S}_2$, $j = 1, 2$,

$$\mathbb{P}_{p_{\mathbf{U}_{[1:2]}\mathbf{Y}_2\mathbf{Z}_{S_2}}} \left(\left(\mathbf{U}_{[1:2]}, \mathbf{Y}_2 \mathbf{Z}_{S_2} \right) \notin \mathcal{D}_j^{S_2} \right) \leq \exp(-\tilde{\beta} n) = \delta^2. \tag{7.90}$$

Note that $\lim_{n\to\infty} \delta^2 = 0$, and hence, for n sufficiently large, $\delta^2 \in (0, \frac{1}{4})$. Thus, the conditions of Lemma 12 are satisfied.

Substituting the choices for $\tilde{W}_1$, $\tilde{W}_2$, $\tilde{F}_1$, $\tilde{F}_2$, $\gamma_2, \gamma_{1,2}$, and $|\mathcal{S}_2||\mathcal{Z}^n| \leq (2(|\mathcal{X}| + 1)|\mathcal{Y}_2|)^n$ in (7.38), we have, for all $\epsilon, \epsilon_1 > 0$, $\tilde{\epsilon} = \epsilon + \epsilon_1$, there exists $n^* \in \mathbb{N}$ and $\kappa_\epsilon, \tilde{\kappa} > 0$ such that for all $n \geq n^*$,

$$\mathbb{P}_{\mathcal{B}} \left(\max_{S_2 \in \mathcal{S}_2} \mathbb{D} \left(\tilde{P}_{W_1 F_{[1:2]} \mathbf{Y}_2 \mathbf{Z}_{S_2}} || p^U_{W_1} p^U_{F_{[1:2]}} p_{\mathbf{Y}_2 \mathbf{Z}_{S_2}} \right) \geq 2\tilde{\epsilon} \right) \leq \exp\left(-\kappa_\epsilon e^{\tilde{\kappa} n} \right), \tag{7.91}$$

as long as,

$$R_1 + \tilde{R}_1 \leq (1 - \epsilon'')[\alpha_2 H(U_1|U_2, X) + (1 - \alpha_2)H(U_1|U_2, Y_2)] \tag{7.92}$$

$$\tilde{R}_2 \leq (1 - \epsilon'')[\alpha_2 H(U_2|X) + (1 - \alpha_2)H(U_2|Y_2)]. \tag{7.93}$$

By the Borel-Cantelli lemma, it follows from (7.91) that

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\max_{S_2 \in \mathcal{S}_2} \mathbb{D}\left(\tilde{P}_{W_1 F_{[1:2]} \mathbf{Y}_2 \mathbf{Z}_{S_2}} || p^U_{W_1} p^U_{F_{[1:2]}} p_{\mathbf{Y}_2 \mathbf{Z}_{S_2}} \right) > 0 \right) = 0 \tag{7.94}$$

Similarly, setting $X_1 = \mathbf{U}_1$, $X_2 = \mathbf{U}_2$, $\tilde{W}_1 = 1$, $\tilde{F}_1 = 2^{n\tilde{R}_1}$, $\tilde{W}_2 = 2^{nR_2}$, $\tilde{F}_2 = 2^{n\tilde{R}_2}$, $\mathcal{S} = \mathcal{S}_1$, and $Z_S = \mathbf{Y}_1 \mathbf{Z}_{S_1}$ in Lemma 12 and using the choices for γ_1, γ_2, $\gamma_{1,2}$, $\gamma_{2,1}$ in (7.87)-(7.89), but with replacing μ_2 and Y_2 by μ_1 and Y_1, gives

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\max_{S_1 \in \mathcal{S}_1} \mathbb{D}\left(\tilde{P}_{W_2 F_{[1:2]} \mathbf{Y}_1 \mathbf{Z}_{S_1}} || p^U_{W_2} p^U_{F_{[1:2]}} p_{\mathbf{Y}_1 \mathbf{Z}_{S_1}} \right) > 0 \right) = 0, \tag{7.95}$$

as long as,

$$R_2 + \tilde{R}_2 \leq \alpha_1 H(U_2|U_1, X) + (1-\alpha_1) H(U_2|U_1, Y_1) \tag{7.96}$$

$$\tilde{R}_1 \leq \alpha_1 H(U_1|X) + (1-\alpha_1) H(U_1|Y_1). \tag{7.97}$$

Remark 22. *Note that we have considered two problems, where in each problem, one source encoder is communicating its key reliably to the corresponding decoder and securely from the other user's decoder, c.f. (7.85), (7.94), and (7.95). In each of these two problems, both the public messages F_1 and F_2 are required to be independent from W_j and $\mathbf{Y}_j$, and $\mathbf{Z}_{S_j}$, cf. (7.94) and (7.95). The reason is that, after converting these conditions to the channel model in protocol B, we need to eliminate the common randomness $F_{[1:2]}$ from the model by conditioning on a certain instance of it while preserving the uniformity of the message W_j, $j = 1, 2$, and its independence from the other receiver's observations.*

7.5.4 Converting reliability and secrecy properties to protocol B

First, for the reliability conditions, using the triangle inequality, it follows from (7.74), (7.77), and (7.85), that, for $j = 1, 2$,

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j} \hat{\mathbf{U}}_j}, P_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j}} \mathbb{1}\{\hat{\mathbf{U}}_j = \mathbf{U}_j\} \right) \right) = 0. \tag{7.98}$$

For the secrecy conditions, by the union bound, (7.81), (7.94),

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\max_{S_2 \in \mathcal{S}_2} \mathbb{D} \left(P_{W_1 F_{[1:2]} \mathbf{Y}_2 \mathbf{Z}_{S_2}} || p^U_{W_1} p^U_{F_{[1:2]}} p_{\mathbf{Y}_2 \mathbf{Z}_{S_2}} \right) > 0 \right) \\ &\quad \le \lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\max_{S_2 \in \mathcal{S}_2} \mathbb{D} \left(\tilde{P}_{W_1 F_{[1:2]} \mathbf{Y}_2 \mathbf{Z}_{S_2}} || p^U_{W_1} p^U_{F_{[1:2]}} p_{\mathbf{Y}_2 \mathbf{Z}_{S_2}} \right) > 0 \right) \\ &\qquad + \lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\mathbb{V} \left(\tilde{P}_{W_1 F_{[1:2]}}, p^U_{W_1} p^U_{F_{[1:2]}} \right) > 0 \right) = 0. \end{aligned} \tag{7.99}$$

Similarly, using the union bound, (7.81) and (7.95), we have,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{B}} \left(\max_{S_1 \in \mathcal{S}_1} \mathbb{D} \left(P_{W_2 F_{[1:2]} \mathbf{Y}_1 \mathbf{Z}_{S_1}} || p^U_{W_2} p^U_{F_{[1:2]}} p_{\mathbf{Y}_1 \mathbf{Z}_{S_1}} \right) > 0 \right) = 0. \tag{7.100}$$

Note that the reliability and secrecy conditions for the original channel model in protocol B, (7.98)-(7.100), are averaged over the random binning of the dual source model in protocol A, where this binning determines the encoders and decoders for the dual source model and hence the encoders and decoders for the original channel model as well, cf., (7.72). By applying the selection lemma to (7.98)-(7.100), there is a binning realization $\mathbf{b}^*$, with a corresponding joint distribution p^* for the original channel model

in protocol B such that

$$\lim_{n\to\infty} \mathbb{V}\left(p^*_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j} \hat{\mathbf{U}}_j}, p^*_{W_j F_{[1:2]} \mathbf{U}_j \mathbf{Y}_j \mathbf{Z}_{S_j}} \mathbb{1}\{\hat{\mathbf{U}}_j = \mathbf{U}_j\} \right) = 0, \qquad j = 1, 2, \quad (7.101)$$

$$\lim_{n\to\infty} \mathbb{1}\left\{ \max_{S_2 \in \mathcal{S}_2} \mathbb{D}\left(p^*_{W_1 F_{[1:2]} \mathbf{Y}_2 \mathbf{Z}_{S_2}} || p^U_{W_1} p^U_{F_{[1:2]}} p_{\mathbf{Y}_2 \mathbf{Z}_{S_2}} \right) > 0 \right\} = 0, \quad (7.102)$$

$$\lim_{n\to\infty} \mathbb{1}\left\{ \max_{S_1 \in \mathcal{S}_1} \mathbb{D}\left(p^*_{W_2 F_{[1:2]} \mathbf{Y}_1 \mathbf{Z}_{S_1}} || p^U_{W_2} p^U_{F_{[1:2]}} p_{\mathbf{Y}_1 \mathbf{Z}_{S_1}} \right) > 0 \right\} = 0, \quad (7.103)$$

where $W_j = b^*_{1j}(\mathbf{U}_j)$ and $F_j = b^*_{2j}(\mathbf{U}_j)$, $j = 1, 2$.

7.5.5 Eliminating the common randomness

By introducing the $\hat{W}$ variables to the distributions in (7.101) as deterministic functions of the $\hat{\mathbf{U}}$ variables, and using (7.103), (7.105), and the union bound, we have

$$\lim_{n\to\infty} \mathbb{E}_{F_{[1:2]}} \left(\mathbb{P}_{p^*}(\hat{W}_j \neq W_j | F_{[1:2]}) \right) = 0, \quad j = 1, 2, \quad (7.104)$$

$$\lim_{n\to\infty} \mathbb{P}_{F_{[1:2]}} \left(\max_{S_2 \in \mathcal{S}_2} \mathbb{D}\left(p^*_{W_1 \mathbf{Y}_2 \mathbf{Z}_{S_2} | F_{[1:2]}} || p^U_{W_1} p^*_{\mathbf{Y}_2 \mathbf{Z}_{S_2} | F_{[1:2]}} \right) > 0 \right) = 0, \quad (7.105)$$

$$\lim_{n\to\infty} \mathbb{P}_{F_{[1:2]}} \left(\max_{S_1 \in \mathcal{S}_1} \mathbb{D}\left(p^*_{W_2 \mathbf{Y}_1 \mathbf{Z}_{S_1} | F_{[1:2]}} || p^U_{W_2} p^*_{\mathbf{Y}_1 \mathbf{Z}_{S_1} | F_{[1:2]}} \right) > 0 \right) = 0. \quad (7.106)$$

Applying the selection lemma to (7.104)-(7.106) results in the existence of $f^*_{[1:2]}$ such that the reliability and secrecy constraints in (7.7)-(7.9) are satisfied. We hence identify the encoder and decoders for the original model as $p(\mathbf{x}|\mathbf{u}_{[1:2]})\tilde{p}^*(\mathbf{u}_{[1:2]}|w_{[1:2]}, f^*_{[1:2]})$ and $(\tilde{p}^*(\hat{\mathbf{u}}_j|\mathbf{y}_j, \mathbf{z}, f^*_j), b^*_{1j}(\hat{\mathbf{u}}_j), j = 1, 2)$; $\tilde{p}^*$ is the induced distribution in protocol A that corresponds to the binning $\mathbf{b}^*$.

Finally, combining the rate conditions in (7.75), (7.76), (7.78), (7.79),(7.86), (7.92), (7.93), (7.96),and (7.97), while taking $\epsilon', \epsilon'' \to \infty$, results in the rate region

in (7.19)-(7.20). The convex hull follows by time sharing independent codes. This completes the proof for Theorem 12.

The proof for Theorem 13 follows similar steps as in the proof for Theorem 12. The difference is that the auxiliary variables U_1 and U_2 are independent, where at the beginning in the proof, we fix the distribution $p_{U_{1:2}X_{[1:2]}}$ which factorizes as $p_{U_1}p_{U_2}p_{X_1|U_1}p_{X_2|U_2}$.

7.6 Conclusion

In this chapter, we have extended our generalized wiretap channel model in Chapter 5 to the three multi-receiver settings. First, we have considered a broadcast wiretap channel with a wiretapper which noiselessly taps into a subset of its choice of the transmitted symbols and observes the remainder through a noisy channel. We have derived an achievable strong secrecy rate region for the model, which extends Marton's inner bound and characterizes the secrecy penalty due to the noiseless observations at the wiretapper. We also have characterized the secrecy capacity for two classes of the generalized broadcast wiretap channel.

Second, we have studied a generalized model for the two-user interference channel with confidential messages, where the receivers, besides their noisy observations, are provided with fixed-length subsets of their choice of noiseless observations for transmitted codeword symbols of the both users. Third, we have proposed a generalized broadcast channel with confidential messages model, with each receiver is provided with a subset of its choice of noiseless observations for the transmitted codeword. We have derived achievable strong secrecy rate regions for the two models. For both models, the size of the subset at each receiver gives rise to a trade-off between the rates of the two receivers,

which is demonstrated in the derived rate regions. We have also highlighted the special case of the generalized broadcast channel with confidential messages, with one receiver's noisy observations are degraded with respect to the other receiver, and only this degraded receiver is provided with a subset of noiseless observations. The achievable rate region for this case shows that the receiver, with the degraded noisy observations, achieves a positive secrecy rate after a certain threshold on the ratio of his noiseless observations.

Chapter 8

The Caching Broadcast Channel with a Wire and Cache Tapping Adversary of Type II

8.1 Introduction

Caching is proposed to efficiently reduce network traffic congestion by storing partial contents at the cache memories of end users earlier during off-peak times, providing local caching gain [4,17,25]. More recently, reference [72] has shown that a careful design of cache contents at the end users in a multi-receiver setting allows the transmitter to send delivery transmissions that are simultaneously useful for many users, providing a *global caching gain*. This gain depends on the aggregate cache memory of the network and demonstrates the ability of coding over delivery transmission and/or cache contents to offset lack of cooperation between the receivers.

In numerous works to date, coded caching has been studied under various modeling assumptions and for various network configurations, including online, decentralized, hierarchical caching [7, 54, 74, 97], non-uniform demands [93], more users than files [101, 117], heterogeneous cache sizes [13, 48], improved bounds [6, 65, 129], interference networks [36, 73, 80], combination networks [52, 131], device-to-device communication [47, 50], and broadcast channels [5, 12, 32, 133].

Coded caching with confidentiality requirements has recently been studied in references [8,53,98,103,130–132]. These references assume secure cache placement, i.e., the

adversary cannot tap into the cache contents or into the communication which performs the cache placement. At the other extreme, if cache placement were to be public, i.e., the adversary were to have perfect access to the cache contents, it follows in a straightforward fashion from [2, 106] that the presence of cache memories cannot increase the secrecy capacity. Given the results of these two extreme settings, this chapter considers an intermediate scenario in which the adversary may have *partial access* to cache placement. The wiretap channel II provides a model for an adversary which has partial access to the legitimate communication, in the form of a threshold on the time fraction during which the adversary is capable of tapping into the communication.

In this chapter, we consider an adversary model of type II as in Sections 4-7, but in a cache-aided communication setting. In particular, the adversary noiselessly observes a partial subset of its choosing of the transmitted symbols over the cache placement and/or delivery phases. Thus we term this model the caching broadcast channel with a *wire and cache tapping* adversary of type II (CBC-WCT II). The legitimate parties do not know whether the cache placement, delivery, or both transmissions are tapped, the relative fractions of tapped symbols in each, or their positions. Only the knowledge of the overall size of the tapped set by the adversary is available to the legitimate terminals.

The challenge in caching stems from the fact that the transmitter, which has access to a library of files, has no knowledge about the future demands of end users when designing their cache contents. This remains to be the case when security against an external adversary is concerned. Additionally, for the adversary model in consideration, the adversary might tap into cache placement, delivery, or both, and where the tapping occurs is *unknown* to the legitimate parties. We show that even under these unfavorable

conditions, strong secrecy guarantees can be provided that are invariant to the positions of the tapped symbols varying between cache placement, delivery, or both.

In coded caching literature up to date, the physical communication which populates the cache memories at end users does not need to be considered in the problem formulation, due to the assumption of secure cache placement. By contrast, in order to model cache placement that is tapped by an adversary, we consider a length-n communication block over a two-user broadcast wiretap II channel, as in Chapter 7. The sizes of cache memories at the receivers are fixed in this model. We note that introducing variable memory sizes for which a rate-memory tradeoff can be characterized, as in the usual setup for caching, requires considering additional communication blocks for cache placement. Being of future interest, we comment on this extension to multiple communication blocks for cache placement in the Discussion section, Section 8.7. We as well provide reasoning for our choice of the broadcast setting for cache placement.

In summary, the contributions of this work are summarized as follows:

1. We introduce the notion of *cache-tapping* into the information theoretic models of coded caching, in which an adversary of type II is able to tap into a fixed-size subset of symbols of its choosing either from cache placement, delivery, or both transmissions.

2. We characterize the strong secrecy capacity of the model, i.e., the maximum achievable file rate which keeps the overall library strongly secure, for the instance of a transmitter's library with two files:

- We devise an achievability scheme which integrates wiretap coding [96], security embedding codes [63, 70], one-time pad keys [106], *coded* cache placement and *uncoded* delivery [72].
- We utilize a genie-aided upper bound, in which a genie provides the transmitter with user demands before cache placement, rendering the model to a two-user broadcast wiretap II channel, in order to establish the converse for this case.

3. We derive lower and upper bounds on the strong secrecy file rate when the transmitter's library has more than two files:
   - We utilize a coding scheme that is similar to the scheme we used for two files. However, the cache placement and delivery schemes we employ to achieve the rates differ from those utilized for two files. In particular, we utilize here *uncoded* cache placement and a *partially coded* delivery.
   - We derive the upper bound in three steps. First, we consider a transformed channel with an adversary which can tap an equal fraction of symbols to our model, but is only allowed to tap into the delivery phase. Since this adversary has a more restricted strategy space than the original one, the corresponding secrecy capacity is at least as large as our original model. Next, we utilize Sanov's theorem in method of types [21, Theorem 11.4.1] to further upper bound the secrecy capacity for the restricted adversary model by the secrecy capacity when the adversary encounters a discrete memoryless binary erasure channel. Finally, the secrecy capacity of the discrete memoryless model is

upper bounded by the secrecy capacity of a single receiver setting in which the receiver requests two files from the library.

The remainder of the chapter is organized as follows. Section 8.2 describes the communication system proposed in this chapter. Section 8.3 presents the main results. The proofs for these results are provided in Sections 8.4, 8.5, and 8.6. Section 8.7 provides a discussion about the communication model in question and the presented results. Section 8.8 concludes the chapter.

8.2 System Model

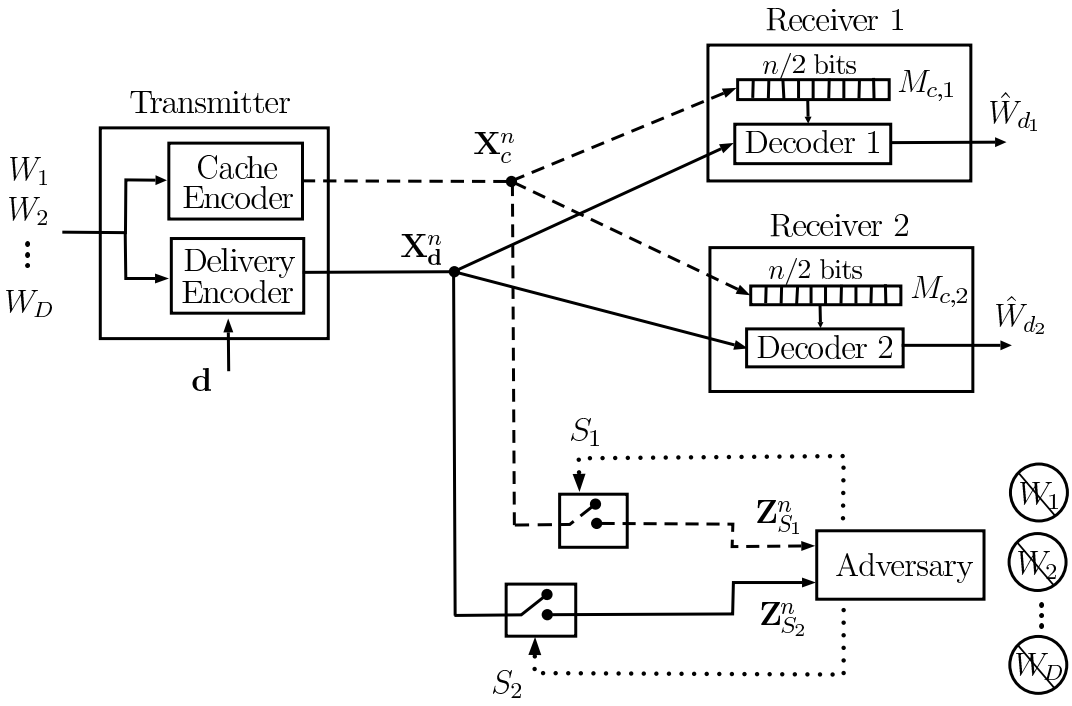


Fig. 8.1. The caching broadcast channel with a wire and cache tapping adversary of type II (CBC-WCT II). The adversary chooses tapping sets S_1 and S_2 in placement and delivery.

Consider the communication system depicted in Fig. 8.1, in which the adversary has the ability to tap into both the cache placement and delivery transmissions. The transmitter observes $D \geq 2$ independent messages (files), $W_1, W_2, \cdots, W_D$, each of which is uniformly distributed over $[1 : 2^{nR_s}]$. Each receiver has a cache memory of size $\frac{n}{2}$ bits. The communication occurs over two phases: placement and delivery. The broadcast channel is noiseless during both phases. The communication model is described as follows:

Cache placement phase: During this phase, the transmitter broadcasts a length-n binary signal, $\mathbf{X}_c^n \in \{0,1\}^n$, to both receivers. The codeword $\mathbf{X}_c^n$ is a function of the library files, i.e., $\mathbf{X}_c^n \triangleq f_c\left(W_{[1:D]}\right)$. The transmitter does not know the receiver demands during cache placement [72]. Each receiver has a cache memory of size $\frac{n}{2}$ bits in which they store a function of $\mathbf{X}_c^n$, $M_{c,j} \triangleq f_{c,j}(\mathbf{X}_c^n)$, where $f_{c,j} : \{0,1\}^n \mapsto [1 : 2^{\frac{n}{2}}]$ and $j = 1, 2$.

Delivery phase: At the beginning of the delivery phase, the two receivers announce their demands $\mathbf{d} \triangleq (d_1, d_2) \in [1 : D]^2$ to the transmitter. The transmitter, in order to satisfy the receiver demands, encodes $W_{[1:D]}$ and $\mathbf{d}$ into the binary codeword $\mathbf{X}_\mathbf{d}^n \in \{0,1\}^n$. In particular, for each $\mathbf{d}$, the transmitter uses the encoder $f_\mathbf{d} : [1 : 2^{nR_s}]^D \mapsto \{0,1\}^n$ and sends $\mathbf{X}_\mathbf{d}^n \triangleq f_\mathbf{d}\left(W_{[1:D]}\right)$.

Decoding: Receiver j utilizes the decoder $g_{\mathbf{d},j} : [1 : 2^{\frac{n}{2}}] \times \{0,1\}^n \mapsto [1 : 2^{nR_s}]$, in order to output the estimate $\hat{W}_{d_j} \triangleq g_{\mathbf{d},j}\left(f_{c,j}(\mathbf{X}_c^n), \mathbf{X}_\mathbf{d}^n\right)$ of its desired message W_{d_j}, where $j = 1, 2$.

Adversary model: The adversary chooses two subsets $S_1, S_2 \subseteq [1 : n]$. The size of the sum of cardinalities of S_1 and S_2 is fixed. That is, for $|S_1| = \mu_1$, $|S_2| = \mu_2$, $\mu_1, \mu_2 \leq n$, we have $\mu_1 + \mu_2 = \mu$. The subsets S_1 and S_2 indicate the positions tapped

by the adversary during the cache placement and delivery phases, respectively. Over the two phases, the adversary observes the length-$2n$ sequence $\mathbf{Z}_S^{2n} = [\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n] \in \mathcal{Z}^{2n}$; $\mathbf{Z}_{S_j}^n \triangleq [Z_{S_j,1}, Z_{S_j,2}, \cdots, Z_{S_j,n}] \in \mathcal{Z}^n$, $j = 1, 2$,

$$Z_{S_1,i} = \begin{cases} X_{c,i},\ i \in S_1 \\ ?,\ i \notin S_1 \end{cases}, \text{ and } \qquad Z_{S_2,i} = \begin{cases} X_{\mathbf{d},i},\ i \in S_2 \\ ?,\ i \notin S_2. \end{cases} \tag{8.1}$$

The alphabet is $\mathcal{Z} = \{0, 1, ?\}$, where "?" denotes an erasure.

The legitimate terminals know neither the realizations of S_1 and S_2, nor the values of μ_1 and μ_2. Only μ is known. Let us define $\alpha_1 = \frac{\mu_1}{n}$ and $\alpha_2 = \frac{\mu_2}{n}$ as the fractions of the tapped symbols in the cache placement and delivery phases, and let $\alpha = \alpha_1 + \alpha_2$ be the overall tapped ratio. Notice that $\alpha_1, \alpha_2 \in [0, 1]$ and $\alpha \in (0, 2]$.

Remark 23. *We consider that α is strictly greater than zero, i.e., the adversary is present. For $\alpha = 0$, i.e., no adversary, the problem considered in this chapter has been extensively studied in the literature, see for example [18, 49, 72, 115].*

A channel code $\mathcal{C}_{2n}$ for this model consists of

- D message sets; $\mathcal{W}_l \triangleq [1 : 2^{nR_s}]$, $l = 1, 2, \cdots, D$,
- Cache encoder; $f_c : [1 : 2^{nR_s}]^D \mapsto \{0, 1\}^n$,
- Cache decoders; $f_{c,j} : \{0, 1\}^n \mapsto [1 : 2^{\frac{n}{2}}]$, $j = 1, 2$,
- Delivery encoders; $\left\{f_{\mathbf{d}} : \mathbf{d} \in [1 : D]^2\right\}$, $f_{\mathbf{d}} : [1 : 2^{nR_s}]^D \mapsto \{0, 1\}^n$,
- Decoders; $\left\{g_{\mathbf{d},j} : j = 1, 2,\ \mathbf{d} \in [1 : D]^2\right\}$, $g_{\mathbf{d},j} : [1 : 2^{\frac{n}{2}}] \times \{0, 1\}^n \mapsto [1 : 2^{nR_s}]$.

The file rate R_s is said to be achievable with strong secrecy if there exists a sequence of channel codes $\{\mathcal{C}_{2n}\}_{n\geq 1}$ satisfying

$$\lim_{n\to\infty} \max_{\mathbf{d}\in[1:D]^2} \mathbb{P}\left(\bigcup_{j=1,2} (\hat{W}_{d_j} \neq W_{d_j})\right) = 0, \tag{8.2}$$

$$\text{and} \quad \lim_{n\to\infty} \max_{\substack{S_1,S_2\subseteq[1:n]:\\ |S_1|+|S_2|\leq\mu}} I\left(W_{[1:D]}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) = 0. \tag{8.3}$$

That is, R_s is the symmetric secure file rate, under any demand vector and adversarial strategy. The strong secrecy capacity C_s is the the supremum of all achievable R_s.

Remark 24. *While we consider the file rate R_s which guarantees reliability for the worst-case demand vector, the average rate for which there exists a prior distribution on the demands has been studied in coded caching literature as well; see for example [51, 65, 93].*

Remark 25. *The condition in (8.3) guarantees strong secrecy against all possible strategies for the adversary, i.e., choices of the subsets S_1 and S_2 which satisfy the condition $|S_1| + |S_2| \leq \mu$.*

8.3 Main Results

For clarity of exposition, we first study the model described in Section 8.2 when the transmitter has two files in its library, i.e., $D = 2$. We then extend the ideas and the analysis to the case of a library with more than two files, i.e., $D > 2$. For $D > 2$, we utilize a channel coding scheme that is similar to the scheme we construct for $D = 2$, but the cache placement and delivery schemes that achieve the best rates are different from those used for $D = 2$.

The following theorem presents the strong secrecy capacity for $D = 2$.

Theorem 14. *For $0 < \alpha \leq 2$ and $D = 2$, the strong secrecy capacity for the caching broadcast channel with a wire and cache tapping adversary of type II (CBC-WCT II), described in Section 8.2, is given by*

$$C_s(\alpha) = 1 - \frac{\alpha}{2}. \tag{8.4}$$

Proof: The proof is provided in Section 8.4. ■

Theorem 15 below presents an achievable strong secrecy file rate for $D > 2$.

Theorem 15. *For $0 < \alpha \leq 2$ and $D > 2$, the achievable strong secrecy file rate for the CBC-WCT II is*

$$R_s(\alpha) \geq \begin{cases} \frac{1}{2} + \frac{3(1-\alpha)}{4D}, & 0 < \alpha < 1 \\ 1 - \frac{\alpha}{2}, & 1 \leq \alpha \leq 2. \end{cases} \tag{8.5}$$

Proof: The proof is provided in Section 8.5. ■

The following theorem upper bounds the secure file rate when $D > 2$.

Theorem 16. *For $0 < \alpha \leq 2$ and $D > 2$, the achievable strong secrecy file rate for the CBC-WCT II is upper bounded as*

$$R_s(\alpha) \leq \begin{cases} \frac{1}{2} + \frac{2D-1}{2D(D-1)}(1-\alpha), & 0 < \alpha < 1 \\ 1 - \frac{\alpha}{2}, & 1 \leq \alpha \leq 2. \end{cases} \tag{8.6}$$

Proof: The proof is provided in Section 8.6. ■

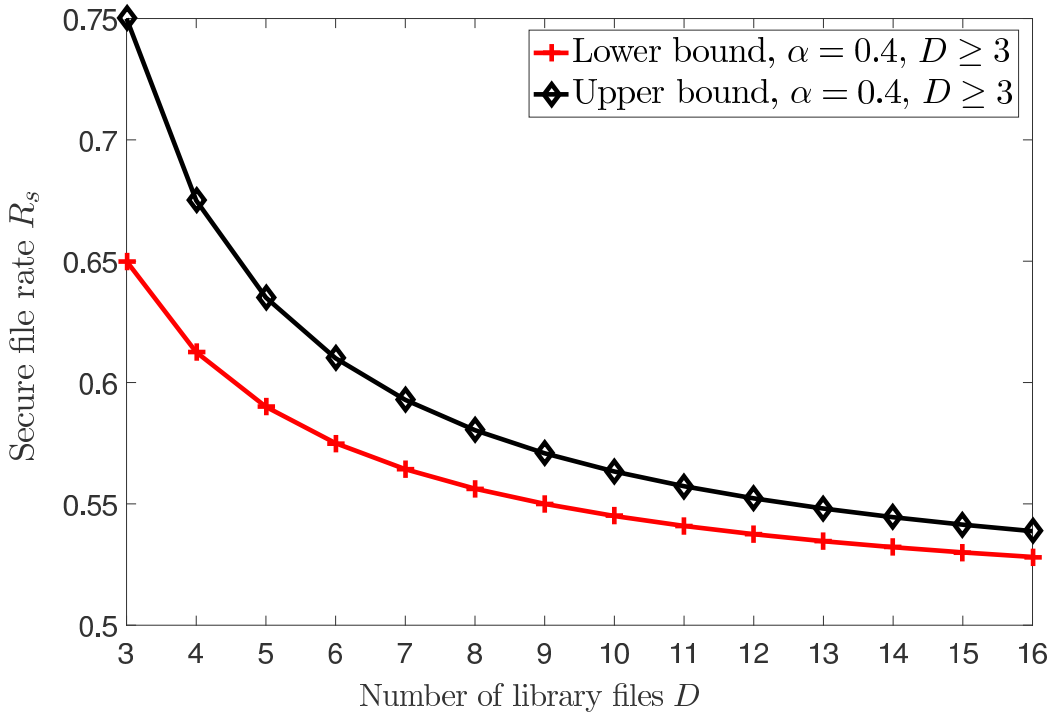


Fig. 8.2. Bounds on the achievable strong secrecy file rate R_s, when $\alpha = 0.4$ and $D \geq 3$.

The following corollary is immediate from Theorems 14, 15, and 16.

Corollary 7. *For $1 \leq \alpha \leq 2$, that is when the adversary can tap longer than one phase of communication, the strong secrecy capacity for the CBC-WCT II is*

$$C_s(\alpha) = 1 - \frac{\alpha}{2}. \tag{8.7}$$

Remark 26. *When $\alpha \in [1, 2]$, i.e., $n \leq \mu \leq 2n$, two possible strategies for the adversary are $\{S_1 = [1:n], S_2 \subset [1:n]\}$ or $\{S_1 \subset [1:n], S_2 = [1:n]\}$. That is, the adversary can tap into all symbols in one phase and a subset of symbols in the other phase. Interestingly,*

a positive strong secrecy capacity is achievable against such an adversary. We elaborate more on the intuition behind this result in the Discussion section.

Unlike for $1 \leq \alpha \leq 2$, for $0 < \alpha < 1$, the lower and upper bounds in (8.5) and (8.6) have a gap. For illustration purposes, these bounds are plotted for $\alpha = 0.4$ in Fig. 8.2.

Remark 27. *When $\alpha = 0$, i.e., no adversary, our achievability scheme for $D > 2$ described in Section 8.5 reduces to the achievability scheme in [72], which is shown to achieve the optimal rate-memory tradeoff for the case of two users and a library size of three or larger [18, 115]. The upper bound for $D > 2$ derived in this work is to address the intricacies of the adversarial model and is useful only when the adversary is present ($\alpha > 0$), i.e., (8.6) is loose when $\alpha = 0$.*

8.4 Proof of Theorem 14

In this section, we prove Theorem 14, which identifies the strong secrecy capacity for the model in Section 8.2 when $D = 2$. Recall that the demand vector is denoted by $\mathbf{d} = (d_1, d_2)$, where $d_1, d_2 \in \{1, 2\}$.

8.4.1 Converse

For the model in Theorem 14, when the demand vector $\mathbf{d}$ is known to the transmitter during cache placement, the model reduces to a broadcast wiretap channel II, over a length-$2n$ communication block. The strong sum secrecy rate for that model, $2R_s$, is

upper bounded by

$$2R_s \leq 2 - \alpha, \tag{8.8}$$

which follows from 9 in Chapter 7. Notice that (8.8) holds for any $\mathbf{d} = (d_1, d_2)$ such that $d_1 \neq d_2$, which represents the worst-case demands. Since the demand vector is unknown for the model in consideration, $1 - \frac{\alpha}{2}$ is an upper bound for its strong secrecy capacity.

8.4.2 Restricted Adversary Models as Building Blocks

Before proceeding with the achievability proof, it is relevant to take a step back and investigate the secrecy capacity when a known fraction of cache placement, a known fraction of delivery, or both, is tapped. In particular, we consider that the adversary taps into (i) cache placement only, (ii) delivery only, or (iii) both and the relative fractions of tapped symbols in each are known. For these three models, we show that the strong secrecy file rate in (8.4), i.e., $1 - \frac{\alpha}{2}$, is achievable, and hence determines their strong secrecy capacities. We then use these models as building blocks for when the relative fractions are *unknown*, and provide the achievability proof in Sections 8.4.3 and 8.4.4.

8.4.2.1 Setting 1: The adversary taps into cache placement only

Let $\alpha_1 = \alpha$ ($\alpha_2 = 0$), and $|S_1| = \mu$ ($S_2 = \varnothing$). That is, the adversary taps into cache placement only. Consider that the transmitter and the receivers know that $\alpha_1 = \alpha$. We show that $1 - \frac{\alpha}{2}$ is an achievable strong secrecy file rate for this setting.

The transmitter divides the message $W_l,\ l = 1, 2$, into three independent messages, $W_l^{(1)}$, $W_l^{(2)}$, and $W_{l,s}$, where $W_l^{(1)}$, $W_l^{(2)}$, are uniform over $\left[1 : 2^{n\frac{1-\alpha-\epsilon_n}{2}}\right]$, and $W_{l,s}$ is uniform over $\left[1 : 2^{n\frac{\alpha+\epsilon_n}{2}}\right]$. Define

$$M_c = \left\{M_{c,1}, M_{c,2}\right\}; \quad M_{c,1} = W_1^{(1)} \oplus W_2^{(1)}, \quad M_{c,2} = W_1^{(2)} \oplus W_2^{(2)}, \tag{8.9}$$

$$M_{\mathbf{d}} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s}, W_{d_2,s}\right\}. \tag{8.10}$$

During cache placement, the transmitter maps M_c into $\mathbf{X}_c^n$ using stochastic encoding, i.e., wiretap coding [121]. Since the rate of M_c is less than $1-\alpha$, M_c is strongly secure from the adversary which observes $n\alpha$ symbols of $\mathbf{X}_c^n$ [33,91]. During the delivery phase, the transmitter sends $\mathbf{X}_{\mathbf{d}}^n$ as the binary representation of $M_{\mathbf{d}}$ which is of length n bits, since the delivery phase is noiseless and secure.

Using $\mathbf{X}_c^n$, noiselessly received during cache placement, receiver j, $j = 1, 2$, recovers $M_{c,j}$ and stores it in its cache memory. Notice that the size of $M_{c,j}$, for $j = 1, 2$, is smaller than $\frac{n}{2}$ bits, i.e., the cache size at each receiver. Using $\mathbf{X}_{\mathbf{d}}^n$, received noiselessly during delivery, both receivers perfectly recover $M_{\mathbf{d}}$. Using $M_{\mathbf{d}}$ along with its cache contents, i.e., $M_{c,j}$, and for n sufficiently large[1], receiver j correctly recovers its desired message W_{d_j}, $j = 1, 2$.

For secrecy, we show in Appendix K that (8.3), which reduces to

$$\lim_{n\to\infty} \max_{S_1\subseteq[1:n]:\ |S_1|=\mu} I(W_1, W_2; \mathbf{Z}_{S_1}^n) = 0, \tag{8.11}$$

[1]Large block-length n is needed in order to ensure a valid subpacketization of the file W_l into the sub-files $\{W_l^{(1)}, W_l^{(2)}, W_{l,s}\}$, for $l = 1, 2$. That is, a bijective map between the file and its sub-files is preserved.

is satisfied. Since $\epsilon_n \to 0$ as $n \to \infty$, the achievable strong secrecy file rate is given by

$$R_s(\alpha) = 2 \times \frac{1-\alpha}{2} + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}. \tag{8.12}$$

8.4.2.2 Setting 2: The adversary taps into the delivery only

This setting corresponds to $\alpha_1 = 0$ and $\alpha_2 = \alpha$, and the transmitter and receivers possess this knowledge. Once again, we show that $1 - \frac{\alpha}{2}$ is an achievable strong secrecy file rate.

The transmitter performs the same division of W_l, $l = 1, 2$, as in Setting 1. In addition, the transmitter generates the keys K_1, K_2, each is uniform over $\left[1 : 2^{n\frac{\alpha+\epsilon_n}{2}}\right]$, independent from one another and from W_1, W_2. In this case, we define M_c, $M_{\mathbf{d}}$, and $\tilde{M}_{\mathbf{d}}$, as follows.

$$M_c = \left\{M_{c,1}, M_{c,2}\right\}; \;\; M_{c,1} = \left\{W_1^{(1)} \oplus W_2^{(1)}, K_1\right\}, \;\; M_{c,2} = \left\{W_1^{(2)} \oplus W_2^{(2)}, K_2\right\}, \tag{8.13}$$

$$M_{\mathbf{d}} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}\right\}, \tag{8.14}$$

$$\tilde{M}_{\mathbf{d}} = \left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\right\}. \tag{8.15}$$

During cache placement, the transmitter sends $\mathbf{X}_c^n$ as the binary representation of M_c, and receiver j, $j = 1, 2$, stores $M_{c,j}$ in its cache memory. During delivery, the transmitter encodes $M_{\mathbf{d}}$ into $\mathbf{X}_{\mathbf{d}}^n$ using wiretap coding, while using $\tilde{M}_{\mathbf{d}}$ as the randomization message. Receiver j recovers $M_{\mathbf{d}}$ and $\tilde{M}_{\mathbf{d}}$, using which, along with $M_{c,j}$, it correctly decodes W_{d_j}, for sufficiently large n. By contrast, the adversary can only retrieve $\tilde{M}_{\mathbf{d}}$ using which it can gain no information about W_1 and W_2. We show in Appendix L that

(8.3), i.e.,

$$\lim_{n\to\infty} \max_{S_2\subseteq[1:n]:\ |S_2|=\mu} I(W_1, W_2; \mathbf{Z}^n_{S_2}) = 0, \tag{8.16}$$

is satisfied. The achievable strong secrecy file rate is again $1 - \frac{\alpha}{2}$.

8.4.2.3 Setting 3: The legitimate terminals know the values of α_1 and α_2

For this setting, neither $\alpha_1 = 0$ nor $\alpha_2 = 0$. However, the transmitter and receivers know the values of α_1 and α_2. Under these assumptions, the scheme which achieves the strong secrecy rate of $1 - \frac{\alpha}{2}$ depends on whether $\alpha_1 \geq \alpha_2$. For $\alpha_1 \geq \alpha_2$, we utilize an achievability scheme similar to Setting 1; for $\alpha_1 < \alpha_2$, we utilize an achievability scheme similar to Setting 2.

Case 1: $\alpha_1 \geq \alpha_2$: The transmitter divides W_l, $l = 1, 2$, into the independent messages $\left\{W_l^{(1)}, W_l^{(2)}, W_{l,s}\right\}$; $W_l^{(1)}, W_l^{(2)}$ are uniform over $\left[1 : 2^{n\frac{1-\alpha_1-\epsilon_n}{2}}\right]$ and $W_{l,s}$ is uniform over $\left[1 : 2^{n\frac{\alpha_1-\alpha_2}{2}}\right]$. The transmitter forms M_c and $M_{\mathbf{d}}$ as in (8.9) and (8.10), and uses wiretap coding to map them into $\mathbf{X}^n_c$ and $\mathbf{X}^n_{\mathbf{d}}$, respectively. As in setting 1, receiver j correctly decodes W_{d_j}.

For the secrecy constraint, notice that M_c and $M_{\mathbf{d}}$ are independent, and their rates are $1-\alpha_1-\epsilon_n$ and $1-\alpha_2-\epsilon_n$, respectively. Thus, wiretap coding strongly secures both M_c and $M_{\mathbf{d}}$ against the adversary. We show in Appendix M that (8.3) is satisfied. The achievable strong secrecy file rate is

$$R_s(\alpha) = 2 \times \frac{1-\alpha_1}{2} + \frac{\alpha_1-\alpha_2}{2} = \frac{2-\alpha_1-\alpha_2}{2} = 1 - \frac{\alpha}{2}. \tag{8.17}$$

Case 2: $\alpha_1 < \alpha_2$: The transmitter (i) divides W_l, $l = 1, 2$, into $\left\{W_l^{(1)}, W_l^{(2)}, W_{l,s}\right\}$; $W_l^{(1)}, W_l^{(2)}$ are uniform over $\left[1 : 2^{n\frac{1-\alpha_2-\epsilon_n}{2}}\right]$ and $W_{l,s}$ is uniform over $\left[1 : 2^{n\frac{\alpha_2-\alpha_1}{2}}\right]$, (ii) generates the keys K_1, K_2, uniform over $\left[1 : 2^{n\frac{\alpha_2-\alpha_1}{2}}\right]$, independently from W_1, W_2, (iii) forms M_c as in (8.13) and encodes it into $\mathbf{X}_c^n$ using wiretap coding, (iv) forms $M_{\mathbf{d}}$ as in (8.14) and forms $\tilde{M}_{\mathbf{d}}$ as

$$\tilde{M}_{\mathbf{d}} = \left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2, \tilde{W}\right\}, \tag{8.18}$$

where $\tilde{W}$ is independent from all other variables and uniform over $\left[1 : 2^{n(\alpha_1+\epsilon_n)}\right]$, and (v) encodes $M_{\mathbf{d}}$ into $\mathbf{X}_{\mathbf{d}}^n$ using wiretap coding, while using $\tilde{M}_{\mathbf{d}}$ as the randomization message.

As in Setting 2, for n sufficiently large, receiver j, $j = 1, 2$, correctly decodes W_{d_j}, and the adversary can only retrieve $\tilde{M}_{\mathbf{d}}$ using which it can gain no information about W_1, W_2. In Appendix N, we show that (8.3) is satisfied. The achievable secrecy rate is

$$R_s(\alpha) = 2 \times \frac{1-\alpha_2}{2} + \frac{\alpha_2 - \alpha_1}{2} = 1 - \frac{\alpha}{2}. \tag{8.19}$$

With the aforementioned settings, we showed that the same secrecy rate, i.e., $1 - \frac{\alpha}{2}$, is achievable irrespective of where the adversary taps as long as α_1 and α_2 are known. The question then arises whether the lack of knowledge about relative fractions of tapped symbols would decrease the secrecy capacity. The following setting we propose provides a hint on the answer.

8.4.2.4 Setting 4: Either $\alpha_1 = 0$ or $\alpha_2 = 0$, the legitimate terminals do not know which is zero

The adversary taps into either cache placement or delivery, but not both. The legitimate terminals *do not* know which phase is tapped. We show that the strong secrecy rate $1-\frac{\alpha}{2}$ is again achievable.

The transmitter performs the same division of W_1, W_2 as in Settings 1, 2, and generates independent keys K_1, K_2 as in Setting 2. Let us define

$$M_c = \left\{M_{c,1}, M_{c,2}\right\}; \quad M_{c,1} = W_1^{(1)} \oplus W_2^{(1)}, \quad M_{c,2} = W_1^{(2)} \oplus W_2^{(2)}, \tag{8.20}$$

$$\tilde{M}_c = \{K_1, K_2\}, \tag{8.21}$$

$$M_{\mathbf{d}} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}\right\}, \tag{8.22}$$

$$\tilde{M}_{\mathbf{d}} = \left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\right\}. \tag{8.23}$$

During cache placement, the transmitter encodes M_c into $\mathbf{X}_c^n$ using wiretap coding, while using $\tilde{M}_c$ as the randomization message. Receiver j, $j = 1, 2$, stores $M_{c,j}$, $\tilde{M}_{c,j}$, in its cache memory. During delivery, the transmitter uses wiretap coding to encode $M_{\mathbf{d}}$ into $\mathbf{X}_{\mathbf{d}}^n$, while using $\tilde{M}_{\mathbf{d}}$ as the randomization message. Using its cache contents, along with $M_{\mathbf{d}}$ and $\tilde{M}_{\mathbf{d}}$, receiver j correctly decodes W_{d_j}. By contrast, the adversary can only retrieve either $\{K_1, K_2\}$ or $\left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\right\}$, but not both, using which it can obtain no information about W_1 and W_2. We show in Appendix O that (8.3) is satisfied. The achievable strong secrecy rate is $1 - \frac{\alpha}{2}$.

The lack of knowledge about which phase is tapped is countered by encrypting pieces of information, $\{W_{d_1,s}, W_{d_2,s}\}$, with one-time pad keys K_1 and K_2, while ensuring that the adversary only retrieves either the keys or the encrypted bits but not both; using which it can gain no information about the messages W_1 and W_2.

In the following subsection, we generalize this idea to tackle the case when the adversary gets to tap into both phases, with no knowledge about the relative fractions of tapped symbols in each, i.e., the model in Fig. 8.1. In particular, similar to [63], in which the uncertainty about the wiretapper's channel is treated by using a security embedding code [70], here, in each phase, we construct an embedding code in which $n\alpha$ single-bit layers are embedded into one another. Doing so, we ensure that, no matter what the values for α_1 and α_2 are, the adversary can retrieve no more than $n\alpha_1$ bits from cache placement, and $n\alpha_2$ bits from delivery. By designing what the adversary retrieves to be either a set of key bits and/or information bits encrypted with a distinct set of key bits, we guarantee no information on the messages is asymptotically leaked to the adversary. We thus prove that the lack of knowledge about relative fractions of tapped symbols *does not decrease* the secrecy capacity.

8.4.3 Achievability for $\alpha \in (0,1)$:

We are now ready to present the achievability for the general model considered in this chapter. Consider first $\alpha \in (0,1)$. For simplicity, assume that $n\frac{\alpha_1}{2} = \frac{\mu_1}{2}$ and $n\frac{\alpha_2}{2} = \frac{\mu_2}{2}$ are integers. A minor modification to the analysis can be adopted otherwise.

The transmitter divides W_l, $l = 1, 2$, into the independent messages $W_l^{(1)}$, $W_l^{(2)}$, $W_{l,s}$; $W_l^{(1)}$, $W_l^{(2)}$ are uniform over $\left[1 : 2^{n\frac{1-\alpha}{2}}\right]$, and $W_{l,s}$ is uniform over $\left[1 : 2^{n\frac{\alpha}{2}}\right]$. The

transmitter generates the independent keys K_1, K_2, uniform over $\left[1 : 2^{n\frac{\alpha}{2}}\right]$, and independent from W_1, W_2. For simplicity of exposition, we have ignored the small rate reduction ϵ_n at this stage, as we will introduce this later into the security analysis. The main ideas of the achievability proof are:

1. The transmitter uses wiretap coding with a randomization message of size $n(\alpha_1 + \alpha_2) = n\alpha$ bits in *both the cache placement and delivery* phases. As the adversary does not tap into more than $n\alpha$ bits in each phase, a secure transmission rate of $1 - \alpha$ is achievable in each phase, as long as the randomization messages in the two phases are independent. Using coded placement for $W_1^{(1)}$, $W_1^{(2)}$, $W_2^{(1)}$, $W_2^{(2)}$, a secure file rate of $1 - \alpha$ can be achieved.

2. The randomization messages over the two phases can deliver additional secure information, of rate $\frac{\alpha}{2}$ per file, via encryption. The overall achievable file rate is thus $R_s = 1 - \frac{\alpha}{2}$. In particular, we utilize the keys K_1, K_2, as the randomization message for cache placement. Along with wiretap coding, we employ a security embedding code [70], by using bits of K_1, K_2, in a manner that allows the adversary to be able to retrieve only the last $n\frac{\alpha_1}{2}$ bits from each. In the delivery phase, we encrypt additional pieces of information, $W_{d_1,s}$ and $W_{d_2,s}$, with the keys K_1 and K_2, and utilize this encrypted information as the randomization message. We employ again a security embedding code, in the *reverse order*, such that the adversary can only retrieve the first $n\frac{\alpha_2}{2}$ bits from each of $W_{d_1,s} \oplus K_1$ and $W_{d_2,s} \oplus K_2$.

3. With the aforementioned construction, the adversary, for any values of α_1 and α_2 it chooses, can only retrieve a set of key bits and/or a set of information bits

encrypted with other key bits. In particular, due to the *reversed embedding order*, the adversary does not obtain, in the delivery phase, any message bits encrypted with key bits it has seen during cache placement. In addition, since $\{K_1, K_2\}$ is independent from $\left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\right\}$, and is an independent sequence, the adversary can not use the revealed key bits in the cache placement to obtain any information about the bits of $W_{d_1,s} \oplus K_1$ and $W_{d_2,s} \oplus K_2$ that need to be securely transmitted in the delivery phase.

We now explain the achievability scheme in more detail. Let us define M_c and $\tilde{M}_c$ as in (8.20) and (8.21). In particular, let

$$M_c = \left\{M_{c,1}, M_{c,2}\right\}; \quad M_{c,1} = W_1^{(1)} \oplus W_2^{(1)}, \quad M_{c,2} = W_1^{(2)} \oplus W_2^{(2)}, \tag{8.24}$$

$$\tilde{M}_c = \left\{\tilde{M}_{c,1}, \tilde{M}_{c,2}\right\}; \qquad \tilde{M}_{c,1} = K_1, \qquad \tilde{M}_{c,2} = K_2. \tag{8.25}$$

M_c in (8.24) represents the message to be securely transmitted during cache placement, regardless of the adversary's choice of α_1. $\tilde{M}_c$ in (8.25) represents the randomization message utilized for wiretap coding in the cache placement.

The transmitter further divides $\tilde{M}_{c,1}$ and $\tilde{M}_{c,2}$ into sequences of independent binary bits, $\left\{\tilde{M}_{c,1}^{(1)}, \tilde{M}_{c,1}^{(2)}, \cdots, \tilde{M}_{c,1}^{\left(n\frac{\alpha}{2}\right)}\right\}$ and $\left\{\tilde{M}_{c,2}^{(1)}, \tilde{M}_{c,2}^{(2)}, \cdots, \tilde{M}_{c,2}^{\left(n\frac{\alpha}{2}\right)}\right\}$, and generates $\mathbf{X}_c^n$ as follows:

Cache Placement Codebook Generation: Let m_c, $\tilde{m}_{c,1} = \left\{\tilde{m}_{c,1}^{(1)}, \tilde{m}_{c,1}^{(2)}, \cdots, \tilde{m}_{c,1}^{\left(n\frac{\alpha}{2}\right)}\right\}$, and $\tilde{m}_{c,2} = \left\{\tilde{m}_{c,2}^{(1)}, \tilde{m}_{c,2}^{(2)}, \cdots, \tilde{m}_{c,2}^{\left(n\frac{\alpha}{2}\right)}\right\}$ be the realizations of M_c, $\tilde{M}_{c,1}$, and $\tilde{M}_{c,2}$ in (8.24) and (8.25). We construct the cache placement codebook $\mathcal{C}_{c,n}$, from which $\mathbf{X}_c^n$ is drawn, as follows. We randomly and independently distribute all the possible 2^n length-n binary

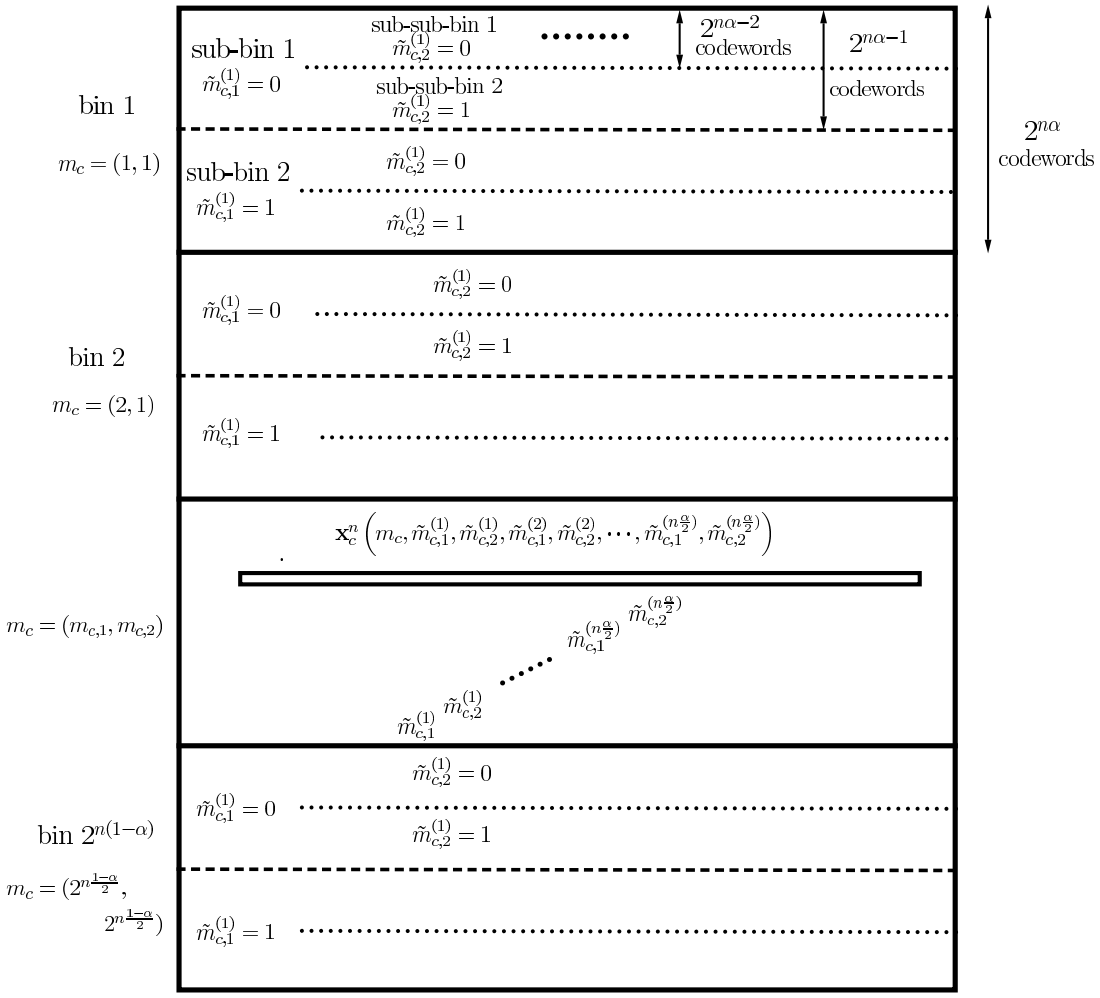


Fig. 8.3. Codebook construction for the cache placement phase, $\mathcal{C}_{c,n}$.

sequences into $2^{n(1-\alpha)}$ bins, indexed by $m_c \in \left[1 : 2^{n\frac{1-\alpha}{2}}\right]^2$. Each bin m_c contains $2^{n\alpha}$ binary sequences (codewords). Further, we randomly and independently divide each bin m_c into two sub-bins, indexed by $\tilde{m}_{c,1}^{(1)}$, and each contains $2^{n\alpha-1}$ codewords. The two sub-bins $\tilde{m}_{c,1}^{(1)}$ are further divided into smaller bins, indexed by $\tilde{m}_{c,2}^{(1)}$, and each contains $2^{n\alpha-2}$ codewords. The process continues, going over $\tilde{m}_{c,1}^{(2)}$, $\tilde{m}_{c,2}^{(2)}$, $\cdots$, $\tilde{m}_{c,1}^{\left(n\frac{\alpha}{2}-1\right)}$, $\tilde{m}_{c,2}^{\left(n\frac{\alpha}{2}-1\right)}$, $\tilde{m}_{c,1}^{\left(n\frac{\alpha}{2}\right)}$, until the remaining two codewords, after each sequence of divisions, are indexed by $\tilde{m}_{c,2}^{\left(n\frac{\alpha}{2}\right)}$. The codebook $\mathcal{C}_{c,n}$ is described in Fig. 8.3.

Remark 28. *An alternative representation of the binning procedure described above is that, each of the $2^{n\alpha}$ binary codewords in the bin m_c, where $m_c \in \left[1 : 2^{n\frac{1-\alpha}{2}}\right]^2$, is randomly assigned to an index $\left\{\tilde{m}_{c,1}^{(1)}, \tilde{m}_{c,2}^{(1)}, \tilde{m}_{c,1}^{(2)}, \tilde{m}_{c,2}^{(2)}, \cdots, \tilde{m}_{c,1}^{\left(n\frac{\alpha}{2}\right)}, \tilde{m}_{c,2}^{\left(n\frac{\alpha}{2}\right)}\right\}$. We chose however to present the former description in order to provide a more detailed explanation of the embedding structure; in particular, the order of embedding, which is a critical component in the achievability scheme.*

Cache Encoder: Given the messages w_1, w_2, i.e., $\left\{w_1^{(1)}, w_1^{(2)}, w_{1,s}\right\}$, $\left\{w_2^{(1)}, w_2^{(2)}, w_{2,s}\right\}$, the transmitter generates m_c, $\tilde{m}_c = \left\{\tilde{m}_{c,1}, \tilde{m}_{c,2}\right\}$ as in (8.24), (8.25). Using the codebook $\mathcal{C}_{c,n}$, the transmitter sends $\mathbf{x}_c^n$ which corresponds to m_c, $\tilde{m}_{c,1}$, $\tilde{m}_{c,2}$, i.e.,
$\mathbf{x}_c^n\left(m_c, \tilde{m}_{c,1}^{(1)}, \tilde{m}_{c,2}^{(1)}, \cdots, \tilde{m}_{c,1}^{\left(n\frac{\alpha}{2}\right)}, \tilde{m}_{c,2}^{\left(n\frac{\alpha}{2}\right)}\right)$.

For the delivery phase, as in (8.22) and (8.23), define

$$M_{\mathbf{d}} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}\right\}, \tag{8.26}$$

$$\tilde{M}_{\mathbf{d}} = \left\{\tilde{M}_{\mathbf{d},1}, \tilde{M}_{\mathbf{d},2}\right\}; \qquad \tilde{M}_{\mathbf{d},1} = W_{d_1,s} \oplus K_1, \qquad \tilde{M}_{\mathbf{d},2} = W_{d_2,s} \oplus K_2. \tag{8.27}$$

$M_{\mathbf{d}}$ in (8.26) represents the message to be securely transmitted during the delivery phase no matter what the adversary's choice of α_2 is. $\tilde{M}_{\mathbf{d}}$ in (8.27) represents the randomization message utilized for the wiretap coding in the delivery phase.

Similar to cache placement, the transmitter further divides $\tilde{M}_{\mathbf{d},1}$, $\tilde{M}_{\mathbf{d},2}$ into sequences of independent binary bits, $\{\tilde{M}_{\mathbf{d},1}^{(1)}, \cdots, \tilde{M}_{\mathbf{d},1}^{\left(n\frac{\alpha}{2}\right)}\}$, $\{\tilde{M}_{\mathbf{d},2}^{(1)}, \cdots, \tilde{M}_{\mathbf{d},2}^{\left(n\frac{\alpha}{2}\right)}\}$, and generates $\mathbf{X}_{\mathbf{d}}^n$ as follows.

Delivery Codebook Generation: Let $m_{\mathbf{d}}$, $\tilde{m}_{\mathbf{d},1} = \{\tilde{m}_{\mathbf{d},1}^{(1)}, \cdots, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha}{2}\right)}\}$, $\tilde{m}_{\mathbf{d},2} = \{\tilde{m}_{\mathbf{d},2}^{(1)}, \cdots, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha}{2}\right)}\}$ be the realizations of $M_{\mathbf{d}}$, $\tilde{M}_{\mathbf{d},1}$, $\tilde{M}_{\mathbf{d},2}$ in (8.26), (8.27). We construct the delivery codebook $\mathcal{C}_{\mathbf{d},n}$, from which $\mathbf{X}_{\mathbf{d}}^n$ is drawn, in a similar fashion as the codebook $\mathcal{C}_{c,n}$, but with a reversed indexing of the sub-bins. In particular, we randomly and independently divide all the 2^n binary sequences into $2^{n(1-\alpha)}$ bins, indexed by $m_{\mathbf{d}} \in \left[1 : 2^{n\frac{1-\alpha}{2}}\right]^2$, and each contains $2^{n\alpha}$ codewords. We further randomly and independently divide each bin $m_{\mathbf{d}}$ into two sub-bins, indexed by $\tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha}{2}\right)}$, and each contains $2^{n\alpha-1}$ codewords. The process continues, going in reverse order over $\tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha}{2}\right)}, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha}{2}-1\right)}$, $\tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha}{2}-1\right)}, \cdots, \tilde{m}_{\mathbf{d},1}^{(1)}$, until the remaining two codewords, after each sequence of divisions, are indexed by $\tilde{m}_{\mathbf{d},2}^{(1)}$. The codebook $\mathcal{C}_{\mathbf{d},n}$ is described in Fig. 8.4.

Delivery Encoder: Given w_1, w_2, i.e., $\{w_1^{(1)}, w_1^{(2)}, w_{1,s}\}$, $\{w_2^{(1)}, w_2^{(2)}, w_{2,s}\}$, and $\mathbf{d} = (d_1, d_2)$, the transmitter generates $m_{\mathbf{d}}$, $\tilde{m}_{\mathbf{d}} = \{\tilde{m}_{\mathbf{d},1}, \tilde{m}_{\mathbf{d},2}\}$ as in (8.26), (8.27). The transmitter sends $\mathbf{x}_{\mathbf{d}}^n$, from $\mathcal{C}_{\mathbf{d},n}$, which corresponds to $m_{\mathbf{d}}$, $\tilde{m}_{\mathbf{d},1}$, and $\tilde{m}_{\mathbf{d},2}$, i.e., $\mathbf{x}_{\mathbf{d}}^n\Big(m_{\mathbf{d}}, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha}{2}\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha}{2}\right)}, \cdots, \tilde{m}_{\mathbf{d},1}^{(1)}, \tilde{m}_{\mathbf{d},2}^{(1)}\Big)$.

Decoding: Using $\mathbf{X}_c^n$, receiver j, $j = 1, 2$, recovers $M_{c,j}$, $\tilde{M}_{c,j}$, and stores them in its cache memory. For $j = 1, 2$, the combined size of $M_{c,j}$ and $\tilde{M}_{c,j}$ does not exceed $\frac{n}{2}$

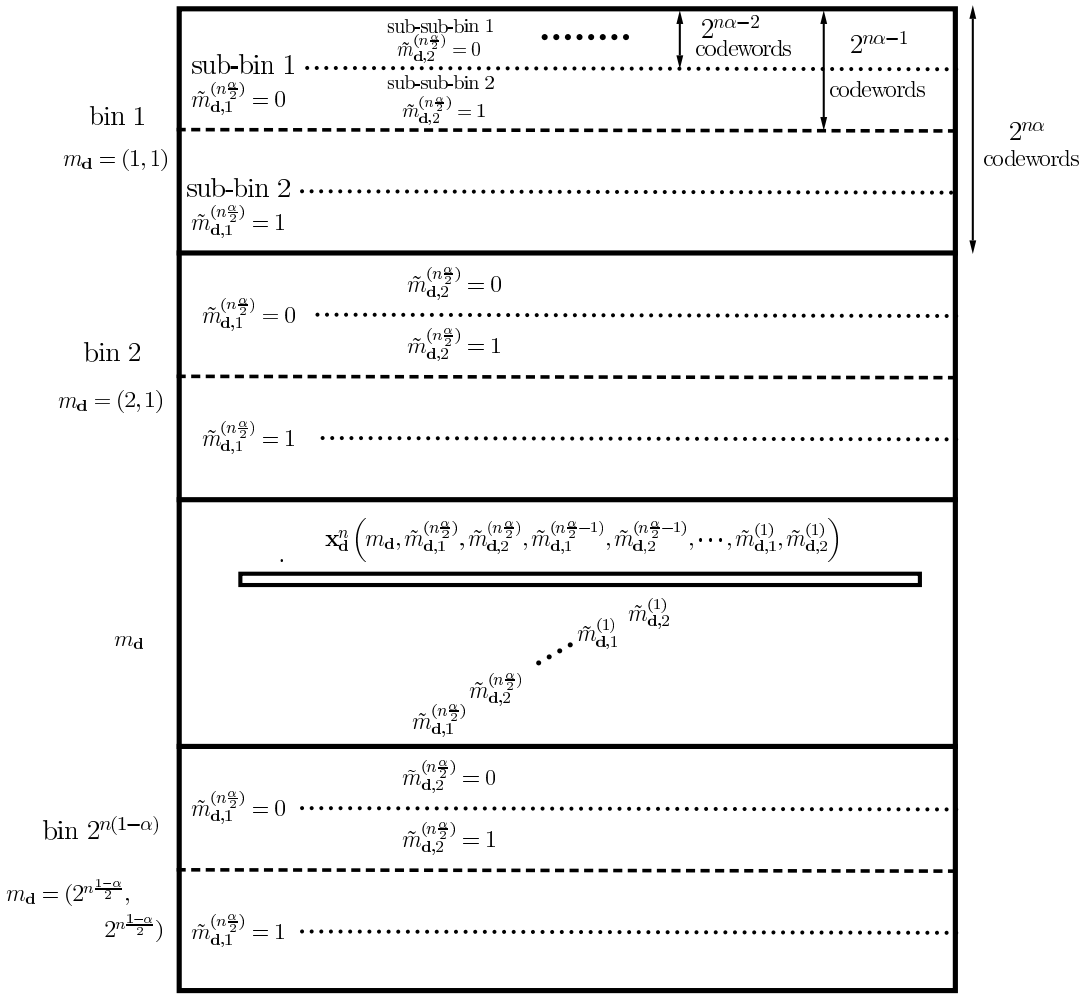


Fig. 8.4. Codebook construction for the delivery phase, $\mathcal{C}_{\mathbf{d},n}$.

bits. Using $\mathbf{X}_{\mathbf{d}}^n$, both receivers recover $M_{\mathbf{d}}$, $\tilde{M}_{\mathbf{d}}$. Using $M_{\mathbf{d}}$, $\tilde{M}_{\mathbf{d}}$, $M_{c,j}$, $\tilde{M}_{c,j}$, and for n sufficiently large, receiver j correctly decodes W_{d_j}.

Security Analysis: Let us first slightly modify the construction above as follows. Recall that $\{\epsilon_n\}_{n\geq 1}$ is a sequence of positive real numbers such that $\epsilon_n \to 0$ as $n \to \infty$. Define

$$\alpha_\epsilon = \alpha + 2\epsilon_n, \qquad \alpha_{1,\epsilon} = \alpha_1 + \epsilon_n, \qquad \alpha_{2,\epsilon} = \alpha_2 + \epsilon_n. \tag{8.28}$$

That is, $\alpha_{1,\epsilon} + \alpha_{2,\epsilon} = \alpha_\epsilon$. We increase the sizes of K_1 and K_2 into $\frac{n\alpha_\epsilon}{2}$ bits, from $n\frac{\alpha}{2}$, and zero-pad the bit strings of $W_{d_1,s}$ and $W_{d_2,s}$ accordingly. Additionally, we decrease the sizes of $W_l^{(1)}$, $W_l^{(2)}$, $l = 1, 2$, to $n\frac{1-\alpha_\epsilon}{2}$ bits, instead of $n\frac{1-\alpha}{2}$. Once again, we assume that $\frac{n\alpha_\epsilon}{2}$ and $\frac{n\alpha_{1,\epsilon}}{2}$ are integers; as minor modifications can be adopted otherwise.

Let us fix the subsets $S_1, S_2 \subseteq [1 : n]$. For the corresponding (fixed) values of α_1 and α_2, the cache placement codebook $\mathcal{C}_{c,n}$ can be viewed as a wiretap code with $2^{n(1-\alpha_{1,\epsilon})}$ bins. Each bin is indexed by the message

$$w_c = \left(m_c, \tilde{m}_{c,1}^{(1)}, \tilde{m}_{c,2}^{(1)}, \tilde{m}_{c,1}^{(2)}, \tilde{m}_{c,2}^{(2)}, \cdots, \tilde{m}_{c,1}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}}}, \tilde{m}_{c,2}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}}} \right). \tag{8.29}$$

Each bin w_c contains $2^{n\alpha_{1,\epsilon}}$ binary codewords which are indexed by the randomization message

$$\tilde{w}_c = \left(\tilde{m}_{c,1}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}+1}}, \tilde{m}_{c,2}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}+1}}, \tilde{m}_{c,1}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}+2}}, \tilde{m}_{c,2}^{\binom{n\frac{\alpha_{2,\epsilon}}{2}+2}}, \cdots, \tilde{m}_{c,1}^{\binom{n\frac{\alpha_\epsilon}{2}}}, \tilde{m}_{c,2}^{\binom{n\frac{\alpha_\epsilon}{2}}} \right). \tag{8.30}$$

Similarly, the delivery codebook $\mathcal{C}_{\mathbf{d},n}$ can be seen as a wiretap code with $2^{n(1-\alpha_{2,\epsilon})}$ bins, each of which is indexed by the message

$$w_{\mathbf{d}} = \left(m_{\mathbf{d}}, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha_\epsilon}{2}\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha_\epsilon}{2}\right)}, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha_\epsilon}{2}-1\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha_\epsilon}{2}-1\right)}, \cdots, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)} \right). \tag{8.31}$$

Each bin $w_{\mathbf{d}}$ contains $2^{n\alpha_{2,\epsilon}}$ codewords, indexed by the randomization message

$$\tilde{w}_{\mathbf{d}} = \left(\tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}, \tilde{m}_{\mathbf{d},1}^{\left(n\frac{\alpha_{2,\epsilon}}{2}-1\right)}, \tilde{m}_{\mathbf{d},2}^{\left(n\frac{\alpha_{2,\epsilon}}{2}-1\right)}, \cdots, \tilde{m}_{\mathbf{d},1}^{(1)}, \tilde{m}_{\mathbf{d},2}^{(1)} \right). \tag{8.32}$$

Let $\left\{\mathcal{B}_{w_c} :\ w_c = 1, 2, \cdots, 2^{n(1-\alpha_{1,\epsilon})}\right\}$ and $\left\{\mathcal{B}_{w_{\mathbf{d}}} :\ w_{\mathbf{d}} = 1, 2, \cdots, 2^{n(1-\alpha_{2,\epsilon})}\right\}$ denote the partition, i.e., bins, of the codebooks $\mathcal{C}_{c,n}$ and $\mathcal{C}_{\mathbf{d},n}$, which correspond to the messages w_c and $w_{\mathbf{d}}$ in (8.29) and (8.31), respectively. Let $\mathbf{x}^{2n} \triangleq (\mathbf{x}_c^n, \mathbf{x}_{\mathbf{d}}^n)$ denote the concatenation of the two length-n binary codewords $\mathbf{x}_c^n$, $\mathbf{x}_{\mathbf{d}}^n$. Define the Cartesian product of the bins $\mathcal{B}_{w_c}$ and $\mathcal{B}_{w_{\mathbf{d}}}$, as

$$\mathcal{B}_{w_c,w_{\mathbf{d}}} \triangleq \left\{ \mathbf{x}^{2n} = (\mathbf{x}_c^n, \mathbf{x}_{\mathbf{d}}^n) :\ \mathbf{x}_c^n \in \mathcal{B}_{w_c},\ \mathbf{x}_{\mathbf{d}}^n \in \mathcal{B}_{w_{\mathbf{d}}} \right\}. \tag{8.33}$$

Since the partitioning of the codebooks $\mathcal{C}_{c,n}$ and $\mathcal{C}_{\mathbf{d},n}$ is random, for every w_c and $w_{\mathbf{d}}$, $\mathcal{B}_{w_c,w_{\mathbf{d}}}$ is a random codebook which results from the Cartesian product of the random bins $\mathcal{B}_{w_c}$, $\mathcal{B}_{w_{\mathbf{d}}}$. Recall that $\mathcal{B}_{w_c}$ contains $2^{n\alpha_{1,\epsilon}}$ and $\mathcal{B}_{w_{\mathbf{d}}}$ contains $2^{n\alpha_{2,\epsilon}}$ length-n binary codewords. Thus, the product $\mathcal{B}_{w_c,w_{\mathbf{d}}}$ contains $2^{n\alpha_\epsilon}$ length-$2n$ binary codewords.

Let $\left\{W_{d_l,s}^{(1)}, W_{d_l,s}^{(2)}, \cdots, W_{d_l,s}^{\left(n\frac{\alpha_\epsilon}{2}\right)}\right\}$ and $\left\{K_l^{(1)}, K_l^{(2)}, \cdots, K_l^{\left(n\frac{\alpha_\epsilon}{2}\right)}\right\}$ denote the binary bit strings of $W_{d_l,s}$ and K_l, $l = 1, 2$. In addition, for notational simplicity, define

$$\mathbf{W}_s^{(1)} = \left\{W_{d_1,s}^{(1)}, W_{d_2,s}^{(1)}, \cdots, W_{d_1,s}^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}, W_{d_2,s}^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}\right\}, \tag{8.34}$$

$$\mathbf{W}_s^{(2)} = \left\{W_{d_1,s}^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)}, W_{d_2,s}^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)}, \cdots, W_{d_1,s}^{\left(n\frac{\alpha_\epsilon}{2}\right)}, W_{d_2,s}^{\left(n\frac{\alpha_\epsilon}{2}\right)}\right\}, \tag{8.35}$$

$$\mathbf{K}^{(1)} = \left\{K_1^{(1)}, K_2^{(1)}, \cdots, K_1^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}, K_2^{\left(n\frac{\alpha_{2,\epsilon}}{2}\right)}\right\}, \tag{8.36}$$

$$\mathbf{K}^{(2)} = \left\{K_1^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)}, K_2^{\left(n\frac{\alpha_{2,\epsilon}}{2}+1\right)}, \cdots, K_1^{\left(n\frac{\alpha_\epsilon}{2}\right)}, K_2^{\left(n\frac{\alpha_\epsilon}{2}\right)}\right\}, \tag{8.37}$$

$$\mathbf{W}_{\oplus\mathbf{K}}^{(1)} = \left\{W_{d_1,s}^{(i)} \oplus K_1^{(i)}, W_{d_2,s}^{(i)} \oplus K_2^{(i)} : \ \ i = 1, 2, \cdots, n\frac{\alpha_{2,\epsilon}}{2}\right\}, \tag{8.38}$$

$$\mathbf{W}_{\oplus\mathbf{K}}^{(2)} = \left\{W_{d_1,s}^{(i)} \oplus K_1^{(i)}, W_{d_2,s}^{(i)} \oplus K_2^{(i)} : \ \ i = n\frac{\alpha_{2,\epsilon}}{2}+1,\ n\frac{\alpha_{2,\epsilon}}{2}+2, \cdots, n\frac{\alpha_\epsilon}{2}\right\}. \tag{8.39}$$

Let W_c, $\tilde{W}_c$, $W_{\mathbf{d}}$, and $\tilde{W}_{\mathbf{d}}$ denote the random variables that correspond to the realizations defined in (8.29)–(8.32). Using (8.24)–(8.27), (8.29)–(8.32), and (8.36)–(8.39), we have

$$W_c = \left\{M_c, \mathbf{K}^{(1)}\right\} = \left\{W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, \mathbf{K}^{(1)}\right\}, \qquad \tilde{W}_c = \mathbf{K}^{(2)} \tag{8.40}$$

$$W_{\mathbf{d}} = \left\{M_{\mathbf{d}}, \mathbf{W}_{\oplus\mathbf{K}}^{(2)}\right\} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}, \mathbf{W}_{\oplus\mathbf{K}}^{(2)}\right\}, \qquad \tilde{W}_{\mathbf{d}} = \mathbf{W}_{\oplus\mathbf{K}}^{(1)}. \tag{8.41}$$

Notice that $\tilde{W}_c$ and $\tilde{W}_{\mathbf{d}}$ are independent, and each is uniformly distributed. $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$ is thus jointly uniform. In addition, $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$ is independent from $\{W_c, W_{\mathbf{d}}\}$. Thus, we can apply the analysis in [33, (94)-(103)] to show that, for every S_1, S_2, w_c, and $w_{\mathbf{d}}$, and

every $\epsilon > 0$, there exists $\gamma(\epsilon) > 0$ such that

$$\mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\mathbb{D}\left(P_{\mathbf{Z}_{S_1}^n\mathbf{Z}_{S_2}^n|W_c=w_c,W_{\mathbf{d}}=w_{\mathbf{d}}}||P_{\mathbf{Z}_{S_1}^n\mathbf{Z}_{S_2}^n}\right)>\epsilon\right)\leq\exp\left(-e^{n\gamma(\epsilon)}\right). \tag{8.42}$$

$P_{\mathbf{Z}_{S_1}^n\mathbf{Z}_{S_2}^n|W_c=w_c,W_{\mathbf{d}}=w_{\mathbf{d}}}$ is the induced distribution at the adversary when $\mathbf{x}_c^n(w_c,\tilde{w}_c)$ and $\mathbf{x}_{\mathbf{d}}^n(w_{\mathbf{d}},\tilde{w}_{\mathbf{d}})$ are the transmitted codewords over cache placement and delivery phases. $P_{\mathbf{Z}_{S_1}^n\mathbf{Z}_{S_2}^n}$ is the output distribution at the adversary.

The number of the messages $\{w_c, w_{\mathbf{d}}\}$ is $2^{n(2-\alpha_\epsilon)}$. Additionally, the number of possible choices for the subsets S_1 and S_2 is $\binom{2n}{\alpha n} < 2^{2n}$. Thus, the combined number of the messages and the subsets is at most exponential in n. Using (8.42) and the union bound, we have

$$\lim_{n\to\infty}\max_{S_1,S_2} I\left(W_c,W_{\mathbf{d}};\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right)=0. \tag{8.43}$$

For the sake of completeness, we provide the full proofs for (8.42) and (8.43) in Appendix P.

We also have, for any $\mathbf{d}=(d_1,d_2)$, $d_1,d_2\in\{1,2\}$,

$$\begin{aligned}
&I\left(W_1,W_2;\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right)\\
&=I\left(W_1^{(1)},W_1^{(2)},W_2^{(1)},W_2^{(2)},W_{1,s},W_{2,s};\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right) &(8.44)\\
&=I\left(W_1^{(1)},W_1^{(2)},W_2^{(1)},W_2^{(2)},\mathbf{W}_s^{(1)},\mathbf{W}_s^{(2)};\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right) &(8.45)\\
&=I\left(W_1^{(1)}\oplus W_2^{(1)},W_1^{(2)}\oplus W_2^{(2)},W_{d_1}^{(2)},W_{d_2}^{(1)},\mathbf{W}_s^{(1)},\mathbf{W}_s^{(2)};\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right) &(8.46)\\
&=I\left(M_c,M_{\mathbf{d}},\mathbf{W}_s^{(1)},\mathbf{W}_s^{(2)};\mathbf{Z}_{S_1}^n,\mathbf{Z}_{S_2}^n\right) &(8.47)
\end{aligned}$$

$$\leq I\left(M_c, M_{\mathbf{d}}, \mathbf{W}_s^{(1)}, \mathbf{W}_{\oplus\mathbf{K}}^{(2)}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.48}$$

$$= I\left(M_c, \mathbf{W}_s^{(1)}, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.49}$$

$$= H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) - H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n \middle| M_c, \mathbf{W}_s^{(1)}, W_{\mathbf{d}}\right). \tag{8.50}$$

Equation (8.45) follows since, for $\mathbf{d} = (d_1, d_2)$, $\mathbf{Z}_{S_1}^n$ and $\mathbf{Z}_{S_2}^n$ depend only on $W_1^{(1)}$, $W_1^{(2)}$, $W_2^{(1)}$, $W_2^{(2)}$, $W_{d_1,s}$, and $W_{d_2,s}$, and by using (8.34) and (8.35). Equation (8.46) follows because there exists a bijection between $\{W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}\}$ and $\{W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, W_{d_1}^{(2)}, W_{d_2}^{(1)}\}$. Equation (8.47) follows from (8.24) and (8.26). The inequality in (8.48) follows due to the Markov chain $\mathbf{W}_s^{(2)} - \left\{M_c, M_{\mathbf{d}}, \mathbf{W}_s^{(1)}, \mathbf{W}_{\oplus\mathbf{K}}^{(2)}\right\} - \left\{\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right\}$, and the data processing inequality. This Markov chain holds because $\left\{M_c, M_{\mathbf{d}}, \mathbf{W}_s^{(1)}\right\}$ are independent from $\left\{\mathbf{W}_s^{(2)}, \mathbf{K}^{(2)}\right\}$, and only the encrypted information $\mathbf{W}_{\oplus\mathbf{K}}^{(2)}$ is transmitted. Equation (8.49) follows from (8.41).

The second term on the right hand side of (8.50) can be lower bounded as

$$H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n \middle| M_c, \mathbf{W}_s^{(1)}, W_{\mathbf{d}}\right) = H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n, \mathbf{W}_s^{(1)} \middle| M_c, W_{\mathbf{d}}\right) - H\left(\mathbf{W}_s^{(1)} | M_c, W_{\mathbf{d}}\right) \tag{8.51}$$

$$= H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n, \mathbf{W}_s^{(1)}, \mathbf{W}_{\oplus K}^{(1)} \middle| M_c, W_{\mathbf{d}}\right) - H\left(\mathbf{W}_{\oplus K}^{(1)} \middle| M_c, W_{\mathbf{d}}, \mathbf{W}_s^{(1)}, \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) - H\left(\mathbf{W}_s^{(1)}\right) \tag{8.52}$$

$$= H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n, \mathbf{K}^{(1)}, \mathbf{W}_{\oplus K}^{(1)} \middle| M_c, W_{\mathbf{d}}\right) - H\left(\mathbf{K}^{(1)} \middle| M_c, W_{\mathbf{d}}, \mathbf{W}_s^{(1)}, \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) - H\left(\mathbf{W}_s^{(1)}\right) \tag{8.53}$$

$$\geq H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n, \mathbf{K}^{(1)}, \mathbf{W}_{\oplus K}^{(1)} \middle| M_c, W_{\mathbf{d}}\right) - H\left(\mathbf{W}_s^{(1)}\right) - \epsilon_n' \tag{8.54}$$

$$\geq H\left(\mathbf{K}^{(1)}\middle|M_c, W_{\mathbf{d}}\right) + H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\middle|M_c, \mathbf{K}^{(1)}, W_{\mathbf{d}}\right) - H\left(\mathbf{W}_s^{(1)}\right) - \epsilon_n' \tag{8.55}$$

$$= H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\middle|W_c, W_{\mathbf{d}}\right) + H\left(\mathbf{K}^{(1)}\right) - H\left(\mathbf{W}_s^{(1)}\right) - \epsilon_n' \tag{8.56}$$

$$= H\left(\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\middle|W_c, W_{\mathbf{d}}\right) - \epsilon_n', \tag{8.57}$$

where $\epsilon_n' \to 0$ as $n \to \infty$. Equation (8.52) follows since $\mathbf{W}_s^{(1)}$ is independent from $\{M_c, W_{\mathbf{d}}\}$. Equation (8.53) follows because there exists a bijection between $\left\{\mathbf{W}_s^{(1)}, \mathbf{W}_{\oplus K}^{(1)}\right\}$ and $\left\{\mathbf{K}^{(1)}, \mathbf{W}_{\oplus K}^{(1)}\right\}$. Equation (8.56) follows from (8.40), and since $\mathbf{K}^{(1)}$ is independent from $\{M_c, W_{\mathbf{d}}\}$. Equation (8.57) follows since $\mathbf{K}^{(1)}$ and $\mathbf{W}_s^{(1)}$ are independent and identically distributed.

The inequality in (8.54) follows because, given $\left\{M_c, \mathbf{W}_s^{(1)}, W_{\mathbf{d}}\right\}$, and for n sufficiently large, the adversary can decode $\mathbf{K}^{(1)}$ using its observations $\mathbf{Z}_{S_1}^n$ and $\mathbf{Z}_{S_2}^n$. In particular, $\left\{M_c, \mathbf{W}_s^{(1)}, W_{\mathbf{d}}\right\}$ determine a partition of the codebook into bins, each of which contains $2^{n\alpha_\epsilon}$ binary codewords. For n sufficiently large, and given the values of M_c, $\mathbf{W}_s^{(1)}$, $W_{\mathbf{d}}$, i.e., the bin index, the adversary can determine the codeword index inside the bin, and hence decode $\mathbf{K}^{(1)}$. We conclude that, $H\left(\mathbf{K}^{(1)}\middle|M_c, W_{\mathbf{d}}, \mathbf{W}_s^{(1)}, \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \leq \epsilon_n'$, where $\epsilon_n' \to 0$ as $n \to \infty$.

Substituting (8.57) in (8.50) gives

$$I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \leq I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) + \epsilon_n'. \tag{8.58}$$

Using (8.43) and (8.58), the secrecy constraint in (8.3) is satisfied. Since $\epsilon_n \to 0$ as $n \to \infty$, we conclude that, the achievable strong secrecy file rate is

$$R_s(\alpha) = 2 \times \frac{1-\alpha}{2} + \frac{\alpha}{2} = 1 - \frac{\alpha}{2}. \tag{8.59}$$

Remark 29. *Although the cache placement and delivery codebooks, $\mathcal{C}_{c,n}$ and $\mathcal{C}_{\mathbf{d},n}$, are designed and generated disjointly, in the security analysis, we have considered the Cartesian products of the individual bins of the two codebooks. We were able to do so since the input distributions for generating the two codebooks are identical, i.e., uniform binary.*

8.4.4 Achievability for $\alpha \in [1,2]$:

For $\alpha \in [1,2]$, we adapt the achievability scheme described in Section 8.4.3 as follows. The messages W_1 and W_2 are uniform over $\left[1 : 2^{n\frac{2-\alpha_\epsilon}{2}}\right]$; α_ϵ is defined in (8.28). The transmitter generates the independent keys K_1, K_2, uniform over $\left[1 : 2^{n\frac{2-\alpha_\epsilon}{2}}\right]$, and independent from W_1, W_2. In addition, the transmitter, independently from W_1, W_2, K_1, K_2, generates the independent randomization messages $\tilde{W}$ and $\tilde{W}_K$, uniformly over $\left[1 : 2^{n(\alpha_\epsilon - 1)}\right]$.

The messages for cache placement at receivers 1 and 2 are

$$M_{c,1} = K_1, \qquad M_{c,2} = K_2. \tag{8.60}$$

That is, receiver j, $j = 1, 2$, stores the key K_j in its cache memory. The message to be securely transmitted during delivery is

$$M_{\mathbf{d}} = \left\{M_{\mathbf{d},1}, M_{\mathbf{d},2}\right\}; \;\; M_{\mathbf{d},1} = W_{d_1} \oplus K_1, \;\; M_{\mathbf{d},2} = W_{d_2} \oplus K_2. \tag{8.61}$$

Let $\left\{W_{d_l}^{(1)}, \cdots, W_{d_l}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}\right\}$, $\left\{K_l^{(1)}, \cdots, K_l^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}\right\}$, and $\left\{M_{\mathbf{d},l}^{(1)}, \cdots, M_{\mathbf{d},l}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}\right\}$ denote the bit strings of W_{d_l}, K_l, and $M_{\mathbf{d},l}$, $l = 1, 2$.

Notice that, for $\alpha \in [1, 2]$, the adversary can see all symbols in at least one of the phases. Hence, unlike Section 8.4.3, we cannot utilize randomization messages, $\tilde{W}$ and $\tilde{W}_K$, to carry any information; only keys are stored in the cache memories of the receivers. Additionally, the cache placement and delivery codebooks for this case have a different embedding structure than for $\alpha \in (0, 1)$ in Section 8.4.3. In particular, the embedding here is performed on the bits of the messages M_c and $M_{\mathbf{d}}$, while the embedding in Section 8.4.3 is performed on the bits of the randomization messages $\tilde{M}_c$ and $\tilde{M}_{\mathbf{d}}$.

Cache Placement Codebook Generation: During cache placement, the transmitter generates $\mathcal{C}_{c,n}$ as follows. The transmitter randomly and independently divides all the 2^n length-n binary sequences into 2 bins, indexed by $K_1^{(1)}$, and each contains 2^{n-1} codewords. These two bins are further randomly and independently divided into two sub-bins, indexed by $K_2^{(1)}$, and each contains 2^{n-2} codewords. The process continues, going over $K_1^{(2)}$, $K_2^{(2)}$, $\cdots$, $K_1^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}$, $K_2^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}$, until the remaining $2^{n(\alpha_\epsilon - 1)}$ codewords, after each sequence of divisions, are indexed by $\tilde{W}_K$.

Cache Encoder: The transmitter sends $\mathbf{X}_c^n$ which corresponds to the keys K_1, K_2, and the randomization message $\tilde{W}_K$, i.e., $\mathbf{X}_c^n\left(K_1^{(1)}, K_2^{(1)}, \cdots, K_1^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, K_2^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, \tilde{W}_K\right)$.

Delivery Codebook Generation: In the delivery phase, the transmitter generates $\mathcal{C}_{\mathbf{d},n}$ as follows. The transmitter randomly and independently divides all the 2^n length-n binary sequences into two bins, indexed by $M_{\mathbf{d},1}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}$, and each contains 2^{n-1} codewords. These two bins are further randomly and independently divided into two sub-bins, indexed by $M_{\mathbf{d},2}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}$, and each contains 2^{n-2} codewords. The process continues, going in reverse order over $M_{\mathbf{d},1}^{\left(n\frac{2-\alpha_\epsilon}{2}-1\right)}$, $M_{\mathbf{d},2}^{\left(n\frac{2-\alpha_\epsilon}{2}-1\right)}$, $\cdots$, $M_{\mathbf{d},1}^{(1)}$, $M_{\mathbf{d},2}^{(1)}$, until the remaining $2^{n(\alpha_\epsilon-1)}$ codewords, after each sequence of divisions, are indexed by the randomization message $\tilde{W}$.

Delivery Encoder: Given W_1, W_2, K_1, K_2, $\tilde{W}$, and $\mathbf{d} = (d_1, d_2)$, the transmitter forms $M_{\mathbf{d},1}$ and $M_{\mathbf{d},2}$ as in (8.61) and sends $\mathbf{X}_{\mathbf{d}}^n$ which corresponds to $M_{\mathbf{d},1}$, $M_{\mathbf{d},2}$, and $\tilde{W}$, i.e.,

$$\mathbf{X}_{\mathbf{d}}^n\Big(M_{\mathbf{d},1}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, M_{\mathbf{d},2}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, \cdots, M_{\mathbf{d},1}^{(1)}, M_{\mathbf{d},2}^{(1)}, \tilde{W}\Big).$$

Decoding: Using $\mathbf{X}_c^n$, receiver j, $j = 1, 2$, recovers $M_{c,j} = K_j$ and stores it in its cache memory. Using $\mathbf{X}_{\mathbf{d}}^n$, both receivers recover $M_{\mathbf{d}} = \{M_{\mathbf{d},1}, M_{\mathbf{d},2}\}$. Using $M_{\mathbf{d},j}$, K_j, and for n sufficiently large, receiver j correctly decodes W_{d_j}.

Security Analysis: Fix the subsets S_1, S_2. Recall that $\alpha_1, \alpha_2 \leq 1$. Since $\alpha \geq 1$, $\alpha_1, \alpha_2 \geq \alpha - 1$. If $\alpha_1 = 1$, then $\alpha_2 = \alpha - 1$, and vice versa. In addition, notice that $1 - \alpha_1, 1 - \alpha_2 \leq 2 - \alpha$.

As in Section 8.4.3, for a fixed value of α_1, the codebook $\mathcal{C}_{c,n}$ is a wiretap code with $2^{n(1-\alpha_{1,\epsilon})}$ bins, indexed by

$$W_c = \left(K_1^{(1)}, K_2^{(1)}, \cdots, K_1^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}\right)}, K_2^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}\right)}\right). \tag{8.62}$$

Each bin W_c contains $2^{n\alpha_{1,\epsilon}}$ binary codewords, indexed by

$$\tilde{W}_c = \left(K_1^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}+1\right)}, K_2^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}+1\right)}, \cdots, K_1^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, K_2^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, \tilde{W}_K \right). \tag{8.63}$$

Similarly, for a fixed value of α_2, the codebook $\mathcal{C}_{\mathbf{d},n}$ is a wiretap code with $2^{n(1-\alpha_{2,\epsilon})}$ bins, each is indexed by

$$W_{\mathbf{d}} = \left(\tilde{M}_{\mathbf{d},1}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, \tilde{M}_{\mathbf{d},2}^{\left(n\frac{2-\alpha_\epsilon}{2}\right)}, \cdots, \tilde{M}_{\mathbf{d},1}^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}+1\right)}, \tilde{M}_{\mathbf{d},2}^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}+1\right)} \right). \tag{8.64}$$

Each bin $W_{\mathbf{d}}$ contains $2^{n\alpha_{2,\epsilon}}$ codewords, indexed by

$$\tilde{W}_{\mathbf{d}} = \left(\tilde{M}_{\mathbf{d},1}^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}\right)}, \tilde{M}_{\mathbf{d},2}^{\left(n\frac{1-\alpha_{1,\epsilon}}{2}\right)}, \cdots, \tilde{M}_{\mathbf{d},1}^{(1)}, \tilde{M}_{\mathbf{d},2}^{(1)}, \tilde{W} \right). \tag{8.65}$$

Let us re-define

$$\mathbf{K}^{(1)} = \left\{ K_1^{(i)}, K_2^{(i)} : \ \ i = 1, \cdots, n\frac{1-\alpha_{1,\epsilon}}{2} \right\}, \tag{8.66}$$

$$\mathbf{K}^{(2)} = \left\{ K_1^{(i)}, K_2^{(i)} : \ \ i = n\frac{1-\alpha_{1,\epsilon}}{2}+1, \cdots, n\frac{2-\alpha_\epsilon}{2} \right\}, \tag{8.67}$$

$$\mathbf{W}_{\oplus\mathbf{K}}^{(1)} = \left\{ W_{d_1}^{(i)} \oplus K_1^{(i)}, W_{d_2}^{(i)} \oplus K_2^{(i)} : \ \ i = 1, \cdots, n\frac{1-\alpha_{1,\epsilon}}{2} \right\}, \tag{8.68}$$

$$\mathbf{W}_{\oplus\mathbf{K}}^{(2)} = \left\{ W_{d_1}^{(i)} \oplus K_1^{(i)}, W_{d_2}^{(i)} \oplus K_2^{(i)} : \ \ i = n\frac{1-\alpha_{1,\epsilon}}{2}+1, \cdots, n\frac{2-\alpha_\epsilon}{2} \right\}, \tag{8.69}$$

and define

$$\mathbf{W}^{(1)} = \left\{ W_{d_1}^{(i)}, W_{d_2}^{(i)} : \ \ i = 1, \cdots, n\frac{1-\alpha_{1,\epsilon}}{2} \right\}, \tag{8.70}$$

$$\mathbf{W}^{(2)} = \left\{ W_{d_1}^{(i)}, W_{d_2}^{(i)} : \ \ i = n\frac{1-\alpha_{1,\epsilon}}{2}+1, \cdots, n\frac{2-\alpha_\epsilon}{2} \right\}, \tag{8.71}$$

From (8.62)-(8.69), we have

$$W_c = \mathbf{K}^{(1)}, \quad \tilde{W}_c = \left\{\mathbf{K}^{(2)}, \tilde{W}_K\right\}, \quad W_{\mathbf{d}} = \mathbf{W}^{(2)}_{\oplus \mathbf{K}}, \quad \tilde{W}_{\mathbf{d}} = \left\{\mathbf{W}^{(1)}_{\oplus \mathbf{K}}, \tilde{W}\right\}. \tag{8.72}$$

Similar to Section 8.4.3, $\tilde{W}_c$, $\tilde{W}_{\mathbf{d}}$, are independent and uniformly distributed, and hence $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$ is jointly uniform. Additionally, $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$ is independent from $\{W_c, W_{\mathbf{d}}\}$. Thus, (8.43) is satisfied.

We also have, for any $\mathbf{d} = (d_1, d_2)$,

$$I\left(W_1, W_2; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) = I\left(\mathbf{W}^{(1)}, \mathbf{W}^{(2)}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{8.73}$$

$$\leq I\left(\mathbf{W}^{(1)}, \mathbf{W}^{(2)}_{\oplus \mathbf{K}}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{8.74}$$

$$= I\left(\mathbf{W}^{(1)}, W_{\mathbf{d}}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{8.75}$$

$$= H\left(\mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) - H\left(\mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\middle|\mathbf{W}^{(1)}, W_{\mathbf{d}}\right) \tag{8.76}$$

$$\leq H\left(\mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) - H\left(\mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\middle|\mathbf{K}^{(1)}, W_{\mathbf{d}}\right) + \epsilon'_n \tag{8.77}$$

$$= I(W_c, W_{\mathbf{d}}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}) + \epsilon'_n. \tag{8.78}$$

The inequality in (8.74) follows due to the Markov chain $\mathbf{W}^{(2)} - \left\{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}_{\oplus \mathbf{K}}\right\} - \left\{\mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right\}$. Equations (8.75) and (8.78) follow from (8.72). The inequality in (8.77) follows by using similar steps as in (8.51)-(8.57). Using (8.43) and (8.78), the secrecy constraint in (8.3) is satisfied. Since $\epsilon_n \to 0$ as $n \to \infty$, the achievable strong secrecy file rate is

$$R_s(\alpha) = \frac{2-\alpha}{2} = 1 - \frac{\alpha}{2}. \tag{8.79}$$

This completes the proof for Theorem 14.

8.5 Proof of Theorem 15

In this section, we extend the achievability scheme presented in Section 8.4 and provide a lower bound on the achievable strong secrecy file rate when $D > 2$. The demand vector is again denoted by $\mathbf{d} = (d_1, d_2)$, where $d_1, d_2 \in [1 : D]$. As in Section 8.4, we divide the proof into two cases for the ranges $\alpha \in (0, 1)$ and $\alpha \in [1, 2]$.

8.5.1 $\alpha \in [1, 2]$

For $\alpha \in [1, 2]$, we utilize the same achievability scheme in Section 8.4.4. The reason behind this is, for this range of α, only the keys K_1, K_2, are transmitted in the cache placement, and stored in receivers 1 and 2 cache memories, respectively. That is, no information messages are stored in the caches, and the user demands are known during the delivery phase. The achievable strong secrecy file rate is $1 - \frac{\alpha}{2}$.

8.5.2 $\alpha \in (0, 1)$

The achievability scheme for this case has the same channel coding structure as in the scheme described in Section 8.4.3. The difference however lies in generating the messages to be securely communicated over cache placement and delivery phases, i.e., M_c and $M_{\mathbf{d}}$. In particular, we utilize here uncoded placement for designing the cache contents, and a partially coded delivery transmission that is simultaneously useful for both receivers.

The transmitter divides W_l, $l \in [1:D]$, into the independent messages $\{W_l^{(1)}, W_l^{(2)}, W_{l,t}, W_{l,s}\}$. $W_l^{(1)}$, $W_l^{(2)}$, are uniform over $\left[1 : 2^{n\frac{1-\alpha_\epsilon}{2D}}\right]$; α_ϵ is defined in (8.28). $W_{l,t}$ is uniform over $\left[1 : 2^{n\frac{(2D-1)(1-\alpha_\epsilon)}{4D}}\right]$, and $W_{l,s}$ is uniform over $\left[1 : 2^{n\frac{\alpha_\epsilon}{2}}\right]$. The transmitter, independently from $W_{[1:D]}$, generates the independent keys K_1, K_2, uniformly distributed over $\left[1 : 2^{n\frac{\alpha_\epsilon}{2}}\right]$.

Let $M_c = \{M_{c,1}, M_{c,2}\}$. Unlike (8.24), we utilize here *uncoded placement* for designing $M_{c,1}$ and $M_{c,2}$. We have,

$$M_{c,1} = \left\{W_1^{(1)}, W_2^{(1)}, \cdots, W_D^{(1)}\right\}, \tag{8.80}$$

$$M_{c,2} = \left\{W_1^{(2)}, W_2^{(2)}, \cdots, W_D^{(2)}\right\}. \tag{8.81}$$

The randomization message for cache placement, $\tilde{M}_c = \{\tilde{M}_{c,1}, \tilde{M}_{c,2}\}$, is identical to (8.25). That is, $\tilde{M}_{c,1} = K_1$ and $\tilde{M}_{c,2} = K_2$. Receiver j stores $M_{c,j}$ and $\tilde{M}_{c,j}$ in its cache memory.

Unlike (8.26), we utilize here *partially coded* delivery. The message to be securely transmitted during the delivery phase is

$$M_{\mathbf{d}} = \left\{W_{d_2}^{(1)} \oplus W_{d_1}^{(2)}, W_{d_1,t}, W_{d_2,t}\right\}. \tag{8.82}$$

The randomization message for delivery, $\tilde{M}_{\mathbf{d}}$, is identical to (8.27).

Remark 30. *Notice that the sizes of M_c, $M_{\mathbf{d}}$, $\tilde{M}_c$, and $\tilde{M}_{\mathbf{d}}$, are the same as in Section 8.4.3. In particular, the sizes of $\tilde{M}_c$ and $\tilde{M}_{\mathbf{d}}$ are both $n\alpha_\epsilon$ bits. The size of M_c is given*

by

$$2 \times D \times n\frac{1-\alpha_\epsilon}{2D} = n(1-\alpha_\epsilon) \quad \textit{bits}, \tag{8.83}$$

and the size of $M_{\mathbf{d}}$ *is given by*

$$n\frac{1-\alpha_\epsilon}{2D} + 2 \times n\frac{(2D-1)(1-\alpha_\epsilon)}{4D} = n(1-\alpha_\epsilon) \quad \textit{bits}. \tag{8.84}$$

Codebooks Generation and Encoders: For the messages M_c, $M_{\mathbf{d}}$, $\tilde{M}_c$, and $\tilde{M}_{\mathbf{d}}$ defined above, the cache placement and delivery codebooks, $\mathcal{C}_{c,n}$ and $\mathcal{C}_{\mathbf{d},n}$, and the cache and delivery encoders, are designed in the same exact manner as in Section 8.4.3, see Figures 8.3 and 8.4.

Decoding: As in Section 8.4.3, using $M_{\mathbf{d}}$, $\tilde{M}_{\mathbf{d}}$, $M_{c,j}$, $\tilde{M}_{c,j}$, and for n sufficiently large, receiver j correctly decodes W_{d_j}, $j = 1, 2$.

Security analysis: Let W_c, $\tilde{W}_c$, $W_{\mathbf{d}}$, and $\tilde{W}_{\mathbf{d}}$, be defined as in (8.29)-(8.32), (8.40), and (8.41). Once again, $\tilde{W}_c$ and $\tilde{W}_{\mathbf{d}}$ are independent and uniform, and hence $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$ is jointly uniform. In addition, $\{W_c, W_{\mathbf{d}}\}$ are independent from $\{\tilde{W}_c, \tilde{W}_{\mathbf{d}}\}$. Thus, (8.43) holds for this case.

For any $\mathbf{d} = (d_1, d_2)$, we have

$$I\left(W_{[1:D]}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) = I\left(\left\{W_l^{(1)}, W_l^{(2)}, W_{l,t}, W_{l,s}\right\}_{l=1}^{D}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.85}$$

$$\leq I\left(M_c, \left\{W_l^{(1)}, W_l^{(2)}, W_{l,t}, W_{l,s}\right\}_{l=1}^{D}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.86}$$

$$\leq I\left(M_c, W_{d_2}^{(1)} \oplus W_{d_1}^{(2)}, W_{d_1,t}, W_{d_2,t}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.87}$$

$$= I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{8.88}$$

$$\leq I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) + \epsilon_n', \tag{8.89}$$

where (8.87) follows form the Markov chain $W_{[1:D]} - \left\{M_c, W_{d_2}^{(1)} \oplus W_{d_1}^{(2)}, W_{d_1,t}, W_{d_2,t}, W_{d_1,s}, W_{d_2,s}\right\} - \left\{\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right\}$; (8.88) follows from (8.82), and (8.89) follows using similar steps as in (8.46)-(8.57). Using (8.43) and (8.89), the secrecy constraint in (8.3) is satisfied.

With $\epsilon_n \to 0$ as $n \to \infty$, the achievable strong secrecy file rate is

$$R_s(\alpha) = 2 \times \frac{1-\alpha}{2D} + \frac{(2D-1)(1-\alpha)}{4D} + \frac{\alpha}{2} \tag{8.90}$$

$$= \frac{1}{2} + \frac{3(1-\alpha)}{4D}. \tag{8.91}$$

This completes the proof for Theorem 15.

Remark 31. *For $D = 2$, the achievable secrecy rate in (8.91) is strictly smaller than the secrecy rate obtained by coded placement and uncoded delivery in Section 8.4.3, i.e., $1 - \frac{\alpha}{2}$.*

Remark 32. *An achievability scheme which utilizes coded placement and uncoded delivery, as in Section 8.4.3, achieves the same secure file rate as (8.91) for $D = 3$. However, this scheme achieves a strictly smaller secure file rate for $D \geq 4$. In this scheme, $W_l^{(1)}$ and $W_l^{(2)}$ are uniform over $\left[1 : 2^{n\frac{1-\alpha_\epsilon}{2(D-1)}}\right]$. $W_{l,t}$ is uniform over $\left[1 : 2^{n\frac{(D-2)(1-\alpha_\epsilon)}{2(D-1)}}\right]$. $W_{l,s}$, K_1 and K_2, are uniform over $\left[1 : 2^{n\frac{\alpha_\epsilon}{2}}\right]$. $M_c = \{M_{c,1}, M_{c,2}\}$ and $M_{\mathbf{d}}$ are given by*

$$M_{c,1} = \left\{W_1^{(1)} \oplus W_2^{(1)}, W_2^{(1)} \oplus W_3^{(1)}, \cdots, W_{D-1}^{(1)} \oplus W_D^{(1)}\right\}, \tag{8.92}$$

$$M_{c,2} = \left\{ W_1^{(2)} \oplus W_2^{(2)}, W_2^{(2)} \oplus W_3^{(2)}, \cdots, W_{D-1}^{(2)} \oplus W_D^{(2)} \right\}, \tag{8.93}$$

$$M_{\mathbf{d}} = \left\{ W_{d_2}^{(1)}, W_{d_1}^{(2)}, W_{d_1,t}, W_{d_2,t} \right\}. \tag{8.94}$$

Without loss of generality, let $d_1 < d_2$. *For any* $\mathbf{d} = (d_1, d_2)$, *using* $M_{c,j}$ *in (8.92) and (8.93), receiver* j, *can restore* $W_{d_1}^{(j)} \oplus W_{d_2}^{(j)}$ *by xor-ing* $\left\{ W_{d_1}^{(j)} \oplus W_{d_1+1}^{(j)} \right\}$, $\left\{ W_{d_1+1}^{(j)} \oplus W_{d_1+2}^{(j)} \right\}$,$\cdots$, $\left\{ W_{d_2-1}^{(j)} \oplus W_{d_2}^{(j)} \right\}$. *The achievable strong secrecy file rate using this scheme is* $R_s(\alpha) = \frac{1}{2} + \frac{1-\alpha}{2(D-1)}$.

8.6 Proof of Theorem 16

When $\alpha \in [1, 2]$, the upper bound on R_s, stated in Theorem 16 for $D > 2$, follows as in Section 8.4.1. Thus, it remains to prove the upper bound for $\alpha \in (0, 1)$. The proof is divided into the three following steps.

Step 1*:* We first upper bound R_s by the secrecy capacity when the adversary is restricted to tap into the delivery transmission only, denoted as C_s^{Res}. That is, C_s^{Res} is the maximum achievable file rate when $\alpha_1 = 0$ and $\alpha_2 = \alpha$. Restricting the adversary to only tap into the delivery phase cannot decrease the secrecy capacity, i.e., $R_s \leq C_s^{\text{Res}}$, since this setting is included in the feasible strategy space for the adversary. The cache placement transmission is secure, and each receiver has a secure cache memory of size $\frac{n}{2}$ bits.

Step 2*:* We upper bound C_s^{Res} by the secrecy capacity, i.e., the maximum achievable file rate, when the delivery channel to the adversary is replaced by a discrete memoryless binary erasure channel, with erasure probability $1-\alpha$, denoted as C_s^{DM}. The proof for this step follows the same lines as in Section 5.5 in Chapter 5.

Step 3*:* From Step 1, each receiver has a secure cache of size $\frac{n}{2}$ bits. Since increasing the cache sizes cannot decrease the achievable file rate, we further upper bound C_s^{DM} with the maximum achievable file rate when each receiver has a cache memory of size n bits, in which it stores $\mathbf{X}_c^n$. Receiver j, $j=1,2$, utilizes both $\mathbf{X}_c^n$ and $\mathbf{X}_{\mathbf{d}}^n$ in order to decode its desired message W_{d_j}, i.e., $\hat{W}_{d_j} = g_{\mathbf{d},j}\left(\mathbf{X}_c^n, \mathbf{X}_{\mathbf{d}}^n\right)$, $\mathbf{d}=(d_1,d_2)$. This setup is thus equivalent to a single receiver, with a cache of size n bits, which demands two files W_{d_1}, W_{d_2}, and utilizes the decoder $g_{\mathbf{d}} \triangleq \{g_{\mathbf{d},1}, g_{\mathbf{d},2}\}$. Let us denote the maximum achievable file rate for this single receiver model as C_s^{SR}. We have $C_s^{\text{DM}} \leq C_s^{\text{SR}}$. In the following, we upper bound C_s^{SR}.

Let M_{D} denote the fraction of the size-n bits cache memory dedicated to store (coded or uncoded) information bits, and let M_{K} denote the fraction dedicated to store key bits. That is, $M_{\text{D}} + M_{\text{K}} = 1$. Let S_{D} denote the information bits stored in this memory, i.e., $S_{\text{D}} = f_{\text{D}}(W_{[1:D]})$ and $H(S_{\text{D}}) = nM_{\text{D}}$. We utilize the following lemma in order to upper bound C_s^{SR}.

Lemma 13. *[132, Lemma 1] For a fixed allocation of M_{D} and M_{K}, and a receiver which demands the files W_{d_1} and W_{d_2}, the secrecy rate for the single receiver model is upper bounded as*

$$2R_s^{\text{SR}} \leq \min\left\{1, 1-\alpha+M_{\text{K}}\right\} + \frac{1}{n} I\left(W_{d_1}, W_{d_2}; S_{\text{D}}\right). \tag{8.95}$$

Notice that (8.95) holds for any demand pair $\mathbf{d} = (d_1, d_2)$ such that $d_1 \neq d_2$, i.e., the worst-case demands. Summing over all such demands, we have

$$2R_s^{\mathrm{SR}} \leq \min\{1, 1-\alpha+M_{\mathrm{K}}\} + \frac{1}{nD(D-1)} \sum_{d_1,d_2\in[1:D],\, d_1\neq d_2} I\left(W_{d_1}, W_{d_2}; S_{\mathrm{D}}\right). \quad (8.96)$$

The second term on the right hand side of (8.96) can be written as

$$\frac{1}{nD(D-1)} \sum_{d_1,d_2\in[1:D],\, d_1\neq d_2} I\left(W_{d_1}, W_{d_2}; S_{\mathrm{D}}\right)$$

$$= \frac{1}{nD} \sum_{d_1\in[1:D]} I\left(W_{d_1}; S_{\mathrm{D}}\right) + \frac{1}{nD(D-1)} \sum_{d_1,d_2\in[1:D],\, d_1\neq d_2} I\left(W_{d_2}; S_{\mathrm{D}}|W_{d_1}\right) \quad (8.97)$$

$$\leq \frac{1}{nD} \sum_{d_1\in[1:D]} I\left(W_{d_1}; S_{\mathrm{D}}\right) + \frac{1}{nD(D-1)} \sum_{d_1\in[1:D]} \left(\sum_{d_2\in[1:D]} I\left(W_{d_2}; S_{\mathrm{D}}|W_{d_1}\right) \right). \quad (8.98)$$

For any $d_1 \in [1:D]$, we have

$$\sum_{d_2\in[1:D]} I\left(W_{d_2}; S_{\mathrm{D}}\middle|W_{d_1}\right) = \sum_{d_2=1}^{D} \left[H\left(W_{d_2}\middle|W_{d_1}\right) - H\left(W_{d_2}\middle|W_{d_1}, S_{\mathrm{D}}\right)\right] \quad (8.99)$$

$$\leq \sum_{d_2=1}^{D} \left[H\left(W_{d_2}\middle|W_1, W_2, \cdots, W_{d_2-1}, W_{d_1}\right) - H\left(W_{d_2}\middle|W_1, W_2, \cdots, W_{d_2-1}, W_{d_1}, S_{\mathrm{D}}\right)\right] \quad (8.100)$$

$$= I\left(W_1, W_2, \cdots, W_{\mathrm{D}}; S_{\mathrm{D}}\middle|W_{d_1}\right) \quad (8.101)$$

$$\leq H(S_{\mathrm{D}}) = nM_{\mathrm{D}}, \quad (8.102)$$

where (8.100) follows because when $d_2 = d_1$, $H\left(W_{d_2}|W_{d_1}\right) = H\left(W_{d_2}|W_1, W_2, \cdots, W_{d_2-1}, W_{d_1}\right) = 0$, and when $d_2 \neq d_1$, $H\left(W_{d_2}|W_{d_1}\right) = H\left(W_{d_2}|W_1, W_2, \cdots, W_{d_2-1}, W_{d_1}\right) = H(W_{d_2})$.

Similarly, we have

$$\sum_{d_1 \in [1:D]} I\left(W_{d_1}; S_{\mathrm{D}}\right) \leq H(S_{\mathrm{D}}) = nM_{\mathrm{D}}. \tag{8.103}$$

Substituting (8.102) and (8.103) in (8.98) gives

$$\frac{1}{nD(D-1)} \sum_{d_1, d_2 \in [1:D],\, d_1 \neq d_2} I\left(W_{d_1}, W_{d_2}; S_{\mathrm{D}}\right) \leq \frac{2D-1}{D(D-1)} M_{\mathrm{D}}. \tag{8.104}$$

Thus, using (8.96) and (8.104), R_s^{SR} is further upper bounded as

$$R_s^{\mathrm{SR}} \leq \frac{1}{2}\left[\min\left\{1, 1-\alpha+M_{\mathrm{K}}\right\} + \frac{2D-1}{D(D-1)} M_{\mathrm{D}}\right]. \tag{8.105}$$

Finally, by maximizing over all possible allocations for M_{D} and M_{K} such that $M_{\mathrm{D}} + M_{\mathrm{K}} = 1$, we obtain

$$C_s^{\mathrm{SR}} \leq \frac{1}{2} \max_{\substack{M_{\mathrm{D}}, M_{\mathrm{K}}: \\ M_{\mathrm{D}}+M_{\mathrm{K}}=1}} \left\{\min\left\{1, 1-\alpha+M_{\mathrm{K}}\right\} + \frac{2D-1}{D(D-1)} M_{\mathrm{D}}\right\} \tag{8.106}$$

$$= \frac{1}{2}\left[1 + \frac{2D-1}{D(D-1)}(1-\alpha)\right]. \tag{8.107}$$

Equation (8.107) follows because, for $D \geq 3$, the maximum occurs at $M_{\mathrm{K}} = \alpha$ and $M_{\mathrm{D}} = 1 - \alpha$. This completes the proof for Theorem 16.

Remark 33. *An upper bound considering uncoded placement only can be derived as follows. The same analysis as in (8.95)-(8.107) carries through with $I(W_{d_2}; S_{\mathrm{D}}|W_{d_1})$ in*

(8.97) is equal to $I(W_{d_2}; S_{\mathrm{D}})$*. Hence the right hand side of (8.104) is replaced by* $\frac{2M_{\mathrm{D}}}{D}$*. The resulting bound* $R_s \leq \frac{1}{2} + \frac{(1-\alpha)}{D}$ *is tighter than (8.107).*

8.7 Discussion

While the fixed-size cache memory setup considered in this chapter can be seen as a clean basic model for the intricate problem in consideration, it also allows us to obtain results and insights that are generalizable to more involved cache memory models. In particular, the extension to variable memory sizes can be done by considering multiple communication blocks for cache placement. Our results and coding scheme readily apply to an adversary model whose tapping capability during the delivery is normalized with respect to tapping during cache placement, i.e., $\mu_1 + B\mu_2 \leq \mu$; B is the number of communication blocks for cache placement. This is a reasonable assumption given that cache placement generally takes place in a longer period than delivery. The problem turns to be more challenging when the adversary optimizes its tapping uniformly over the multiple blocks for cache placement as well as the delivery phase. This is left for future investigation.

It is typical to model the cache placement as a noiseless channel since placement is assumed to occur when networks are not congested and their rates are assumed to be large enough. Here however we model the cache placement as a broadcast channel communication. The broadcast model avails a clean and tractable solution without compromising its generalizability. A time division multiple access (TDMA) model for cache placement is a special case by imposing an additional constraint in which each receiver

has to decode its desired file using only one half of the transmitted codeword. Additionally, the broadcast model is in line with the network information theory literature and it does not limit the cache placement to occur over low rate traffic. With the ever-growing user demands, placement and delivery occurring in less asymmetric network loads is likely to be expected in the near future.

Corollary 7 demonstrates that, for the model considered in this chapter, when $\alpha \in [1,2]$, the strong secrecy capacity is equal to $1-\frac{\alpha}{2}$ for any library size. For $\alpha \in [1,2]$, $\{S_1 = [1:n], S_2 \subset [1:n]\}$ and $\{S_1 \subset [1:n], S_2 = [1:n]\}$ are two possible strategies for the adversary. In other words, the adversary can tap into either all transmitted symbols in cache placement and a subset of symbols in the delivery, or all transmitted symbols in the delivery and a subset of symbols in cache placement. Such an adversary limits the communication for cache placement, i.e., the use of cache memories, to exchanging additional randomness (key bits) that allows for communicating a positive secure rate over the two phases. In other words, the cache memories are not utilized to store any data bits, and hence the lack of knowledge of user demands during cache placement is immaterial.

For a library with two files, if the receivers were to have cache memories of size n bits in which they store the transmitted signal during cache placement, the strong secrecy file rate in Theorem 14 is achievable using a simple wiretap code. In particular, the transmitter encodes $W = (W_1, W_2) \in [1 : 2^{n2R}]$ into a length-$2n$ binary codeword using a wiretap code, and sends the first n bits of this codeword during cache placement and the last n bits during delivery. Each receiver can thus decode both files, and the secrecy of W_1 and W_2 against the adversary follows by the results in [33, 91]. In caching

problems, the relevant setup however is when the receivers have cache memories of limited size with respect to the overall transmission during cache placement. This calls for the limited size cache memory model considered in this chapter, which in turn necessitates the use of the more elaborate coding scheme in Section 8.4.

8.8 Conclusion

We have introduced the caching broadcast channel with a *a wire and cache* tapping adversary of type II. In this broadcast model, each receiver is equipped with a fixed-size cache memory, and the adversary is able to tap into a subset of its choosing of the transmitted symbols during cache placement, or delivery, or both. The legitimate terminals have no knowledge about the fractions of the tapped symbols in each phase, or their positions. Only the size of the overall tapped set is known. We have identified the strong secrecy capacity of this model, i.e., the maximum achievable file rate while keeping the overall library secure, when the transmitter's library has two files. We have derived lower and upper bounds for the strong secrecy file rate when the transmitter has more than two files in its library. We have devised an achievability scheme which combines wiretap coding, security embedding codes, one-time pad keys, and coded caching techniques. The results presented in this chapter highlight the robustness of (stochastic) coding against a smart adversary which performs a chosen attack, jointly optimized over both cache placement and delivery phases. Future directions include investigating a tighter upper bound for a library with more than two files, and exploring the extensions of this work to variable cache memory sizes, more than two users, and a noisy legitimate channel.

Chapter 9

Conclusion

9.1 Thesis Summary

In this thesis, we have investigated more capable models for the adversary against which we showed that information theoretic security guarantees are possible. The thesis is divided into three major themes. In the first theme, we have considered an adversary with potentially more resources than the legitimate terminals, in terms of the number of its antennas. We have shown that a positive secure degrees of freedom, i.e., a secrecy capacity that scales with the transmit power, is attainable against such an adversary by employing a helper terminal in the network such that the combined number of transmit antennas exceeds the number of adversary's antennas. To do so, we have utilized a variety of beamforming, alignment, and signaling schemes for the different antenna configurations of the model. Among these schemes, we have devised a projection and cancellation decoding scheme for structured transmitted signals, which is optimal for certain antenna configurations.

In the second theme of the thesis, we have addressed adversarial models in which the adversary performs a chosen codeword (cipher-text) attack. In particular, we have considered the wiretap II channel model, introduced three decades ago, and generalized the model outside its original special communication setting. We have introduced a *noisy* legitimate channel to the model, and derived inner and outer bounds on its secrecy

capacity, which match for certain instances. Further, we have introduced a *generalized* wiretap channel which subsumes both the classical wiretap and the wiretap II models as its special cases. The adversary chooses a subset of the transmitted symbols to noiselessly observe, while observing the remainder through a noisy channel. We have derived the strong secrecy capacity of the model. Achievability is established by utilizing concentration of measures to prove a super-exponential convergence rate for the security measure, which dominates the exponentially many possible strategies for the adversary. We have explored several multi-terminal extensions of our generalized wiretap model, including the multiple access wiretap channel, the broadcast wiretap channel with common and private messages, and the interference and broadcast channels with confidential messages, and highlighted some insightful remarks.

In the third theme, we have studied the impact of an adversary that designs an attack in a multi-phase communication system. In particular, we have studied the problem of coded caching in the presence of an adversary which taps into a subset of symbols of its choice either from cache placement, delivery, or both transmissions. The legitimate terminals know neither the fractions of tapped symbols in each phase, nor their positions. Introducing an adversary which taps into cache placement is original and requires challenging standard assumptions in caching literature. We have derived the strong secrecy capacity for the instance of two receivers and two library files, and inner and outer bounds for the case of library size of three or larger. We have shown that information theoretic security is possible against a powerful adversary which optimizes its chosen attack over both phases of a cache-aided communication system.

9.2 Future Directions

The models we have introduced and studied in this thesis can be viewed as examples of adversaries that are stronger than those encountered previously in the information theoretic security literature, and against which strong information theoretic security guarantees are still possible. Adopting a similar line of thought to what we have presented is important for information theoretic security to move forward and its vision to get closer to be implemented in real and practical systems.

Several generalizations and practical concerns about the models we investigated are of interest for future research. Among those, the extension of the generalized wiretap model to Gaussian legitimate and wiretapper channels seems tractable. Additionally, generalizing the caching broadcast channel with a type II adversary to the cases of multiple cache placement blocks, more than two users, and noisy communication channels are open interesting directions.

We note that the type II adversary we have extensively studied in this thesis gives rise to an interesting modeling direction, in which the transmitter learns the existence of an adversary through feedback signals, caused by the changes in the surrounding environment, and aborts the communication protocol whenever positive. In this model, the threshold of the learning process corresponds to the threshold on the time during which the adversary can co-exist in the communication medium without being detected, and hence tap into the communication.

Appendix A

Proof of Lemma 4

Let $U_1, U_2, \cdots, U_n$ be a sequence of non-negative independent random variables, which satisfy the conditions of the Lemma. For any $\theta > 0$, we have

$$\mathbb{P}\Big(\sum_{i=1}^{n} U_i \geq (1+\epsilon)\bar{m} \Big) = \mathbb{P}\left(e^{\theta \sum_{i=1}^{n} U_i} \geq e^{\theta(1+\epsilon)\bar{m}} \right) \tag{A.1}$$

$$\leq \frac{\mathbb{E}\left(e^{\theta \sum_{i=1}^{n} U_i} \right)}{e^{\theta(1+\epsilon)\bar{m}}} \tag{A.2}$$

$$= \frac{\prod_{i=1}^{n} \mathbb{E}\left(e^{\theta U_i} \right)}{e^{\theta(1+\epsilon)\bar{m}}} \tag{A.3}$$

$$\leq \frac{\prod_{i=1}^{n} \left(1 + \frac{e^{\theta b}-1}{b}\, \mathbb{E}(U_i) \right)}{e^{\theta(1+\epsilon)\bar{m}}} \tag{A.4}$$

$$\leq \frac{\prod_{i=1}^{n} e^{\frac{e^{\theta b}-1}{b}\bar{m}_i}}{e^{\theta(1+\epsilon)\bar{m}}} \tag{A.5}$$

$$\leq \frac{e^{\frac{e^{\theta b}-1}{b}\bar{m}}}{e^{\theta(1+\epsilon)\bar{m}}} \tag{A.6}$$

$$= \exp\left(- \left[\theta(1+\epsilon) - \frac{e^{\theta b}-1}{b} \right] \bar{m} \right), \tag{A.7}$$

where (A.2) follows from Markov's inequality. (A.4) follows because $e^{\theta x} \leq 1 + \frac{e^{\theta b}-1}{b} x$ for $x \in [0, b]$, as e^x is a convex function in x, (A.5) follows because $1 + x \leq e^x$ for all $x \geq 0$, and (A.6) follows because $\sum_{i=1}^{n} \bar{m}_i \leq \bar{m}$.

The value of θ which maximizes the right hand side of (A.7) is $\theta^* = \frac{1}{b}\ln(1+\epsilon) > 0$, for which we have

$$\mathbb{P}\left(\sum_{i=1}^{n} U_i \geq \ (1+\epsilon)\bar{m}\right) \leq \exp\left(-\frac{\bar{m}}{b}\left[(1+\epsilon)\left(\ln(1+\epsilon)-1\right)+1\right]\right). \tag{A.8}$$

By considering Taylor's expansion of $x[ln(x)-1]$ around $x=1$, we have, for all $x \geq 1$, that

$$x[\ln(x)-1]+1 \geq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3. \tag{A.9}$$

We also have, for $x \in [1,2]$, that

$$\frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3 \geq \frac{1}{3}(x-1)^2. \tag{A.10}$$

Thus, for all $x \in [1,2]$, we have

$$x[\ln(x)-1]+1 \geq \frac{1}{3}(x-1)^2. \tag{A.11}$$

By applying (A.11), with $x = (1+\epsilon)$, to the right hand side of (A.8), we have, for $\epsilon \in [0,1]$, that

$$\mathbb{P}\left(\sum_{i=1}^{n} U_i \geq \ (1+\epsilon)\bar{m}\right) \leq \exp\left(-\frac{\bar{m}}{3b}\epsilon^2\right). \tag{A.12}$$

Appendix B

Choice of $\mathbf{K}_t$ and $\mathbf{K}_c$

The covariance matrices $\mathbf{K}_t$ and $\mathbf{K}_c$ are chosen so that they are *positive definite*, i.e., $\mathbf{K}_t, \mathbf{K}_c \succ \mathbf{z}$, and hence non-singular, in order to guarantee the finiteness of $h(\tilde{\mathbf{Z}}_t)$ and $h(\tilde{\mathbf{Z}}_c)$ in (3.26). In addition, positive definite $\mathbf{K}_t$ and $\mathbf{K}_c$ result in positive definite $\mathbf{\Sigma}_{\tilde{\mathbf{Z}}_1}$ and $\mathbf{\Sigma}_{\tilde{\mathbf{Z}}_2}$, and hence, $h(\tilde{\mathbf{Z}}_1)$ and $h(\tilde{\mathbf{Z}}_2)$ in (3.28) are also finite.

For $\mathbf{I}_{N_e} - \mathbf{G}_t\mathbf{K}_t\mathbf{G}_t^H$ to be a valid covariance matrix for $\tilde{\mathbf{Z}}_e$ in (3.30), $\mathbf{K}_t$ has to satisfy $\mathbf{G}_t\mathbf{K}_t\mathbf{G}_t^H \preceq \mathbf{I}_{N_e}$, which is equivalent to

$$||\mathbf{K}_t^{\frac{1}{2}}\mathbf{G}_t^H|| \leq 1. \tag{B.1}$$

Recall that $||\mathbf{K}_t^{\frac{1}{2}}\mathbf{G}_t^H||$ is the induced norm for the matrix $\mathbf{K}_t^{\frac{1}{2}}\mathbf{G}_t^H$.

Similarly, for $\mathbf{I}_N - \mathbf{H}_c\mathbf{K}_c\mathbf{H}_c^H$, $\mathbf{I}_{N_e} - \mathbf{G}_t\mathbf{K}_t\mathbf{G}_t^H - \mathbf{G}_c\mathbf{K}_c\mathbf{G}_c^H$, and $\mathbf{I}_N - \mathbf{H}'_{c_2}\mathbf{K}'_c\mathbf{H}'^H_{c_2}$ to be valid covariance matrices for $\tilde{\mathbf{Z}}_r, \tilde{\mathbf{Z}}'_e$, and $\tilde{\mathbf{Z}}'_r$, in (3.40), (3.53), (3.62), $\mathbf{K}_t, \mathbf{K}_c, \mathbf{K}'_c$ have to satisfy

$$||\mathbf{K}_c^{\frac{1}{2}}\mathbf{H}_c^H|| \leq 1, \quad ||\mathbf{K}_t^{\frac{1}{2}}\mathbf{G}_t^H||^2 + ||\mathbf{K}_c^{\frac{1}{2}}\mathbf{G}_c^H||^2 \leq 1,$$

$$\text{and } ||\mathbf{K}'^{\frac{1}{2}}_c\mathbf{H}'^H_{c_2}|| \leq 1. \tag{B.2}$$

In order to satisfy the conditions (B.1) and (B.2), we choose $\mathbf{K}_t = \rho^2 \mathbf{I}_N$, $\mathbf{K}_c = \rho^2 \mathbf{I}_K$, where

$$0 < \rho \leq \frac{1}{\max\left\{||\mathbf{G}_t^H||, ||\mathbf{H}_c^H||, ||\mathbf{H}_{c_2}'^H||, \sqrt{||\mathbf{G}_t^H||^2 + ||\mathbf{G}_c^H||^2}\right\}} \tag{B.3}$$

$$= \frac{1}{\max\left\{||\mathbf{H}_c^H||, \sqrt{||\mathbf{G}_t^H||^2 + ||\mathbf{G}_c^H||^2}\right\}}. \tag{B.4}$$

Appendix C

Derivation of (3.48), (3.49), and (3.66)

In order to upper bound $h(Y_{r,k}(i))$, for all $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, N$, we first upper bound the variance of $Y_{r,k}(i)$, denoted by $\mathbb{V}\text{ar}\left(Y_{r,k}(i)\right)$. Let $\mathbf{h}_{t,k}^{r}$ and $\mathbf{h}_{c,k}^{r}$ denote the transpose of the kth row vectors of $\mathbf{H}_t$ and $\mathbf{H}_c$, respectively. Let $\mathbf{Z}_r(i) = \left[Z_{r,1}(i) \cdots Z_{r,N}(i)\right]^T$. Using (3.1), $Y_{r,k}(i)$ is expressed as

$$Y_{r,k}(i) = \mathbf{h}_{t,k}^{r^T}\mathbf{X}_t(i) + \mathbf{h}_{c,k}^{r^T}\mathbf{X}_c(i) + Z_{r,k}(i). \tag{C.1}$$

Thus, $\mathbb{V}\text{ar}\left(Y_{r,k}(i)\right)$ can be bounded as

$$\mathbb{V}\text{ar}\left(Y_{r,k}(i)\right) \leq \mathbb{E}\left(Y_{r,k}(i)Y_{r,k}^{*}(i)\right) \tag{C.2}$$

$$= \mathbb{E}\left(|\mathbf{h}_{t,k}^{r^T}\mathbf{X}_t(i)|^2\right) + \mathbb{E}\left(|\mathbf{h}_{c,k}^{r^T}\mathbf{X}_c(i)|^2\right) + \mathbb{E}\left(|Z_{r,k}(i)|^2\right) \tag{C.3}$$

$$\leq ||\mathbf{h}_{t,k}^{r}||^2\, \mathbb{E}\left(||\mathbf{X}_t(i)||^2\right) + ||\mathbf{h}_{c,k}^{r}||^2\, \mathbb{E}\left(||\mathbf{X}_c(i)||^2\right) + 1 \tag{C.4}$$

$$\leq 1 + \left(||\mathbf{h}_{t,k}^{r}||^2 + ||\mathbf{h}_{c,k}^{r}||^2\right) P, \tag{C.5}$$

where (C.4) follows from Cauchy-Schwarz inequality and monotonicity of expectation, and (C.5) follows from the power constraints at the transmitter and cooperative jammer.

Define $h^2 = \max\limits_{k}\ \left(||\mathbf{h}_{t,k}^{r}||^2 + ||\mathbf{h}_{c,k}^{r}||^2\right)$. Since $h(Y_{r,k}(i))$ is upper bounded by the entropy of a complex Gaussian random variable with the same variance, we have, for all

$i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, N$,

$$h(Y_{r,k}(i)) \leq \log 2\pi e \left(1 + \left(||\mathbf{h}^r_{t,k}||^2 + ||\mathbf{h}^r_{c,k}||^2\right) P\right) \tag{C.6}$$

$$\leq \log 2\pi e + \log(1 + h^2 P). \tag{C.7}$$

Similarly, we have

$$\bar{Y}_{r,k}(i) = \mathbf{h}^{r^T}_{t,k}\mathbf{X}_t(i) + \mathbf{h}'^{r^T}_{c,k}\mathbf{X}'_{c_2}(i) + Z_{r,k}(i), \tag{C.8}$$

where $\mathbf{h}'^r_{c,k}$ is the transpose of the k-th row vector of $\mathbf{H}'_{c_2}$. Thus, we have,

$$h(\bar{Y}_{r,k}(i)) \leq \log 2\pi e + \log(1 + \bar{h}^2 P), \tag{C.9}$$

where $\bar{h}^2 = \max\limits_k \left(||\mathbf{h}^r_{t,k}||^2 + ||\mathbf{h}'^r_{c,k}||^2\right)$.

Next, we upper bound $h(\tilde{X}_{t,k}(i))$. The power constraint at the transmitter, for $i = 1, 2, \cdots, n$, is $\mathbb{E}\left(\mathbf{X}^H_t(i)\ \mathbf{X}_t(i)\right) = \sum_{k=1}^N \mathbb{E}\left(|X_{t,k}(i)|^2\right) \leq P$. Thus, $\mathbb{E}\left(|X_{t,k}(i)|^2\right) \leq P$ for all $i = 1, 2, \cdots, n$, and $k = 1, 2, \cdots, N$. Recall that $\tilde{X}_{t,k}(i) = X_{t,k}(i) + \tilde{Z}_{t,k}(i)$, where $X_{t,k}(i)$ and $\tilde{Z}_{t,k}(i)$ are independent, and the covariance matrix of $\tilde{\mathbf{Z}}_t$ is $\mathbf{K}_t = \rho^2\mathbf{I}_N$, where $0 < \rho \leq \min\left\{\frac{1}{||\mathbf{H}^H_c||}, \frac{1}{\sqrt{||\mathbf{G}^H_t||^2 + ||\mathbf{G}^H_c||^2}}\right\}$. Thus, $\mathbb{V}\text{ar}\left(\tilde{X}_{t,k}(i)\right)$ is upper bounded as

$$\mathbb{V}\text{ar}\left(\tilde{X}_{t,k}(i)\right) = \mathbb{V}\text{ar}\left(X_{t,k}(i)\right) + \mathbb{V}\text{ar}\left(\tilde{Z}_{t,k}(i)\right) \tag{C.10}$$

$$\leq \mathbb{E}\left(|X_{t,k}(i)|^2\right) + \rho^2 \leq P + \rho^2. \tag{C.11}$$

Thus, for $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, N$, we have

$$h(\tilde{X}_{t,k}(i)) \leq \log 2\pi e + \log(\rho^2 + P). \tag{C.12}$$

Similarly, using the power constraint at the cooperative jammer, we have, for $i = 1, \cdots, n$ and $m = 1, \cdots, K$,

$$h(\tilde{X}_{c,m}(i)) \leq \log 2\pi e + \log(\rho^2 + P). \tag{C.13}$$

Appendix D

Proof of Lemma 6

Consider two matrices $\mathbf{Q} \in \mathbb{C}^{M\times K}$ and $\mathbf{W} \in \mathbb{C}^{K\times N}$ such that $\mathbf{Q}$ is full row-rank and $\mathbf{W}$ has all of its entries independently drawn from a continuous distribution, where $K > N, M$. Let $L = \min\{N, M\}$. We show that $\mathbf{QW}$ has a rank L a.s. The matrices $\mathbf{Q}$ and $\mathbf{W}$ can be written as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_K \end{bmatrix}, \tag{D.1}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_N \end{bmatrix}, \tag{D.2}$$

where $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_K$ are the K length-M column vectors of $\mathbf{Q}$, and $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N$ are the N length-K column vectors of $\mathbf{W}$.

Let $w_{m,i}$ denotes the entry in the mth row and ith column of $\mathbf{W}$. Let $\mathbf{QW} = [\mathbf{s}_1\ \mathbf{s}_2\ \cdots\ \mathbf{s}_N]$, where $\mathbf{s}_i$ is a length-M column vector, $i = 1, 2, \cdots, N$. When $M \geq N$, $\mathbf{QW} = [\mathbf{s}_1\ \mathbf{s}_2\ \cdots\ \mathbf{s}_L]$, and when $M < N$, $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L\}$ are the first L columns of $\mathbf{QW}$. In order to show that the matrix $\mathbf{QW}$ has rank L, we show that, in either case, $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L\}$ are a.s. linearly independent, i.e.,

$$\sum_{i=1}^{L} \lambda_i \mathbf{s}_i = \mathbf{z}_{M\times 1} \tag{D.3}$$

if and only if $\lambda_i = 0$ for all $i = 1, 2, \cdots, L$.

Each $\mathbf{s}_i$, for $i = 1, 2, \cdots, L$, can be viewed as a linear combination of the K columns of $\mathbf{Q}$ with coefficients that are the entries of the ith column of $\mathbf{W}$, i.e.,

$$\mathbf{s}_i = \sum_{m=1}^{K} w_{m,i}\mathbf{q}_m. \tag{D.4}$$

Using (D.4), we can rewrite (D.3) as

$$\sum_{m=1}^{K} \varphi_m \mathbf{q}_m = \mathbf{z}_{M\times 1} \tag{D.5}$$

where, for $m = 1, 2, \cdots, K$,

$$\varphi_m = \sum_{i=1}^{L} \lambda_i w_{m,i}. \tag{D.6}$$

The K columns of $\mathbf{Q}$ are linearly dependent since each of them is of length M and $K > M$. Thus, equation (D.5) has infinitely many solutions for $\{\varphi_m\}_{m=1}^{K}$.

Each of these solutions for φ_m's constitutes a system of K linear equations $\{\varphi_m = \sum_{i=1}^{L} \lambda_i w_{m,i}, m = 1, 2, \cdots, K\}$. The number of unknowns in this system, i.e. λ's, is L. Since the number of equations in this system, K, is greater than the number of unknowns, L, this system has a solution for $\{\lambda_i\}_{i=1}^{L}$ only if the elements $\{w_{m,i} : m = 1, 2, \cdots, K, \text{ and } i = 1, 2, \cdots, L\}$ are dependent. Since the entries of $\mathbf{W}$ are all randomly and independently drawn from some continuous distribution, the probability that these entries are dependent is zero.

Moreover, consider the set with infinite cardinality, where each element in this set is a structured $\mathbf{W}$ that causes the system of equations in (D.6) to have a solution

for $\{\lambda_i\}_{i=1}^{L}$ for one of the infinitely many solutions of $\{\varphi_m\}_{m=1}^{K}$ to (D.5). This set with infinite cardinality has a measure zero in the space $\mathbb{C}^{K\times L}$, since this set is a subspace of $\mathbb{C}^{K\times L}$ with a dimension strictly less than $K \times L$. We conclude that (D.3) a.s. has no non-zero solution for $\{\lambda_i\}_{i=1}^{L}$. Thus, $\mathbf{QW}$ has rank L a.s.

If $\mathbf{QW}$ has rank L a.s. , then so does $(\mathbf{QW})^T = \mathbf{W}^T\mathbf{Q}^T$. Setting $\mathbf{E}_1 = \mathbf{W}^T$ and $\mathbf{E}_2 = \mathbf{Q}^T$, we have $\mathbf{E}_1 \in \mathbb{C}^{N\times K}$ has all of its entries independently drawn from some continuous distribution, $\mathbf{E}_2 \in \mathbb{C}^{K\times M}$ is full column-rank, $K > N, M$, and $\mathbf{E}_1\mathbf{E}_2$ has rank $L = \min\{N, M\}$ a.s. Thus, Lemma 6 is proved.

Appendix E

Derivation of (3.89) and (3.90)

The power constraints at the transmitter and cooperative jammer are $\mathbb{E}\left(\mathbf{X}_t^H\mathbf{X}_t\right) \leq P$ and $\mathbb{E}\left(\mathbf{X}_c^H\mathbf{X}_c\right) \leq P$. Using (3.72), we have

$$\mathbb{E}\left(\mathbf{X}_t^H\mathbf{X}_t\right) = \mathbb{E}\left(\mathbf{U}_t^H\mathbf{P}_t^H\mathbf{P}_t\mathbf{U}_t\right) \tag{E.1}$$

$$= \sum_{i=1}^{d}\sum_{m=1}^{d}\mathbf{p}_{t,m}^H\mathbf{p}_{t,i}\mathbb{E}\left(U_m^*U_i\right) \tag{E.2}$$

$$= \sum_{i=1}^{d}||\mathbf{p}_{t,i}||^2\mathbb{E}\left(|U_i|^2\right) \tag{E.3}$$

$$= ||\mathbf{p}_{t,1}||^2\mathbb{E}\left(|U_1|^2\right) + \sum_{i=2}^{d}||\mathbf{p}_{t,i}||^2\left(\mathbb{E}\left(U_{i,\mathrm{Re}}^2\right) + \mathbb{E}\left(U_{i,\mathrm{Im}}^2\right)\right) \tag{E.4}$$

$$\leq \left(||\mathbf{p}_{t,1}||^2 + 2\sum_{i=2}^{d}||\mathbf{p}_{t,i}||^2\right)a^2Q^2, \tag{E.5}$$

where (E.3) follows since U_i and U_m, for $i \neq m$, are independent, and (E.5) follows since $\mathbb{E}\left(U_1^2\right)$, $\mathbb{E}\left(U_{i,\mathrm{Re}}^2\right)$, $\mathbb{E}\left(U_{i,\mathrm{Im}}^2\right) \leq a^2Q^2$, for $i = 2, 3, \cdots, d$.

Similarly, using (3.72) and (3.87), we have

$$\mathbb{E}\left(\mathbf{X}_c^H\mathbf{X}_c\right) = \mathbb{E}\left(\mathbf{V}_c^H\mathbf{P}_c^H\mathbf{P}_c\mathbf{V}_c\right) = \sum_{i=1}^{l}\mathbb{E}\left(|V_i|^2\right) \tag{E.6}$$

$$= \mathbb{E}\left(V_1^2\right) + \sum_{i=2}^{l}\left(\mathbb{E}\left(V_{i,\mathrm{Re}}^2\right) + \mathbb{E}\left(V_{i,\mathrm{Im}}^2\right)\right) \tag{E.7}$$

$$\leq (2l-1)a^2Q^2. \tag{E.8}$$

From (E.5) and (E.8), in order to satisfy the power constraints, we need that

$$a^2 Q^2 \leq \gamma^2 P, \tag{E.9}$$

where,

$$\gamma^2 = \frac{1}{\max\left\{2l-1, ||\mathbf{p}_{t,1}||^2 + 2\sum_{i=2}^{d} ||\mathbf{p}_{t,i}||^2\right\}}. \tag{E.10}$$

Let us choose the integer Q as

$$Q = \left\lfloor P^{\frac{1-\epsilon}{2+\epsilon}} \right\rfloor = P^{\frac{1-\epsilon}{2+\epsilon}} - \nu, \tag{E.11}$$

where ν is a constant which does not depend on the power P. Thus,

$$a = \gamma P^{\frac{3\epsilon}{2(2+\epsilon)}}, \tag{E.12}$$

satisfies the condition in (E.9). Thus, the power constraints at the transmitter and cooperative jammer are satisfied.

Appendix F

Proof of Lemma 7

For $w, f \in [1 : \tilde{W}] \times [1 : \tilde{F}]$, we have

$$P_{WF}(w, f) = \sum_{x \in \mathcal{X}} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}. \tag{F.1}$$

We also have that, for all $x \in \mathcal{X}$,

$$\begin{aligned} &\mathbb{E}_{\mathcal{B}}\Big(\mathbb{1}\left\{\mathcal{B}_1(x) = w\right\} \mathbb{1}\left\{\mathcal{B}_2(x) = f\right\} \Big) \\ &\quad = \mathbb{P}(\mathcal{B}_1(x) = w)\mathbb{P}(\mathcal{B}_2(x) = f) = \frac{1}{\tilde{W}\tilde{F}}. \end{aligned} \tag{F.2}$$

Thus, we have $\mathbb{E}_{\mathcal{B}}(P_{WF}) = \frac{1}{\tilde{W}\tilde{F}} = p_W^U p_F^U$. For all w and f, define the random variables

$$P_1(w, f) = \sum_{x \notin \mathcal{D}_\gamma} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \tag{F.3}$$

$$P_2(w, f) = \sum_{x \in \mathcal{D}_\gamma} p_X(x) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}. \tag{F.4}$$

Note that $P_{WF}(w, f) = P_1(w, f) + P_2(w, f)$. Thus, we have

$$\begin{aligned} &\mathbb{E}_{\mathcal{B}} \left(\mathbb{V} \left(P_{WF}, p_W^U p_F^U \right) \right) \\ &\quad = \frac{1}{2} \mathbb{E}_{\mathcal{B}} \left(\sum_{w,f} \Big| P_{WF}(w, f) - \mathbb{E}_{\mathcal{B}} \left(P_{WF}(w, f) \right) \Big| \right) \end{aligned} \tag{F.5}$$

$$= \frac{1}{2}\mathbb{E}_{\mathcal{B}}\left(\sum_{w,f}\left|\sum_{i=1}^{2}\left(P_i(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_i(w,f)\right)\right)\right|\right) \tag{F.6}$$

$$\leq \frac{1}{2}\sum_{w,f}\mathbb{E}_{\mathcal{B}}\left|P_1(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_1(w,f)\right)\right|$$

$$+ \frac{1}{2}\sum_{w,f}\mathbb{E}_{\mathcal{B}}\left|P_2(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_2(w,f)\right)\right|, \tag{F.7}$$

where (F.7) follows from the triangle inequality. We now upper bound each term on the right hand side of (F.7). For the first term, we have

$$\frac{1}{2}\sum_{w,f}\mathbb{E}_{\mathcal{B}}\Big|P_1(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_1(w,f)\right)\Big| \leq \sum_{w,f}\mathbb{E}_{\mathcal{B}}\left(P_1(w,f)\right) \tag{F.8}$$

$$= \sum_{w,f}\sum_{x\notin\mathcal{D}_\gamma} p_X(x)\,\mathbb{E}_{\mathcal{B}}\Big(\mathbb{1}\left\{\mathcal{B}_1(x) = w\right\}\mathbb{1}\left\{\mathcal{B}_2(x) = f\right\}\Big) \tag{F.9}$$

$$= \sum_{x\notin\mathcal{D}_\gamma} p_X(x) = \mathbb{P}(X \notin \mathcal{D}_\gamma), \tag{F.10}$$

where (F.8) follows from the triangle inequality.

For the second term in the right hand side of (F.7), we have

$$\frac{1}{2}\sum_{w,f}\mathbb{E}_{\mathcal{B}}\Big|P_2(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_2(w,f)\right)\Big|$$

$$= \frac{1}{2}\sum_{w,f}\mathbb{E}_{\mathcal{B}}\sqrt{\left(P_2(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_2(w,f)\right)\right)^2} \tag{F.11}$$

$$\leq \frac{1}{2}\sum_{w,f}\sqrt{\mathbb{E}_{\mathcal{B}}\left(P_2(w,f) - \mathbb{E}_{\mathcal{B}}\left(P_2(w,f)\right)\right)^2} \tag{F.12}$$

$$= \frac{1}{2}\sum_{w,f}\sqrt{\mathbb{V}\mathrm{ar}_{\mathcal{B}}\left(P_2(w,f)\right)} \leq \frac{1}{2}\sqrt{\frac{\tilde{W}\tilde{F}}{2^\gamma}}, \tag{F.13}$$

where (F.12) follows from Jensen's inequality and the concavity of square root. The inequality in (F.13) follows because, for all w and f, we have

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}_{\mathcal{B}}\left(P_2(w,f)\right) \\
&= \mathbb{V}\mathrm{ar}_{\mathcal{B}}\left(\sum_{x\in\mathcal{D}_\gamma} p_X(x)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\right) && \text{(F.14)}\\
&= \sum_{x\in\mathcal{D}_\gamma} \mathbb{V}\mathrm{ar}_{\mathcal{B}}\Big(p_X(x)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\Big) && \text{(F.15)}\\
&\le \sum_{x\in\mathcal{D}_\gamma} p_X^2(x)\,\mathbb{E}_{\mathcal{B}}\Big(\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\Big) && \text{(F.16)}\\
&= \frac{1}{\tilde{W}\tilde{F}}\sum_{x\in\mathcal{D}_\gamma} p_X^2(x) && \text{(F.17)}\\
&\le \frac{2^{-\gamma}}{\tilde{W}\tilde{F}}\sum_{x\in\mathcal{D}_\gamma} p_X(x) \le \frac{2^{-\gamma}}{\tilde{W}\tilde{F}}, && \text{(F.18)}
\end{aligned}$$

where (F.15) follows since the random variables $\Big\{p_X(x)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\Big\}_{x\in\mathcal{D}_\gamma}$ are independent due to the structure of the random binning, and (F.18) follows because $p_X(x)\le 2^{-\gamma}$ for all $x\in\mathcal{D}_\gamma$. Lemma 7 follows from substituting (F.10) and (F.13) in (F.7).

Appendix G

Proof of Lemma 8

G.1 High probability $\mathcal{Z}$-set:

For all $S \in \mathcal{S}$, define the set

$$\mathcal{A}_S \triangleq \left\{ z \in \mathcal{Z} : \mathbb{P}_{p_{X|Z_S}} \left((X, z) \in \mathcal{D}_\gamma^S \right) \geq 1 - \delta \right\}. \tag{G.1}$$

Recall that $\mathbb{P}_{p_{XZ_S}} \left((X, Z_S) \in \mathcal{D}_\gamma^S \right) \geq 1 - \delta^2$ by assumption. Using Markov inequality, we have

$$\mathbb{P}_{p_{Z_S}}(\mathcal{A}_S^c) = \mathbb{P}_{p_{Z_S}} \left(\mathbb{P}_{p_{X|Z_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \geq \delta \right) \tag{G.2}$$

$$\leq \frac{1}{\delta} \mathbb{E}_{p_{Z_S}} \left(\mathbb{P}_{p_{X|Z_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \right) \tag{G.3}$$

$$= \frac{1}{\delta} \mathbb{P}_{p_{XZ_S}} \left((X, Z_S) \notin \mathcal{D}_\gamma^S \right) \tag{G.4}$$

$$\leq \frac{\delta^2}{\delta} = \delta. \tag{G.5}$$

G.2 Typical and non-typical events:

For all $w, f \in [1 : \tilde{W}] \times [1 : \tilde{F}]$, $z \in \mathcal{Z}$, and $S \in \mathcal{S}$, define the random variables

$$P_1^S(w, f|z) = \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\} \mathbb{1}\left\{ (x, z) \in \mathcal{D}_\gamma^S \right\} \tag{G.6}$$

$$P_2^S(w,f|z) = \sum_{x\in\mathcal{X}} p_{X|Z_S}(x|z)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\mathbb{1}\left\{(x,z)\notin\mathcal{D}_\gamma^S\right\}. \tag{G.7}$$

Thus, we have, for all w, f, z, and S, that

$$P_{WF|Z_S}(w,f|z) = \sum_{x\in\mathcal{X}} p_{X|Z_S}(x|z)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\} \tag{G.8}$$

$$= P_1^S(w,f|z) + P_2^S(w,f|z). \tag{G.9}$$

Note that, for fixed $z\in\mathcal{Z}$ and $S\in\mathcal{S}$, each of the the random variables $P_i^S(w,f|z)$, $i=1,2$, is identically distributed for all $w,f\in[1:\tilde{W}]\times[1:\tilde{F}]$ due to the symmetry in the random binning. We then fix $z\in\mathcal{Z}$ and $S\in\mathcal{S}$, and let $P_1^S(w,f|z) = \sum_{x\in\mathcal{X}} U_x(w,f,z,S)$, where

$$U_x(w,f,z,S) = p_{X|Z_S}(x|z)\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\mathbb{1}\left\{(x,z)\in\mathcal{D}_\gamma^S\right\}. \tag{G.10}$$

The random variables $\{U_x(w,f,z,S)\}_{x\in\mathcal{X}}$ are non-negative and independent, and for all $x\in\mathcal{X}$,

$$U_x(w,f,z,S) \le p_{X|Z_S}(x|z)\ \mathbb{1}\left\{(x,z)\in\mathcal{D}_\gamma^S\right\} < 2^{-\gamma}, \tag{G.11}$$

where $p_{X|Z_S}(x|z) < 2^{-\gamma}$, for all $(x,z)\in\mathcal{D}_\gamma^S$. Also, we have

$$\sum_{x\in\mathcal{X}} \mathbb{E}_{\mathcal{B}}(U_x(w,f,z,S)) = \sum_{x\in\mathcal{X}} p_{X|Z_S}(x|z)\ \mathbb{E}_{\mathcal{B}}\left(\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}\right)\mathbb{1}\left\{(x,z)\in\mathcal{D}_\gamma^S\right\} \tag{G.12}$$

$$= \frac{1}{\tilde{W}\tilde{F}} \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\left\{(x,z) \in \mathcal{D}_\gamma^S\right\} \tag{G.13}$$

$$= \frac{1}{\tilde{W}\tilde{F}} \, \mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right). \tag{G.14}$$

By applying the variation on Chernoff's bound in Lemma 2.7 to the random variables $\{U_x(w,f,z,S)\}_{x \in \mathcal{X}}$, with $\bar{m} = \frac{\mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right)}{\tilde{W}\tilde{F}}$ and $b = 2^{-\gamma}$, we have, for every $\epsilon_1 \in [0,1]$ and $z \in \mathcal{A}_S$, that

$$\mathbb{P}_{\mathcal{B}}\left(P_1^S(w,f|z) \geq \frac{1+\epsilon_1}{\tilde{W}\tilde{F}}\right) \leq \mathbb{P}\left(\sum_{x \in \mathcal{X}} U_x(w,f,z,S) \geq \frac{1+\epsilon_1}{\tilde{W}\tilde{F}} \, \mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right)\right) \tag{G.15}$$

$$\leq \exp\left(\frac{-\epsilon_1^2 \, \mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right) 2^\gamma}{3\tilde{W}\tilde{F}}\right) \tag{G.16}$$

$$\leq \exp\left(\frac{-\epsilon_1^2(1-\delta)2^\gamma}{3\tilde{W}\tilde{F}}\right), \tag{G.17}$$

where (G.15) follows since $\mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right) \leq 1$, and (G.17) follows because, for all $z \in \mathcal{A}_S$, we have $\mathbb{P}_{p_{X|Z_S}}\left((X,z) \in \mathcal{D}_\gamma^S\right) \geq 1-\delta$.

We also have have that,

$$\mathbb{E}_{p_{Z_S}}\left(\sum_{w,f} P_2^S(w,f|Z_S)\right)$$

$$= \mathbb{E}_{p_{Z_S}}\left(\sum_{x \in \mathcal{X}} p_{X|Z_S}(x|Z_S) \mathbb{1}\left\{(x,Z_S) \notin \mathcal{D}_\gamma^S\right\} \sum_{w,f} \mathbb{1}\{\mathcal{B}_1(x) = w\} \mathbb{1}\{\mathcal{B}_2(x) = f\}\right) \tag{G.18}$$

$$= \sum_{z \in \mathcal{Z}} p_{Z_S}(z) \sum_{x \in \mathcal{X}} p_{X|Z_S}(x|z) \mathbb{1}\left\{(x,z) \notin \mathcal{D}_\gamma^S\right\} \tag{G.19}$$

$$= \sum_{(x,z)\notin\mathcal{D}_\gamma^S} p_{XZ_S}(x,z) \tag{G.20}$$

$$= \mathbb{P}_{p_{XZ_S}}\left((X,Z_S)\notin\mathcal{D}_\gamma^S\right) \leq \delta^2, \tag{G.21}$$

where (G.19) follows since every $x\in\mathcal{X}$ is assigned to only one pair (w,f), and hence,

$$\sum_{w,f}\mathbb{1}\{\mathcal{B}_1(x)=w\}\mathbb{1}\{\mathcal{B}_2(x)=f\}=1. \tag{G.22}$$

G.3 Good binning functions:

Let $\mathbf{b}\triangleq(b_1,b_2):\mathcal{X}\mapsto[1:\tilde{W}]\times[1:\tilde{F}]$ be a realization of the random binning $\mathcal{B}$. Recall that the random variable $P_1^S(w,f|z)$ is identically distributed for every w and f. We then define the class $\mathcal{G}$ of binning functions $\mathbf{b}$ as

$$\mathcal{G}\triangleq\left\{\mathbf{b}: P_1^S(w,f|z)<\frac{1+\epsilon_1}{\tilde{W}\tilde{F}},\quad \text{for all } S\in\mathcal{S} \text{ and } z\in\mathcal{A}_S\right\}. \tag{G.23}$$

Using the union bound and (G.17), we have that

$$\mathbb{P}_{\mathcal{B}}\left(\mathcal{G}^c\right)=\mathbb{P}_{\mathcal{B}}\left(P_1^S(w,f|z)\geq\frac{1+\epsilon_1}{\tilde{W}\tilde{F}},\ \text{for some } S\in\mathcal{S},\ \text{or } z\in\mathcal{A}_S\right) \tag{G.24}$$

$$=\mathbb{P}_{\mathcal{B}}\left(\bigcup_{S\in\mathcal{S}}\bigcup_{z\in\mathcal{A}_S}P_1^S(w,f|z)\geq\frac{1+\epsilon_1}{\tilde{W}\tilde{F}}\right) \tag{G.25}$$

$$\leq\sum_{S\in\mathcal{S}}\sum_{z\in\mathcal{A}_S}\mathbb{P}_{\mathcal{B}}\left(P_1^S(w,f|z)\geq\frac{1+\epsilon_1}{\tilde{W}\tilde{F}}\right) \tag{G.26}$$

$$\leq\sum_{S\in\mathcal{S}}|\mathcal{A}_S|\exp\left(\frac{-\epsilon_1^2(1-\delta)2^\gamma}{3\tilde{W}\tilde{F}}\right) \tag{G.27}$$

$$\leq |\mathcal{S}||\mathcal{Z}| \exp\left(\frac{-\epsilon_1^2(1-\delta)2^\gamma}{3\tilde{W}\tilde{F}}\right). \tag{G.28}$$

Take $\mathbf{b}$ such that $\mathbf{b} \in \mathcal{G}$, and set $W = b_1(X)$ and $F = b_2(X)$. For every $S \in \mathcal{S}$, we have

$$\mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right) = \mathbb{E}_{p_{Z_S}}\left(\mathbb{D}\left(P_{WF|Z_S}||p_W^U p_F^U\right)\right) \tag{G.29}$$

$$= \mathbb{E}_{p_{Z_S}}\left(\sum_{w,f} P_{WF|Z_S}(w,f|Z_S) \log \frac{P_{WF|Z_S}(w,f|Z_S)}{p_W^U p_F^U}\right) \tag{G.30}$$

$$= \mathbb{E}_{p_{Z_S}}\left(\sum_{w,f}\sum_{i=1}^{2} P_i^S(w,f|Z_S)\ \log\left(\tilde{W}\tilde{F}\sum_{i=1}^{2} P_i^S(w,f|Z_S)\right)\right) \tag{G.31}$$

$$= \mathbb{E}_{p_{Z_S}}\left(\sum_{w,f}\sum_{i=1}^{2} P_i^S(w,f|Z_S)\ \log \frac{\sum_{i=1}^{2} P_i^S(w,f|Z_S)}{\frac{1}{\tilde{W}\tilde{F}}\sum_{i=1}^{2}\sum_{w,f} P_i^S(w,f|Z_S)}\right) \tag{G.32}$$

$$\leq \mathbb{E}_{p_{Z_S}}\left(\sum_{i=1}^{2}\sum_{w,f} P_i^S(w,f|Z_S) \log \frac{\tilde{W}\tilde{F}P_i^S(w,f|Z_S)}{\sum_{w,f} P_i^S(w,f|Z_S)}\right) \tag{G.33}$$

$$= \sum_{i=1}^{2} \mathbb{E}_{p_{Z_S}}\left(\sum_{w,f} P_i^S(w,f|Z_S) \log\left(\tilde{W}\tilde{F}P_i^S(w,f|Z_S)\right)\right)$$

$$+ \mathbb{E}_{p_{Z_S}}\left(\sum_{i=1}^{2}\sum_{w,f} P_i^S(w,f|Z_S) \log \frac{1}{\sum_{w,f} P_i^S(w,f|Z_S)}\right) \tag{G.34}$$

where (G.32) follows because

$$\sum_{w,f}\sum_{i=1}^{2} P_i^S(w,f|Z_S) = \sum_{w,f} P_{WF|Z_S}(w,f|Z_S) = 1, \tag{G.35}$$

and (G.33) follows from the log-sum inequality.

Now, we upper bound each term in (G.34). For $\mathbf{b} \in \mathcal{G}$ and every $S \in \mathcal{S}$, we have

$$\begin{aligned}
&\mathbb{E}_{p_{Z_S}} \left(\sum_{w,f} P_1^S(w,f|Z_S) \log \left(\tilde{W}\tilde{F} P_1^S(w,f|Z_S) \right) \right) \\
&\quad = \mathbb{E}_{p_{Z_S}} \left(\sum_{w,f} P_1^S(w,f|Z_S) \ \log \left(\tilde{W}\tilde{F} P_1^S(w,f|Z_S) \right) \mathbb{1}\left\{ Z_S \in \mathcal{A}_S \right\} \right) \\
&\qquad\qquad + \mathbb{E}_{p_{Z_S}} \left(\sum_{w,f} P_1^S(w,f|Z_S) \ \log \left(\tilde{W}\tilde{F} P_1^S(w,f|Z_S) \right) \mathbb{1}\left\{ Z_S \notin \mathcal{A}_S \right\} \right) \qquad \text{(G.36)} \\
&\quad < \log(1+\epsilon_1) + \sum_{x,z} p_{XZ_S}(x,z) \log \left(\tilde{W}\tilde{F} P_1^S(w,f|z) \right) \mathbb{1}\left\{ z \notin \mathcal{A}_S \right\} \qquad \text{(G.37)} \\
&\quad \leq \log(1+\epsilon_1) + \log(\tilde{W}\tilde{F}) \, \mathbb{P}_{p_{Z_S}}(Z_S \notin \mathcal{A}_S) \qquad \text{(G.38)} \\
&\quad \leq \epsilon_1 + \delta \log(\tilde{W}\tilde{F}), \qquad \text{(G.39)}
\end{aligned}$$

where (G.37) follows because, for every $\mathbf{b} \in \mathcal{G}$ and $S \in \mathcal{S}$, we have $\tilde{W}\tilde{F} P_1^S(w,f|Z_S) < (1+\epsilon)$ for $Z_S \in \mathcal{A}_S$ and every w, f, and (G.39) follows from (G.5).

Using (G.21), we have, for every $S \in \mathcal{S}$, that

$$\begin{aligned}
&\mathbb{E}_{p_{Z_S}} \left(\sum_{w,f} P_2^S(w,f|Z_S) \log \left(\tilde{W}\tilde{F} P_2^S(w,f|Z_S) \right) \right) \\
&\qquad\qquad \leq \log(\tilde{W}\tilde{F}) \, \mathbb{E}_{p_{Z_S}} \left(\sum_{w,f} P_2^S(w,f|Z_S) \right) \leq \delta^2 \log(\tilde{W}\tilde{F}). \qquad \text{(G.40)}
\end{aligned}$$

We also have, for every $S \in \mathcal{S}$, that

$$\mathbb{E}_{p_{Z_S}}\left(\sum_{i=1}^{2}\sum_{w,f} P_i^S(w,f|Z_S)\log\frac{1}{\sum_{w,f}P_i^S(w,f|Z_S)}\right) = \mathbb{E}_{p_{Z_S}}\left(H_b\left(\mathbb{P}_{p_{X|Z_S}}\left((X,Z_S)\in\mathcal{D}_\gamma^S\right)\right)\right) \tag{G.41}$$

$$\leq H_b\left(\mathbb{E}_{p_{Z_S}}\left(\mathbb{P}_{p_{X|Z_S}}\left((X,Z_S)\in\mathcal{D}_\gamma^S\right)\right)\right) \tag{G.42}$$

$$= H_b\left(\mathbb{P}_{p_{XZ_S}}\left((X,Z_S)\in\mathcal{D}_\gamma^S\right)\right) \tag{G.43}$$

$$\leq H_b(1-\delta^2) = H_b(\delta^2), \tag{G.44}$$

where (G.42) follows from Jensen's inequality and the concavity of H_b, and (G.44) follows since $H_b(x)$ is monotonically decreasing in $x \in \left(\frac{1}{2},1\right)$. Equation (G.41) follows since $\sum_{i=1}^{2}\sum_{w,f} P_i^S(w,f|Z_S) = 1$, and $\sum_{w,f} P_1^S(w,f|Z_S) = \mathbb{P}_{p_{X|Z_S}}\left((X,Z_S)\in\mathcal{D}_\gamma^S\right)$.

By substituting (G.39), (G.40), and (G.44) in (G.34), we have, for every $\mathbf{b}\in\mathcal{G}$ and $S\in\mathcal{S}$, that

$$\mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right) < \epsilon_1 + (\delta+\delta^2)\log(\tilde{W}\tilde{F}) + H_b(\delta^2) = \tilde{\epsilon}. \tag{G.45}$$

Thus, we have

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\ \mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right)\geq\tilde{\epsilon}\right) = 1-\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\ \mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right)<\tilde{\epsilon}\right) \tag{G.46}$$

$$= 1-\mathbb{P}_{\mathcal{B}}\left(\mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right)<\tilde{\epsilon},\ \text{ for all } S\in\mathcal{S}\right) \tag{G.47}$$

$$\leq 1-\mathbb{P}_{\mathcal{B}}(\mathcal{G}) = \mathbb{P}_{\mathcal{B}}(\mathcal{G}^c) \tag{G.48}$$

$$\leq |\mathcal{S}||\mathcal{Z}| \exp\left(\frac{\epsilon_1^2(1-\delta)2^{\gamma}}{3\tilde{W}\tilde{F}}\right), \tag{G.49}$$

where the inequality in (G.48) follows because (G.45) implies that

$$\mathbb{P}_{\mathcal{B}}\left(\mathbb{D}\left(P_{WFZ_S}||p_W^U p_F^U p_{Z_S}\right) < \tilde{\epsilon}, \text{ for all } S \in \mathcal{S}\right) \geq \mathbb{P}_{\mathcal{B}}(\mathcal{G}). \tag{G.50}$$

This completes the proof for Lemma 8. The analysis in this proof is adapted from [3, Appendix].

Appendix H

Proof of Lemma 10

First, we rewrite the relative entropy in (6.41) as follows:

$$\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}Z_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{Z_S}\right)$$

$$= \sum_{w_{[1:2]},f_{[1:2]},z} P_{W_{[1:2]}F_{[1:2]}Z_S}(w_{[1:2]},f_{[1:2]},z)\log\frac{P_{W_{[1:2]}F_{[1:2]}Z_S}(w_{[1:2]},f_{[1:2]},z)}{p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{Z_S}(z)} \tag{H.1}$$

$$= \sum_{w_{[1:2]},f_{[1:2]},z} P_{W_{[1:2]}F_{[1:2]}Z_S}(w_{[1:2]},f_{[1:2]},z)$$

$$\times\log\left(\frac{P_{W_{[1:2]}F_{[1:2]}Z_S}(w_{[1:2]},f_{[1:2]},z)}{P_{W_1F_1Z_S}(w_1,f_1,z)p^U_{W_2}p^U_{F_2}}\cdot\frac{P_{W_1F_1Z_S}(w_1,f_1,z)}{p^U_{W_1}p^U_{F_1}p_{Z_S}(z)}\right) \tag{H.2}$$

$$= \mathbb{E}_{p_{Z_S}}\left(\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}|Z_S}||P_{W_1F_1|Z_S}p^U_{W_2}p^U_{F_2}\right)\right)+\mathbb{D}\left(P_{W_1F_1Z_S}||p^U_{W_1}p^U_{F_1}p_{Z_S}\right). \tag{H.3}$$

Thus, the probability in (6.41) is upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}Z_S}||p^U_{W_{[1:2]}}p^U_{F_{[1:2]}}p_{Z_S}\right)\geq 2\tilde{\epsilon}\right)$$

$$\leq \mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_1F_1Z_S}||p^U_{W_1}p^U_{F_1}p_{Z_S}\right)>\tilde{\epsilon}\right)$$

$$+\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{E}_{p_{Z_S}}\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}|Z_S}||P_{W_1F_1|Z_S}p^U_{W_2}p^U_{F_2}\right)>\tilde{\epsilon}\right). \tag{H.4}$$

We upper bound each term on the right hand side of (H.4). Using Lemma 8, the first term is upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_1F_1Z_S}||p^U_{W_1}p^U_{F_1}p_{Z_S}\right)>\tilde{\epsilon}\right)\leq|\mathcal{S}||\mathcal{Z}|\exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_1}}{3\tilde{W}_1\tilde{F}_1}\right). \tag{H.5}$$

Next, we upper bound the second term in (H.4). For all $S\in\mathcal{S}$, let us define

$$\mathcal{A}_S\triangleq\left\{z\in\mathcal{Z}:\mathbb{P}_{p_{X_{[1:2]}|Z_S}}\left((X_{[1:2]},z)\in\mathcal{D}_1^S\right)\geq 1-\delta\right\}, \tag{H.6}$$

where $\mathcal{D}_1^S$ is defined in (6.37). As in (G.2)-(G.5), we have

$$\mathbb{P}_{p_{Z_S}}(\mathcal{A}_S^c)=\mathbb{P}_{p_{Z_S}}\left(\mathbb{P}_{p_{X_{[1:2]}|Z_S}}\left((X_{[1:2]},z)\notin\mathcal{D}_1^S\right)\geq\delta\right)\leq\delta. \tag{H.7}$$

For all $w_{[1:2]},f_{[1:2]}\in[1:\tilde{W}]\times[1:\tilde{F}]$, $z\in\mathcal{Z}$, and $S\in\mathcal{S}$, define

$$\begin{aligned}P_1^S(w_{[1:2]},f_{[1:2]}|z)=&\sum_{x_{[1:2]}\in\mathcal{X}_1\times\mathcal{X}_2}p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z)\mathbb{1}\left\{(x_{[1:2]},z)\in\mathcal{D}_1^S\right\}\\&\times\mathbb{1}\left\{\mathcal{B}_1^{(j)}(x_j)=w_j,\mathcal{B}_2^{(j)}(x_j)=f_j,\forall j=1,2\right\}\end{aligned} \tag{H.8}$$

$$\begin{aligned}P_2^S(w_{[1:2]},f_{[1:2]}|z)=&\sum_{x_{[1:2]}\in\mathcal{X}_1\times\mathcal{X}_2}p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z)\mathbb{1}\left\{(x_{[1:2]},z)\notin\mathcal{D}_1^S\right\}\\&\times\mathbb{1}\left\{\mathcal{B}_1^{(j)}(x_j)=w_j,\mathcal{B}_2^{(j)}(x_j)=f_j,\forall j=1,2\right\}.\end{aligned} \tag{H.9}$$

Thus, we have $P_{W_{[1:2]}F_{[1:2]}|Z_S}(w_{[1:2]},f_{[1:2]}|z)=P_1^S(w_{[1:2]},f_{[1:2]}|z)+P_2^S(w_{[1:2]},f_{[1:2]}|z)$.

Now, for every $x_2 \in \mathcal{X}_2$, define

$$U_{x_2} = \sum_{x_1 \in \mathcal{X}_1} p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z) \mathbb{1}\left\{\mathcal{B}_1^{(2)}(x_2) = w_2, \mathcal{B}_2^{(2)}(x_2) = f_2\right\} \mathbb{1}\left\{(x_{[1:2]}, z) \in \mathcal{D}_1^S\right\}. \tag{H.10}$$

The random variables $\left\{U_{x_2}\right\}_{x_2 \in \mathcal{X}_2}$ are non-negative and independent since the random variables $\left\{\mathcal{B}_1^{(2)}(x_2), \mathcal{B}_2^{(2)}(x_2)\right\}_{x_2 \in \mathcal{X}_2}$ are independent. From the definition of $\mathcal{D}_1^S$ in (6.37), we have for $(x_{[1:2]}, z) \in \mathcal{D}_1^S$ that $(x_{[1:2]}, z) \in \mathcal{D}_{\gamma_{21}}^S$. Additionally, from the definition of $\mathcal{D}_{\gamma_{21}}$ in (6.39), we have that $p(x_2|x_1, z) \leq 2^{-\gamma_{21}}$. From (H.10), we have

$$U_{x_2} \leq \sum_{x_1} p_{X_1|Z_S}(x_1|z) p_{X_2|X_1,Z_S}(x_2|x_1, z) \mathbb{1}\left\{(x_{[1:2]}, z) \in \mathcal{D}_1^S\right\} \tag{H.11}$$

$$\leq 2^{-\gamma_{21}} \sum_{x_1} p_{X_1|Z_S}(x_1|z) \mathbb{1}\left\{(x_{[1:2]}, z) \in \mathcal{D}_1^S\right\} \tag{H.12}$$

$$\leq 2^{-\gamma_{21}}. \tag{H.13}$$

Since for all $x_2 \in \mathcal{X}_2$,

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{1}\left\{\mathcal{B}_1^{(2)}(x_2) = w_2, \mathcal{B}_2^{(2)}(x_2) = f_2\right\}\right) = \frac{1}{\tilde{W}_2 \tilde{F}_2}, \tag{H.14}$$

we have,

$$\sum_{x_2 \in \mathcal{X}_2} \mathbb{E}_{\mathcal{B}}(U_{x_2}) = \frac{1}{\tilde{W}_2 \tilde{F}_2} \sum_{x_{[1:2]} \in \mathcal{X}_1 \times \mathcal{X}_2} p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z) \mathbb{1}\left\{(x_{[1:2]}, z) \in \mathcal{D}_1^S\right\} \tag{H.15}$$

$$= \frac{\mathbb{P}_{p_{X_{[1:2]}|Z_S}}\left(\left(X_{[1:2]}, z\right) \in \mathcal{D}_1^S\right)}{\tilde{W}_2 \tilde{F}_2}. \tag{H.16}$$

In addition, notice that

$$\sum_{w_1,f_1} P_1^S(w_{[1:2]},f_{[1:2]}|z) = \sum_{x_{[1:2]}} p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z)\mathbb{1}\left\{\left(x_{[1:2]},z\right)\in\mathcal{D}_1^S\right\}$$
$$\times \sum_{w_1,f_1} \mathbb{1}\left\{\mathcal{B}_1^{(j)}(x_j)=w_j,\mathcal{B}_2^{(j)}(x_j)=f_j,\forall j=1,2\right\} \tag{H.17}$$

$$= \sum_{x_2}\sum_{x_1} p_{X_{[1:2]}|Z_S}(x_{[1:2]}|z)\mathbb{1}\left\{\mathcal{B}_1^{(2)}(x_2)=w_2,\mathcal{B}_2^{(2)}(x_2)=f_2\right\}\mathbb{1}\left\{\left(x_{[1:2]},z\right)\in\mathcal{D}_1^S\right\} \tag{H.18}$$

$$= \sum_{x_2} U_{x_2} \tag{H.19}$$

The random variables $\left\{U_{x_2}\right\}_{x_2\in\mathcal{X}_2}$ are non-negative, independent, and $U_{x_2}\in[0,2^{-\gamma_{21}}]$ for all $x_2\in\mathcal{X}_2$. By applying Lemma 4 to the random variables $\{U_{x_2}\}_{x_2\in\mathcal{X}_2}$, we have,

$$\mathbb{P}_{\mathcal{B}}\left(P_1^S(w_{[1:2]},f_{[1:2]}|z)\geq\frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2}P_{W_1F_1|Z_S}(w_1,f_1|z)\right)$$
$$\leq \mathbb{P}_{\mathcal{B}}\left(\sum_{w_1,f_1}P_1^S(w_{[1:2]},f_{[1:2]}|z)\geq\frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2}\sum_{w_1,f_1}P_{W_1F_1|Z_S}(w_1,f_1|z)\right) \tag{H.20}$$
$$= \mathbb{P}\left(\sum_{x_2}U_{x_2}\geq\frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2}\right) \tag{H.21}$$
$$\leq \mathbb{P}\left(\sum_{x_2}U_{x_2}\geq\frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2}\mathbb{P}_{p_{X_{[1:2]}|Z_S}}\left(\left(X_{[1:2]},z\right)\in\mathcal{D}_1^S\right)\right) \tag{H.22}$$
$$= \mathbb{P}\left(\sum_{x_2}U_{x_2}\geq(1+\epsilon)\sum_{x_2}\mathbb{E}_{\mathcal{B}}(U_{x_2})\right) \tag{H.23}$$
$$\leq \exp\left(\frac{-\epsilon^2 2^{\gamma_{21}}}{3\tilde{W}_2\tilde{F}_2}\mathbb{P}_{p_{X_{[1:2]}|Z_S}}\left(\left(X_{[1:2]},z\right)\in\mathcal{D}_1^S\right)\right). \tag{H.24}$$

where (H.21) follows from (H.19), (H.23) follows from (H.16), and (H.24) follows from Lemma 4.

From the definition of $\mathcal{A}_S$ in (H.6), we have, for all $z \in \mathcal{A}_S$, that $\mathbb{P}_{p_{X_{[1:2]}|Z_S}}((X_{[1:2]}, z) \in \mathcal{D}_1^S) \geq 1 - \delta$. Thus, for all $z \in \mathcal{A}_S$,

$$\mathbb{P}_{\mathcal{B}}\left(P_1^S(w_{[1:2]}, f_{[1:2]}|z) \geq \frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2} P_{W_1F_1|Z_S}(w_1, f_1|z)\right) \leq \exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_{21}}}{3\tilde{W}_2\tilde{F}_2}\right). \quad \text{(H.25)}$$

Note that, for fixed $z \in \mathcal{Z}$ and $S \in \mathcal{S}$, the random variables $\left\{P_1^S(w_{[1:2]}, f_{[1:2]}|z)\right\}$ are identically distributed for all $w_{[1:2]}, f_{[1:2]}$ due to the symmetry in the random binning. Let $\mathbf{b} \triangleq \{b_1^{(j)}, b_2^{(j)}, j = 1, 2\}$ be a realization of the random binning $\mathcal{B}$. We define the class $\mathcal{G}$ of binning functions $\mathbf{b}$ as

$$\mathcal{G} \triangleq \left\{\mathbf{b} : P_1^S(w_{[1:2]}, f_{[1:2]}|z) < \frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2} P_{W_1F_1|Z_S}(w_1, f_1|z), \text{ for all } S \in \mathcal{S} \text{ and } z \in \mathcal{A}_S\right\}. \quad \text{(H.26)}$$

Using similar steps as in (G.24)-(G.28), we have

$$\mathbb{P}_{\mathcal{B}}(\mathcal{G}^c) = \mathbb{P}_{\mathcal{B}}\left(P_1^S(w_{[1:2]}, f_{[1:2]}|z) \geq \frac{1+\epsilon}{\tilde{W}_2\tilde{F}_2} P_{W_1F_1|Z_S}(w_1, f_1|z), \text{ for some } S \in \mathcal{S} \text{ or } z \in \mathcal{A}_S\right) \quad \text{(H.27)}$$

$$\leq |S||\mathcal{Z}| \exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_{21}}}{3\tilde{W}_2\tilde{F}_2}\right). \quad \text{(H.28)}$$

Take $\mathbf{b}$ such that $\mathbf{b} \in \mathcal{G}$, and set $W_j = b_1^{(j)}(X_j)$ and $F_j = b_2^{(j)}(X_j)$ for $j = 1, 2$. Using similar steps as in (G.29)-(G.34), we have, for all $S \in \mathcal{S}$

$$\begin{aligned}
&\mathbb{E}_{p_{Z_S}}\left(\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}|Z_S}||P_{W_1F_1|Z_S}p^U_{W_2}p^U_{F_2}\right)\right) \\
&\leq \mathbb{E}_{p_{Z_S}}\left(\sum_{i=1}^{2}\sum_{w_{[1:2]},f_{[1:2]}} P_i^S(w_{[1:2]},f_{[1:2]}|Z_S)\log\frac{1}{\sum\limits_{w_{[1:2]},f_{[1:2]}} P_i^S(w_{[1:2]},f_{[1:2]}|Z_S)}\right) \\
&\quad + \mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)\log\frac{\tilde{W}_2\tilde{F}_2P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1,f_1|Z_S)}\right) \\
&\quad + \mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_2^S(w_{[1:2]},f_{[1:2]}|Z_S)\log\frac{\tilde{W}_2\tilde{F}_2P_2^S(w_{[1:2]},f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1,f_1|Z_S)}\right). \qquad \text{(H.29)}
\end{aligned}$$

Now, we upper bound each term in the right hand side of (H.29) for $\mathbf{b} \in \mathcal{G}$. The second term in the right hand side of (H.29) is upper bounded as follows:

$$\begin{aligned}
&\mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)\log\frac{\tilde{W}_2\tilde{F}_2P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1,f_1|Z_S)}\right) \\
&\leq \log(\tilde{W}_2\tilde{F}_2)\,\mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)\mathbb{1}\left\{Z_S \notin \mathcal{A}_S\right\}\right) \\
&\quad + \mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)\log\frac{\tilde{W}_2\tilde{F}_2P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1,f_1|Z_S)}\mathbb{1}\left\{Z_S \in \mathcal{A}_S\right\}\right) \qquad \text{(H.30)} \\
&\leq \log(\tilde{W}_2\tilde{F}_2)\sum_{z\in\mathcal{Z}} p_{Z_S}(z)\mathbb{1}\{z \notin \mathcal{A}_S\}\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|z) \\
&\quad + \log(1+\epsilon)\,\mathbb{E}_{p_{Z_S}}\left(\sum_{w_{[1:2]},f_{[1:2]}} P_1^S(w_{[1:2]},f_{[1:2]}|Z_S)\right) \qquad \text{(H.31)} \\
&\leq \mathbb{P}_{p_{Z_S}}\left(Z_S \notin \mathcal{A}_S\right)\log(\tilde{W}_2\tilde{F}_2) + \log(1+\epsilon) \qquad \text{(H.32)}
\end{aligned}$$

$$\leq \delta \log(\tilde{W}_2 \tilde{F}_2) + \epsilon, \tag{H.33}$$

where (H.30) follows because, for $i = 1, 2$,

$$P_i^S(w_{[1:2]}, f_{[1:2]}|Z_S) \leq P_{W_{[1:2]}F_{[1:2]}|Z_S}(w_{[1:2]}, f_{[1:2]}|Z_S) \tag{H.34}$$

$$= P_{W_1F_1|Z_S}(w_1, f_1|z) P_{W_2F_2|W_1F_1Z_S}(w_2, f_2|w_1, f_1, z) \tag{H.35}$$

$$\leq P_{W_1F_1|Z_S}(w_1, f_1|z), \tag{H.36}$$

and hence $\frac{P_i^S(w_{[1:2]}, f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1, f_1|z)} \leq 1$ for all $w_{[1:2]}, f_{[1:2]}$ and $i = 1, 2$. Equation (H.31) follows because, from (H.26), we have for all $\mathbf{b} \in \mathcal{G}$ and $z \in \mathcal{A}_S$ that $\frac{\tilde{W}_2\tilde{F}_2 P_1^S(w_{[1:2]}, f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1, f_1|Z_S)} < (1 + \epsilon)$.

Next, we upper bound the third term in the right hand side of (H.29). Using similar steps as in (G.18)-(G.22), we have

$$\mathbb{E}_{p_{Z_S}} \left(\sum_{w_{[1:2]}, f_{[1:2]}} P_2^S(w_{[1:2]}, f_{[1:2]}|Z_S) \right) \leq \delta^2. \tag{H.37}$$

Using (H.36) and (H.37), we have

$$\mathbb{E}_{p_{Z_S}} \left(\sum_{w_{[1:2]}, f_{[1:2]}} P_2^S(w_{[1:2]}, f_{[1:2]}|Z_S) \log \frac{\tilde{W}_2\tilde{F}_2 P_2^S(w_{[1:2]}, f_{[1:2]}|Z_S)}{P_{W_1F_1|Z_S}(w_1, f_1|Z_S)} \right)$$

$$\leq \log(\tilde{W}_2\tilde{F}_2)\, \mathbb{E}_{p_{Z_S}} \left(\sum_{w_{[1:2]}, f_{[1:2]}} P_2^S(w_{[1:2]}, f_{[1:2]}|Z_S) \right) \tag{H.38}$$

$$\leq \delta^2 \log(\tilde{W}_2\tilde{F}_2). \tag{H.39}$$

Since we have

$$\sum_{i=1}^{2} \sum_{w_{[1:2]}, f_{[1:2]}} P_i^S \left(w_{[1:2]}, f_{[1:2]} | Z_S\right) = 1, \tag{H.40}$$

$$\sum_{w_{[1:2]}, f_{[1:2]}} P_1^S \left(w_{[1:2]}, f_{[1:2]} | Z_S\right) = \mathbb{P}_{p_{X_{[1:2]}|Z_S}} \left(\left(X_{[1:2]}, Z_S\right) \in \mathcal{D}_1^S\right), \tag{H.41}$$

$$\text{and} \sum_{w_{[1:2]}, f_{[1:2]}} P_2^S \left(w_{[1:2]}, f_{[1:2]} | Z_S\right) = 1 - \mathbb{P}_{p_{X_{[1:2]}|Z_S}} \left(\left(X_{[1:2]}, Z_S\right) \in \mathcal{D}_1^S\right), \tag{H.42}$$

the first term on the right hand side of (H.29) is upper bounded as follows:

$$\begin{aligned}
&\mathbb{E}_{p_{Z_S}} \left(\sum_{i=1}^{2} \sum_{w_{[1:2]}, f_{[1:2]}} P_i^S(w_{[1:2]}, f_{[1:2]} | Z_S) \log \frac{1}{\sum\limits_{w_{[1:2]}, f_{[1:2]}} P_i^S(w_{[1:2]}, f_{[1:2]} | Z_S)} \right) \\
&\quad = \mathbb{E}_{p_{Z_S}} \left(H_b \left(\mathbb{P}_{p_{X_{[1:2]}|Z_S}} \left(\left(X_{[1:2]}, Z_S\right) \in \mathcal{D}_1^S \right) \right) \right) \qquad \text{(H.43)} \\
&\quad \leq H_b(1 - \delta^2) = H_b(\delta^2), \qquad \text{(H.44)}
\end{aligned}$$

where (H.44) follows as in (G.43) and (G.44).

Using (H.33), (H.39), and (H.44), for any $\mathbf{b} \in \mathcal{G}$ and for all $S \in \mathcal{S}$, the left hand side of (H.29) is upper bounded as

$$\mathbb{E}_{p_{Z_S}} \left(\mathbb{D} \left(P_{W_{[1:2]} F_{[1:2]} | Z_S} || P_{W_1 F_1 | Z_S} p_{W_2}^U p_{F_2}^U \right) \right) \leq \epsilon + (\delta + \delta^2) \log(\tilde{W}_2 \tilde{F}_2) + H_b(\delta^2) \leq \tilde{\epsilon}. \tag{H.45}$$

Thus, the second probability on the right hand side of (H.4) is upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{E}_{p_{Z_S}}\mathbb{D}\left(P_{W_{[1:2]}F_{[1:2]}|Z_S}||P_{W_1F_1|Z_S}p^U_{W_2}p^U_{F_2}\right)>\tilde{\epsilon}\right)\leq|S||\mathcal{Z}|\exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_{21}}}{3\tilde{W}_2\tilde{F}_2}\right). \tag{H.46}$$

Finally, by rewriting (H.3) with switching the roles of (W_1, F_1) and (W_2, F_2) and repeating the whole proof, we obtain the second term in the minimum in (6.41), which completes the proof for Lemma 10.

Appendix I

Proof of Lemma 11

Recall that $\mathcal{J} \triangleq \{J : J \subseteq [1:T], J \neq \emptyset\}$. For notational simplicity, define, for all $J \in \mathcal{J}$,

$$\mathbb{1}\{x, w, f, J\} = \mathbb{1}\left\{\mathcal{B}_{1t}(x_t) = w_t, \mathcal{B}_{2t}(x_t) = f_t, \forall t \in J\right\}. \tag{I.1}$$

We have,

$$P_{W_{[1:T]}F_{[1:T]}}\left(w_{[1:T]}, f_{[1:T]}\right) = \sum_{x_{[1:T]} \in \mathcal{X}_{[1:T]}} p_{X_{[1:T]}}(x_{[1:T]}) \mathbb{1}\{x, w, f, [1:T]\}. \tag{I.2}$$

Also, for $J \in \mathcal{J}$, we have

$$\mathbb{E}_{\mathcal{B}}\left(\mathbb{1}\{x, w, f, J\}\right) = \prod_{t \in J} \frac{1}{\tilde{W}_t \tilde{F}_t} = \frac{1}{\tilde{W}_J \tilde{F}_J}. \tag{I.3}$$

Let $P_{W_{[1:T]}F_{[1:T]}}\left(w_{[1:T]}, f_{[1:T]}\right) = P_1\left(w_{[1:T]}, f_{[1:T]}\right) + P_2\left(w_{[1:T]}, f_{[1:T]}\right)$, where

$$P_1\left(w_{[1:T]}, f_{[1:T]}\right) = \sum_{x_{[1:T]}} p_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\{x, w, f, [1:T]\}\ \mathbb{1}\left\{x_{[1:T]} \notin \mathcal{D}\right\} \tag{I.4}$$

$$P_2\left(w_{[1:T]}, f_{[1:T]}\right) = \sum_{x_{[1:T]}} p_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\{x, w, f, [1:T]\}\ \mathbb{1}\left\{x_{[1:T]} \in \mathcal{D}\right\}. \tag{I.5}$$

Using similar steps as in Appendix F, we have

$$2\mathbb{E}_{\mathcal{B}}\left(\mathbb{V}\left(P_{W_{[1:T]}F_{[1:T]}}, p^{U}_{W_{[1:T]}} p^{U}_{F_{[1:T]}}\right)\right) \leq 2\mathbb{P}\left(X_{[1:T]} \notin \mathcal{D}\right) + \sum_{w_{[1:T]}, f_{[1:T]}} \mathbb{E}_{\mathcal{B}}\left|P_2\left(w_{[1:T]}, f_{[1:T]}\right) - \mathbb{E}_{\mathcal{B}}\left(P_2\left(w_{[1:T]}, f_{[1:T]}\right)\right)\right|. \tag{I.6}$$

We partition $\mathcal{X}_{[1:T]}$ as follows:

- At the first iteration, $s = 1$, for all $J \in \mathcal{J}$, pick the largest possible set $\mathcal{N}_{J,1}$ of sequences $x_{[1:T]}$ that have different coordinates in each position of J, and at least one other position. That is, $\mathcal{N}_{J,1}$ is on the form

$$\left\{x_{[1:T]} : \bar{x}_{[1:T]} \in \mathcal{N}_{J,1} \Rightarrow x_{J^c} \neq \bar{x}_{J^c}, \text{ and } \forall t \in J, x_t \neq \bar{x}_t\right\}. \tag{I.7}$$

  Notice that, for $J \in \mathcal{J}$, the largest set $\mathcal{N}_{J,1}$ is not unique. Choose the sets $\{\mathcal{N}_{J,1}\}_{J\in\mathcal{J}}$ such that they do not overlap.

- We repeat the process, such that $\mathcal{N}_{J,s} \cap \mathcal{N}_{J',s'} = \emptyset$ for $s \neq s'$ or $J \neq J'$, and for $x_{[1:T]} \in \mathcal{N}_{J,s}, x'_{[1:T]} \in \mathcal{N}_{J,s'}, x_{J^c} \neq x'_{J^c}$, until we run out of sequences in $\mathcal{X}_{[1:T]}$.

Let N be the number of iterations. Thus $\mathcal{X}_{[1:T]} = \cup_{s=1}^{N} \cup_{J\in\mathcal{J}} \mathcal{N}_{J,s}$. We thus have

$$P_2\left(w_{[1:T]}, f_{[1:T]}\right) = \sum_{s=1}^{N} \sum_{J\in\mathcal{J}} \bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}}, \tag{I.8}$$

where

$$\bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} = \sum_{x_{[1:T]}\in\mathcal{N}_{J,s}} p_{X_{[1:T]}}\left(x_{[1:T]}\right) \mathbb{1}\{x, w, f, [1:T]\} \mathbb{1}\left\{x_{[1:T]} \in \mathcal{D}\right\}. \tag{I.9}$$

Using the triangle inequality, we have

$$\sum_{w_{[1:T]},f_{[1:T]}} \mathbb{E}_{\mathcal{B}} \left| P_2\left(w_{[1:T]},f_{[1:T]}\right) - \mathbb{E}_{\mathcal{B}}\left(P_2\left(w_{[1:T]},f_{[1:T]}\right)\right)\right|$$

$$\leq \sum_{s,J} \sum_{w_{[1:T]},f_{[1:T]}} \mathbb{E}_{\mathcal{B}} \left| \bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} - \mathbb{E}_{\mathcal{B}} \bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} \right|. \tag{I.10}$$

Notice that $\sum_{w_{J^c},f_{J^c}} \mathbb{1}\{x,w,f,[1:T]\} = \mathbb{1}\{x,w,f,J\}$. Define

$$P^{J,s}_{2,w_J,f_J} \triangleq \sum_{x_{[1:T]}\in\mathcal{N}_{J,s}} p_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\{x,w,f,J\}\ \mathbb{1}\left\{x_{[1:T]}\in\mathcal{D}\right\}, \tag{I.11}$$

and hence,

$$\sum_{w_{J^c},f_{J^c}} \bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} = P^{J,s}_{2,w_J,f_J}. \tag{I.12}$$

We also have

$$\mathbb{E}_{\mathcal{B}}\left(\bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}}\right) = \frac{1}{\tilde{W}_{J^c}\tilde{F}_{J^c}}\ \mathbb{E}_{\mathcal{B}_J}\left(P^{J,s}_{2,w_J,f_J}\right), \tag{I.13}$$

$$\mathbb{P}_{\mathcal{B}}\left(\bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} > \frac{1}{\tilde{W}_{J^c}\tilde{F}_{J^c}}\ P^{J,s}_{2,w_J,f_J}\right) \leq \mathbb{P}_{\mathcal{B}}\left(\sum_{w_{J^c},f_{J^c}} \bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} > P^{J,s}_{2,w_J,f_J}\right) = 0, \tag{I.14}$$

where $\mathcal{B}_J \triangleq \{\mathcal{B}_{1t}(x_t),\mathcal{B}_{2t}(x_t),\forall x_t\in\mathcal{X}_t,t\in J\}$. Using (I.14), we have

$$\mathbb{P}_{\mathcal{B}}\left(\bar{P}^{J,s}_{2,w_{[1:T]},f_{[1:T]}} \leq \frac{1}{\tilde{W}_{J^c}\tilde{F}_{J^c}} P^{J,s}_{2,w_J,f_J}\right) = 1. \tag{I.15}$$

Using the law of total expectation and (I.13), (I.10) is further upper bounded as

$$\sum_{w_{[1:T]},f_{[1:T]}} \mathbb{E}_{\mathcal{B}} \left| P_2\left(w_{[1:T]}, f_{[1:T]}\right) - \mathbb{E}_{\mathcal{B}}\left(P_2\left(w_{[1:T]}, f_{[1:T]}\right)\right)\right|$$

$$= \sum_{w_{[1:T]},f_{[1:T]}} \mathbb{E}_{\mathcal{B}} \left(\sqrt{\left(P_2\left(w_{[1:T]}, f_{[1:T]}\right) - \mathbb{E}_{\mathcal{B}}\left(P_2\left(w_{[1:T]}, f_{[1:T]}\right)\right)\right)^2} \right) \tag{I.16}$$

$$\leq \sum_{s,J} \sum_{w_J,f_J} \mathbb{E}_{\mathcal{B}_J} \sqrt{\left(P^{J,s}_{2,w_J,f_J} - \mathbb{E}_{\mathcal{B}_J} P^{J,s}_{2,w_J,f_J}\right)^2}$$

$$\leq \sum_{s,J} \sum_{w_J,f_J} \sqrt{\mathbb{V}\mathrm{ar}_{\mathcal{B}_J} P^{J,s}_{2,w_J,f_J}}, \tag{I.17}$$

where (I.17) follows from Jensen's inequality and the concavity of square root.

For any s and $J \in \mathcal{J}$, we have

$$\mathbb{V}\mathrm{ar}_{\mathcal{B}_J}\left(P^{J,s}_{2,w_J,f_J}\right) = \mathbb{V}\mathrm{ar}_{\mathcal{B}_J} \sum_{x_{[1:T]} \in \mathcal{N}_{J,s}} p_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\{x,w,f,J\}\ \mathbb{1}\left\{x_{[1:T]} \in \mathcal{D}\right\} \tag{I.18}$$

$$= \sum_{x_{[1:T]} \in \mathcal{N}_{J,s}} \mathbb{V}\mathrm{ar}_{\mathcal{B}_J}\left(p_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\{x,w,f,J\}\ \mathbb{1}\left\{x_{[1:T]} \in \mathcal{D}\right\}\right) \tag{I.19}$$

$$\leq \sum_{x_{[1:T]} \in \mathcal{N}_{J,s}} p^2_{X_{[1:T]}}\left(x_{[1:T]}\right)\ \mathbb{1}\left\{x_{[1:T]} \in \mathcal{D}\right\}\ \mathbb{E}_{\mathcal{B}_J}\left(\mathbb{1}\{x,w,f,J\}\right) \tag{I.20}$$

$$\leq \frac{1}{\tilde{W}_J \tilde{F}_J} \sum_{x_J \in \mathcal{N}_{J,s}} p^2_{X_J}(x_J)\ \mathbb{1}\left\{x_J \in \mathcal{D}_{\gamma^{(J)}}\right\} \sum_{x_{J^c} \in \mathcal{N}_{J,s}} p^2_{X_{J^c}|X_J}(x_{J^c}|x_J) \tag{I.21}$$

$$\leq 2^{-\gamma^{(J)}} \frac{1}{\tilde{W}_J \tilde{F}_J} \sum_{x_{J^c} \in \mathcal{N}_{J,s}} p^2_{X_{J^c}|X_J}\left(x_{J^c}|x_J\right), \tag{I.22}$$

where (I.19) follows since $\left\{\mathbb{1}\{x,w,f,J\} : x_{[1:T]} \in \mathcal{N}_{J,s}\right\}$ are independent random variables due to the structure of the set $\mathcal{N}_{J,s}$ and the random binning. The inequality in (I.22) follows since for all $x_J \in \mathcal{D}_{\gamma^{(J)}}$, we have $p_{X_J}(x_J) \leq 2^{-\gamma^{(J)}}$.

Lemma 11 follows by substituting (I.22) in (I.17), and noticing that

$$\sum_s \sqrt{\sum_{x_{J^c} \in \mathcal{N}_{J,s}} p^2_{X_{J^c}|X_J}(x_{J^c}|x_J)}$$

$$\leq \sum_s \sum_{x_{J^c} \in \mathcal{N}_{J,s}} p_{X_{J^c}|X_J}(x_{J^c}|x_J) \leq \sum_{x_{J^c} \in \mathcal{X}_{J^c}} p_{X_{J^c}|X_J}(x_{J^c}|x_J) = 1. \tag{I.23}$$

Appendix J

Proof of Lemma 12

Let us first consider $\bar{\mathbf{p}} \in \mathcal{P}$ that is the natural ordering of $[1:k]$. We first prove the inequality in (7.38) for $\bar{\mathbf{p}}$, i.e., by replacing the minimum in (7.38) with $\mathbf{p} = \bar{\mathbf{p}}$. The proof for (7.38) then follows from a similar proof for all $\mathbf{p} \in \mathcal{P}$. For $\mathbf{p} = \bar{\mathbf{p}}$, we prove Lemma 12 by induction. For the base of induction, $T = 1$, (7.38) reduces to the assertion in Lemma 8, i.e., (5.19). We now show that if the assertion in the lemma holds for $T = k - 1$, then it holds for $T = k$.

We rewrite the relative entropy in (7.38) as follows:

$$\mathbb{D}\left(P_{W_{[1:k]}F_{[1:k]}Z_S}||p^U_{W_{[1:k]}}p^U_{F_{[1:k]}}p_{Z_S}\right) = \sum_{w_{[1:k]},f_{[1:k]},z} P(w_{[1:k]},f_{[1:k]},z)\log\frac{P(w_{[1:k]},f_{[1:k]},z)}{p^U_{W_{[1:k]}}p^U_{F_{[1:k]}}p(z)} \tag{J.1}$$

$$= \sum_{w_{[1:k]},f_{[1:k]},z} P(w_{[1:k]},f_{[1:k]},z)\log\left(\frac{P(w_{[1:k]},f_{[1:k]},z)}{P(w_{[1:k-1]},f_{[1:k-1]},z)p^U_{W_k}p^U_{F_k}}\frac{P(w_{[1:k-1]},f_{[1:k-1]},z)}{p^U_{W_{[1:k-1]}}p^U_{F_{[1:k-1]}}p(z)}\right) \tag{J.2}$$

$$= \mathbb{D}\left(P_{W_{[1:k]}F_{[1:k]}Z_S}||P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}p^U_{W_k}p^U_{F_k}\right) + \mathbb{D}\left(P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}||p^U_{W_{[1:k-1]}}p^U_{F_{[1:k-1]}}p_{Z_S}\right). \tag{J.3}$$

Thus, the probability in (7.38) can be upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:k]}F_{[1:k]}Z_S}||p^U_{W_{[1:k]}}p^U_{D_{[1:k]}}p_{Z_S}\right) \geq k\tilde{\epsilon}\right) \leq$$

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}||p^U_{W_{[1:k-1]}}p^U_{F_{[1:k-1]}}p_{Z_S}\right)>(k-1)\tilde{\epsilon}\right)$$

$$+\,\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:k]}F_{[1:k]}Z_S}||P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}p^U_{W_k}p^U_{F_k}\right)>\tilde{\epsilon}\right). \tag{J.4}$$

By the induction hypothesis, the first probability in (J.4) is upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}||p^U_{W_{[1:k-1]}}p^U_{F_{[1:k-1]}}p_{Z_S}\right)>(k-1)\tilde{\epsilon}\right)$$

$$\leq|\mathcal{S}||\mathcal{Z}|\sum_{t=1}^{k-1}\exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_t^{\bar{\mathbf{p}}}}}{3\tilde{W}_t\tilde{F}_t}\right). \tag{J.5}$$

Using similar analysis as in Appendix H, we can show that the second probability in (J.4) is be upper bounded as

$$\mathbb{P}_{\mathcal{B}}\left(\max_{S\in\mathcal{S}}\mathbb{D}\left(P_{W_{[1:k]}F_{[1:k]}Z_S}||P_{W_{[1:k-1]}F_{[1:k-1]}Z_S}p^U_{W_k}p^U_{F_k}\right)>\tilde{\epsilon}\right)$$

$$\leq|\mathcal{S}||\mathcal{Z}|\exp\left(\frac{-\epsilon^2(1-\delta)2^{\gamma_k^{\bar{\mathbf{p}}}}}{3\tilde{W}_k\tilde{F}_k}\right). \tag{J.6}$$

We conclude that (7.38) holds for $\mathbf{p}=\bar{\mathbf{p}}$. By rewriting (J.3) with the different permutations of $[1:k]$ and repeating the proof, the minimum over $\mathbf{p}\in\mathcal{P}$ in (7.38) follows, hence Lemma 12.

Appendix K

Secrecy Constraint for Setting 1: Proof of (8.11)

For every $S_1 \subseteq [1:n]$ satisfying $|S_1| = \mu$, we have

$$\lim_{n\to\infty} I\left(W_1, W_2; \mathbf{Z}^n_{S_1}\right) = \lim_{n\to\infty} I\left(W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}, W_{1,s}, W_{2,s}; \mathbf{Z}^n_{S_1}\right) \tag{K.1}$$

$$= \lim_{n\to\infty} I\left(W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}; \mathbf{Z}^n_{S_1}\right) \tag{K.2}$$

$$\leq \lim_{n\to\infty} I\left(W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}; \mathbf{Z}^n_{S_1}\right) \tag{K.3}$$

$$= \lim_{n\to\infty} I\left(M_c; \mathbf{Z}^n_{S_1}\right) = 0. \tag{K.4}$$

Recall that the adversary's observation over cache placement, $\mathbf{Z}^n_{S_1}$, results from sending $M_c = \{M_{c,1}, M_{c,2}\}$, where $M_{c,1} = W_1^{(1)} \oplus W_2^{(1)}$ and $M_{c,2} = W_1^{(2)} \oplus W_2^{(2)}$. Thus, (K.2) follows because $\mathbf{Z}^n_{S_1}$ does not depend on $\{W_{1,s}, W_{2,s}\}$ and (K.3) follows due to the Markov chain $\left\{W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}\right\} - \left\{W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}\right\} - \mathbf{Z}^n_{S_1}$. The second equality in (K.4) follows from [33, Theorem 2], and the fact that the rate of M_c is less than $1 - \alpha$.

Appendix L

Secrecy Constraint for Setting 2: Proof of (8.16)

For every $S_2 \subseteq [1:n]$ satisfying $|S_2| = \mu$, and any $\mathbf{d} = (d_1, d_2)$, $d_1, d_2 \in \{1, 2\}$, we have

$$I\big(W_1,W_2;\mathbf{Z}_{S_2}^n\big) = I\left(W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_2}^n\right) \tag{L.1}$$

$$= I\left(W_{d_1}^{(2)}, W_{d_2}^{(1)}; \mathbf{Z}_{S_2}^n\right) + I\left(W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_2}^n \middle| W_{d_1}^{(2)}, W_{d_2}^{(1)}\right) \tag{L.2}$$

$$\leq I\left(W_{d_1}^{(2)}, W_{d_2}^{(1)}; \mathbf{Z}_{S_2}^n\right) + I\left(W_{d_1,s}, W_{d_2,s}; W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2 \middle| W_{d_1}^{(2)}, W_{d_2}^{(1)}\right) \tag{L.3}$$

$$= I\left(W_{d_1}^{(2)}, W_{d_2}^{(1)}; \mathbf{Z}_{S_2}^n\right) + I(W_{d_1,s}, W_{d_2,s}; W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\big) \tag{L.4}$$

$$= I\left(W_{d_1}^{(2)}, W_{d_2}^{(1)}; \mathbf{Z}_{S_2}^n\right) \tag{L.5}$$

$$= I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right). \tag{L.6}$$

The adversary's observation over the delivery phase, $\mathbf{Z}_{S_2}^n$, results from sending $M_{\mathbf{d}} = \{W_{d_1}^{(2)}, W_{d_2}^{(1)}\}$ and the randomization message $\tilde{M}_{\mathbf{d}} = \big\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\big\}$. Equation (L.1) thus follows because $\mathbf{Z}_{S_2}^n$ depends only on $W_{d_1}^{(2)}$, $W_{d_2}^{(1)}$, $W_{d_1,s}$, and $W_{d_2,s}$. The inequality in (L.3) follows from the Markov chain $\{W_{d_1,s}, W_{d_2,s}\} - \{W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\} - \mathbf{Z}_{S_2}^n$. Equation (L.4) follows because $\{W_{d_1}^{(2)}, W_{d_2}^{(1)}\}$ are independent from $\big\{W_{d_1,s}, W_{d_2,s}, K_1, K_2\big\}$.

The randomization message of the wiretap code in the delivery phase, $\tilde{M}_{\mathbf{d}}$, is independent from the message $M_{\mathbf{d}}$. Thus, using (L.6) and [33, Theorem 2], we have

$$\lim_{n\to\infty} \max_{S_2\subseteq[1:n]:|S_2|=\mu} I(W_1, W_2; \mathbf{Z}^n_{S_2})$$

$$\leq \lim_{n\to\infty} \max_{S_2\subseteq[1:n]:\ |S_2|=\mu} I(M_{\mathbf{d}}; \mathbf{Z}^n_{S_2}) = 0. \tag{L.7}$$

Appendix M

Secrecy Constraint for Setting 3 When $\alpha_1 \geq \alpha_2$

Recall that M_c and $M_{\mathbf{d}}$ are defined as in (8.9) and (8.10), respectively. For a fixed choice of the subsets $S_1, S_2 \subseteq [1:n]$ such that $|S_1| + |S_2| = \mu$, and any $\mathbf{d} = (d_1, d_2)$, $d_1, d_2 \in \{1, 2\}$, we have

$$I(W_1,W_2;\mathbf{Z}^n_{S_1},\mathbf{Z}^n_{S_2}) = I\left(W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}, W_{1,s}, W_{2,s}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{M.1}$$

$$= I\left(W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{M.2}$$

$$= I\left(M_c, M_{\mathbf{d}}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) \tag{M.3}$$

$$= I\left(M_c; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2}\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}^n_{S_1}, \mathbf{Z}^n_{S_2} | M_c\right) \tag{M.4}$$

$$= I\left(M_c; \mathbf{Z}^n_{S_1}\right) + I\left(M_c; \mathbf{Z}^n_{S_2} | \mathbf{Z}^n_{S_1}\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}^n_{S_2} | M_c\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}^n_{S_1} | M_c, \mathbf{Z}^n_{S_2}\right), \tag{M.5}$$

where (M.2) follows because, for any $d_1, d_2 \in \{1, 2\}$, there exists a bijective map between $\left\{W_1^{(1)}, W_1^{(2)}, W_2^{(1)}, W_2^{(2)}\right\}$ and $\left\{W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, W_{d_1}^{(2)}, W_{d_2}^{(1)}\right\}$.

From (8.9) and (8.10), M_c and $M_{\mathbf{d}}$ are independent. The adversary's observation in cache placement, $\mathbf{Z}^n_{S_1}$, results from sending M_c, while its observation in the delivery phase, $\mathbf{Z}^n_{S_2}$, results from sending $M_{\mathbf{d}}$. Thus, for a fixed choice of the subsets S_1 and S_2,

$\left\{M_c, \mathbf{Z}_{S_1}^n\right\}$ are independent from $\mathbf{Z}_{S_2}^n$. We thus have

$$I\left(M_c; \mathbf{Z}_{S_2}^n | \mathbf{Z}_{S_1}^n\right) = 0. \tag{M.6}$$

In addition, $\left\{M_{\mathbf{d}}, \mathbf{Z}_{S_2}^n\right\}$ are independent from M_c. Thus,

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n | M_c\right) = H\left(\mathbf{Z}_{S_2}^n | M_c\right) - H\left(\mathbf{Z}_{S_2}^n | M_c, M_{\mathbf{d}}\right) \tag{M.7}$$

$$= H\left(\mathbf{Z}_{S_2}^n | M_c\right) - H\left(\mathbf{Z}_{S_2}^n | M_{\mathbf{d}}\right) \tag{M.8}$$

$$\leq I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right). \tag{M.9}$$

Finally, using the Markov chain $\left\{M_{\mathbf{d}}, \mathbf{Z}_{S_2}^n\right\} - M_c - \mathbf{Z}_{S_1}^n$, we have

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n | M_c, \mathbf{Z}_{S_2}^n\right) = H\left(\mathbf{Z}_{S_1}^n | M_c, \mathbf{Z}_{S_2}^n\right) - H\left(\mathbf{Z}_{S_1}^n | M_c, \mathbf{Z}_{S_2}^n, M_{\mathbf{d}}\right) \tag{M.10}$$

$$\leq H\left(\mathbf{Z}_{S_1}^n\right) - H\left(\mathbf{Z}_{S_1}^n | M_c\right) = I\left(M_c; \mathbf{Z}_{S_1}^n\right). \tag{M.11}$$

Substituting (M.6), (M.9), and (M.11) in (M.5),

$$I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \leq 2I\left(M_c; \mathbf{Z}_{S_1}^n\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right). \tag{M.12}$$

The rates of M_c and $M_{\mathbf{d}}$ are $1 - \alpha_1 - \epsilon_n$ and $1 - \alpha_2 - \epsilon_n$, respectively. By applying [33, Theorem 2] to (M.12), we have

$$\lim_{n\to\infty} \max_{\substack{S_1, S_2 \subseteq [1:n]: \\ |S_1|+|S_2|=\mu}} I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right)$$

$$\leq 2 \lim_{n\to\infty} \max_{S_1\subseteq[1:n]:\ |S_1|=\mu_1} I(M_c; \mathbf{Z}^n_{S_1}) + \lim_{n\to\infty} \max_{S_2\subseteq[1:n]:\ |S_2|=\mu_2} I\left(M_{\mathbf{d}}; \mathbf{Z}^n_{S_2}\right) \quad \text{(M.13)}$$

$$= 0. \quad \text{(M.14)}$$

Appendix N

Secrecy Constraint for Setting 3 When $\alpha_1 < \alpha_2$

For this case, M_c and $M_{\mathbf{d}}$ are defined in (8.13) and (8.14) and the randomization message $\tilde{M}_{\mathbf{d}}$ is defined in (8.15). For notational simplicity, let us define

$$M_{c,1\backslash K_1} = W_1^{(1)} \oplus W_2^{(1)}, \qquad M_{c,2\backslash K_2} = W_1^{(2)} \oplus W_2^{(2)} \tag{N.1}$$

$$M_{c\backslash K} = \left\{M_{c,1\backslash K_1}, M_{c,2\backslash K_2}\right\}. \tag{N.2}$$

For a fixed choice of $S_1, S_2 \subseteq [1:n]$ such that $|S_1|+|S_2| = \mu$, and any $\mathbf{d} = (d_1, d_2)$, $d_1, d_2 \in \{1,2\}$, we have

$$I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) = I\left(W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{N.3}$$

$$= I\left(M_{c\backslash K}, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{N.4}$$

$$= I\left(M_{c\backslash K}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n | M_{c\backslash K}\right) + I\left(W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n | M_{\mathbf{d}}, M_{c\backslash K}\right). \tag{N.5}$$

From (8.13), (8.14), and (8.15), M_c is independent from $\{M_{\mathbf{d}}, \tilde{M}_{\mathbf{d}}\}$. The adversary's observation in cache placement, $\mathbf{Z}_{S_1}^n$, results from sending $M_c = \left\{M_{c\backslash K}, K_1, K_2\right\}$, and its observation in the delivery results from sending $M_{\mathbf{d}} = \left\{W_{d_1}^{(2)}, W_{d_2}^{(1)}\right\}$ and the randomization message $\tilde{M}_{\mathbf{d}} = \left\{W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus W_2\right\}$. We now upper bound each

term on the right hand side of (N.5). For the third term, we have

$$I\Big(W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n | M_{\mathbf{d}}, M_{c\backslash K}\Big) \leq I\left(W_{d_1,s}, W_{d_2,s}; \tilde{M}_{\mathbf{d}} | M_{\mathbf{d}}, M_{c\backslash K}\right) \tag{N.6}$$

$$= I\left(W_{d_1,s}, W_{d_2,s}; W_{d_1,s} \oplus K_1, W_{d_2,s} \oplus K_2\right) = 0, \tag{N.7}$$

where (N.6) follows due to the Markov chain $\{W_{d_1,s}, W_{d_2,s}\} - \{M_{c\backslash K}, M_{\mathbf{d}}, \tilde{M}_{\mathbf{d}}\} - \{\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\}$, and (N.7) follows because $\tilde{M}_{\mathbf{d}}$ is independent from $\{W_{d_1,s}, W_{d_2,s}, M_{\mathbf{d}}, M_{c\backslash K}\}$.

For a fixed choice of the subsets S_1 and S_2, $\mathbf{Z}_{S_2}^n$ is independent from $\left\{M_c, \mathbf{Z}_{S_1}^n\right\}$. Thus, the first term on the right hand side of (N.5) is bounded as

$$I\left(M_{c\backslash K}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \leq I\left(M_c; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{N.8}$$

$$= I\left(M_c; \mathbf{Z}_{S_1}^n\right) + I\left(M_c; \mathbf{Z}_{S_2}^n | \mathbf{Z}_{S_1}^n\right) = I\left(M_c; \mathbf{Z}_{S_1}^n\right). \tag{N.9}$$

For the second term on the right hand side of (N.5), we have

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n | M_{c\backslash K}\right) = I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n | M_{c\backslash K}\right) + I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n | M_{c\backslash K}, \mathbf{Z}_{S_2}^n\right). \tag{N.10}$$

Notice that $M_{c\backslash K}$ and $\mathbf{Z}_{S_2}^n$ are conditionally independent given $M_{\mathbf{d}}$. Thus,

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n | M_{c\backslash K}\right) = H\left(\mathbf{Z}_{S_2}^n | M_{c\backslash K}\right) - H\left(\mathbf{Z}_{S_2}^n | M_{\mathbf{d}}\right) \leq I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right). \tag{N.11}$$

In addition, using the independence between $\left\{M_{\mathbf{d}}, \mathbf{Z}_{S_2}^n\right\}$ and $\left\{M_c, \mathbf{Z}_{S_1}^n\right\}$, we have

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n | M_{c\backslash K}, \mathbf{Z}_{S_2}^n\right) = H\left(\mathbf{Z}_{S_1}^n | M_{c\backslash K}, \mathbf{Z}_{S_2}^n\right) - H\left(\mathbf{Z}_{S_1}^n | M_{c\backslash K}, M_{\mathbf{d}}, \mathbf{Z}_{S_2}^n\right) \tag{N.12}$$

$$\leq H\left(\mathbf{Z}_{S_1}^n\right) - H\left(\mathbf{Z}_{S_1}^n \middle| M_{c\backslash K}, K_1, K_2, M_{\mathbf{d}}, \mathbf{Z}_{S_2}^n\right) \tag{N.13}$$

$$= H\left(\mathbf{Z}_{S_1}^n\right) - H\left(\mathbf{Z}_{S_1}^n \middle| M_c\right) = I\left(M_c; \mathbf{Z}_{S_1}^n\right). \tag{N.14}$$

Substituting (N.11) and (N.14) in (N.10) gives

$$I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n \middle| M_{c\backslash K}\right) \leq I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right) + I\left(M_c; \mathbf{Z}_{S_1}^n\right). \tag{N.15}$$

Finally, substituting (N.7), (N.9), (N.15) in (N.5), and applying [33, Theorem 2], we have

$$\lim_{n\to\infty} \max_{\substack{S_1, S_2 \subseteq [1:n]: \\ |S_1|+|S_2|=\mu}} I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) = 0, \tag{N.16}$$

since the rates of M_c and $M_{\mathbf{d}}$ are $1-\alpha_1-\epsilon_n$ and $1-\alpha_2-\epsilon_n$, respectively.

Appendix O

Secrecy Constraint for Setting 4

For this setting, M_c and $M_{\mathbf{d}}$ are defined in (8.20) and (8.22), and the randomization messages $\tilde{M}_c$ and $\tilde{M}_{\mathbf{d}}$ are defined in (8.21) and (8.23). Notice that, M_c is independent from $\tilde{M}_c$; $M_{\mathbf{d}}$ is independent from $\tilde{M}_{\mathbf{d}}$, and $\{M_c, \tilde{M}_c\}$ are independent from $\{M_{\mathbf{d}}, \tilde{M}_{\mathbf{d}}\}$.

Conditioned on a fixed choice of the subsets S_1 and S_2, which satisfies the conditions for this setting, i.e., either $\{|S_1| = \mu,\ |S_2| = 0\}$ or $\{|S_1| = 0,\ |S_2| = \mu\}$, define the random variable

$$\mathbf{Z}_S^n \triangleq \mathbf{Z}_{S_1}^n\ \mathbb{1}_{\{|S_2|=0\}} + \mathbf{Z}_{S_2}^n\ \mathbb{1}_{\{|S_1|=0\}}. \tag{O.1}$$

Notice that the random variable $\mathbf{Z}_S^n$ only have a well-defined probability distribution when conditioned on the event $\{S_1, S_2\}$, since a prior distribution on these subsets is not defined. For this fixed choice of the subsets, and any $\mathbf{d} = (d_1, d_2)$, $d_1, d_2 \in \{1, 2\}$, we have

$$I\big(W_1,W_2;\mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\big) = I\left(W_1^{(1)} \oplus W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}, W_{d_1}^{(2)}, W_{d_2}^{(1)}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{O.2}$$

$$= I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_S^n\right) \tag{O.3}$$

$$= \mathbb{1}_{\{|S_2|=0\}}\ I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_S^n \middle| \{|S_2| = 0\}\right)$$

$$+ \mathbb{1}_{\{|S_1|=0\}} \, I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_S^n \middle| \{|S_1|=0\}\right) \tag{O.4}$$

$$= \mathbb{1}_{\{|S_2|=0\}} \, I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_1}^n\right) + \mathbb{1}_{\{|S_1|=0\}} \, I\left(M_c, M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_2}^n\right) \tag{O.5}$$

$$= \mathbb{1}_{\{|S_2|=0\}} \, I\left(M_c; \mathbf{Z}_{S_1}^n\right) + \mathbb{1}_{\{|S_1|=0\}} \, I\left(M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}; \mathbf{Z}_{S_2}^n\right) \tag{O.6}$$

$$\leq \mathbb{1}_{\{|S_2|=0\}} \, I\left(M_c; \mathbf{Z}_{S_1}^n\right) + \mathbb{1}_{\{|S_1|=0\}} \, I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right). \tag{O.7}$$

Equation (O.6) follows because (i) $\mathbf{Z}_{S_1}^n$ results from $\{M_c, \tilde{M}_c\}$ which are independent from $\{M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}\}$, and (ii) $\mathbf{Z}_{S_2}^n$ is conditionally independent from M_c given $\{M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}\}$, due to the Markov chain $M_c - \{M_{\mathbf{d}}, W_{d_1,s}, W_{d_2,s}\} - \{M_{\mathbf{d}}, \tilde{M}_{\mathbf{d}}\} - \mathbf{Z}_{S_2}^n$. The inequality in (O.7) follows using the same steps in (L.1)–(L.6), in Appendix L.

Finally, since $\tilde{M}_c$ is independent from M_c; $\tilde{M}_{\mathbf{d}}$ is independent from $M_{\mathbf{d}}$, and the rates of M_c and $\tilde{M}_{\mathbf{d}}$ are both equal to $1 - \alpha - \epsilon_n$, we have

$$\lim_{n\to\infty} \max_{\substack{S_1,S_2\subseteq[1:n]:\\ |S_1|+|S_2|=\mu}} I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) = \lim_{n\to\infty} \max_{\substack{S_1,S_2\subseteq[1:n]:\\ |S_i|=0,\, |S_j|=\mu\\ i,j\in\{1,2\},\, i\neq j}} I\left(W_1, W_2; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \tag{O.8}$$

$$\leq \lim_{n\to\infty} \max\left\{ \max_{S_1\subseteq[1:n]:\, |S_1|=\mu} I(M_c; \mathbf{Z}_{S_1}^n), \max_{S_2\subseteq[1:n]:\, |S_2|=\mu} I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right) \right\} \tag{O.9}$$

$$= \max\left\{ \lim_{n\to\infty} \max_{S_1\subseteq[1:n]:\, |S_1|=\mu} I(M_c; \mathbf{Z}_{S_1}^n), \lim_{n\to\infty} \max_{S_2\subseteq[1:n]:\, |S_2|=\mu} I\left(M_{\mathbf{d}}; \mathbf{Z}_{S_2}^n\right) \right\} \tag{O.10}$$

$$= 0, \tag{O.11}$$

where (O.9) follows from (O.7), and (O.10) follows because both limits exist and equal to zero; by using [33, Theorem 2].

Appendix P

Proofs of (8.42) and (8.43)

Let us fix the subsets S_1 and S_2, and the messages w_c and $w_{\mathbf{d}}$. Consider the Cartesian product of the random bins $\mathcal{B}_{w_c}$ and $\mathcal{B}_{w_{\mathbf{d}}}$, i.e., $\mathcal{B}_{w_c,w_{\mathbf{d}}}$, defined in (8.33). Recall that $P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n \left| W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}} \right.}$ denotes the induced distribution at the adversary's output when the transmitted codewords over cache placement and delivery phases are $\mathbf{x}_c^n\left(w_c, \tilde{w}_c\right)$ and $\mathbf{x}_{\mathbf{d}}^n\left(w_{\mathbf{d}}, \tilde{w}_{\mathbf{d}}\right)$, i.e., when $\left(\mathbf{x}_c^n, \mathbf{x}_{\mathbf{d}}^n\right)$ belongs to $\mathcal{B}_{w_c,w_{\mathbf{d}}}$. In addition, $P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}$ denotes the output distribution at the adversary, induced by the cache placement and delivery codebooks, $\mathcal{C}_{c,n}$ and $\mathcal{C}_{\mathbf{d},n}$, defined in Figures 8.3 and 8.4.

Let $\mathbf{z}_1^n, \mathbf{z}_2^n \in \mathcal{Z}^n$, where $\mathcal{Z} \triangleq \{0,1\} \cup \{?\}$. Define the distribution $Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}$ as follows:

$$Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(\mathbf{z}_1^n, \mathbf{z}_2^n) = \prod_{i \notin S_1, j \notin S_2} \mathbb{1}\{z_{1,i} =?, z_{2,j} =?\} \prod_{i \in S_1, j \in S_2} U_X(z_{1,i})\, U_X(z_{2,i}), \tag{P.1}$$

where $U_X(z)$ is a uniform binary distribution when $z = 0, 1$, and $U_X(z) = 0$ when $z =?$.

We thus have

$$I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) = \mathbb{D}\left(P_{W_c W_{\mathbf{d}} \mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2^n}} \middle|\middle| P_{W_c W_{\mathbf{d}}} P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2^n}}\right) \tag{P.2}$$

$$= \sum_{w_c, w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}}) \sum_{\mathbf{z}_1^n, \mathbf{z}_2^n} P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n \left| W_c W_{\mathbf{d}} \right.}(\mathbf{z}_1^n, \mathbf{z}_2^n | w_c, w_{\mathbf{d}}) \log\left(\frac{P_{W_c W_{\mathbf{d}} \mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(w_c, w_{\mathbf{d}}, \mathbf{z}_1^n, \mathbf{z}_2^n)}{P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(\mathbf{z}_1^n, \mathbf{z}_2^n) P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}})}\right) \tag{P.3}$$

$$= \sum_{w_c, w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}}) \sum_{\mathbf{z}_1^n, \mathbf{z}_2^n} P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n | W_c W_{\mathbf{d}}}(\mathbf{z}_1^n, \mathbf{z}_2^n | w_c, w_{\mathbf{d}})$$

$$\times \log \left(\frac{P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n | W_c W_{\mathbf{d}}}(\mathbf{z}_1^n, \mathbf{z}_2^n | w_c, w_{\mathbf{d}})}{Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(\mathbf{z}_1^n, \mathbf{z}_2^n)} \times \frac{Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(\mathbf{z}_1^n, \mathbf{z}_2^n)}{P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n}(\mathbf{z}_1^n, \mathbf{z}_2^n)} \right) \tag{P.4}$$

$$= \sum_{w_c, w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}}) \left[\mathbb{D}\left(P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n | W_c = w_c, W_{\mathbf{d}} = w_{\mathbf{d}}} \middle|\middle| Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n} \right) - \mathbb{D}\left(P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n} \middle|\middle| Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n} \right) \right] \tag{P.5}$$

$$\leq \sum_{w_c, w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}}) \, \mathbb{D}\left(P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n | W_c = w_c, W_{\mathbf{d}} = w_{\mathbf{d}}} \middle|\middle| Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n} \right). \tag{P.6}$$

Define $\mathbf{Z}^{S_1} \triangleq \{Z_{S_1,i} : i \in S_1\}$, $\mathbf{Z}^{S_2} \triangleq \{Z_{S_2,i} : i \in S_2\}$, $\mathbf{Z}^{S_1^c} \triangleq \{Z_{S_1,i} : i \notin S_1\}$, $\mathbf{Z}^{S_2^c} \triangleq \{Z_{S_2,i} : i \notin S_2\}$, and let $\mathbf{z}^{S_1}$, $\mathbf{z}^{S_2}$, $\mathbf{z}^{S_1^c}$, $\mathbf{z}^{S_2^c}$ be the corresponding realizations. Note that $\mathbf{Z}_{S_1}^n = \{\mathbf{Z}^{S_1}, \mathbf{Z}^{S_1^c}\}$ and $\mathbf{Z}_{S_2}^n = \{\mathbf{Z}^{S_2}, \mathbf{Z}^{S_2^c}\}$. For each S_1, S_2, w_c, and $w_{\mathbf{d}}$, we have

$$\mathbb{D}\left(P_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n | W_c = w_c, W_{\mathbf{d}} = w_{\mathbf{d}}} \middle|\middle| Q_{\mathbf{Z}_{S_1}^n \mathbf{Z}_{S_2}^n} \right) = \mathbb{D}\left(P_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2} \mathbf{Z}^{S_2^c} | W_c = w_c, W_{\mathbf{d}} = w_{\mathbf{d}}} \middle|\middle| Q_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2} \mathbf{Z}^{S_2^c}} \right) \tag{P.7}$$

$$= \sum_{\mathbf{z}^{S_1}, \mathbf{z}^{S_1^c}, \mathbf{z}^{S_2}, \mathbf{z}^{S_2^c}} P_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2} \mathbf{Z}^{S_2^c} | W_c W_{\mathbf{d}}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_1^c}, \mathbf{z}^{S_2}, \mathbf{z}^{S_2^c} | w_c, w_{\mathbf{d}} \right)$$

$$\times \log \left(\frac{P_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2} \mathbf{Z}^{S_2^c} | W_c W_{\mathbf{d}}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_1^c}, \mathbf{z}^{S_2}, \mathbf{z}^{S_2^c} | w_c, w_{\mathbf{d}} \right)}{Q_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2} \mathbf{Z}^{S_2^c}}(\mathbf{z}^{S_1}, \mathbf{z}^{S_1^c}, \mathbf{z}^{S_2}, \mathbf{z}^{S_2^c})} \right) \tag{P.8}$$

$$= \sum_{\mathbf{z}^{S_1}, \mathbf{z}^{S_1^c}, \mathbf{z}^{S_2}, \mathbf{z}^{S_2^c}} P_{\mathbf{Z}^{S_1} \mathbf{Z}^{S_2} | W_c W_{\mathbf{d}}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_2} | w_c, w_{\mathbf{d}} \right) P_{\mathbf{Z}^{S_1^c} \mathbf{Z}^{S_2^c} | W_c W_{\mathbf{d}} \mathbf{Z}^{S_1} \mathbf{Z}^{S_2}}\left(\mathbf{z}^{S_1^c}, \mathbf{z}^{S_2^c} | w_c, w_{\mathbf{d}}, \mathbf{z}^{S_1}, \mathbf{z}^{S_2} \right)$$

$$\times \log \left(\frac{P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}|W_c W_{\mathbf{d}}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_2} | w_c, w_{\mathbf{d}}\right) P_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|W_c W_{\mathbf{d}}\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\left(\mathbf{z}^{S_1^c}, \mathbf{z}^{S_2^c} | w_c, w_{\mathbf{d}}, \mathbf{z}^{S_1}, \mathbf{z}^{S_2}\right)}{Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_2}\right) Q_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\left(\mathbf{z}^{S_1^c}, \mathbf{z}^{S_2^c} | \mathbf{z}^{S_1}, \mathbf{z}^{S_2}\right)} \right) \tag{P.9}$$

$$= \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}} || Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}} \right) + \sum_{\mathbf{z}^{S_1}, \mathbf{z}^{S_2}} P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}}\left(\mathbf{z}^{S_1}, \mathbf{z}^{S_2}\right)$$

$$\times \mathbb{D}\left(P_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}, \mathbf{Z}^{S_1}=\mathbf{z}^{S_1}, \mathbf{Z}^{S_2}=\mathbf{z}^{S_2}} || Q_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|\mathbf{Z}^{S_1}=\mathbf{z}^{S_1}, \mathbf{Z}^{S_2}=\mathbf{z}^{S_2}} \right) \tag{P.10}$$

$$= \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}} || Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}} \right), \tag{P.11}$$

where (P.11) follows because

$$P_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}, \mathbf{Z}^{S_1}=\mathbf{z}^{S_1}, \mathbf{Z}^{S_2}=\mathbf{z}^{S_2}} = Q_{\mathbf{Z}^{S_1^c}\mathbf{Z}^{S_2^c}|\mathbf{Z}^{S_1}=\mathbf{z}^{S_1}, \mathbf{Z}^{S_2}=\mathbf{z}^{S_2}}$$

$$= \prod_{i \notin S_1, j \notin S_2} \mathbb{1}\{z_{1,i} = ?, z_{2,j} = ?\}. \tag{P.12}$$

By applying the stronger version of Wyner's soft covering lemma in [33, Lemma 1] to (P.11), for every $\epsilon > 0$, there exists a $\gamma(\epsilon) > 0$ such that

$$\mathbb{P}_{\mathcal{B}_{w_c, w_{\mathbf{d}}}}\left(\mathbb{D}\left(P_{\mathbf{Z}^n_{S_1}\mathbf{Z}^n_{S_2}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}} || Q_{\mathbf{Z}^n_{S_1}\mathbf{Z}^n_{S_2}} \right) > \epsilon \right)$$

$$= \mathbb{P}_{\mathcal{B}_{w_c, w_{\mathbf{d}}}}\left(\mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}} || Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}} \right) > \epsilon \right) \leq \exp\left(-e^{n\gamma(\epsilon)} \right), \tag{P.13}$$

since the rate of $\mathcal{B}_{w_c, w_{\mathbf{d}}}$ is slightly greater than α, i.e., $\mathcal{B}_{w_c, w_{\mathbf{d}}}$ contains $2^{n\alpha_\epsilon}$ codewords.

Using (P.6) and (P.11), we have

$$I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \leq \sum_{w_c, w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}})\, \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}\left|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}\right.} \middle|\middle| Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\right). \tag{P.14}$$

Thus,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\max_{\substack{S_1,S_2\subseteq[1:n]:\\ |S_1|+|S_2|=\mu}} I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right) \geq \epsilon\right) \\
&\leq \mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\max_{S_1,S_2} \sum_{w_c,w_{\mathbf{d}}} P_{W_c W_{\mathbf{d}}}(w_c, w_{\mathbf{d}})\, \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}\left|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}\right.} \middle|\middle| Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\right) > \epsilon\right) && \text{(P.15)} \\
&\leq \mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\max_{\substack{w_c,w_{\mathbf{d}}\\ S_1,S_2}} \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}\left|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}\right.} \middle|\middle| Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\right) > \epsilon\right) && \text{(P.16)} \\
&= \mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\bigcup_{\substack{w_c,w_{\mathbf{d}}\\ S_1,S_2}} \mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}\left|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}\right.} \middle|\middle| Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\right) > \epsilon\right) && \text{(P.17)} \\
&\leq \sum_{\substack{w_c,w_{\mathbf{d}}\\ S_1,S_2}} \mathbb{P}_{\mathcal{B}_{w_c,w_{\mathbf{d}}}}\left(\mathbb{D}\left(P_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}\left|W_c=w_c, W_{\mathbf{d}}=w_{\mathbf{d}}\right.} \middle|\middle| Q_{\mathbf{Z}^{S_1}\mathbf{Z}^{S_2}}\right) > \epsilon\right) && \text{(P.18)}
\end{aligned}$$

where (P.15) follows from (P.14), and (P.18) follows from the union bound. Since the combined number of the messages $w_c, w_{\mathbf{d}}$, and the subsets S_1, S_2 is at most exponential in n, using the super-exponential decay rate in (P.13), the probability term on the right hand side of (P.15) goes to zero as n goes to infinity. Thus, $\max_{S_1,S_2} I\left(W_c, W_{\mathbf{d}}; \mathbf{Z}_{S_1}^n, \mathbf{Z}_{S_2}^n\right)$ converges to zero almost surely. This completes the proof of (8.43).

Bibliography

[1] V. Aggarwal, L. Lai, A. R. Calderbank, and H. V. Poor. Wiretap channel type II with an active eavesdropper. In *IEEE International Symposium on Information Theory*, pages 1944–1948, July 2009.

[2] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–Part I: Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993.

[3] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–Part II: CR capacity. *IEEE Transactions on Information Theory*, 44(1):225–240, 1998.

[4] K. C. Almeroth and M. H. Ammar. The use of multicast delivery to provide a scalable and interactive video-on-demand service. *IEEE Journal on Selected Areas in Communications*, 14(6):1110–1122, 1996.

[5] M. Amiri and D. Gündüz. Cache-aided content delivery over erasure broadcast channels. *IEEE Transactions on Communications*, 66(1):370–381, 2018.

[6] M. M. Amiri and D. Gündüz. Fundamental limits of coded caching: Improved delivery rate-cache capacity tradeoff. *IEEE Transactions on Communications*, 65(2):806–815, 2017.

[7] M. M. Amiri, Q. Yang, and D. Gündüz. Decentralized caching and coded delivery with distinct cache capacities. *IEEE Transactions on Communications*, 65(11):4657–4669, 2017.

[8] Z. H. Awan and A. Sezgin. Fundamental limits of caching in D2D networks with secure delivery. In *IEEE International Conference on Communication Workshop*, pages 464–469, June 2015.

[9] M. Benammar and P. Piantanida. Secrecy capacity region of some classes of wiretap broadcast channels. *IEEE Transactions on Information Theory*, 61(10):5564–5582, 2015.

[10] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, December 1984.

[11] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM journal on Computing*, 17(2):210–229, 1988.

[12] S. S. Bidokhti, M. Wigger, and R. Timo. Noisy broadcast networks with receiver caching. *Submitted to IEEE Transactions on Information Theory*, 2016. arXiv preprint arXiv:1605.02317.

[13] S. S. Bidokhti, M. Wigger, and A. Yener. Benefits of cache assignment on degraded broadcast channels. *Submitted to IEEE Transactions on Information Theory*, 2017. arXiv preprint arXiv:1702.08044.

[14] M. Bloch and J. Barros. *Physical-layer security: From information theory to security engineering*. Cambridge University Press, 2011.

[15] M. Bloch, M. Hayashi, and A. Thangaraj. Error-control coding for physical-layer secrecy. *Proceedings of the IEEE*, 103(10):1725–1746, 2015.

[16] M. R. Bloch and J. N. Laneman. Strong secrecy from channel resolvability. *IEEE Transactions on Information Theory*, 59(12):8077–8098, 2013.

[17] S. Borst, V. Gupta, and A. Walid. Distributed caching algorithms for content distribution networks. In *IEEE International Conference on Computer Communications*, pages 1–9, March 2010.

[18] D. Cao, D. Zhang, P. Chen, N. Liu, W. Kang, and D. Gündüz. Coded caching with heterogeneous cache sizes and link qualities: The two-user case. In *IEEE International Symposium on Information Theory*, pages 1545–1549, June 2018.

[19] Y.-K. Chia and A. El Gamal. Three-receiver broadcast channels with common and confidential messages. *IEEE Transactions on Information Theory*, 58(5):2748–2765, 2012.

[20] T. M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.). *IEEE Transactions on Information Theory*, 21(2):226–228, 1975.

[21] T. M. Cover and J. A. Thomas. *Elements of information theory 2nd edition*. New York, NY, USA: Wiley, 2006.

[22] I. Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):40–47, 1996.

[23] I. Csiszar and J. Korner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–3487, 1978.

[24] P. Cuff. Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096, 2013.

[25] L. W. Dowdy and D. V. Foster. Comparative models of the file assignment problem. *ACM Computing Surveys (CSUR)*, 14(2):287–313, 1982.

[26] E. Ekrem and S. Ulukus. Secrecy capacity of a class of broadcast channels with an eavesdropper. *EURASIP Journal on Wireless Communications and Networking, Special Issue on Wireless Physical Layer Security*, 2009, Article ID 824235, 29 pages, 2009.

[27] E. Ekrem and S. Ulukus. The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel. *IEEE Transactions on Information Theory*, 57(4):2083–2114, 2011.

[28] E. Ekrem and S. Ulukus. Secrecy in cooperative relay broadcast channels. *IEEE Transactions on Information Theory*, 57(1):137–155, 2011.

[29] E. Ekrem and S. Ulukus. Capacity region of Gaussian MIMO broadcast channels with common and confidential messages. *IEEE Transactions on Information Theory*, 58(9):5669–5680, 2012.

[30] A. El Gamal and Y.-H. Kim. *Network information theory*. Cambridge university press, 2011.

[31] E. Ersen and U. Sennur. Secrecy capacity of a class of broadcast channels with an eavesdropper. *EURASIP Journal on Wireless Communications and Networking, Special Issue Wireless Physical Layer Security*, March 2009.

[32] A. Ghorbel, M. Kobayashi, and S. Yang. Content delivery in erasure broadcast channels with cache and feedback. *IEEE Transactions on Information Theory*, 62(11):6407–6422, 2016.

[33] Z. Goldfeld, P. Cuff, and H. H. Permuter. Semantic-security capacity for wiretap channels of type II. *IEEE Transactions on Information Theory*, 62(7):3863–3879, 2016.

[34] P. K. Gopala, L. Lai, and H. El Gamal. On the secrecy capacity of fading channels. *IEEE Transactions on Information Theory*, 54(10):4687–4698, 2008.

[35] A. Gupta and S. Verdú. Operational duality between lossy compression and channel coding. *IEEE Transactions on Information Theory*, 57(6):3171–3179, 2011.

[36] J. Hachem, U. Niesen, and S. Diggavi. Degrees of freedom of cache-aided wireless interference networks. *arXiv preprint arXiv:1606.03175*, 2016.

[37] T. S. Han and S. Verdu. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, 1993.

[38] X. He, A. Khisti, and A. Yener. MIMO multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom. *IEEE Transactions on Information Theory*, 59(8):4733–4745, 2013.

[39] X. He and A. Yener. K-user interference channels: Achievable secrecy rate and degrees of freedom. In *IEEE Information Theory Workshop*, pages 336–340, June 2009.

[40] X. He and A. Yener. Cooperation with an untrusted relay: A secrecy perspective. *IEEE Transactions on Information Theory*, 56(8):3807–3827, 2010.

[41] X. He and A. Yener. The Gaussian many-to-one interference channel with confidential messages. *IEEE Transactions on Information Theory*, 57(5):2730–2745, 2011.

[42] X. He and A. Yener. The role of feedback in two-way secure communications. *IEEE Transactions on Information Theory*, 59(12):8115–8130, 2013.

[43] X. He and A. Yener. Strong secrecy and reliable Byzantine detection in the presence of an untrusted relay. *IEEE Transactions on Information Theory*, 59(1):177–192, 2013.

[44] X. He and A. Yener. MIMO wiretap channels with unknown and varying eavesdropper channel states. *IEEE Transactions on Information Theory*, 60(11):6844–6869, 2014.

[45] X. He and A. Yener. Providing secrecy with structured codes: Two-user Gaussian channels. *IEEE Transactions on Information Theory*, 60(4):2121–2138, 2014.

[46] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[47] A. Ibrahim, A. A. Zewail, and A. Yener. Device-to-device coded caching with heterogeneous cache sizes. In *IEEE International Conference on Communications*, May 2018.

[48] A. M. Ibrahim, A. A. Zewail, and A. Yener. Centralized coded caching with heterogeneous cache sizes. In *IEEE Wireless Communications and Networking Conference*, pages 1–6. IEEE, March 2017.

[49] A. M. Ibrahim, A. A. Zewail, and A. Yener. Optimization of heterogeneous caching systems with rate limited links. In *IEEE International Conference on Communications*, pages 1–6, May 2017.

[50] M. Ji, G. Caire, and A. F. Molisch. Fundamental limits of caching in wireless D2D networks. *IEEE Transactions on Information Theory*, 62(2):849–869, 2016.

[51] M. Ji, A. M. Tulino, J. Llorca, and G. Caire. Order-optimal rate of caching and coded multicasting with random demands. *IEEE Transactions on Information Theory*, 63(6):3923–3949, 2017.

[52] M. Ji, M. F. Wong, A. M. Tulino, J. Llorca, G. Caire, M. Effros, and M. Langberg. On the fundamental limits of caching in combination networks. In *IEEE International Workshop on Signal Processing Advances in Wireless Communications*, pages 695–699, June 2015.

[53] S. Kamel, M. Sarkiss, M. Wigger, and G. R.-B. Othman. Secrecy capacity-memory tradeoff of erasure broadcast channels. *arXiv preprint arXiv:1801.00606*, 2018.

[54] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. N. Diggavi. Hierarchical coded caching. *IEEE Transactions on Information Theory*, 62(6):3212–3229, 2016.

[55] A. Khandani, G. Bagherikaram, and A. Motahary. The secrecy capacity region of the Gaussian MIMO broadcast channel. *IEEE Transactions on Information Theory*, 59(5):2673–2682, 2013.

[56] A. Khisti. Interference alignment for the multiantenna compound wiretap channel. *IEEE Transactions on Information Theory*, 57(5):2976–2993, 2011.

[57] A. Khisti, A. Tchamkerten, and G. W. Wornell. Secure broadcasting over fading channels. *IEEE Transactions on Information Theory*, 54(6):2453–2469, 2008.

[58] A. Khisti and G. W. Wornell. Secure transmission with multiple antennas-Part II: The MIMOME wiretap channel. *IEEE Transactions on Information Theory*, 56(11):5515–5532, 2010.

[59] D. Kleinbock. Baker-Sprindzhuk conjectures for complex analytic manifolds. 2002. arXiv preprint math/0210369.

[60] L. Lai and H. El Gamal. The relay-eavesdropper channel: Cooperation for secrecy. *IEEE Transactions on Information Theory*, 54(9):4005–4019, 2008.

[61] S. Leung-Yan-Cheong and M. Hellman. The Gaussian wire-tap channel. *IEEE Transactions on Information Theory*, 24(4):451–456, 1978.

[62] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai. Compound wiretap channels. *EURASIP Journal on Wireless Communications and Networking, Special Issue Wireless Physical Layer Security*, 2009, Article ID 142374, 12 pages, 2009.

[63] Y. Liang, L. Lai, H. V. Poor, and S. Shamai. A broadcast approach for fading wiretap channels. *IEEE Transactions on Information Theory*, 60(2):842–858, 2014.

[64] Y. Liang and H. V. Poor. Multiple-access channels with confidential messages. *IEEE Transactions on Information Theory*, 54(3):976–1002, 2008.

[65] S. H. Lim, C.-Y. Wang, and M. Gastpar. Information-theoretic caching: The multi-user case. *IEEE Transactions on Information Theory*, 63(11):7018–7037, 2017.

[66] R. Liu, Y. Liang, H. V. Poor, and P. Spasojevic. Secure nested codes for type II wiretap channels. In *IEEE Information Theory Workshop*, pages 337–342, September 2007.

[67] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates. Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions. *IEEE Transactions on Information Theory*, 54(6):2493–2507, 2008.

[68] R. Liu and H. V. Poor. Secrecy capacity region of a multiple-antenna Gaussian broadcast channel with confidential messages. *IEEE Transactions on Information Theory*, 55(3):1235–1249, 2009.

[69] T. Liu and S. Shamai. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Transactions on Information Theory*, 55(6):2547–2553, 2009.

[70] H. D. Ly, T. Liu, and Y. Blankenship. Security embedding codes. *IEEE Transactions on Information Forensics and Security*, 7(1):148–159, 2012.

[71] M. A. Maddah-Ali. On the degrees of freedom of the compound MIMO broadcast channels with finite states. 2009. arXiv preprint arXiv:0909.5006.

[72] M. A. Maddah-Ali and U. Niesen. Fundamental limits of caching. *IEEE Transactions on Information Theory*, 60(5):2856–2867, 2014.

[73] M. A. Maddah-Ali and U. Niesen. Cache-aided interference channels. In *IEEE International Symposium on Information Theory*, pages 809–813, June 2015.

[74] M. A. Maddah-Ali and U. Niesen. Decentralized coded caching attains order-optimal memory-rate tradeoff. *IEEE/ACM Transactions on Networking*, 23(4):1029–1040, 2015.

[75] U. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 351–368, May 2000.

[76] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[77] U. M. Maurer. The strong secret key rate of discrete random triples. *Communications and Cryptography: Two Sides of One Tapestry. Norwell, MA, USA: Kluwer*, pages 271–285, 1994.

[78] A. S. Motahari, S. O. Gharan, M.-A. Maddah-Ali, and A. K. Khandani. Real interference alignment: Exploiting the potential of single antenna systems. *IEEE Transactions on Information Theory*, 60(8):4799–4810, 2014.

[79] P. Mukherjee, J. Xie, and S. Ulukus. Secure degrees of freedom of one-hop wireless networks with no eavesdropper CSIT. *IEEE Transactions on Information Theory*, 63(3):1898–1922, 2017.

[80] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr. Fundamental limits of cache-aided interference management. *IEEE Transactions on Information Theory*, 63(5):3092–3107, 2017.

[81] M. Nafea and A. Yener. Degrees of freedom of the single antenna Gaussian wiretap channel with a helper irrespective of the number of antennas at the eavesdropper. *IEEE GlobalSIP Symposium on Cyber-Security and Privacy*, December 2013.

[82] M. Nafea and A. Yener. How many antennas does a cooperative jammer need for achieving the degrees of freedom of multiple antenna Gaussian channels in the presence of an eavesdropper. *51st Annual Allerton Conference on Communication, Control, and Computing*, October 2013.

[83] M. Nafea and A. Yener. Secure degrees of freedom for the MIMO wiretap channel with a multiantenna cooperative jammer. *IEEE Information Theory Workshop*, pages 626–630, November 2014.

[84] M. Nafea and A. Yener. Secure degrees of freedom of $N \times N \times M$ wiretap channel with a K-antenna cooperative jammer. *IEEE International Conference on Communications*, June 2015.

[85] M. Nafea and A. Yener. Wiretap channel II with a noisy main channel. In *IEEE International Symposium on Information Theory*, pages 1159–1163, June 2015.

[86] M. Nafea and A. Yener. The multiple access wiretap channel II with a noisy main channel. In *IEEE International Symposium on Information Theory*, pages 2983–2987, July 2016.

[87] M. Nafea and A. Yener. A new multiple access wiretap channel model. In *IEEE Information Theory Workshop*, pages 349–353, September 2016.

[88] M. Nafea and A. Yener. A new wiretap channel model and its strong secrecy capacity. In *IEEE International Symposium on Information Theory*, pages 2804–2808, July 2016.

[89] M. Nafea and A. Yener. Secure degrees of freedom for the MIMO wire-tap channel with a multi-antenna cooperative jammer. *IEEE Transactions on Information Theory*, 63(11):7420–7441, 2017.

[90] M. Nafea and A. Yener. Generalizing multiple access wiretap and wiretap II channel models: Achievable rates and cost of strong secrecy. *Submitted to IEEE Transactions on Information Theory*, January 2018. arXiv preprint arXiv:1802.02131.

[91] M. Nafea and A. Yener. A new wiretap channel model and its strong secrecy capacity. *IEEE Transactions on Information Theory*, 64(3):2077–2092, 2018.

[92] C. Nair and A. El Gamal. The capacity region of a class of three-receiver broadcast channels with degraded message sets. *IEEE Transactions on Information Theory*, 55(10):4479–4493, 2009.

[93] U. Niesen and M. A. Maddah-Ali. Coded caching with nonuniform demands. *IEEE Transactions on Information Theory*, 63(2):1146–1158, 2017.

[94] F. Oggier and B. Hassibi. The secrecy capacity of the MIMO wiretap channel. *IEEE Transactions on Information Theory*, 57(8):4961–4972, 2011.

[95] Y. Oohama. Capacity theorems for relay channels with confidential messages. In *IEEE International Symposium on Information Theory*, pages 926–930, June 2007.

[96] L. Ozarow and A. D. Wyner. Wire-tap channel II. *Bell System Technical Journal*, 63(10):2135–2157, 1984.

[97] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen. Online coded caching. *IEEE/ACM Transactions on Networking*, 24(2):836–845, 2016.

[98] V. Ravindrakumar, P. Panda, N. Karamchandani, and V. M. Prabhakaran. Private coded caching. *IEEE Transactions on Information Forensics and Security*, 13(3):685–694, 2018.

[99] J. M. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Transactions on Information Theory*, 57(11):7377–7385, 2011.

[100] R. Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

[101] S. Sahraei and M. Gastpar. K users caching two files: An improved achievable rate. In *IEEE Conference on Information Sciences and Systems*, pages 620–624, March 2016.

[102] W. M. Schmidt. *Diophantine approximation*. Berlin, Germany: Springer-Verlag, 1980.

[103] A. Sengupta, R. Tandon, and T. C. Clancy. Fundamental limits of caching with secure delivery. *IEEE Transactions on Information Forensics and Security*, 10(2):355–370, 2015.

[104] S. Shafiee, N. Liu, and S. Ulukus. Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel. *IEEE Transactions on Information Theory*, 55(9):4033–4039, 2009.

[105] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[106] C. E. Shannon. Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4):656–715, 1949.

[107] O. Simeone and A. Yener. The cognitive multiple access wire-tap channel. In *IEEE Conference on Information Sciences and Systems*, pages 158–163, March 2009.

[108] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, 1973.

[109] V. Sprindzuk. More on Mahler's conjecture. *Doklady Akademii Nauk SSSR*, 155:54–56, 1964. (in Russian); English translation in Soviet Math. Dokl. 5, (1964), 361-363.

[110] V. Sprindzuk. On Mahler's conjecture. *Doklady Akademii Nauk SSSR*, 154:783–786, 1964. (in Russian); English translation in Soviet Math. Dokl. 5, (1964), 183-186.

[111] E. Tekin and A. Yener. Achievable rates for the general Gaussian multiple access wire-tap channel with collective secrecy. *44th Annual Allerton Conference On Communication, Control, and Computing*, September 2006.

[112] E. Tekin and A. Yener. The Gaussian multiple access wire-tap channel. *IEEE Transactions on Information Theory*, 54(12):5747–5755, 2008.

[113] E. Tekin and A. Yener. The general Gaussian multiple-access and two-way wire-tap channels: Achievable rates and cooperative jamming. *IEEE Transactions on Information Theory*, 54(6):2735–2751, 2008.

[114] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J.-M. Merolla. Applications of LDPC codes to the wiretap channel. *IEEE Transactions on Information Theory*, 53(8):2933–2945, 2007.

[115] C. Tian. Symmetry, demand types and outer bounds in caching systems. In *IEEE International Symposium on Information Theory*, pages 825–829, July 2016.

[116] M. Tomamichel. *Quantum information processing with finite resources: Mathematical foundations*, volume 5. Springer-Verlag, 2015. Springer Briefs in Mathematical Physics.

[117] K. Wan, D. Tuninetti, and P. Piantanida. On caching with more users than files. In *IEEE International Symposium on Information Theory*, pages 809–813, July 2016.

[118] V. K. Wei. Generalized Hamming weights for linear codes. *IEEE Transactions on Information Theory*, 37(5):1412–1418, 1991.

[119] M. M. Wilde. *Quantum information theory*. Cambridge University Press, 2013.

[120] A. D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975.

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

[122] J. Xie and S. Ulukus. Secure degrees of freedom of one-hop wireless networks. *IEEE Transactions on Information Theory*, 60(6):3359–3378, 2014.

[123] J. Xie and S. Ulukus. Secure degrees of freedom regions of multiple access and interference channels: The polytope structure. *Submitted to IEEE Transactions on Information Theory*, 2014. arXiv preprint arXiv:1404.7478.

[124] J. Xie and S. Ulukus. Secure degrees of freedom of K-user Gaussian interference channels: A unified view. *IEEE Transactions on Information Theory*, 61(5):2647–2661, 2015.

[125] M. H. Yassaee and M. R. Aref. Multiple access wiretap channels with strong secrecy. In *IEEE Information Theory Workshop*, pages 1–5, September 2010.

[126] M. H. Yassaee, M. R. Aref, and A. Gohari. Achievability proof via output statistics of random binning. *IEEE Transactions on Information Theory*, 60(11):6760–6786, 2014.

[127] A. Yener. New directions in information theoretic security: Benefits of bidirectional signaling. In *IEEE Information Theory Workshop*, pages 1–5, April 2015.

[128] A. Yener and S. Ulukus. Wireless physical-layer security: Lessons learned from information theory. *Proceedings of the IEEE*, 103(10):1814–1825, 2015.

[129] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr. The exact rate-memory tradeoff for caching with uncoded prefetching. *IEEE Transactions on Information Theory*, 64(2):1281–1296, 2018.

[130] A. A. Zewail and A. Yener. Fundamental limits of secure device-to-device coded caching. In *Asilomar Conference on Signals, Systems and Computers*, pages 1414–1418, November 2016.

[131] A. A. Zewail and A. Yener. Combination networks with or without secrecy constraints: The impact of caching relays. *IEEE Journal on Selected Areas in Communications*, 36(7):1–13, 2018.

[132] A. A. Zewail and A. Yener. The wiretap channel with a cache. In *IEEE International Symposium on Information Theory*, pages 1720–1724, June 2018.

[133] J. Zhang and P. Elia. Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback. *IEEE Transactions on Information Theory*, 63(5):3142–3160, 2017.

Vita

Mohamed Nafea received the B.Sc. degree in electrical engineering from Alexandria University, Alexandria, Egypt, in 2010; the M.Sc. degree in wireless communications from Wireless Intelligent Networks Center (WINC), Nile University, Giza, Egypt, in 2012, and the M.A. degree in mathematics from the Department of Mathematics, The Pennsylvania State University, University Park, PA, USA, in 2017. He is currently pursuing his Ph.D. degree in electrical engineering, and has been a Graduate Research Assistant with the Wireless Communications and Networking (WCAN) Laboratory, at The Pennsylvania State University, University Park, PA, USA, since 2012. His research interests include network information theory, information theoretic security, communication theory, statistical learning, mathematical logic and modeling theory.

Selected Publications

1. **M. Nafea** and A. Yener, MIMO Wiretap Channels, Information Theoretic Security and Privacy of Information Systems, Editors: R. Schaefer, H. Boche, A. Khisti, and V. Poor, Cambridge Univ. Press 2017.

2. **M. Nafea** and A. Yener, The Caching Broadcast Channel with a Wire and Cache Tapping Adversary of Type II, submitted to IEEE Transaction on Information Theory, Aug. 2018, pre-print arXiv:1808.02477

3. **M. Nafea** and A. Yener, Generalizing Multiple Access Wiretap and Wiretap II Channel Models: Achievable Rates and Cost of Strong Secrecy, Submitted to IEEE Transaction on Information Theory, Jan. 2018, pre-print arXiv:1802.02131

4. **M. Nafea** and A. Yener, A New Wiretap Channel Model and its Strong Secrecy Capacity, IEEE Transaction on Information Theory, 64(3), pp. 2077-2092, Mar. 2018.

5. **M. Nafea** and A. Yener, Secure Degrees of Freedom for the MIMO Wire-tap Channel with a Multi-antenna Cooperative Jammer, IEEE Transaction on Information Theory, 63(11), pp. 7420-7441, Nov. 2017.

6. **M. Nafea** and A. Yener, A New Broadcast Wiretap Channel Model, IEEE ISIT'17, Aachen, Germany, Jun. 2017.

7. **M. Nafea** and A. Yener, New Models for Interference and Broadcast Channels with Confidential Messages, IEEE ISIT'17, Aachen, Germany, Jun. 2017.

8. **M. Nafea** and A. Yener, Wiretap Channel II with a Noisy Main Channel, IEEE ISIT'15, Hong Kong, Jun. 2015.