Multiterminal Secrecy by Public Discussion

Multiterminal Secrecy by Public Discussion

Prakash Narayan
University of Maryland, College Park
prakash@umd.edu

Himanshu Tyagi
Indian Institute of Science, Bangalore
htyagi@ece.iisc.ernet.in

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

P. Narayan and H. Tyagi. *Multiterminal Secrecy by Public Discussion*. Foundations and Trends® in Communications and Information Theory, vol. 13, no. 2-3, pp. 129–275, 2016.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-186-3

Foundations and Trends® in Communications and Information Theory
Volume 13, Issue 2-3, 2016

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2016, Volume 13, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Foundations and Trends® in Communications and
Information Theory
Vol. 13, No. 2-3 (2016) 129–275

DOI: 10.1561/0100000072

now
the essence of knowledge

Multiterminal Secrecy by Public Discussion

Prakash Narayan
University of Maryland, College Park
prakash@umd.edu

Himanshu Tyagi
Indian Institute of Science, Bangalore
htyagi@ece.iisc.ernet.in

Contents

Abstract

This monograph describes principles of information theoretic secrecy generation by legitimate parties with public discussion in the presence of an eavesdropper. The parties are guaranteed secrecy in the form of independence from the eavesdropper's observation of the communication.

Part I develops basic technical tools for secrecy generation, many of which are potentially of independent interest beyond secrecy settings. Various information theoretic and cryptographic notions of secrecy are compared. Emphasis is placed on central themes of interactive communication and common randomness as well as on core methods of balanced coloring and leftover hash for extracting secret uniform randomness. Achievability and converse results are shown to emerge from "single shot" incarnations that serve to explain essential structure.

Part II applies the methods of Part I to secrecy generation in two settings: a multiterminal source model and a multiterminal channel model, in both of which the legitimate parties are afforded privileged access to correlated observations of which the eavesdropper has only partial knowledge. Characterizations of secret key capacity bring out inherent connections to the data compression concept of omniscience and, for a specialized source model, to a combinatorial problem of maximal spanning tree packing in a multigraph. Interactive common information is seen to govern the minimum rate of communication needed to achieve secret key capacity in the two-terminal source model. Furthermore, necessary and sufficient conditions are analyzed for the secure computation of a given function in the multiterminal source model.

Based largely on known recent results, this self-contained monograph also includes new formulations with associated new proofs. Supplementing each chapter in Part II are descriptions of several open problems.

P. Narayan and H. Tyagi. *Multiterminal Secrecy by Public Discussion*. Foundations and Trends® in Communications and Information Theory, vol. 13, no. 2-3, pp. 129–275, 2016.
DOI: 10.1561/0100000072.

1

Introduction

Information theoretic cryptography is founded on the principle of guaranteeing legitimate users provable data security from an adversary with unlimited computational power. Such an unconditional guarantee of security assures secrecy in the form of statistical independence (or near-independence) from the adversary's observations. This is accomplished, however, by giving the legitimate users a hearty leg up. By comparison, most existing cryptosystems for data security are based on the concept of computational complexity. The latter form of security rests on the infeasibility of existing mathematical and computational techniques in solving "hard" underlying computational problems, for instance, inverting specific functions.

Information theoretic perfect secrecy, introduced by Claude Shannon [72], constitutes the strongest definition of data security. It requires independence of a secret from the adversary's observations. A practically acceptable relaxation to near-independence ensures negligible information leakage to the adversary. Taken together with resources for the legitimate parties that lend them a decided advantage over the adversary, it leads to a rich theory raring for application.

In this monograph, we consider secrecy generation with public communication by multiple legitimate parties in two settings: a multiterminal source model and a multiterminal channel model. In both models, the legitimate parties are given privileged access to correlated observations that are only partially available to the eavesdropper. Our primary focus is on the former model.

The multiterminal source model consists of $m \geqslant 2$ terminals with prior access to correlated observations, and the means to communicate interactively among themselves over a public and noiseless broadcast medium of unlimited capacity. In the multiterminal channel model, a subset of k terminals, $1 \leqslant k \leqslant m-1$, govern the inputs of a noisy but secure transmission channel with the remaining $m-k$ terminals receiving the channel outputs. In between transmissions over the secure channel, all the terminals additionally can communicate among themselves publicly as in the source model. In both models, a passive adversary can eavesdrop on the communication among the terminals but cannot tamper with it, *i.e.*, the communication is authenticated. In the setting of each model, the primary goal is to generate a secret key of optimal length for all the m terminals under the requirement of information theoretic secrecy from the eavesdropped communication. We also consider secure function computation by trusted computing parties for a multiterminal source model under a similar secrecy constraint.

We do not address "wiretap channel" secrecy, launched in seminal works [98, 17], that entails secure transmission of messages over insecure channels which are wiretapped by an adversary; this is chronicled in [49, 19, 65]. Also, the classical multiterminal (information theoretically) secure function computation problem where the parties themselves are not trusted is not considered here; it has a substantial literature (cf. [46, 15, 95, 25, 58, 40, 96, 93, 94, 3, 87]).

This self-contained monograph is written in the language of information theory and aims to appeal as well to the cryptographer. To this end, we have strived to emphasize its following distinctive features: Comparison of various information theoretic and cryptographic notions of secrecy; bringing out of the significance – in distributed cooperative secrecy generation – of central themes of interactive commu-

nication and the common randomness or shared bits thereby created; and a presentation of "single-shot" results with a minimum of statistical assumptions (beyond knowledge of a joint distribution of pertinent random variables). Such a single-shot analysis, redolent of standard practice in cryptography, lies at the heart of information theoretic coding theorems. Also, by virtue of their lean and not mean but essential form, these results are of potential significance for models beyond those considered here.

Although this monograph largely treats known recent results, adherence to a consistency of themes has engendered also new formulations with associated new proofs. Our effort is to be viewed as a complement to the rich chapter on information theoretic security in [19] as well as jaunts in new directions.

Organization

Part I consists of Chapters 2 - 5 that deal with basic technical tools for secrecy generation. Many of these tools are potentially of independent interest beyond secrecy applications. Part II contains Chapters 6 - 9 that apply the methods of Part I to secrecy generation for the multiterminal source and channel models. In order to maintain a smooth flow of presentation, credits are provided only at the end of each chapter in a story of results a la [19]. The list of references is representative but not exhaustive. Supplementing the credits in Chapters 6 - 9 are descriptions of open problems.

Beginning with rudiments, Chapter 2 describes secrecy indices for a key with their operational meanings, as well as secrecy indices for a message and relationships among the latter. Turning to basic methods, Chapter 3 deals with the central concepts of interactive communication among multiple terminals and the common randomness generated thereby; a fundamental structural property of interactive communication and single-shot converse upper bounds for the ensuing common randomness are derived. The concept of a secret key is introduced formally in Chapter 4, and suitable upper bounds on its length are obtained by means of two different converse techniques: bounding the entropy of common randomness and through the error exponent of conditional independence hypothesis testing. The notion of shared in-

formation is introduced as an upper bound for the length of a secret key; shared information has a potential role as a measure of mutual dependence among $m \geqslant 2$ random variables. Chapter 5 describes two achievability approaches – balanced coloring and leftover hash – for extracting uniform randomness from a given random variable with near independence from another random variable. These methods pave the way for extracting a secret key from common randomness by means of public communication.

Chapter 6 addresses secret key generation for the multiterminal source model in which each terminal observes one component of a discrete memoryless multiple source. A single-letter characterization of secret key capacity is obtained on the strength of an inherent link to a data compression problem of "omniscience" without secrecy constraints. This capacity is seen as being equal to shared information, thereby imbuing the latter with an operational meaning. Secret key generation for a special "pairwise independent network" model reveals connections to a combinatorial problem of maximal packing of spanning trees in a multigraph. For the two-terminal source model, the minimum rate of interactive communication needed to generate an optimal rate secret key is addressed in Chapter 7, and is shown to be related to a new interactive variant of Wyner's common information. Chapter 8 examines conditions that enable a special form of secrecy generation for the multiterminal source model: secure function computation in which multiple terminals compute a given function of the collective data at the terminals using public communication that does not reveal the function value. The closing Chapter 9 studies secret key generation for the multiterminal channel model in which one subset of the terminals are connected to the remaining terminals by a secure discrete memoryless multiaccess channel. While a general single-letter characterization of secret key capacity remains open, in the special case of a channel with a single output terminal, interesting connections are shown between secrecy capacity and the transmission capacity region of the multiple access channel with and without feedback.

A note: All the random variables (rvs) throughout this monograph take values in finite sets, with known joint probability mass functions (pmfs). Probabilities of events involving rvs X, Y will be denoted by P_{XY}, $P_{X|Y}$, *etc.*, and by a general $\mathbb{P}$ when appropriate.

Part I

Basic Tools

2

Notions of Secrecy and Their Relationships

We set the stage with a description of measures of information theoretic secrecy and their properties. Variational secrecy and divergence secrecy are defined in §2.1. Corresponding secrecy indices for a key and a useful relationship between them are described in §2.2. The connection, which sandwiches divergence secrecy index between functions of variational secrecy index, will enable us in later chapters to establish secrecy results that involve switching back and forth between the two indices. The operational secrecy of a key in terms of its resistance to querying by an eavesdropper possessing related side information is also explained. In a different setting of message security, relationships among various message secrecy indices: variational or divergence, semantic and distinguishing, are considered in §2.3. It is established that these indices are, in effect, in agreement with each other. The use of a secret key in encryption by a one-time pad is shown in §2.4.

2.1 Information theoretic secrecy

Let the $\mathcal{K}$-valued rv K denote a secret for legitimate parties, and let the $\mathcal{Z}$-valued rv Z denote the observation of an eavesdropper. A secret

K is *information theoretically* secure if K and Z are "almost independent." Formally, it is required that the joint pmf P_{KZ} of K, Z be close to $P_K \times P_Z$. Two different measures of closeness are used, yielding corresponding notions of secrecy.

1. **Variational secrecy.** Given $\epsilon \geqslant 0$, an rv K is ϵ-secure from the rv Z in variational distance if
$$\|P_{KZ} - P_K \times P_Z\| \leqslant \epsilon,$$
where $\|P - Q\|$ denotes the variational distance between pmfs P and Q on $\mathcal{X}$, given by
$$\begin{aligned}\|P - Q\| &= \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| \\ &= \max_{A\subseteq\mathcal{X}} P(A) - Q(A).\end{aligned}$$

2. **Divergence secrecy.** An rv K is ϵ-secure from an rv Z in divergence if
$$D\big(P_{KZ}\big\|P_K \times P_Z\big) \leqslant \epsilon,$$
where $D\big(P\big\|Q\big)$ denotes the Kullback-Leibler divergence between the pmfs P and Q on $\mathcal{X}$, given by[1]
$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}.$$
Note that the expression on the left-side of inequality above is simply the mutual information $I(K \wedge Z)$ between K and Z.

As we shall see below, divergence secrecy is more stringent than variational secrecy. On the other hand, variational secrecy provides ready operational meanings to be seen in this and subsequent chapters. The following relation plays the role of a "chain rule" for variational distance.

Lemma 2.1. For two pmfs P_{U^m} and Q_{U^m} on a finite set $\mathcal{U}_1 \times ... \times \mathcal{U}_m$, it holds that
$$\|P_{U^m} - Q_{U^m}\| \leqslant \sum_{i=1}^{m} \left\|P_{U^i} - P_{U^{i-1}}Q_{U_i|U^{i-1}}\right\|.$$

[1] The standard convention $0\log(0/0) = 0$ is followed.

Proof. For $1 \leqslant i \leqslant m-1$, denote by $P(i)$ the pmf $P_{U^i} Q_{U_{i+1}^m|U^i}$, and let $P(0) = Q_{U^m}$ and $P(m) = P_{U^m}$. Then,

$$\begin{aligned}
\|P_{U^m} - Q_{U^m}\| &= \left\| \sum_{i=1}^{m} P(i) - \sum_{i=1}^{m} P(i-1) \right\| \\
&\leqslant \sum_{i=1}^{m} \|P(i) - P(i-1)\| \\
&= \sum_{i=1}^{m} \left\| P_{U^i} Q_{U_{i+1}^m|U^i} - P_{U^{i-1}} Q_{U_i^m|U^{i-1}} \right\| \\
&= \sum_{i=1}^{m} \left\| P_{U^i} - P_{U^{i-1}} Q_{U_i|U^{i-1}} \right\| .
\end{aligned}$$

□

2.2 Secrecy of a key

In the secrecy requirements above, we allow the secret K to have an arbitrary pmf. In applications involving a secret key, it is required that the secret key K additionally must be close to a uniform rv. Indeed, the security of many cryptographic primitives relies on the uniformity of the underlying secret key.[2]

2.2.1 Secrecy indices for a key

We define two secrecy indices that incorporate a uniformity requirement for K into the notions of variational secrecy and divergence secrecy.

Definition 2.2. The *variational secrecy index* for K given Z is

$$\sigma_{\mathtt{var}}(K; Z) \triangleq \left\| P_{KZ} - P^{\mathcal{K}}_{\mathtt{unif}} \times P_Z \right\| ,$$

where $P^{\mathcal{K}}_{\mathtt{unif}}$ is the uniform pmf on the range $\mathcal{K}$ of K. The *divergence*

[2] If only computational secrecy is required, it suffices to use a pseudorandom uniform key.

secrecy index for K given Z is

$$\begin{aligned}\sigma_{\mathtt{div}}(K;Z) &\triangleq D\big(P_{KZ}\big\|P^{\mathcal{K}}_{\mathtt{unif}}\times P_Z\big)\\ &= \log|\mathcal{K}| - H(K) + I(K\wedge Z)\\ &= \log|\mathcal{K}| - H(K\mid Z).\end{aligned}$$

In fact, the two secrecy indices are closely related; the following technical lemma is used to relate the two indices.

Lemma 2.3. For pmfs P and Q on a finite set $\mathcal{X}$,

$$|H(P)-H(Q)| \leqslant \|P-Q\|\log(|\mathcal{X}|-1) + h\left(\|P-Q\|\right),$$

where $h(x) = -x\log x - (1-x)\log(1-x)$, $0\leqslant x\leqslant 1$, is the binary entropy function.

Proof. For any pmf P_{XY} on $\mathcal{X}\times\mathcal{X}$ with $P_X = P$, $P_Y = Q$,

$$\begin{aligned}H(P) - H(Q) &= H(X) - H(Y)\\ &= H(X|Y) - H(Y|X)\end{aligned}$$

so that

$$\begin{aligned}|H(P)-H(Q)| &= |H(X\mid Y) - H(Y\mid X)|\\ &\leqslant \max\{H(X\mid Y), H(Y\mid X)\}.\\ &\leqslant P_{XY}\left(X\neq Y\right)\log(|\mathcal{X}|-1) + h(P_{XY}\left(X\neq Y\right)),\end{aligned}$$

where the last step uses Fano's inequality. The claim follows by choosing P_{XY} to be a a maximal coupling of P and Q, *i.e.*, with $P_X = P$, $P_Y = Q$ and $P_{XY}(X=Y)$ being maximal, whence $P_{XY}\left(X\neq Y\right) = \|P-Q\|$. □

Lemma 2.4. The secrecy indices $\sigma_{\mathtt{var}}(K;Z)$ and $\sigma_{\mathtt{div}}(K;Z)$ satisfy

$$\begin{aligned}(2\log e)\,\sigma_{\mathtt{var}}(K;Z)^2 &\leqslant \sigma_{\mathtt{div}}(K;Z)\\ &\leqslant \sigma_{\mathtt{var}}(K;Z)\log(|\mathcal{K}|-1) + h\left(\sigma_{\mathtt{var}}(K;Z)\right)\end{aligned}$$

Proof. The first claim follows from Pinsker's inequality

$$2\log e\,\|P-Q\|^2 \leqslant D\big(P\|Q\big),$$

with $P = P_{KZ}$ and $Q = P^{\mathcal{K}}_{\mathtt{unif}} \times P_Z$. For the second, we have

$$\begin{aligned}
&\sigma_{\mathtt{div}}(K;Z) \\
&\quad = \log \mathcal{K} - H(K \mid Z) \\
&\quad = \sum_z P_Z(z) \left[H\left(P^{\mathcal{K}}_{\mathtt{unif}}\right) - H\left(P_{K|Z=z}\right) \right] \\
&\quad \leqslant \sum_z P_Z(z) \left[\left\| P_{K|Z=z} - P^{\mathcal{K}}_{\mathtt{unif}} \right\| \log(|\mathcal{K}| - 1) \right. \\
&\qquad\qquad \left. + h\left(\left\| P_{K|Z=z} - P^{\mathcal{K}}_{\mathtt{unif}} \right\| \right) \right],
\end{aligned}$$

where the previous step is by Lemma 2.3. The second claim follows since $h(\cdot)$ is concave. □

2.2.2 Operational notion of secrecy

Often it is required that we "extract" a secret key K as a function of another rv U, with the former being ϵ-secure from Z. This is captured by the following definition of secrecy.

Definition 2.5. Given $k \in \mathbb{N}$ and $0 \leqslant \epsilon < 1$, a $\mathcal{U}$-valued rv U is (k, ϵ)-secure from Z if there exists a mapping $\kappa \colon \mathcal{U} \to \{1, ..., k\}$ such that the rv $K = \kappa(U)$ satisfies $\sigma_{\mathtt{var}}(K; Z) \leqslant \epsilon$.

While the definition of secrecy above is mathematically appealing, its operational significance is not apparent immediately. In this section, we shall present a heuristic definition of secrecy and establish its equivalence with the (k, ϵ) secrecy of Definition 2.5.

Specifically, an eavesdropper observing Z wishes to ascertain the value of the rv U by asking questions of the form "Is $U = u$?" with yes-no answers. A *query strategy* q for U given $Z = z$ is a bijection $q(\cdot|z) : \mathcal{U} \to \{1, ..., |\mathcal{U}|\}$, where the querier, upon observing $Z = z$, asks the question "Is $U = u$?" in the $q(u|z)^{\text{th}}$ query. A natural secrecy requirement is to force the task of the querying eavesdropper to be as onerous as possible.

Definition 2.6. Given $0 \leqslant \epsilon < 1$ and $\lambda > 0$, a rv U is (λ, ϵ)-secure from a querying eavesdropper with Z if for every query strategy q for U given Z,

$$\mathbb{P}\left(q(U|Z) > \lambda\right) \geqslant 1 - \epsilon,$$

i.e., with probability exceeding $1-\epsilon$, the querying eavesdropper must make more than λ queries to ascertain the value of U.

The (k,ϵ) secrecy of Definition 2.5 is, in effect, equivalent to the (λ,ϵ) secrecy from a querying eavesdropper, in that both secrecy criteria are tantamount to requiring *large probability upper bounds on* $P_{U|Z}$. The next two lemmas bring out this correspondence.

Lemma 2.7. Let $0 \leqslant \epsilon < 1$ and $\eta > 0$ with $\epsilon + \eta < 1$ be given. If U is (k,ϵ)-secure from Z *a la* Definition 2.5, then

$$\mathbb{P}\left(P_{U|Z}\left(U \mid Z\right) \leqslant (\eta k)^{-1}\right) \geqslant 1-\epsilon-\eta,$$

where (U,Z) has pmf P_{UZ}. Furthermore, if

$$\mathbb{P}\left(P_{U|Z}\left(U \mid Z\right) \leqslant k^{-1}\right) \geqslant 1-\epsilon, \tag{2.1}$$

then U is $(\eta k, \epsilon+\eta)$-secure from Z.

Proof. To show the first claim, suppose there exists a mapping $\kappa : \mathcal{U} \to \{1,...,k\}$ such that $K = \kappa(U)$ satisfies

$$\sigma_{\mathtt{var}}(K;Z) \leqslant \epsilon. \tag{2.2}$$

Since for every u and z with $\kappa(u) = l$,

$$P_{K|Z}\left(l \mid z\right) \geqslant P_{U|Z}\left(u \mid z\right),$$

it suffices to show that

$$P_{KZ}\left(\left\{(l,z)\colon P_{K|Z}\left(l \mid z\right) \leqslant (\eta k)^{-1}\right\}\right) \geqslant 1-\epsilon-\eta.$$

To this end, denoting by $\mathcal{A}$ the complement of the set in $\{\cdot\}$ above, we have by (2.2) that

$$P_{KZ}\left(\mathcal{A}\right) - \left(P_{\mathtt{unif}}^{[1,k]} \times P_Z\right)\left(\mathcal{A}\right) \leqslant \epsilon$$

whereby

$$P_{KZ}\left(\mathcal{A}\right) \leqslant \epsilon + \sum_z P_Z\left(z\right)\frac{|\mathcal{A}_z|}{k}, \tag{2.3}$$

where $\mathcal{A}_z = \{v\colon (v,z) \in \mathcal{A}\}$. Since by the definition of the set $\mathcal{A}$, $|\mathcal{A}_z| \leqslant \eta k$, it follows by (2.3) that

$$P_{KZ}(\mathcal{A}) \leqslant \epsilon + \eta,$$

which completes the proof of the first claim.

To prove the second claim, assuming (2.1), we need to show the existence of a mapping $\kappa\colon \mathcal{U} \to \{1,...,\eta k\}$ with $\sigma_{\mathtt{var}}(\kappa(U);Z) \leqslant \epsilon + \eta$. In fact, all our secrecy generation schemes in this monograph rely on the existence of such mappings. We defer the proof of this part to Chapter 5, where general results establishing the existence of such mappings will be developed. Specifically, we refer to Lemma 5.17. □

Lemma 2.8. Given $0 \leqslant \epsilon < 1$ and $\lambda > 0$, if U is (λ,ϵ)-secure from Z *a la* Definition 2.6, then

$$\mathbb{P}\left(P_{U|Z}(U \mid Z) \leqslant \lambda^{-1}\right) > 1 - \epsilon. \tag{2.4}$$

Furthermore, if (2.4) holds, then U is $(\eta\lambda, \epsilon+\eta)$-secure from Z for every $0 < \eta < 1 - \epsilon$.

Proof. Fix $0 \leqslant \epsilon < 1$ and $\lambda > 0$, and suppose that (2.4) does not hold, *i.e.*,

$$\mathbb{P}\left(\left\{(u,z)\colon P_{U|Z}(u \mid z) > \lambda^{-1}\right\}\right) > \epsilon. \tag{2.5}$$

Letting $\mathcal{U}_z \triangleq \{u\colon P_{U|Z}(u \mid z) > \lambda^{-1}\}$ for each z, we have

$$\frac{|\mathcal{U}_z|}{\lambda} \leqslant \sum_{u\in\mathcal{U}_z} P_{U|Z}(u \mid z) \leqslant 1. \tag{2.6}$$

Consider a query strategy q that upon observing $Z = z$ first makes queries "Is $U = u$?" for $u \in \mathcal{U}_z$ (in any order) and then, for $u \notin \mathcal{U}_z$ (in any order). It follows from (2.5) and (2.6) that

$$\begin{aligned}\mathbb{P}(q(U|Z) \leqslant \lambda) &\geqslant \sum_z P_Z(z)\, P_{U|Z}(\mathcal{U}_z \mid z)\\ &= \mathbb{P}\left(\left\{(u,z)\colon P_{U|Z}(u \mid z) > \lambda^{-1}\right\}\right)\\ &> \epsilon,\end{aligned}$$

which is the same as $\mathbb{P}\left(q(U|Z) > \lambda\right) < 1 - \epsilon$, so that U is not (λ, ϵ)-secure from Z.

Conversely, suppose that (2.4) holds but there exists a query strategy q with

$$\mathbb{P}\left(q(U \mid Z) > \eta\lambda\right) < 1 - \epsilon - \eta.$$

Then, by (2.4),

$$\begin{aligned}\eta &< \mathbb{P}\left(\left\{(u,z)\colon P_{U|Z}\left(u \mid z\right) \leqslant \lambda^{-1},\, q(u|z) \leqslant \eta\lambda\right\}\right)\\ &\leqslant \sum_z P_Z\left(z\right) \sum_{u\colon\, q(u|z)\leqslant\eta\lambda} \lambda^{-1}.\end{aligned}$$

Note that $q(\cdot|z)$ is a bijection for each fixed z, so that

$$\left|\{u\colon q(u|z) \leqslant \eta\lambda\}\right| \leqslant \eta\lambda.$$

Hence,

$$\eta < \sum_z P_Z\left(z\right) \sum_{u\colon\, q(u|z)\leqslant\eta\lambda} \lambda^{-1} \leqslant \eta,$$

which is inconsistent. Thus, for every query strategy q for U given Z, it must hold that $\mathbb{P}\left(q(U \mid Z) > \eta\lambda\right) \geqslant 1 - \epsilon - \eta$. □

2.3 Secrecy of a message

In a different setting from that of a secret key, a secret K takes the form of a secret message M that must be concealed from an eavesdropper observing Z. For each transmitted message $M = m$ in $\mathcal{M}$, the eavesdropper's observation Z has (conditional) pmf $P_{Z|M=m}$. Note that the conditional pmf $P_{Z|M}$, denoted by a stochastic matrix $W\colon \mathcal{M} \to \mathcal{Z}$, is determined by an encryption or a transmission protocol and is assumed to be known to the eavesdropper. In contrast with a secret key, a secret message M has a pmf P_M that is allowed to vary with application or the eavesdropper's prior belief. Accordingly, the secrecy indices below guarantee security of the message for all P_M.

2.3.1 Secrecy indices for a message

The secrecy indices for a message, $\sigma_{\mathtt{var}}(\mathcal{M}; W)$ and $\sigma_{\mathtt{div}}(\mathcal{M}; W)$, are natural counterparts of those for a secret key in §2.2. For a random

message M with pmf P_M, the rv Z with $P_{Z|M} = W$ denotes the eavesdropper's observation.

Definition 2.9. For a message set $\mathcal{M}$ and eavesdropper's channel W, the *variational secrecy index* is

$$\sigma_{\mathtt{var}}(\mathcal{M}; W) \triangleq \max_{P_M} \|P_{MZ} - P_M \times P_Z\|,$$

and the *divergence secrecy index* is

$$\begin{aligned}\sigma_{\mathtt{div}}(\mathcal{M}; W) &\triangleq \max_{P_M} D\left(P_{MZ} \| P_M \times P_Z\right) \\ &= \max_{P_M} I(P_M, W). \end{aligned} \tag{2.7}$$

Traditionally, in problems of secure message transmission $P_M = P^{\mathcal{M}}_{\mathtt{unif}}$ is assumed, and $Z = Z^n = (Z_1, ..., Z_n)$ denotes the n-length observation of the eavesdropper. The *strong and weak secrecy indices*, respectively, are given by $I(M \wedge Z^n)$ and $(1/n)I(M \wedge Z^n)$ and correspond to fixing $P_M = P^{\mathcal{M}}_{\mathtt{unif}}$ and $Z = Z^n$ in (2.7).

A more stringent notion, often adopted in cryptography, is semantic secrecy which requires that the eavesdropper does no better than random guessing in inferring any (nontrivial) function f of the message and for any P_M.

Definition 2.10. The *semantic secrecy index* for $\mathcal{M}$ and eavesdropper's channel W is

$$\sigma_{\mathtt{sem}}(\mathcal{M}; W) \triangleq \min_{G} \max_{P_M, f, \hat{f}} \mathbb{P}\left(\hat{f}(Z) = f(M)\right) - \mathbb{P}\left(\hat{f}(G) = f(M)\right),$$

where the estimate[3] $\hat{f}(Z)$ of $f(M)$ can depend on P_M and f, and the rv G is independent of (M, Z) and does not depend on P_M or f, *i.e.*, $\hat{f}(G)$ is tantamount a random guess[4].

A relaxation of semantic secrecy considers only pmfs P_M of support-size 2. Termed distinguishing secrecy, the latter is seen in Lemma 2.14 to be nonetheless equivalent to the former, while affording a reduced search space for P_M. In contrast, this equivalence does not hold under computational secrecy, in general.

[3] There is no loss of generality in a restriction to deterministic estimators.

[4] Allowing G to depend additionally on P_M would result in a weaker notion of secrecy.

Definition 2.11. The *distinguishing secrecy index* for $\mathcal{M}$ and W is

$$\sigma_{\mathtt{dis}}(\mathcal{M};W) \triangleq \max_{m_0,m_1\in\mathcal{M}} \left(\max_{\hat{b}} \mathbb{P}\left(\hat{b}\left(Z_{m_B}\right)=B\right) - \frac{1}{2}\right),$$

where B is uniformly distributed rv on $\{0,1\}$, Z_{m_B} denotes the eavesdropper's observation corresponding to the random message m_B, and the estimate $\hat{b}\left(Z_{m_B}\right)$ of B depends on $\{m_0,m_1\}$.

The following alternative expression for $\sigma_{\mathtt{dis}}(\mathcal{M};W)$ is useful.

Lemma 2.12.

$$\sigma_{\mathtt{dis}}(\mathcal{M};W) = \max_{m_0,m_1\in\mathcal{M}} \frac{1}{2}\left\|P_{Z_{m_0}} - P_{Z_{m_1}}\right\|,$$

where

$$P_{Z_m} = W(\cdot|m), \quad m\in\mathcal{M}.$$

Proof. We show for each $\{m_0,m_1\}$ that

$$\max_{\hat{b}} \mathbb{P}\left(\hat{b}\left(Z_{m_B}\right)=B\right) - \frac{1}{2} = \frac{1}{2}\left\|P_{Z_{m_0}} - P_{Z_{m_1}}\right\|.$$

For each $\hat{b}$, denoting $\mathcal{Z}_0 = \hat{b}^{-1}(0)$, we have

$$\begin{aligned}\mathbb{P}\left(\hat{b}\left(Z_{m_B}\right)=B\right) &= \frac{1}{2}\left[\mathbb{P}\left(\hat{b}\left(Z_{m_0}\right)=0\right) + \mathbb{P}\left(\hat{b}\left(Z_{m_1}\right)=1\right)\right]\\ &= \frac{1}{2}\left[W(\mathcal{Z}_0|m_0) + W(\mathcal{Z}_0^c|m_1)\right]\\ &= \frac{1}{2}\left[W(\mathcal{Z}_0|m_0) - W(\mathcal{Z}_0|m_1)\right] + \frac{1}{2}.\end{aligned}$$

The proof is completed upon noting that

$$\max_{\mathcal{Z}_0}\left[W(\mathcal{Z}_0|m_0) - W(\mathcal{Z}_0|m_1)\right] = \left\|P_{Z_{m_0}} - P_{Z_{m_1}}\right\|.$$

□

2.3.2 Relationships between secrecy indices for a message

Surprisingly the three secrecy indices above are, in essence, equivalent.

First, analogous to Lemma 2.4, the following relationship holds between $\sigma_{\mathtt{var}}(\mathcal{M};W)$ and $\sigma_{\mathtt{div}}(\mathcal{M};W)$.

Lemma 2.13.

$$\begin{aligned}&\frac{\log e}{2}\,\sigma_{\mathtt{var}}(\mathcal{M};W)^2\\&\leqslant \sigma_{\mathtt{div}}(\mathcal{M};W)\\&\leqslant \sigma_{\mathtt{var}}(\mathcal{M};W)\log(|\mathcal{M}|-1)+h\left(\min\{\sigma_{\mathtt{var}}(\mathcal{M};W),1/2\}\right).\end{aligned}$$

Next, $\sigma_{\mathtt{sem}}(\mathcal{M};W)$ and $\sigma_{\mathtt{dis}}(\mathcal{M};W)$ are equivalent up to a multiplicative constant.

Lemma 2.14.

$$\sigma_{\mathtt{dis}}(\mathcal{M};W)\leqslant\sigma_{\mathtt{sem}}(\mathcal{M};W)\leqslant 2\sigma_{\mathtt{dis}}(\mathcal{M};W).$$

Proof. The first inequality obtains from the definition of $\sigma_{\mathtt{sem}}$ upon fixing $f(m)=m$, $m\in\mathcal{M}$, and choosing $P_M=P_{\mathtt{unif}}^{\{m_0,m_1\}}$ for every pair of messages in $\mathcal{M}$.

For the second inequality, given P_M, f, and $\hat{f}$, let $G=Z_{m_0}$ for an arbitrary but fixed message m_0 in $\mathcal{M}$. Then, there exists $m_1\in\mathcal{M}$ such that

$$\begin{aligned}&\mathbb{P}\left(\hat{f}(Z_M)=f(M)\right)-\mathbb{P}\left(\hat{f}(G)=f(M)\right)\\&=\sum_m P_M(m)\left[\mathbb{P}\left(\hat{f}(Z_m)=f(m)\right)-\mathbb{P}\left(\hat{f}(Z_{m_0})=f(m)\right)\right]\\&\leqslant\mathbb{P}\left(\hat{f}(Z_{m_1})=f(m_1)\right)-\mathbb{P}\left(\hat{f}(Z_{m_0})=f(m_1)\right).\end{aligned}\tag{2.8}$$

Furthermore, for $\hat{b}(z)=\mathbb{1}\left(\hat{f}(z)=f(m_1)\right)$, we have

$$\begin{aligned}&\mathbb{P}\left(\hat{b}(Z_{m_B})=B\right)\\&=\frac{1}{2}\left[\mathbb{P}\left(\hat{b}(Z_{m_1})=1\right)+\mathbb{P}\left(\hat{b}(Z_{m_0})=0\right)\right]\\&=\frac{1}{2}\left[\mathbb{P}\left(\hat{f}(Z_{m_1})=f(m_1)\right)+\mathbb{P}\left(\hat{f}(Z_{m_0})\neq f(m_1)\right)\right]\\&=\frac{1}{2}+\frac{1}{2}\left[\mathbb{P}\left(\hat{f}(Z_{m_1})=f(m_1)\right)-\mathbb{P}\left(\hat{f}(Z_{m_0})=f(m_1)\right)\right]\\&\geqslant\frac{1}{2}+\frac{1}{2}\left[\mathbb{P}\left(\hat{f}(Z_M)=f(M)\right)-\mathbb{P}\left(\hat{f}(G)=f(M)\right)\right],\end{aligned}$$

where the last inequality is by (2.8), and the claim follows since P_M, f, and $\hat{f}$ are arbitrary. □

Finally, $\sigma_{\mathtt{dis}}(\mathcal{M}; W)$ and $\sigma_{\mathtt{var}}(\mathcal{M}; W)$ are also equivalent up to a multiplicative constant.

Lemma 2.15.

$$\sigma_{\mathtt{dis}}(\mathcal{M}; W) \leqslant \sigma_{\mathtt{var}}(\mathcal{M}; W) \leqslant 2\sigma_{\mathtt{dis}}(\mathcal{M}; W).$$

Proof. We shall use the alternative expression for $\sigma_{\mathtt{dis}}(\mathcal{M}; W)$ given in Lemma 2.12. For $m_0, m_1 \in \mathcal{M}$ and $P_M = P_{\mathtt{unif}}^{\{m_0, m_1\}}$, it holds that

$$\begin{aligned}
\sigma_{\mathtt{var}}(\mathcal{M}; W) &\geqslant \left\| P_{MZ} - P_M \times P_Z \right\| \\
&= \frac{1}{2} \left\| P_{Z_{m_0}} - P_Z \right\| + \frac{1}{2} \left\| P_{Z_{m_1}} - P_Z \right\| \\
&\geqslant \frac{1}{2} \left\| P_{Z_{m_0}} - P_{Z_{m_1}} \right\|.
\end{aligned}$$

Since $m_0, m_1 \in \mathcal{M}$ are arbitrary, the first inequality follows. For the second inequality, observe that for each P_M,

$$\begin{aligned}
\left\| P_{MZ} - P_M \times P_Z \right\| &= \sum_{m \in \mathcal{M}} P_M(m) \left\| P_{Z_m} - P_Z \right\| \\
&= \sum_{m \in \mathcal{M}} P_M(m) \left\| P_{Z_m} - \sum_{m' \in \mathcal{M}} P_M\left(m'\right) P_{Z_{m'}} \right\| \\
&\leqslant \sum_{m, m' \in \mathcal{M}} P_M(m) P_M\left(m'\right) \left\| P_{Z_m} - P_{Z_{m'}} \right\| \\
&\leqslant \max_{m, m' \in \mathcal{M}} \left\| P_{Z_m} - P_{Z_{m'}} \right\| \\
&= 2\sigma_{\mathtt{dis}}(\mathcal{M}; Z).
\end{aligned}$$

□

2.4 Secure transmission with one-time pad

We close this chapter with a composability result showing that secrecy under $\sigma_{\mathtt{var}}$ and $\sigma_{\mathtt{div}}$ is preserved for message transmission using a one-time pad with a secret key. Such a result serves to guarantee the overall secrecy for an encryption system when its components are individually secure.

Let $(\mathbb{G}, +)$ be a commutative group. Let K be a $\mathbb{G}$-valued secret key with $\sigma_{\mathtt{var}}(K; Z) \leqslant \epsilon$ or $\sigma_{\mathtt{div}}(K; Z) \leqslant \epsilon$, where Z is the eavesdropper's

side information. Let M be a message with values in $\mathbb{G}$ and pmf P_M, and suppose that M is independent of (K, Z).

A transmitter "encrypts" the message M as $M + K$ and sends it over an insecure public channel which is observed by a receiver as well as by the eavesdropper. The receiver, also knowing K, can decrypt the message by subtracting K from $M + K$. However, the message M remains concealed from the eavesdropper's observations of $(Z, M + K)$. Specifically, for a random message M with pmf P_M, the eavesdropper observes $(Z, M + K)$ with $P_{Z,M+K|M} = \tilde{W}$ where the stochastic matrix $\tilde{W} : \mathbb{G} \to \mathcal{Z} \times \mathbb{G}$ is

$$\begin{aligned}\tilde{W}(z, l|m) &= P_{Z,M+K|M}\,(z, l|m) \\ &= P_{Z|M}\,(z|m)\,P_{M+K|M,Z}\,(l|m, z) \\ &= P_Z\,(z)\,P_{K|Z}\,(l - m|z)\,,\end{aligned}$$

by the assumed independence of M and (K, Z).

Proposition 2.16. Let M be independent of (K, Z). Then,

$$\sigma_{\mathtt{var}}(K; Z) \leqslant \epsilon \text{ and } \sigma_{\mathtt{div}}(K; Z) \leqslant \epsilon$$

imply, respectively, that

$$\sigma_{\mathtt{var}}(\mathbb{G}; \tilde{W}) \leqslant 2\epsilon \text{ and } \sigma_{\mathtt{div}}(\mathbb{G}; \tilde{W}) \leqslant \epsilon.$$

Proof. We show that

$$\sigma_{\mathtt{var}}(\mathbb{G}; \tilde{W}) \leqslant 2\sigma_{\mathtt{var}}(K; Z) \text{ and } \sigma_{\mathtt{div}}(\mathbb{G}; \tilde{W}) \leqslant \sigma_{\mathtt{div}}(K; Z).$$

First, for every pmf P_M,

$$\begin{aligned}
&\|P_{M,Z,M+K} - P_M \times P_{Z,M+K}\| \\
&\quad = \frac{1}{2} \sum_{m\in\mathbb{G}} P_M(m) \sum_{z\in\mathcal{Z}} P_Z(z) \sum_{l\in\mathbb{G}} \left|P_{K|Z}(l-m|z) - P_{M+K|Z}(l|z)\right| \\
&\quad \leqslant \frac{1}{2} \sum_{m\in\mathbb{G}} P_M(m) \sum_{z\in\mathcal{Z}} P_Z(z) \sum_{l\in\mathbb{G}} \left[\left|P_{K|Z}(l-m|z) - \frac{1}{|\mathcal{K}|}\right| \right. \\
&\qquad\qquad \left. + \left| \sum_{m'\in\mathbb{G}} P_M(m') P_{K|Z}(l-m'|z) - \frac{1}{|\mathcal{K}|}\right| \right] \\
&\quad \leqslant 2 \cdot \frac{1}{2} \sum_{m\in\mathbb{G}} P_M(m) \sum_{z\in\mathcal{Z}} P_Z(z) \sum_{l\in\mathbb{G}} \left|P_{K|Z}(l|z) - \frac{1}{|\mathcal{K}|}\right| \\
&\quad = 2\sigma_{\mathtt{var}}(K;Z).
\end{aligned}$$

Next, for every pmf P_M,

$$\begin{aligned}
&D\big(P_{M,Z,M+K} \| P_M \times P_{Z,M+K}\big) \\
&\quad = I(M \wedge Z, M+K) \\
&\quad = I(M \wedge M+K|Z) \\
&\quad = H(M+K|Z) - H(M+K|Z,M) \\
&\quad \leqslant \log|\mathbb{G}| - H(K|Z,M) \\
&\quad = \log|\mathbb{G}| - H(K|Z) \\
&\quad = \sigma_{\mathtt{div}}(K;Z).
\end{aligned}$$

□

2.5 Story of results

The notion of information theoretic perfect secrecy with $I(K \wedge Z) = 0 = \|P_{KZ} - P_K \times P_Z\|$ was introduced by Shannon in his seminal work [72] where it was used to provide the first formal security analysis of a one-time pad. See also [52] for a review in the context of cryptology. The secrecy criteria $\sigma_{\mathtt{var}}(K;Z)$ and $\sigma_{\mathtt{div}}(K;Z)$ that combine independence and uniformity requirements are due to [21]. A precursor to Lemma 2.4 relating $\sigma_{\mathtt{var}}(K;Z)$ and $\sigma_{\mathtt{div}}(K;Z)$ was proved first in

[21]. The slightly stronger version here is obtained by using Lemma 2.3, which is a stronger bound for the difference $|H(P) - H(Q)|$ than that used in the proof of the original version in [21]. Lemma 2.3 is from [4, 104] (see, also, [19, Problem 3.10]). The operational notion of query-based secrecy treated in §2.2.2 was defined in [79]. The equivalence of variational secrecy in Definition 2.5 and query-based secrecy in Definition 2.6, obtained by connecting them to large probability bounds for the conditional probability $P_{U|Z}$, was established in [79]. The clearer treatment presented here, along the lines of the information spectrum approach [35, 33], is new and was developed in [86]; see also [43]. §2.3, which extends the well-known notions of semantic and distinguishing secrecy [30], is based on [5] with slightly modified proofs. Lastly, a simpler form of Proposition 2.16 was proved by Shannon in [72]; the part on $\sigma_{\mathtt{div}}$ given here is from [19, Chapter 17] (see also [1, Lemma 2.1]).

3

Interactive Communication and Common Randomness

The central concepts of interactive communication and common randomness for multiple terminals, of independent interest beyond applications to secrecy problems, are studied in this chapter. These concepts will permeate the monograph. Characteristics of interactive communication, including a fundamental structural property, are described in §3.1. Common randomness generated by multiple terminals using interactive communication is addressed in §3.2. A "single-shot" upper bound for the entropy of common randomness conditioned on interactive communication is derived, using the mentioned special property of the latter. As an application, it is shown to provide a converse lower bound for communication in a data compression problem of omniscience (*sans* secrecy constraints). The mentioned upper bound, termed shared information and particularizing to mutual information for two terminals, will be applied in later chapters for proving converse results in secrecy settings.

3.1 Interactive communication and properties

Consider a set of terminals $\mathcal{M} = \{1, ..., m\}$ that observe, respectively, finite-valued rvs $X_1, ..., X_m$ with joint pmf $P_{X_\mathcal{M}} = P_{X_1 \cdots X_m}$. The terminals cooperate to accomplish a given task in a distributed manner, using *interactive communication* over an unrestricted and noiseless network. It is assumed that the communication is in broadcast mode, *i.e.*, all the terminals receive instantaneously all the communication.

Randomization is permitted at terminals, and the rv U_i denotes the local randomness at Terminal i, where $U_1, ..., U_m$ are mutually independent[1]. It is assumed that $U_\mathcal{M}$ is independent of $X_\mathcal{M}$. For notational simplicity, we shall use $Y_i = (U_i, X_i)$, $i \in \mathcal{M}$.

Definition 3.1. Assume without any loss of generality that the communication of the terminals in $\mathcal{M}$ occurs in consecutive time slots in r rounds; such communication is described in terms of the mappings

$$f_{11}, \ldots, f_{1m}, f_{21}, \ldots, f_{2m}, \ldots, f_{r1}, \ldots, f_{rm},$$

with f_{ji} corresponding to a message in round j from Terminal i, $1 \leqslant j \leqslant r$, $1 \leqslant i \leqslant m$; in general, f_{ji} is allowed to yield any function of Y_i and of previous communication

$$\phi_{ji} = \{f_{kl} \colon k < j,\ l \in \mathcal{M} \text{ or } k = j,\ l < i\}.$$

The corresponding rvs representing the communication will be depicted collectively as

$$\mathbf{F} = \{F_{11}, \ldots, F_{1m}, F_{21}, \ldots, F_{2m}, \ldots, F_{r1}, \ldots, F_{rm}\},$$

where $\mathbf{F} = \mathbf{F}(Y_\mathcal{M})$, namely a deterministic function of $Y_\mathcal{M}$; the rv corresponding to ϕ_{ji} is denoted by Φ_{ji}. A special form of such communication will be termed *simple communication* if $\mathbf{F} = (F_1, ..., F_m)$, where $F_i = F_i(Y_i)$, $i \in \mathcal{M}$.

An interactive communication possesses several distinguishing properties that will play an important role. We begin by considering the case $m = 2$.

[1] A more general model would include an additional shared randomness U_0 known to all the terminals in $\mathcal{M}$.

Lemma 3.2. Given rvs Y_1, Y_2, for interactive communication $\mathbf{F}$ of the terminals in $\mathcal{M} = \{1, 2\}$, it holds that

$$I\left(Y_1 \wedge Y_2 \mid \mathbf{F}\right) \leqslant I\left(Y_1 \wedge Y_2\right). \tag{3.1}$$

In particular, independent rvs Y_1, Y_2 remain so upon conditioning on an interactive communication.

Proof. For interactive communication

$$\mathbf{F} = (F_{11}, F_{12}, ..., F_{r1}, F_{r2}),$$

we have

$$\begin{aligned} I\left(Y_1 \wedge Y_2\right) &= I\left(Y_1, F_{11} \wedge Y_2\right) \\ &\geqslant I\left(Y_1 \wedge Y_2 \mid F_{11}\right) \\ &= I\left(Y_1 \wedge Y_2, F_{12} \mid F_{11}\right) \\ &\geqslant I\left(Y_1 \wedge Y_2 \mid F_{11}, F_{12}\right). \end{aligned}$$

The first claim follows by iterating the steps above. The second claim is immediate. □

In general, the lemma above does not hold for every function $F = F(Y_1, Y_2)$, as shown by the following example.

Example 3.3. Let $(Y_1, Y_2) = (X_1, X_2)$ be binary symmetric rvs with

$$\begin{aligned} P_{X_1X_2}(0,0) = P_{X_1X_2}(1,1) &= \frac{(1-p)}{2}, \\ P_{X_1X_2}(0,1) = P_{X_1X_2}(1,0) &= \frac{p}{2}, \end{aligned}$$

where $0 < p < 1$. The *noninteractive* function $F = Y_1 \oplus Y_2$ violates Lemma 3.2 since

$$I\left(Y_1 \wedge Y_2 \mid F\right) = 1 > 1 - h(p) = I\left(Y_1 \wedge Y_2\right),$$

where $h(p) = p \log 1/p + (1-p) \log 1/(1-p)$ is the binary entropy function.

The inequality (3.1) can be expressed equivalently as

$$I\left(\mathbf{F} \wedge Y_1, Y_2\right) \geqslant I\left(\mathbf{F} \wedge Y_1 \mid Y_2\right) + I\left(\mathbf{F} \wedge Y_2 \mid Y_1\right),$$

where the left- and right-sides are referred to as the *extrinsic* and *intrinsic* information, respectively. Another useful form of (3.1) is

$$H(\mathbf{F}) \geqslant H(\mathbf{F} \mid Y_1) + H(\mathbf{F} \mid Y_2), \tag{3.2}$$

which holds with equality if Y_1 and Y_2 are independent. In fact, the previous form generalizes to $m \geqslant 2$.

Definition 3.4. For the family $\mathcal{S}(\mathcal{M}) = \{S\colon S \subsetneq \mathcal{M}, S \neq \varnothing\}$ of subsets of $\mathcal{M}$, a *fractional partition* $\lambda = \{\lambda_S, S \in \mathcal{S}(\mathcal{M}))\}$ is a collection of weights $0 \leqslant \lambda_S \leqslant 1$ satisfying

$$\sum_{S \in \mathcal{S}(\mathcal{M})\colon\ i \in S} \lambda_S = 1, \quad \text{for all } i \in \mathcal{M}.$$

Lemma 3.5. Given rvs $Y_1, ..., Y_m$ and an interactive communication $\mathbf{F}$, it holds that for every fractional partition λ of $\mathcal{M}$

$$H(\mathbf{F}) \geqslant \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(\mathbf{F} \mid Y_{S^c}),$$

with equality if $Y_1, ..., Y_m$ are mutually independent.

Remark 3.6. For $m = 2$, the choice $\lambda_{\{1\}} = \lambda_{\{2\}} = 1$ leads to (3.2).

Proof. Since F_{ji} is a function of Y_i and the previous communication Φ_{ji}, we have

$$\begin{aligned} H(\mathbf{F} \mid Y_{S^c}) &= \sum_{j=1}^{r} \sum_{i=1}^{m} H(F_{ji} \mid Y_{S^c}, \Phi_{ji}) \\ &= \sum_{j=1}^{r} \sum_{i \in S} H(F_{ji} \mid Y_{S^c}, \Phi_{ji}), \\ &\leqslant \sum_{j=1}^{r} \sum_{i \in S} H(F_{ji} \mid \Phi_{ji}), \end{aligned}$$

with equality holding iff

$$I\left(F_{ji} \wedge Y_{S^c} \mid \Phi_{ji}\right) = 0, \quad \text{for all } j = 1, ..., r,\ i \in S. \tag{3.3}$$

Then,

$$\begin{aligned}
&\sum_S \lambda_S H(\mathbf{F} \mid Y_{S^c}) \\
&\quad \leqslant \sum_S \sum_{j=1}^{r} \sum_{i \in S} \lambda_S H(F_{ji} \mid \Phi_{ji}) \\
&\quad = \sum_{j=1}^{r} \sum_{i=1}^{m} \left(\sum_{S:\, i \in S} \lambda_S \right) H(F_{ji} \mid \Phi_{ji}) \\
&\quad = H(\mathbf{F}),
\end{aligned}$$

where the last equality holds since λ is a fractional partition of $\mathcal{M}$. Equality holds for mutually independent $Y_1, ..., Y_m$, since then (3.3) holds for all S. □

3.2 Common randomness

The concept of common randomness with interactive communication is pivotal in information theoretic secrecy.

Definition 3.7. For $0 \leqslant \epsilon < 1$, given interactive communication $\mathbf{F}$, an rv $L = L(Y_{\mathcal{M}})$ is an *ϵ-common randomness* (ϵ-CR) for $\mathcal{M}$ using $\mathbf{F}$ if there exist local estimates $L_i = L_i(Y_i, \mathbf{F})$, $i \in \mathcal{M}$, of L satisfying

$$\mathbb{P}\left(L_i = L,\, i \in \mathcal{M}\right) \geqslant 1 - \epsilon.$$

We shall say that L is ϵ-recoverable from $\mathbf{F}$.

Distributed processing tasks with interactive communication entail the generation of CR, and bounds on the amount of such CR are needed for establishing converse results.

Theorem 3.8. Assume that $H(U_{\mathcal{M}}) < \infty$. Given $0 \leqslant \epsilon < 1$, for an ϵ-CR L for $\mathcal{M}$ using interactive communication $\mathbf{F}$,

$$H(L \mid \mathbf{F}) \leqslant H(X_{\mathcal{M}}) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + \nu,$$

for every fractional partition λ of $\mathcal{M}$, where $\nu = m(\epsilon \log |\mathcal{L}| + h(\epsilon))$.

Proof. Since L is recoverable from $(Y_{S^c}, \mathbf{F})$ with probability exceeding $1-\epsilon$, by Fano's inequality

$$\begin{aligned} H(Y_S \mid Y_{S^c}, \mathbf{F}) &= H(Y_S \mid Y_{S^c}, \mathbf{F}, L) + I(L \wedge Y_S \mid Y_{S^c}, \mathbf{F}) \\ &\leqslant H(Y_S \mid Y_{S^c}, \mathbf{F}, L) + \nu_0, \end{aligned}$$

where $\nu_0 = \epsilon \log |\mathcal{L}| + h(\epsilon)$. This, with

$$\begin{aligned} H(Y_S \mid Y_{S^c}, \mathbf{F}, L) &= \sum_{i \in S} H(Y_i \mid Y_{\{1,\ldots,i-1\} \cup S^c}, \mathbf{F}, L) \\ &\leqslant \sum_{i \in S} H(Y_i \mid Y^{i-1}, \mathbf{F}, L) \end{aligned}$$

yields

$$\begin{aligned} &\sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(Y_S \mid Y_{S^c}, \mathbf{F}) \\ &\quad\leqslant \sum_{S \in \mathcal{S}(\mathcal{M})} \sum_{i \in S} \lambda_S \left[H(Y_i \mid Y^{i-1}, \mathbf{F}, L) + \nu_0 \right] \\ &\quad= \sum_{i=1}^{m} \left(\sum_{S \in \mathcal{S}(\mathcal{M}):\ i \in S} \lambda_S \right) \left[H(Y_i \mid Y^{i-1}, \mathbf{F}, L) + \nu_0 \right] \\ &\quad= \sum_{i=1}^{m} \left[H(Y_i \mid Y^{i-1}, \mathbf{F}, L) + \nu_0 \right] \\ &\quad= H(Y_{\mathcal{M}} \mid \mathbf{F}, L) + \nu \\ &\quad= H(Y_{\mathcal{M}} \mid \mathbf{F}) - H(L \mid \mathbf{F}) + \nu. \end{aligned}$$

Therefore,

$$\begin{aligned} &H(L \mid \mathbf{F}) \\ &\quad\leqslant H(Y_{\mathcal{M}} \mid \mathbf{F}) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(Y_S \mid Y_{S^c}, \mathbf{F}) + \nu \\ &\quad= H(Y_{\mathcal{M}}) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(Y_S \mid Y_{S^c}) \\ &\qquad\qquad - \left[H(\mathbf{F}) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(\mathbf{F} \mid Y_{S^c}) \right] + \nu \\ &\quad\leqslant H(Y_{\mathcal{M}}) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(Y_S \mid Y_{S^c}) + \nu, \end{aligned}$$

where the previous inequality is by Lemma 3.5. Finally, the claim follows upon canceling the terms $H(U_i)$, $i \in \mathcal{M}$, using the independence of $U_{\mathcal{M}}$ and $X_{\mathcal{M}}$ and the mutual independence of U_i, $i \in \mathcal{M}$, since

$$H(Y_{\mathcal{M}}) = H(X_{\mathcal{M}}) + H(U_{\mathcal{M}})$$

and

$$\begin{aligned}
&\sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(Y_S \mid Y_{S^c}) \\
&\quad = \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(U_S) \\
&\quad = \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S \sum_{i \in S} H(U_i) \\
&\quad = \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + \sum_{i \in \mathcal{M}} \left(\sum_{S \in \mathcal{S}(\mathcal{M}):\ i \in S} \lambda_S \right) H(U_i) \\
&\quad = \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + H(U_{\mathcal{M}}).
\end{aligned}$$

□

As a first application of Theorem 3.8, consider the problem of achieving *omniscience*, namely when each terminal in $\mathcal{M}$ wishes to recover the observations of every other terminal. Specifically, with $Y_i = (U_i, X_i)$ consisting of local randomness U_i and observation X_i, $i \in \mathcal{M}$, the terminals in $\mathcal{M}$ use interactive communication $\mathbf{F}$ to form ϵ-CR $L = X_{\mathcal{M}}$. Theorem 3.8 leads to a lower bound on the entropy $H(\mathbf{F})$ which will be seen to be tight in special cases in Chapter 4 and 6.

Lemma 3.9. For ϵ-CR $L = X_{\mathcal{M}}$ using interactive communication $\mathbf{F}$,

$$H(\mathbf{F}) \geqslant \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) - \nu,$$

for every fractional partition λ of $\mathcal{M}$, where $\nu = m(\epsilon \log |\mathcal{L}| + h(\epsilon))$.

Proof. With $L = X_{\mathcal{M}}$, the claim follows from Theorem 3.8 upon noting that

$$H(X_{\mathcal{M}}) - H(\mathbf{F}) \leqslant H(X_{\mathcal{M}} \mid \mathbf{F}).$$

□

We close this chapter with facile extensions of Lemma 3.5 and Theorem 3.8 that entail additional conditioning on a rv Z.

Theorem 3.10. For rvs $U_{\mathcal{M}}, X_{\mathcal{M}}, Z$, where $U_1, ..., U_m$ are mutually independent and $U_{\mathcal{M}}$ is independent of $(X_{\mathcal{M}}, Z)$, and ϵ-CR L for $\mathcal{M}$ using interactive communication $\mathbf{F}$, we have for every fractional partition λ of $\mathcal{M}$ that

$$H(\mathbf{F} \mid Z) \geqslant \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(\mathbf{F} \mid Y_{S^c}, Z),$$

and

$$H(L \mid \mathbf{F}, Z) \leqslant H(X_{\mathcal{M}} \mid Z) - \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}, Z) + \nu,$$

where $\nu = m(\epsilon \log |\mathcal{L}| + h(\epsilon))$.

Remark 3.11. (Shared information). Defining *shared information* for $X_1, \ldots, X_m$ by

$$SI(X_{\mathcal{M}}) \triangleq H(X_{\mathcal{M}}) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}), \tag{3.4}$$

Theorem 3.10 says that $H(L \mid \mathbf{F}) \lesssim SI(X_1, \ldots, X_m)$, and affords the useful interpretation that the maximum CR generated by the terminals in $\mathcal{M}$ that is distinct from any interactive communication $\mathbf{F}$ used to generate it, as measured by $H(L \mid \mathbf{F})$, cannot exceed $SI(X_{\mathcal{M}})$. We shall return to the concept of shared information in Chapters 4 and 6 where its significance will be seen.

3.3 Story of results

The first formal definition of an interactive protocol for distributed computing appeared in [100], where it was restricted to a tree-based structure. It is perhaps the most popular model for protocols in the computer science literature. More elaborate definitions have been considered, for instance, in [26, 62]. The general description here is a multiterminal extension from [21] of the two-terminal versions in [44, 54, 1].

Compared with its predecessors, it does not specify the structure of the set of possible transcripts. However, it suffices for our purpose of showing worst-case converse results, *e.g.*, worst-case lower bounds on the number of communication bits. The achievability schemes presented in this monograph consist of either simple protocols or bounded-round protocols with the terminals communicating in a fixed order. The fundamental structural property in Lemma 3.2 is from [54, 1], but has been rediscovered in other contexts; see, for instance, the "intrinsic-extrinsic information bound" (cf. [8]). The multiterminal generalization involving fractional partitions in Lemma 3.5 is from [22], and also can be seen as a special case of a more general bound for submodular functions in [51]. Common randomness was defined in a two-terminal setting in [1, 2], where its connection to secret key agreement was explored together with communication requirements for common randomness generation. The general treatment here, specifically the bounds involving fractional partitions, follows [21, 22].

4

Secret Key Generation

The concept of a secret key and different approaches for deriving upper limits on its length are examined in this chapter. A secret key is defined formally in §4.1 under the variational secrecy index and the more stringent divergence secrecy index; it is instructive as well as convenient to deal with both forms of secrecy, as will be seen in this and subsequent chapters. §4.2 describes three methods for deriving upper bounds for secret key length, with emphasis placed on the first and second means which will be the mainstay of converse proofs in later chapters. First, the upper bound for the conditional entropy of common randomness conditioned on interactive communication, obtained in Chapter 3, leads directly to upper bounds for secret key length under the variational and divergence secrecy indices. The second method relates secret key agreement under variational secrecy to an appropriate binary hypothesis testing problem; the needed upper bound then is elicited from the error exponent of the latter. A third approach that analyzes secret key monotones is illustrated for the case of $m = 2$ terminals.

4.1 Multiterminal secret key

Each terminal i in $\mathcal{M}$ observes rv X_i and possesses local randomness U_i. The terminals in $\mathcal{M}$ cooperate, using interactive public communication $\mathbf{F}$, to generate a secret key which is concealed from an eavesdropper with access to $\mathbf{F}$ and additional side information Z. We assume that $U_{\mathcal{M}}$ is independent of $(X_{\mathcal{M}}, Z)$, and that all the parties know $P_{U_{\mathcal{M}} X_{\mathcal{M}} Z}$.

Definition 4.1. Given $0 \leqslant \epsilon < 1$, $\delta \geqslant 0$, an rv K with values in $\mathcal{K}$ constitutes an (ϵ, δ)-*secret key* ((ϵ, δ)-SK) for $\mathcal{M}$ if K is an ϵ-CR for $\mathcal{M}$ using interactive communication $\mathbf{F}$, which satisfies the secrecy requirement $\sigma(K; \mathbf{F}, Z) \leqslant \delta$, where σ can either be $\sigma_{\mathtt{var}}$ or $\sigma_{\mathtt{div}}$ as in §2.2. The corresponding largest size $\log |\mathcal{K}|$ of an (ϵ, δ)-SK is denoted by $S^{\mathtt{var}}_{\epsilon,\delta}$ or $S^{\mathtt{div}}_{\epsilon,\delta}$.

An SK with $\epsilon = \delta = 0$ is termed a *perfect* SK.

Remark 4.2. The only interesting case in Definition 4.1 is when $\epsilon + \delta < 1$ since otherwise $S^{\mathtt{var}}_{\epsilon,\delta}$ is unbounded. This is illustrated for $m = 2$. Terminal 1 generates a (trivial) SK K_1 uniformly on an arbitrary set $\mathcal{K}$ using local randomness U_1, and sends K_1 to Terminal 2. Then K_1 constitutes a $(0, 1-1/|\mathcal{K}|)$-SK and, therefore, also a $(0, 1)$-SK. Terminal 2 generates K_2 uniformly on $\mathcal{K}$ using U_2 (and does not communicate it publicly). Note that K_2 constitutes trivially a $(1 - 1/|\mathcal{K}|, 0)$-SK and, therefore, also a $(1, 0)$-SK. If $\epsilon + \delta \geqslant 1$, an rv K that equals K_1 with probability $(1-\epsilon)$ and K_2 with probability ϵ constitutes an $(\epsilon, 1-\epsilon)$-SK of length $\log |\mathcal{K}|$ and, therefore, also an (ϵ, δ)-SK of the same length. Since $\mathcal{K}$ was arbitrary, $S^{\mathtt{var}}_{\epsilon,\delta}$ is unbounded.

The following simple examples of SK generation assume that $Z = constant$.

Example 4.3. For $m = 3$, let $X_1 = (B_{11}, B_{12})$, $X_2 = (B_{21}, B_{22})$, and

$$X_3 = (B_{11} \oplus B_{21}, B_{12} \oplus B_{22}),$$

where B_{ij}, $1 \leqslant i, j \leqslant 2$, are mutually independent random bits. Using interactive communication $\mathbf{F} = (F_{11}, F_{12}, F_{13})$, where

$$F_{11} = B_{11},\ F_{12} = B_{22},\ F_{13} = B_{11} \oplus B_{21} \oplus B_{12} \oplus B_{22},$$

the terminals become omniscient with a perfect recovery of (X_1, X_2, X_3). Furthermore, they generate a perfect SK $K = B_{12}$ (or B_{21}) of length 1.

The next two examples illustrate how mutually independent SKs between pairs of terminals in $\mathcal{M}$ can be used to generate a SK for $\mathcal{M}$. The initial pairwise SKs can be represented conveniently by a suitable undirected graph $\mathcal{G}$ with vertex set $\mathcal{M}$, edge set $\mathcal{E}$ and no self loops. For each edge $(i, j) \in \mathcal{E}$, the terminals i and j have access to a shared random bit B_{ij}, where $B_{ij} = B_{ji}$ and B_{ij}, $1 \leqslant i < j \leqslant m$, are mutually independent. In particular,

$$X_i = \{B_{ij}, (i, j) \in \mathcal{E}\}, \quad i \in \mathcal{M},$$

i.e., each terminal $i \in \mathcal{M}$ observes all the bits corresponding to the edges incident on it.

Example 4.4. Let $\mathcal{G}$ be a tree with vertex set $\mathcal{M}$. In the first step, the root node selects a bit B_e from an edge e incident on it and broadcasts the modulo 2 sums of B_e with every other incident bit. This enables all the child nodes of the root to recover B_e. The protocol proceeds with each child node repeating the previous step for propagating B_e further to its children, terminating when the leaf nodes recover B_e. The random bit B_e constitutes a perfect SK for $\mathcal{M}$ since it is independent of

$$\mathbf{F} = \{B_e \oplus B_{ij}, (i, j) \in \mathcal{E} \backslash \{e\}\}.$$

Example 4.5. Let $\mathcal{G}$ be a complete graph with vertex set $\mathcal{M}$ where m is even. By the previous example, each spanning tree of $\mathcal{G}$ gives rise to 1 bit of perfect SK for $\mathcal{M}$. A repeated application of this protocol to d edge-disjoint spanning trees yields a d-bit perfect SK. Clearly,

$$d \leqslant \frac{\text{Total number of edges in } \mathcal{G}}{m-1} = \frac{m}{2}.$$

On the other hand, there exists an edge-disjoint spanning tree packing of $\mathcal{G}$ of size $m/2$. Specifically, for each edge e in a matching of size $m/2$, the spanning trees given in Figure 4.1 are disjoint, thereby leading to a $m/2$-bit SK.

In the next section, we shall see that the SK lengths in the examples above cannot be bettered.

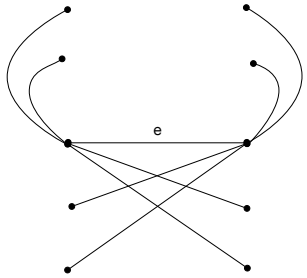

Figure 4.1: Spanning tree corresponding to an edge e of the matching.

4.2 Upper bounds for secret key length

We present three different approaches for obtaining upper bounds for the maximum length of an (ϵ, δ)-SK, with relative merits in different regimes of ϵ, δ.

4.2.1 Common randomness entropy bound

An upper bound on $S^{\mathtt{div}}_{\epsilon,\delta}$ follows directly from Theorem 3.10.

Theorem 4.6. Given $0 \leqslant \epsilon < 1/m$ and $\delta \geqslant 0$,

$$S^{\mathtt{div}}_{\epsilon,\delta} \leqslant \frac{1}{1-m\epsilon}\left[H(X_{\mathcal{M}} \mid Z) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}, Z)\right] + \frac{mh(\epsilon)+\delta}{1-m\epsilon}, \quad (4.1)$$

where the maximum if over all fractional partitions λ for $\mathcal{S}(\mathcal{M})$ (see Definition 3.4).

Remark 4.7. A corresponding bound for $S^{\mathtt{var}}_{\epsilon,\delta}$ follows from Lemma 2.4:

$$S^{\mathtt{var}}_{\epsilon,\delta} \leqslant \frac{1}{1-m\epsilon-\delta}\left[H(X_{\mathcal{M}} \mid Z) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}, Z)\right] + \frac{mh(\epsilon)+h(\delta)}{1-m\epsilon-\delta}.$$

Proof. Let K be an (ϵ, δ)-SK for $\mathcal{M}$ using interactive communication $\mathbf{F}$. By Theorem 3.10, for every fractional partition λ of $\mathcal{M}$,

$$\begin{aligned} H(K \mid \mathbf{F}, Z) \leqslant H(X_{\mathcal{M}} \mid Z) \quad &- \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}, Z) \\ &+ \quad m(\epsilon \log |\mathcal{K}| + h(\epsilon)). \end{aligned}$$

Since $\sigma_{\mathtt{div}}(K; \mathbf{F}, Z) = \log |\mathcal{K}| - H(K \mid \mathbf{F}, Z)$, the claim follows. □

The expression

$$SI(X_{\mathcal{M}} \mid Z) \triangleq H(X_{\mathcal{M}} \mid Z) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}, Z), \quad (4.2)$$

in (4.1), which plays a central role in the maximum length of a SK, has an appealing equivalent form in terms of the Kullback-Leibler divergence. Let $\pi = (\pi_1, ..., \pi_l)$ be a nontrivial partition of $\mathcal{M}$ with $|\pi| = l$ atoms, $2 \leqslant l \leqslant m$. Consider the corresponding fractional partition $\lambda = \lambda(\pi)$ given by

$$\lambda_S = \begin{cases} \frac{1}{l-1}, & \text{if } S = \pi_i^c,\ 1 \leqslant i \leqslant l, \\ 0, & \text{otherwise.} \end{cases}$$

Then for each π,

$$\begin{aligned} SI(X_{\mathcal{M}} \mid Z) &\leqslant H(X_{\mathcal{M}} \mid Z) - \frac{1}{|\pi| - 1} \sum_{i=1}^{|\pi|} H(X_{\pi_i^c} \mid X_{\pi_i}, Z) \\ &= \frac{1}{|\pi| - 1} \left[\sum_{i=1}^{|\pi|} H(X_{\pi_i} \mid Z) - H(X_{\mathcal{M}} \mid Z) \right] \\ &= \frac{1}{|\pi| - 1} D\big(P_{X_{\mathcal{M}}|Z} \big\| \prod_{i=1}^{|\pi|} P_{X_{\pi_i}|Z} \,\big|\, P_Z\big), \end{aligned}$$

so that

$$SI(X_{\mathcal{M}}|Z) \leqslant \min_{\pi} \frac{1}{|\pi| - 1} D\big(P_{X_{\mathcal{M}}|Z} \big\| \prod_{i=1}^{|\pi|} P_{X_{\pi_i}|Z} \,\big|\, P_Z\big).$$

In fact, the previous inequality can be shown to hold with equality.

Theorem 4.8. It holds that

$$SI(X_{\mathcal{M}}|Z) = \min_{\pi} \frac{1}{|\pi|-1} D\big(P_{X_{\mathcal{M}}|Z} \| \prod_{i=1}^{|\pi|} P_{X_{\pi_i}|Z} \,\big|\, P_Z\big). \tag{4.3}$$

Next, we apply the bound of Theorem 4.6 to show that the SKs generated in Examples 4.3-4.5 are of maximum lengths.

Example 4.9. Let X_1, X_2, X_3 be as in Example 4.3 and let $Z =$ constant. For a perfect SK K, upon using Theorem 4.6, (4.2), and (4.3) with $\pi = (\{1\}, \{2\}, \{3\})$, we get

$$\log|\mathcal{K}| \leqslant \frac{1}{2}\left[\sum_{i=1}^{3} H(X_i) - H(X_1, X_2, X_3)\right] = \frac{1}{2}[6-4] = 1.$$

Example 4.10. Consider an undirected graph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$ with no loops and each edge corresponding to a random bit as in §4.1. Then, for every partition π of $\mathcal{M}$,

$$H(X_{\pi_i}) = \text{ number of edges incident on vertices in } \pi_i, \quad 1 \leqslant i \leqslant |\pi|,$$

and

$$H(X_{\mathcal{M}}) = |\mathcal{E}|.$$

Denoting by $\mathcal{E}_\pi$ the edge-cut of π, using Theorem 4.6, (4.2) and (4.3) the length of a perfect SK K is bounded above by $|\mathcal{E}_\pi|/(|\pi|-1)$, and for the particular choice $\pi = (\{i\}, i \in \mathcal{M})$, by $|\mathcal{E}|/(m-1)$. Therefore, for the tree and complete graph of Examples 4.4 and 4.5, respectively, the length of a perfect SK is bounded above by 1 and $m/2$.

4.2.2 Conditional independence testing bound

Another upper bound for SK length $S_{\epsilon,\delta}^{\mathtt{var}}$ can be obtained by relating SK agreement to simple binary hypothesis testing. First, we consolidate the recovery and secrecy conditions for an SK K under $\sigma_{\mathtt{var}}$ into a single convenient form which involves the local estimates $K_1, ..., K_m$ of K (*cf.* Definition 3.7). Denote $K_{\mathcal{M}} = (K_1, ..., K_m)$, and for a pmf P on $\mathcal{K}$ let $P^{(m)}$ denote its extension to $\mathcal{K}^m$ given by

$$P^{(m)}(k_1, ..., k_m) = P(k)\, \mathbb{1}(k_1 = ... = k_m), \quad (k_1, ..., k_m) \in \mathcal{K}^m.$$

Lemma 4.11. Given $0 \leqslant \epsilon, \delta \leqslant 1$ and an (ϵ, δ)-SK K under $\sigma_{\mathtt{var}}$ using interactive communication $\mathbf{F}$, the local estimates $K_1, ..., K_m$ satisfy

$$\left\| P_{K_{\mathcal{M}} \mathbf{F} Z} - {P_{\mathtt{unif}}}^{(m)} \times P_{\mathbf{F}Z} \right\| \leqslant \epsilon + \delta. \tag{4.4}$$

Conversely, if $K_{\mathcal{M}}$ satisfies (4.4) with ϵ in lieu of $\epsilon + \delta$, each K_i, $i \in \mathcal{M}$, constitutes an (ϵ, ϵ)-SK under $\sigma_{\mathtt{var}}$.

Proof. For an (ϵ, δ)-SK K,

$$\begin{aligned}
&\left\| P_{K_{\mathcal{M}} \mathbf{F} Z} - {P_{\mathtt{unif}}}^{(m)} \times P_{\mathbf{F}Z} \right\| \\
&\quad \leqslant \left\| P_{K K_{\mathcal{M}} \mathbf{F} Z} - {P_{\mathtt{unif}}}^{(m+1)} \times P_{\mathbf{F}Z} \right\| \\
&\quad \leqslant \left\| P_{K K_{\mathcal{M}} \mathbf{F} Z} - {P_{K|\mathbf{F}Z}}^{(m+1)} \times P_{\mathbf{F}Z} \right\| \\
&\qquad + \left\| {P_{K|\mathbf{F}Z}}^{(m+1)} \times P_{\mathbf{F}Z} - {P_{\mathtt{unif}}}^{(m+1)} \times P_{\mathbf{F}Z} \right\|.
\end{aligned}$$

Since

$$\|P - Q\| = P(\{x \colon P(x) > Q(x)\}) - Q(\{x \colon P(x) > Q(x)\}),$$

the first term on the right-side above satisfies

$$\begin{aligned}
\left\| P_{K K_{\mathcal{M}} \mathbf{F} Z} - {P_{K|\mathbf{F}Z}}^{(m+1)} \times P_{\mathbf{F}Z} \right\| &= 1 - \mathbb{P}\left(K = K_1 = ...K_m\right) \\
&\leqslant \epsilon,
\end{aligned} \tag{4.5}$$

and the second term equals

$$\left\| P_{K|\mathbf{F}Z} \times P_{\mathbf{F}Z} - P_{\mathtt{unif}} \times P_{\mathbf{F}Z} \right\| \leqslant \delta,$$

which gives (4.4).

For the second claim, ϵ-secrecy is immediate and ϵ-recoverability follows as in (4.5). □

Next, consider a simple binary hypothesis testing problem with null hypothesis P and alternative hypothesis Q, where P and Q are pmfs on a finite set $\mathcal{X}$. An observer of $x \in \mathcal{X}$ decides if x were generated by P or by Q. To this end, a randomized test T, *i.e.*, a conditional pmf on $\{0, 1\}$ given x in $\mathcal{X}$, is used. For each x in $\mathcal{X}$, the test T decides P with probability $T(0|x)$ and Q with probability $T(1|x) = 1 - T(0|x)$.

For $0 \leqslant \epsilon < 1$, denote by $\beta_\epsilon(P, Q)$ the infimum over tests T of the probability of error of type II given that the probability of error of type I is less than ϵ, *i.e.*,

$$\beta_\epsilon(P, Q) \triangleq \inf_{T\,:\;(P\circ T)(0)\geqslant 1-\epsilon} (Q \circ T)(0), \tag{4.6}$$

where $(P\circ T)(0) = \sum_{x\in\mathcal{X}} P(x)T(0|x)$ and $(Q\circ T)(0)$ is defined similarly.

For any stochastic matrix $W\colon \mathcal{X} \to \mathcal{Y}$, the following *data processing inequality* holds:

$$\beta_\epsilon(P, Q) \leqslant \beta_\epsilon(P \circ W, Q \circ W). \tag{4.7}$$

Consider a partition $\pi = \{\pi_1, ..., \pi_l\}$ of $\mathcal{M}$ with $l \geqslant 2$ atoms. Heuristically, if $P_{X_\mathcal{M} Z}$ is such that $X_{\pi_1}, ..., X_{\pi_l}$ are conditionally independent given Z, the length of a SK that can be generated is 0. This suggests a bound for the length of an SK in terms of "how far" the pmf $P_{X_\mathcal{M} Z}$ is from another pmf $Q^\pi_{X_\mathcal{M} Z}$ with the conditional independence property. The closeness of the two pmfs is measured by $\beta_\epsilon\big(P_{X_\mathcal{M} Z}, Q^\pi_{X_\mathcal{M} Z}\big)$.

Specifically, let $\mathcal{Q}(\pi)$ be the set of all pmfs $Q^\pi_{X_\mathcal{M} Z}$ that factorize as follows:

$$Q^\pi_{X_\mathcal{M}|Z} = \prod_{i=1}^{|\pi|} Q^\pi_{X_{\pi_i}|Z}. \tag{4.8}$$

A repeated application of Lemma 3.2 gives for each $Q^\pi_{X_\mathcal{M} Z} \in \mathcal{Q}(\pi)$ and an interactive communication $\mathbf{F}$ that

$$Q^\pi_{X_\mathcal{M}|\mathbf{F}Z} = \prod_{i=1}^{|\pi|} Q^\pi_{X_{\pi_i}|\mathbf{F}Z}. \tag{4.9}$$

Theorem 4.12. Given $0 \leqslant \epsilon + \delta < 1$, $0 < \eta < 1 - \epsilon - \delta$, and a partition π of $\mathcal{M}$, it holds that

$$S^{\mathtt{var}}_{\epsilon,\delta} \leqslant \frac{1}{|\pi| - 1}\bigg[- \log \beta_{\epsilon+\delta+\eta}\big(P_{X_\mathcal{M} Z}, Q^\pi_{X_\mathcal{M} Z}\big) + |\pi| \log(1/\eta)\bigg]$$

for all $Q^\pi_{X_\mathcal{M} Z} \in \mathcal{Q}(\pi)$.

Proof. Let K be an (ϵ, δ)-SK using $\mathbf{F}$. By Lemma 4.11, the local estimates $K_\mathcal{M} = (K_1, ..., K_m)$ satisfy (4.4). Denote by $W_{K_\mathcal{M}\mathbf{F}|X_\mathcal{M} Z}$

the stochastic matrix corresponding to the SK generation protocol $(K_{\mathcal{M}}, \mathbf{F})$. Using the data processing inequality (4.7) with $P = P_{X_{\mathcal{M}}Z}$ and $Q = Q^{\pi}_{X_{\mathcal{M}}Z}$, and $W = W_{K_{\mathcal{M}}\mathbf{F}|X_{\mathcal{M}}Z}$, we get

$$\beta_{\epsilon+\delta+\eta}\big(P_{X_{\mathcal{M}}Z}, Q^{\pi}_{X_{\mathcal{M}}Z}\big) \leqslant \beta_{\epsilon+\delta+\eta}\big(P_{K_{\mathcal{M}}\mathbf{F}Z}, Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}\big).$$

The main step of the proof entails showing next that

$$\log|\mathcal{K}| \leqslant \frac{1}{|\pi|-1}\bigg[-\log\beta_{\epsilon+\delta+\eta}\big(P_{K_{\mathcal{M}}\mathbf{F}Z}, Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}\big) + |\pi|\log(1/\eta)\bigg].$$

□

Lemma 4.13. Let $K_{\mathcal{M}} = (K_1, ..., K_m)$ be the local estimates of an (ϵ, δ)-SK K using an interactive communication $\mathbf{F}$. Then, for $0 \leqslant \epsilon+\delta < 1$, $0 < \eta < 1-\epsilon-\delta$ and every $Q^{\pi}_{X_{\mathcal{M}}Z} \in \mathcal{Q}(\pi)$, we have

$$\log|\mathcal{K}| \leqslant \frac{1}{|\pi|-1}\bigg[-\log\beta_{\epsilon+\delta+\eta}\big(P_{K_{\mathcal{M}}\mathbf{F}Z}, Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}\big) + |\pi|\log(1/\eta)\bigg],$$

where $P_{K_{\mathcal{M}}\mathbf{F}Z}$ is the marginal pmf of $(K_{\mathcal{M}}, \mathbf{F}, Z)$ from the joint pmf

$$P_{K_{\mathcal{M}}\mathbf{F}X_{\mathcal{M}}Z} = P_{X_{\mathcal{M}}Z}W_{K_{\mathcal{M}}\mathbf{F}|X_{\mathcal{M}}Z},$$

and $Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}$ is the corresponding marginal pmf from the joint pmf

$$Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}X_{\mathcal{M}}Z} = Q^{\pi}_{X_{\mathcal{M}}Z}W_{K_{\mathcal{M}}\mathbf{F}|X_{\mathcal{M}}Z}.$$

Proof. We construct a test for the hypothesis testing problem with null hypothesis $P = P_{K_{\mathcal{M}}\mathbf{F}Z}$ and alternative hypothesis $Q = Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}$. Specifically, we use a deterministic test[1] with the following acceptance region (for the null hypothesis)[2]:

$$\mathcal{A} := \left\{(k_{\mathcal{M}}, f, z)\colon \log\frac{P^{(m)}_{\text{unif}}(k_{\mathcal{M}})}{Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(k_{\mathcal{M}}|f,z)} \geqslant \lambda_{\pi}\right\},$$

where

$$\lambda_{\pi} = (|\pi|-1)\log|\mathcal{K}| - |\pi|\log(1/\eta).$$

[1] In fact, we use a simple threshold test on the loglikelihood ratio but with $P_{\texttt{unif}}{}^{(m)} \times P_{\mathbf{F}Z}$ in place of $P_{K_{\mathcal{M}}\mathbf{F}Z}$, since the two distributions are close to each other by (4.4).

[2] Those $(k_{\mathcal{M}}, f, z)$ for which $Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(k_{\mathcal{M}}|f,z) = 0$ are included in $\mathcal{A}$.

For this test, the probability of error of type II is bounded above as

$$\begin{aligned} Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}(\mathcal{A}) &= \sum_{f,z} Q^{\pi}_{\mathbf{F}Z}(f,z) \sum_{\substack{k_{\mathcal{M}}:\\ (k_{\mathcal{M}},f,z)\in\mathcal{A}}} Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(k_{\mathcal{M}}|f,z) \\ &\leqslant 2^{-\lambda_\pi} \sum_{f,z} Q^{\pi}_{\mathbf{F}Z}(f,z) \sum_{k_{\mathcal{M}}} P^{(m)}_{\text{unif}}(k_{\mathcal{M}}) \\ &= |\mathcal{K}|^{1-|\pi|}\eta^{-|\pi|}. \end{aligned} \tag{4.10}$$

On the other hand, the probability of error of type I is bounded above as

$$\begin{aligned} P_{K_{\mathcal{M}}\mathbf{F}Z}(\mathcal{A}^c) &\leqslant \left\| P_{K_{\mathcal{M}}\mathbf{F}Z} - P^{(m)}_{\text{unif}} \times P_{\mathbf{F}Z} \right\| + {P_{\text{unif}}}^{(m)} \times P_{Z\mathbf{F}}(\mathcal{A}^c) \\ &\leqslant \epsilon + \delta + P^{(m)}_{\text{unif}} \times P_{\mathbf{F}Z}(\mathcal{A}^c), \end{aligned} \tag{4.11}$$

where the first inequality is by the definition of variational distance and the second is by (4.4). The second term above can be expressed as follows:

$$\begin{aligned} &{P_{\text{unif}}}^{(m)} \times P_{\mathbf{F}Z}(\mathcal{A}^c) \\ &\quad = \sum_{f,z} P_{\mathbf{F}Z}(f,z) \frac{1}{|\mathcal{K}|} \sum_{k} \mathbb{1}((\mathbf{k},f,z)\in\mathcal{A}^c) \\ &\quad = \sum_{f,z} P_{\mathbf{F}Z}(f,z) \frac{1}{|\mathcal{K}|} \sum_{k} \mathbb{1}\left(Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(\mathbf{k}|f,z)|\mathcal{K}|^{|\pi|}\eta^{|\pi|} > 1\right), \end{aligned} \tag{4.12}$$

where $\mathbf{k} = (k,\ldots,k)$. The inner sum can be further bounded above as

$$\begin{aligned} &\sum_{k} \mathbb{1}\left(Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(\mathbf{k}|f,z)|\mathcal{K}|^{|\pi|}\eta^{|\pi|} > 1\right) \\ &\qquad \leqslant \sum_{k} \left(Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(\mathbf{k}|f,z)|\mathcal{K}|^{|\pi|}\eta^{|\pi|}\right)^{\frac{1}{|\pi|}} \\ &\qquad = |\mathcal{K}|\eta \sum_{k} Q^{\pi}_{K_{\mathcal{M}}|\mathbf{F}Z}(\mathbf{k}|f,z)^{\frac{1}{|\pi|}} \\ &\qquad = |\mathcal{K}|\eta \sum_{k} \prod_{i=1}^{|\pi|} Q^{\pi}_{K_{\pi_i}|\mathbf{F}Z}(\mathbf{k}|f,z)^{\frac{1}{|\pi|}}, \end{aligned} \tag{4.13}$$

where the previous equality uses (4.9) and the fact that given $\mathbf{F}$, K_{π_i} is a function of (U_{π_i}, X_{π_i}). Next, an application of Hölder's inequality

to the sum on the right-side of (4.13) yields

$$\begin{aligned}\sum_{k}\prod_{i=1}^{|\pi|}Q^{\pi}_{K_{\pi_i}|\mathbf{F}Z}(\mathbf{k}|f,z)^{\frac{1}{|\pi|}} &\leqslant \prod_{i=1}^{|\pi|}\left(\sum_{k}Q^{\pi}_{K_{\pi_i}|\mathbf{F}Z}(\mathbf{k}|f,z)\right)^{\frac{1}{|\pi|}}\\ &\leqslant \prod_{i=1}^{|\pi|}\left(\sum_{k_{\pi_i}}Q^{\pi}_{K_{\pi_i}|\mathbf{F}Z}(k_{\pi_i}|f,z)\right)^{\frac{1}{|\pi|}}\\ &= 1. \qquad (4.14)\end{aligned}$$

Upon combining (4.12)-(4.14) we obtain

$$P^{(m)}_{\text{unif}}\times P_{\mathbf{F}Z}(\mathcal{A}^c)\leqslant\eta,$$

which along with (4.11) limits the probability of error of type I as

$$P_{K_{\mathcal{M}}\mathbf{F}Z}\left(\mathcal{A}^c\right)\leqslant\epsilon+\delta+\eta.$$

It follows from (4.10) that

$$\beta_{\epsilon+\delta+\eta}\big(P_{K_{\mathcal{M}}\mathbf{F}Z},Q^{\pi}_{K_{\mathcal{M}}\mathbf{F}Z}\big)\leqslant|\mathcal{K}|^{1-|\pi|}\eta^{-|\pi|},$$

which completes the proof. □

Note that the converse bound above involves ϵ and δ in the form $\epsilon+\delta$ and holds when $\epsilon+\delta<1$, which is the interesting case by Remark 4.2. In fact, $S^{\mathtt{var}}_{\epsilon,\delta}$ is determined by $\epsilon+\delta$ alone and not ϵ,δ separately. This is shown next for $m=2$ and can be extended to $m\geqslant 2$ in a straightforward manner.

Lemma 4.14. Given an (ϵ,δ)-SK for $\mathcal{M}=\{1,2\}$, there exists an $(\epsilon+\delta,0)$-SK of the same length.

Proof. Let K be an (ϵ,δ)-SK for $\mathcal{M}=\{1,2\}$ using interactive communication $\mathbf{F}$, with local estimates K_1 and K_2. We construct a new $(\epsilon+\delta,0)$-SK K' using the maximal coupling lemma referred to in the proof of Lemma 2.3.

For each fixed realization of $(\mathbf{F},Z)$, let $P_{KK'|\mathbf{F},Z}$ be the maximal coupling of $P_{K|\mathbf{F}Z}$ and $P_{\mathtt{unif}}$. Then $P_{K'\mathbf{F}Z}=P_{\mathtt{unif}}\times P_{\mathbf{F}Z}$, and since K is an (ϵ,δ)-SK, we get by the maximal coupling property that

$$\mathbb{P}\left(K\neq K'\right)=\|P_{K\mathbf{F}Z}-P_{\mathtt{unif}}\times P_{\mathbf{F}Z}\|\leqslant\delta.$$

Define the joint pmf

$$P_{K'K\mathbf{F}ZK_1K_2U_1X_1U_2X_2} = P_{K'|K\mathbf{F}Z}P_{K\mathbf{F}ZK_1K_2U_1X_1U_2X_2}. \tag{4.15}$$

Since $\mathbb{P}\left(K = K_1 = K_2\right) \geqslant 1 - \epsilon$ under the joint pmf (4.15), we have

$$\mathbb{P}\left(K_1 = K_2 = K'\right) \geqslant 1 - \epsilon - \delta.$$

Thus, K' constitutes[3] an $(\epsilon + \delta, 0)$-SK. □

The evaluation of the conditional independence testing upper bound of Theorem 4.12 relies on evaluating $\beta_\epsilon(P, Q)$. A direct computation of $\beta_\epsilon(P, Q)$, which is a linear program, is not feasible for large alphabets. However, we can find upper bounds for $-\log \beta_\epsilon(P, Q)$ which may be evaluated easily, especially when P and Q have a product structure.

We begin with an *information spectrum* upper bound for $-\log \beta_\epsilon(P, Q)$. Denote by P_γ the tail probability

$$P_\gamma \triangleq \mathbb{P}\left(\log \frac{P(X)}{Q(X)} \leqslant \gamma\right),$$

where X has pmf P.

Lemma 4.15. For pmfs P and Q on $\mathcal{X}$ and every $0 \leqslant \epsilon < P_\lambda$,

$$-\log \beta_\epsilon(P, Q) \leqslant \inf_\gamma \gamma - \log\left(P_\gamma - \epsilon\right).$$

Remark 4.16. The term $E_\gamma \triangleq -\log P_\gamma$ is of fundamental importance in information spectrum methods. For the case when $P = P^n$ and $Q = Q^n$ correspond to i.i.d. rvs, asymptotically tight bounds for E_γ can be found using large deviations analysis.

Proof. Let T be a test for P versus Q with $(P \circ \mathrm{T})(0) \geqslant 1 - \epsilon$. Denoting

$$\mathcal{A}_\gamma = \left\{x \in \mathcal{X} \colon \log \frac{P(x)}{Q(x)} \leqslant \gamma\right\},$$

[3]The additional randomness for generating K' can be provided as local randomness to one of the terminals.

we get

$$\begin{aligned}(Q \circ \mathrm{T})(0) &\geqslant \sum_{x \in \mathcal{A}_\gamma} Q(x)\mathrm{T}(0|x) \\ &\geqslant 2^{-\gamma} \sum_{x \in \mathcal{A}_\gamma} P(x)\mathrm{T}(0|x) \\ &\geqslant 2^{-\gamma} \left[(P \circ \mathrm{T})(0) - P(\mathcal{A}_\gamma^c)\right] \\ &= 2^{-\gamma} \left[P_\gamma - (P \circ \mathrm{T})(1)\right] \\ &\geqslant 2^{-\gamma} \left[P_\gamma - \epsilon\right],\end{aligned}$$

which upon taking logarithm yields

$$-\log(Q \circ \mathrm{T})(0) \leqslant \gamma - \log(P_\gamma - \epsilon).$$

The claim of the lemma follows since T is arbitrary. □

Thus, finding γ such that P_γ is bounded away from 0 will result in upper bounds for $-\log \beta_\epsilon(P, Q)$. Interestingly, the Rényi divergence constitutes such a choice of γ.

Definition 4.17. For pmfs P and Q on $\mathcal{X}$, the *Rényi divergence of order* $\alpha \neq 1$, $\alpha \geqslant 0$, is given by

$$D_\alpha(P, Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}.$$

Lemma 4.18. Given $0 < \epsilon' < 1$ and $\alpha > 1$, for the choice

$$\gamma = D_\alpha(P, Q) + \frac{1}{\alpha - 1} \log \frac{1}{\epsilon'},$$

we have $P_\gamma \geqslant 1 - \epsilon'$.

Proof. For this choice of γ, the set $\mathcal{A}_\gamma$ satisfies

$$\begin{aligned}1 &= P(\mathcal{A}_\gamma) + \sum_{x \in \mathcal{A}_\gamma^c} P(x) \left(\frac{P(x)}{Q(x)}\right)^{\alpha-1} \left(\frac{P(x)}{Q(x)}\right)^{1-\alpha} \\ &\leqslant P(\mathcal{A}_\gamma) + \epsilon' 2^{(1-\alpha)D_\alpha(P,Q)} \sum_{x \in \mathcal{A}_\gamma^c} P(x) \left(\frac{P(x)}{Q(x)}\right)^{\alpha-1} \\ &\leqslant P(\mathcal{A}_\gamma) + \epsilon',\end{aligned}$$

which completes the proof. □

As a corollary, the following bound holds.

Lemma 4.19. Given $0 \leqslant \epsilon < 1$ and pmfs P and Q on $\mathcal{X}$, for every $\alpha > 1$ and $0 < \epsilon' < 1 - \epsilon$,

$$-\log \beta_\epsilon(P, Q) \leqslant D_\alpha(P, Q) + \frac{1}{\alpha - 1} \log \frac{1}{\epsilon'} + \log \frac{1}{1 - \epsilon - \epsilon'}.$$

4.2.3 Secret key monotone bound

A *monotone* is an attribute of an SK generation protocol that evolves "monotonically" as the protocol proceeds. A suitable choice of a monotone for a given SK generation problem leads to upper bounds for the resulting SK length. This approach is illustrated for the case $m = 2$.

Definition 4.20. Given rvs X_1, X_2, Z, a nonnegative-valued function $M_{\epsilon,\delta} = M_{\epsilon,\delta}(X_1, X_2|Z)$, $0 \leqslant \epsilon, \delta < 1$, constitutes a monotone for SK generation under $\sigma_{\mathtt{var}}$ or $\sigma_{\mathtt{div}}$ if it satisfies the following properties:

1. $M_{\epsilon,\delta}(X_1, X_2|Z)$ does not increase if X_1 or X_2 are replaced by their degraded version, *i.e.*, for the Markov chain $X_1' \multimap X_1 \multimap X_2 Z$

$$M_{\epsilon,\delta}(X_1, X_2|Z) \geqslant M_{\epsilon,\delta}(X_1', X_2|Z),$$

and for the Markov chain $X_2' \multimap X_2 \multimap X_1, Z$

$$M_{\epsilon,\delta}(X_1, X_2|Z) \geqslant M_{\epsilon,\delta}(X_1, X_2'|Z);$$

2. $M_{\epsilon,\delta}(X_1, X_2|Z)$ does not increase on revealing degraded versions of X_1 or X_2, *i.e.*, for X_1' and X_2' as in Property 1,

$$M_{\epsilon,\delta}(X_1, X_2|Z) \geqslant M_{\epsilon,\delta}(X_1, (X_2, X_1')|Z, X_1'),$$

and

$$M_{\epsilon,\delta}(X_1, X_2|Z) \geqslant M_{\epsilon,\delta}(X_2', X_1, X_2|Z, X_2');$$

3. $M_{\epsilon,\delta}(X_1, X_2|Z)$ does not decrease upon replacing Z by its degraded version $Z' \multimap Z \multimap X_1, X_2$:

$$M_{\epsilon,\delta}(X_1, X_2|Z) \leqslant M_{\epsilon,\delta}(X_1, X_2|Z');$$

4. for an (ϵ, δ)-SK K using $\mathbf{F}$ under $\sigma_{\mathtt{var}}$ or $\sigma_{\mathtt{div}}$, with local estimates K_1 and K_2, it holds that

$$\log |\mathcal{K}| \leqslant M_{\epsilon,\delta}((K_1, \mathbf{F}), (K_2, \mathbf{F})|Z, \mathbf{F}) + \Delta(\epsilon, \delta),$$

for a suitable $\Delta(\epsilon, \delta) > 0$.

Lemma 4.21. For $0 \leqslant \epsilon, \delta < 1$ and a monotone $M_{\epsilon,\delta}$ under $\sigma = \sigma_{\mathtt{var}}$ or $\sigma = \sigma_{\mathtt{div}}$,

$$S^{\sigma}_{\epsilon,\delta} \leqslant \inf_{Z' \multimap Z \multimap X_1, X_2} M_{\epsilon,\delta}(X_1, X_2|Z') + \Delta(\epsilon, \delta).$$

Proof. Since K_i is recoverable from $(X_i, \mathbf{F}, U_i)$, $i = 1, 2$, Properties 1 and 4 of a monotone imply that

$$\log |\mathcal{K}| \leqslant M_{\epsilon,\delta}((X_1, \mathbf{F}), (X_2, \mathbf{F})|Z, \mathbf{F}) + \Delta(\epsilon, \delta).$$

Also, a repeated application of Property 2 for communication in each time-slot gives

$$M_{\epsilon,\delta}((X_1, \mathbf{F}), (X_2, \mathbf{F})|Z, \mathbf{F}) \leqslant M_{\epsilon,\delta}(X_1, X_2|Z),$$

and so

$$\log |\mathcal{K}| \leqslant M_{\epsilon,\delta}(X_1, X_2|Z) + \Delta(\epsilon, \delta).$$

The claim follows upon using Property 3. □

Example 4.22. Consider

$$M_{\epsilon,\delta}(X_1, X_2|Z) = \frac{1}{1-\epsilon} I\left(X_1 \wedge X_2 \mid Z\right),$$

and let

$$\Delta(\epsilon, \delta) = \frac{h(\epsilon) + \delta}{1 - \epsilon}.$$

It can be verified that $M_{\epsilon,\delta}$ is a monotone under $\sigma_{\mathtt{div}}$; in particular, Property 3 holds since $I\left(X_1 \wedge X_2 \mid Z\right) \leqslant I\left(X_1 \wedge X_2 \mid Z'\right)$ by the convexity of divergence. Therefore, by Lemma 4.21

$$S^{\mathtt{div}}_{\epsilon,\delta} \leqslant \min_{Z' \multimap Z \multimap X_1, X_2} \frac{1}{1-\epsilon} I\left(X_1 \wedge X_2 \mid Z'\right) + \Delta(\epsilon, \delta).$$

4.3 Story of results

A preliminary incarnation of information theoretic SK generation by two terminals using public discussion was considered in [7]. The two-terminal version of the formulation addressed in this chapter was introduced in [53], a forerunner of [54]. For this model, the largest asymptotic rate of a SK for i.i.d. observations, a question we return to in

Chapter 6, was characterized in [54, 1]. Multiterminal SK generation as formulated in this chapter was studied first in [21] (see also [9]) for i.i.d. observations. The treatment here and the common randomness entropy upper bound of §4.2.1 are based on [21, 22], paraphrased in single-shot form. The expression for *shared information* $SI\left(X_{\mathcal{M}} \mid Z\right)$ and its upper bound involving only partitions were identified in [21]; the tightness of the bound in Theorem 4.8 was shown in [13]. The conditional independence testing bound of §4.2.2 is from [85, 87] and the secret key monotone bound of §4.2.3 from [69]. (Secret key monotones appeared earlier in [9].) Example 4.3 is from [21]. Examples 4.4 and 4.5 evoke the pairwise independent network model introduced in [61, 59]; the maximal spanning tree packing depicted in Figure 4.1 is from [78]. Lemma 4.14, which converts an (ϵ, δ)-SK to an $(\epsilon + \delta, 0)$-SK, is from [39].

5

Extracting Uniform Randomness

Two approaches – balanced coloring and leftover hash – are developed side-by-side for extracting uniform and independent random bits from a given source rv, while keeping near independence from another rv. The extractions employ random mappings. The two methods are at the heart of achievability proofs in later chapters for generating secret common randomness for multiple terminals. Elemental forms of the balanced coloring and leftover hash lemmas are presented in §5.1 and 5.2, respectively, where the eavesdropper can be assumed to have access to the realization of the mentioned random mappings. §5.3 extends both lemmas to the setting in which the eavesdropper has additional access to side information that is correlated with the source. Further refinements in §5.4 make for improved efficiency of extraction and enable an asymptotic performance analysis of the extraction methods for sequences of rvs. We note that the basic bounds presented below are adequate for our purposes, but are not necessarily the best that are known in general.

We begin with pertinent definitions. For a pmf P on a countable set $\mathcal{U}$, the minentropy $H_{\min}(P)$ of P is defined as

$$H_{\min}(P) \triangleq \inf_{u\in\mathcal{U}} (-\log P(u)) = \log \frac{1}{\sup_{u\in\mathcal{U}} P(u)}.$$

For a joint pmf P_{UV} on a countable set $\mathcal{U}\times\mathcal{V}$, the conditional minentropy of U given V is defined as

$$H_{\min}(P_{UV}|V) \triangleq \sup_{Q_V\,:\,\mathtt{supp}(Q_V)\supseteq\mathtt{supp}(P_V)} H_{\min}(P_{UV}|Q_V),$$

where

$$H_{\min}(P_{UV}|Q_V) \triangleq \inf_{u\in\mathcal{U},\ v\in\mathtt{supp}(Q_V)} -\log \frac{P_{UV}(u,v)}{Q_V(v)}.$$

In fact, it can be shown that

$$H_{\min}(P_{UV}|V) = -\log \sum_{v\in\mathcal{V}} P_V(v) \sup_{u\in\mathcal{U}} P_{U|V}(u|v) \tag{5.1}$$

and operationally is the negative logarithm of the average maximum probability of guessing U from V.

We shall show that roughly as many uniformly distributed bits from an rv with pmf P can be extracted as its minentropy $H_{\min}(P)$. Furthermore, for rvs U,V with joint pmf P_{UV}, we can extract roughly as many uniformly distributed bits from U independent of V as the conditional minentropy $H_{\min}(P_{UV}|V)$. In fact, we can replace these bounds with more tractable ones by achieving *smooth versions* of these entropies (to be defined below).

5.1 Balanced coloring lemma

Given a rv U with pmf P on a finite set $\mathcal{U}$, we first show that most of the mappings $\phi\colon \mathcal{U}\to\{1,...,k\}$ can be used to extract $\log k$ random bits from U, provided that $\log k$ is smaller than approximately $H_{\min}(P)$.

Specifically, for $k\geqslant 1$, consider the family $\mathcal{F}_{\mathtt{all}} = \{\phi\colon \mathcal{U}\to\{1,...,k\}\}$ of all mappings from $\mathcal{U}$ to $\{1,...,k\}$. Let Φ be a rv that is distributed uniformly on $\mathcal{F}_{\mathtt{all}}$, i.e.,

$$P_\Phi(\phi) = \frac{1}{|\mathcal{F}_{\mathtt{all}}|} = \frac{1}{k^{|\mathcal{U}|}}, \quad \phi\in\mathcal{F}_{\mathtt{all}},$$

and is independent of U. Then the rvs $\Phi(u), u \in \mathcal{U}$, with the obvious connotation, are i.i.d., with each distributed uniformly on $\{1, \ldots, k\}$.

For a mapping $\phi \in \mathcal{F}_{\mathtt{all}}$, the rv $\phi(U)$ has pmf $P_{\phi(U)} = \big(P_{\phi(U)}(i),\ i = 1, \ldots, k\big)$, where

$$P_{\phi(U)}\,(i) = \sum_{u\colon\, \phi(u)=i} P_U\,(u)\,, \quad i = 1, \ldots, k.$$

With an abuse of notation, let $P_{\Phi(U)} = \big(P_{\Phi(U)}(i),\ i = 1, \ldots, k\big)$ denote the random pmf taking values in the set of pmfs $\{P_{\phi(U)},\ \phi \in \mathcal{F}_{\mathtt{all}}\}$ on $\{1, ..., k\}$.

Lemma 5.1 (Balanced coloring)**.** Let U be a $\mathcal{U}$-valued rv, $|\mathcal{U}| < \infty$, with pmf P. For $0 < \epsilon < 1$, the random mapping Φ distributed uniformly on $\mathcal{F}_{\mathtt{all}}$ and independent of U, satisfies

$$\mathbb{P}\Big(\big\|P_{\Phi(U)} - P_{\mathtt{unif}}\big\| \leqslant \epsilon\Big) \geqslant 1 - 2k \exp\Big(-\epsilon^2 2^{H_{\min}(P) - \log k - 1}\Big)\,,$$

where $P_{\mathtt{unif}}$ is the uniform pmf on $\{1, .., k\}$.

Remark 5.2. It follows from the result above that for $\log k = (1 - \eta)H_{\min}\,(P) - 1$, $\eta > 0$, all but a fraction

$$\exp\Big((1 - \eta)H_{\min}\,(P) - \epsilon^2 2^{\eta H_{\min}(P)}\Big)$$

of mappings $\phi \in \mathcal{F}_{\mathtt{all}}$ satisfy

$$\big\|P_{\phi(U)} - P_{\mathtt{unif}}\big\| \leqslant \epsilon.$$

In particular, for a large $H_{\min}\,(P)$, most of the mappings $\phi \in \mathcal{F}_{\mathtt{all}}$ extract close to $H_{\min}\,(P)$ random bits.

Proof. Fix $i \in \{1, \ldots, k\}$. It suffices to show that

$$\mathbb{P}\Big(\Big|P_{\Phi(U)}(i) - \frac{1}{k}\Big| > \frac{2\epsilon}{k}\Big) \;\leqslant\; 2\exp\Big(-\frac{\epsilon^2 2^{H_{\min}(P)}}{2k}\Big) \tag{5.2}$$

which implies the assertion of the lemma. Observe that

$$P_{\Phi(U)}(i) = \sum_{u \in \mathcal{U}} P(u) \mathbb{1}\big(\Phi(u) = i\big)$$

is a sum of independent rvs since the rvs $\mathbb{1}\big(\Phi(u) = i\big),\ u \in \mathcal{U}$, are i.i.d. Also, $\mathbb{P}\big(\Phi(u) = i\big) = \frac{1}{k},\ u \in \mathcal{U}$, whereby $\sum_{u\in\mathcal{U}} \mathbb{E}\big[P(u)\mathbb{1}\big(\Phi(u) = i\big)\big] = \frac{1}{k}$. The tail inequality (5.2) is obtained by applying Chernoff bounding in the form of a slightly paraphrased version of Bernstein's inequality, stated next, to the sum of the independent rvs $\mathbb{1}\big(\Phi(u) = i\big),\ u \in \mathcal{U}$.

Specifically, let $X_1, \ldots, X_n$ be independent $\mathbb{R}$-valued rvs satisfying the conditions

$$\sum_{j=1}^{n} \mathbb{E}[X_j^2] \leqslant \nu, \qquad \sum_{j=1}^{n} \mathbb{E}[\max\{X_j^l, 0\}] \leqslant \frac{l!\nu c^{l-2}}{2}$$

for all integers $l \geqslant 3$ and for some $0 \leqslant \nu, c < \infty$. Then for every $t > 0$,

$$\mathbb{P}\Big(\big|\sum_{j=1}^{n} \big(X_j - \mathbb{E}[X_j]\big)\big| > \sqrt{2\nu t} + ct\Big) \ \leqslant\ 2\exp(-t). \tag{5.3}$$

We apply Bernstein's inequality (5.3) with $P(u)\mathbb{1}\big(\Phi(u) = i\big),\ u \in \mathcal{U}$, as $X_j,\ j = 1, \ldots, n$, noting that the conditions in its hypothesis are met according to

$$\begin{aligned}\sum_{u\in\mathcal{U}} \mathbb{E}\big[\big(P(u)\mathbb{1}\big(\Phi(u) = i\big)\big)^2\big] &= \sum_{u\in\mathcal{U}} P(u)^2\mathbb{P}\big(\Phi(u) = i\big) \\ &\leqslant \sum_{u\in\mathcal{U}} 2^{-H_{\min}(P)}P(u)\frac{1}{k} \ =\ \frac{2^{-H_{\min}(P)}}{k},\end{aligned}$$

and for $l \geqslant 3$,

$$\begin{aligned}\sum_{u\in\mathcal{U}} \mathbb{E}\big[\big(P(u)\mathbb{1}\big(\Phi(u) = i\big)\big)^l\big] &= \sum_{u\in\mathcal{U}} P(u)^2P(u)^{l-2}\mathbb{P}\big(\Phi(u) = i\big) \\ &\leqslant 2^{-(l-2)H_{\min}(P)}\frac{2^{-H_{\min}(P)}}{k}\end{aligned}$$

with $\nu = \frac{2^{-H_{\min}(P)}}{k}$ and $c = 2^{-H_{\min}(P)}$. By (5.3) for the choice $t = \frac{2\epsilon}{k}2^{H_{\min}(P)}$,

$$\mathbb{P}\Big(\Big|P_{\Phi(U)}(i) - \frac{1}{k}\Big| > \frac{2\sqrt{\epsilon}}{k} + \frac{2\epsilon}{k}\Big) \ \leqslant\ 2\exp\Big(-\frac{2\epsilon}{k}2^{H_{\min}(P)}\Big)$$

so that

$$\mathbb{P}\Big(\Big|P_{\Phi(U)}(i) - \frac{1}{k}\Big| > \frac{4\sqrt{\epsilon}}{k}\Big) \ \leqslant\ 2\exp\Big(-\frac{2\epsilon}{k}2^{H_{\min}(P)}\Big)$$

since $0 \leqslant \epsilon \leqslant \sqrt{\epsilon} \leqslant 1$. The claim in (5.2) follows by a simple change of variables. □

5.2 Leftover hash lemma

The balanced coloring lemma of the previous section shows that for a large $H_{\min}(P)$, most of the mappings $\phi\colon \mathcal{U} \to \{1, ..., k\}$ with range sizes close to $H_{\min}(P)$ will enable the extraction, from a rv U with pmf P, of roughly $H_{\min}(P)$ almost unbiased and independent bits. However, its proof does not suggest a method for obtaining even a single, deterministic extractor. Of course, a randomized extractor can be effected by using a randomly selected map as in the proof of Lemma 5.1; however, this calls for additional uniform randomness of size exponential in $\mathcal{U}$. In this section, we present an alternative mechanism for a randomized extractor. It only requires additional uniform randomness of size roughly $|\mathcal{U}|$; however, the specifics of the construction, while simple, are beyond the scope of this monograph.

A key element of this construction is a 2-*universal hash family* (UHF).

Definition 5.3. Given a finite set $\mathcal{U}$ and $k \geqslant 1$, a family $\mathcal{F}$ of mappings $\phi\colon \mathcal{U} \to \{1, \ldots, k\}$ constitutes a UHF of length k if for every $u \neq u'$ in $\mathcal{U}$,

$$\frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \mathbb{1}\left(\phi(u) = \phi(u')\right) \leqslant \frac{1}{k}. \tag{5.4}$$

Heuristically, random selection over a UHF yields an "almost invertible" mapping from $\mathcal{U}$ to $\{1, \ldots, k\}$ which, for sufficiently small k, will be seen below to enable uniform randomness extraction. Note that the family $\mathcal{F}_{\mathtt{all}}$ of all mappings in the previous section constitutes a UHF with equality in (5.4).

The main result of this section is the *leftover hash lemma* which shows that a randomly selected member of a UHF will enable the extraction of roughly $H_{\min}(P)$ random bits from a rv U with pmf P.

Lemma 5.4 (Leftover hash)**.** Let U be a $\mathcal{U}$-valued rv, $|\mathcal{U}| < \infty$, with pmf P, and let $\mathcal{F}$ be a UHF of length k. It holds that

$$\frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \left\| P_{\phi(U)} - P_{\mathtt{unif}} \right\| \leqslant \frac{1}{2} \sqrt{2^{\log k - H_{\min}(P)}},$$

or, equivalently, a random mapping Φ distributed uniformly on $\mathcal{F}$ and independent of U, satisfies

$$\sigma_{\mathtt{var}}(\Phi(U);\Phi) \leqslant \frac{1}{2}\sqrt{2^{\log k - H_{\min}(P)}}. \tag{5.5}$$

Remark 5.5. The equivalent form in (5.5) ensures the secrecy of the output of the extractor from an eavesdropper with access to the random selection of Φ. Thus, the randomization over a UHF can be implemented using "public randomness."

Proof. For every mapping $\phi\colon \mathcal{U} \to \{1,...,k\}$,

$$\begin{aligned}\left\|P_{\phi(U)} - P_{\mathtt{unif}}\right\| &= \frac{1}{2}\sum_{i=1}^{k}\left|\sum_{u\colon \phi(u)=i} P(u) - \frac{1}{k}\right| \\ &\leqslant \frac{1}{2}\sqrt{k\sum_{i=1}^{k}\left(\sum_{u\colon \phi(u)=i} P(u) - \frac{1}{k}\right)^2}\end{aligned}$$

by the Cauchy-Schwarz inequality. To bound further the right-side, observe that

$$\begin{aligned}&\sum_{i=1}^{k}\left(\sum_{u\colon \phi(u)=i} P(u) - \frac{1}{k}\right)^2 \\ &= \sum_{i=1}^{k}\sum_{u,u'} P(u)P(u')\mathbb{1}\left(\phi(u)=\phi(u')=i\right) - \frac{1}{k} \\ &= \sum_{u} P(u)^2 + \sum_{i=1}^{k}\sum_{u,u'\colon u\neq u'} P(u)P(u')\mathbb{1}\left(\phi(u)=\phi(u')=i\right) - \frac{1}{k} \\ &= \sum_{u} P(u)^2 + \sum_{u,u'\colon u\neq u'} P(u)P(u')\mathbb{1}\left(\phi(u)=\phi(u')\right) - \frac{1}{k} \\ &= 2^{-H_2(P)} + \sum_{u,u'\colon u\neq u'} P(u)P(u')\mathbb{1}\left(\phi(u)=\phi(u')\right) - \frac{1}{k},\end{aligned}$$

where $H_2(P)$ denotes the Rényi entropy of order-2 given by

$$H_2(P) = -\log\sum_{u\in\mathcal{U}} P(u)^2.$$

Furthermore, noting that $H_2(P) \geqslant H_{\min}(P)$ and applying Jensen's inequality to the concave function $f(x) = \sqrt{x},\ x \geqslant 0$, we get

$$\begin{aligned}
&\frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \sqrt{\sum_{i=1}^{k} \left(\sum_{u\colon \phi(u)=i} P(u) - \frac{1}{k} \right)^2} \\
&\leqslant \sqrt{2^{-H_{\min}(P)} + \sum_{u,u'\colon u \neq u'} P(u)P(u') \frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \mathbb{1}\left(\phi(u) = \phi(u')\right) - \frac{1}{k}} \\
&\leqslant \sqrt{2^{-H_{\min}(P)} + \sum_{u,u'\colon u \neq u'} P(u)P(u') \frac{1}{k} - \frac{1}{k}} \\
&\leqslant \sqrt{2^{-H_{\min}(P)}},
\end{aligned}$$

where the second inequality holds by the definition of a UHF. Thus, by combining the steps above, we get

$$\frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \left\| P_{\phi(U)} - P_{\mathtt{unif}} \right\| \leqslant \frac{1}{2} \sqrt{k 2^{-H_{\min}(P)}}.$$

The equivalent form follows upon checking that

$$\left\| P_{\Phi(U)\Phi} - P_{\mathtt{unif}} \times P_{\Phi} \right\| \ = \ \frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \left\| P_{\phi(U)} - P_{\mathtt{unif}} \right\|,$$

using the independence of Φ and U. □

Remark 5.6. The proof of Lemma 5.4 establishes the stronger result

$$\frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \left\| P_{\phi(U)} - P_{\mathtt{unif}} \right\| \leqslant \frac{1}{2} \sqrt{2^{\log k - H_2(P)}}.$$

Indeed, analogous strengthened results can be had below. However, the forms with $H_{\min}$ are apt for our purpose.

5.3 Extractor lemmas with side information

The basic extractor lemmas of the previous two sections treat the case where an eavesdropper has access to the realization of the random

extractor but has no additional side information. For applications involving secrecy generation considered in this monograph, we require extractor lemmas when, in addition to the public randomness, the eavesdropper observes additional side information V that is correlated with U. Simple extensions of the balanced coloring lemma and the leftover hash lemma enable the extraction of roughly $H_{\min}(P_{UV}|P_V)$ random bits independent of V.

Formally, consider rvs U and V taking values in finite sets $\mathcal{U}$ and $\mathcal{V}$, respectively, and with joint pmf P_{UV}. Note that

$$H_{\min}(P_{UV} \mid P_V) \leqslant H_{\min}\left(P_{U|V=v}\right), \quad v \in \mathcal{V}. \tag{5.6}$$

For a mapping $\phi\colon \mathcal{U} \to \{1, ..., k\}$, let $P_{\phi(U)V}$ denote the joint pmf

$$P_{\phi(U)V}(i, v) = \sum_{u\colon \phi(u)=i} P_{UV}(u, v), \quad i \in \{1, ..., k\},\ v \in \mathcal{V}.$$

Furthermore,

$$\begin{aligned}\sigma_{\mathtt{var}}(\phi(U); V) &= \left\|P_{\phi(U)V} - P_{\mathtt{unif}} \times P_V\right\| \\ &= \sum_{v\in\mathcal{V}} P_V(v) \left\|P_{\phi(U)|V=v} - P_{\mathtt{unif}}\right\|.\end{aligned}$$

Consider a random mapping Φ distributed uniformly on $\mathcal{F}_{\mathtt{all}}$ and independent of U and V. The rvs $\Phi(u), u \in \mathcal{U}$, are conditionally i.i.d., with each conditionally distributed uniformly on $\{1, \ldots, k\}$, conditioned on $V = v,\ v \in \mathcal{V}$. Then

$$\begin{aligned}\mathbb{P}\Big(\sigma_{\mathtt{var}}(\Phi(U); V) > \epsilon\Big) &= \mathbb{P}\Big(\sum_{v\in\mathcal{V}} P_V(v) \left\|P_{\Phi(U)|V=v} - P_{\mathtt{unif}}\right\| > \epsilon\Big) \\ &\leqslant \mathbb{P}\Big(\bigcup_{v\in\mathcal{V}} \Big\{\left\|P_{\Phi(U)|V=v} - P_{\mathtt{unif}}\right\| > \epsilon\Big\}\Big) \\ &\leqslant \sum_{v\in\mathcal{V}} \mathbb{P}\Big(\left\|P_{\Phi(U)|V=v} - P_{\mathtt{unif}}\right\| > \epsilon\Big) \\ &\leqslant 2|\mathcal{V}|k \exp\left(-\epsilon^2 2^{H_{\min}(P_{UV}|P_V) - \log k - 1}\right),\end{aligned}$$

where the last inequality is by Lemma 5.1 and (5.6). We have obtained the following extension of the balanced coloring lemma.

Lemma 5.7 (Balanced coloring with side information)**.** Let U and V be $\mathcal{U}$- and $\mathcal{V}$-valued rvs, $|\mathcal{U}| < \infty$, $|\mathcal{V}| < \infty$, with joint pmf P_{UV}. For $0 < \epsilon < 1$, the random mapping Φ distributed uniformly on $\mathcal{F}_{\mathtt{all}}$ and independent of U, V, satisfies

$$\mathbb{P}\Big(\sigma_{\mathtt{var}}(\Phi(U); V) \leqslant \epsilon\Big) \geqslant 1 - 2|\mathcal{V}|k \exp\left(-\epsilon^2 2^{H_{\min}(P_{UV}|P_V) - \log k - 1}\right).$$

Remark 5.8. As in Remark 5.2, with $H = H_{\min}(P_{UV}|P_V)$, for $\log k = (1-\eta)H - 1$, $\eta > 0$, all but a fraction

$$\exp\left((1-\eta)H + \log|\mathcal{V}| - \epsilon^2 2^{\eta H}\right)$$

of mappings $\phi \in \mathcal{F}_{\mathtt{all}}$ satisfy

$$\sigma_{\mathtt{var}}(\phi(U); V) \leqslant \epsilon.$$

In particular, when H is sufficiently large and $\log|\mathcal{V}|$ is of the order of H, most mappings $\phi \in \mathcal{F}_{\mathtt{all}}$ extract close to H random bits independent of V.

The leftover hash Lemma 5.4 admits a similar extension to the case of an eavesdropper with side information. Unlike its balanced coloring counterpart, however, it does not require $\log|\mathcal{V}|$ to be of the order of H in order to be effective. Specifically, let $\mathcal{F}$ be a UHF of length k of mappings $\phi\colon \mathcal{U} \to \{1, ..., k\}$, and let Φ be distributed uniformly on $\mathcal{F}$ and independent of U and V. Then,

$$\begin{aligned}\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) &= \sum_{v \in \mathcal{V}} P_V(v) \frac{1}{|\mathcal{F}|} \sum_{\phi \in \mathcal{F}} \left\| P_{\phi(U)|V=v} - P_{\mathtt{unif}} \right\| \\ &\leqslant \frac{1}{2}\sqrt{k 2^{-H_{\min}(P_{UV}|P_V)}},\end{aligned}$$

where the equality uses the independence of Φ and U, V, and the inequality is by Lemma 5.4 and (5.6). Thus, we have the following extension of Lemma 5.4.

Lemma 5.9 (Leftover hash with side information)**.** Let U and V be $\mathcal{U}$- and $\mathcal{V}$-valued rvs, $|\mathcal{U}| < \infty$, $|\mathcal{V}| < \infty$, with joint pmf P_{UV}. Let $\mathcal{F}$ be a UHF of length k consisting of mappings $\phi\colon \mathcal{U} \to \{1, ..., k\}$. Then, for

a random mapping Φ distributed uniformly on $\mathcal{F}$ and independent of U, V, it holds that

$$\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) \leqslant \frac{1}{2}\sqrt{2^{\log k - H_{\min}(P_{UV}|P_V)}}.$$

Remark 5.10. As in Remark 5.5, here, too, the random mapping Φ can be implemented with public randomness that is available to the eavesdropper.

5.4 Extracting smooth minentropies

A shortcoming of the extractor lemmas discussed above is that the size of the extracted uniform randomness is governed by minentropies, which can diminish significantly even if there were a single element in $\mathcal{U}$ with large probability. In this section, we present the notion of smoothing which not only affords a remedy, but also provides a convenient way to quantify the asymptotic performance of the resulting extractors for a sequence of pmfs $\{P_{U_nV_n}\}_{n=1}^{\infty}$ with suitable concentration behavior.

Given a pmf P_{UV} on a finite set $\mathcal{U} \times \mathcal{V}$, consider another pmf Q_{UV} on $\mathcal{U} \times \mathcal{V}$. In applications below, Q_{UV} will be obtained from P_{UV}. For a mapping $\phi\colon \mathcal{U} \to \{1, ..., k\}$, let $Q_{\phi(U)V}$ be defined analogously as $P_{\phi(U)V}$ with Q_{UV} replacing P_{UV}. Then, for every $\phi \in \mathcal{F}_{\mathtt{all}}$,

$$\begin{aligned}\left\|P_{\phi(U)V} - P_{\mathtt{unif}} \times P_V\right\| &\leqslant \left\|Q_{\phi(U)V} - P_{\mathtt{unif}} \times P_V\right\| + \|P_{UV} - Q_{UV}\| \\ &\leqslant \left\|Q_{\phi(U)V} - P_{\mathtt{unif}} \times Q_V\right\| + 2\,\|P_{UV} - Q_{UV}\|\,.\end{aligned}$$

Application of the earlier extractor lemmas to the pmf Q_{UV} in lieu of P_{UV} incurs an extra leakage of variational secrecy not exceeding $2\,\|P_{UV} - Q_{UV}\|$. In particular, restricting Q_{UV} according to

$$\|P_{UV} - Q_{UV}\| \leqslant \eta, \tag{5.7}$$

will incur an additional leakage of at most 2η. Thus, Lemmas 5.7 and 5.9 can be revised accordingly to enable the extraction of $\sup H_{\min}(Q_{UV}|Q_V)$ random bits, where the supremum is over all Q_{UV} for which (5.7) holds.

For a function $g(P_{UV})$ of P_{UV}, the quantity

$$g^{\eta}(P_{UV}) \triangleq \sup_{Q_{UV}\colon \|P_{UV} - Q_{UV}\| \leqslant \eta} g(Q_{UV})$$

is called the η-*smooth* version[1] of g. The salient observation above is that the number of bits extracted can be increased to the η-smooth versions of the earlier amounts by incurring an additional secrecy leakage not exceeding 2η. In the next two subsections, we derive extensions of Lemmas 5.7 and 5.9 using the idea of smoothing, and then note their consequence for the case of an i.i.d. sequence of pmfs $\{P_{U_nV_n}\}_{n=1}^{\infty}$.

5.4.1 Extractor lemmas with smooth minentropies

As a direct consequence of the foregoing discussion, we obtain the following extensions of Lemmas 5.7 and 5.9. The underlying probability is with respect to P_{UV}.

Lemma 5.11. Let U and V be $\mathcal{U}$- and $\mathcal{V}$-valued rvs, $|\mathcal{U}| < \infty,\ |\mathcal{V}| < \infty$, with joint pmf P_{UV}. For $0 < \eta < 1$, let H^η denote the η-smooth minentropy

$$H^\eta = \sup_{Q_{UV}:\, \|P_{UV}-Q_{UV}\| \leqslant \eta} H_{\min}\left(Q_{UV}|Q_V\right).$$

Then for $0 < \epsilon < 1$, the random mapping Φ distributed uniformly on $\mathcal{F}_{\mathtt{all}}$ and independent of U, V, satisfies

$$\mathbb{P}\Big(\sigma_{\mathtt{var}}(\Phi(U); V) \leqslant \epsilon + 2\eta\Big) \geqslant 1 - 2|\mathcal{V}|k \exp\left(-\epsilon^2 2^{H^\eta - \log k - 1}\right).$$

Remark 5.12. Clearly, $H^\eta \geqslant H\left(P_{UV}|P_V\right),\ 0 < \eta < 1$.

Lemma 5.13. Let U and V be $\mathcal{U}$- and $\mathcal{V}$-valued rvs, $|\mathcal{U}| < \infty,\ |\mathcal{V}| < \infty$, with joint pmf P_{UV}. Let $\mathcal{F}$ be a UHF of length k of mappings $\phi : \mathcal{U} \to \{1, ..., k\}$. Then, for a random mapping Φ distributed uniformly on $\mathcal{F}$ and independent of U, V, it holds that

$$\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) \leqslant \frac{1}{2}\sqrt{2^{\log k - H^\eta}} + 2\eta.$$

Thus, approximately H^η random bits can be extracted from U that are nearly independent of V. Often, it is more convenient to use weakened versions of the extractor results above, obtained by deriving lower

[1] Smoothing can be defined with a supremum or an infimum over an η-ball depending on context.

bounds for H^η. We provide a few popular versions below, each of which can be deduced as an instantiation of the following general lower bound for H^η. In particular, we shall use this bound to show that approximately $H_{\min}(P_{UV}|V)$ unbiased bits can be extracted from U that are independent of V. Furthermore, it will be used to obtain estimates of the number of extracted bits in the important case when $\{(U_i, V_i)\}_{i=1}^n$ have certain i.i.d. attributes. Also, we note that the mentioned lower bound applies, in general, to any smooth function $g^\eta(P_{UV})$ of P_{UV}.

Proposition 5.14. For $0 < \eta < 1$, given a subset $\mathcal{A} \subseteq \mathcal{U} \times \mathcal{V}$ such that

$$P_{UV}(\mathcal{A}) \geqslant 1 - \eta,$$

let $P_{UV}^{\mathcal{A}}$ denote the joint conditional pmf of U, V, given that $(U, V) \in \mathcal{A}$. Then the η-smooth version $g^\eta(P_{UV})$ of a function $g(P_{UV})$ of P_{UV} satisfies

$$g^\eta(P_{UV}) \geqslant g(P_{UV}^{\mathcal{A}}).$$

In particular,

$$H^\eta \geqslant H_{\min}\left(P_{UV}^{\mathcal{A}}|P_V^{\mathcal{A}}\right).$$

Proof. It suffices to show that

$$\left\|P_{UV} - P_{UV}^{\mathcal{A}}\right\| \leqslant \eta.$$

Indeed, for $B = \mathtt{supp}\,(P_{UV}) \cap A^c$, we have

$$B = \left\{(u, v)\colon P_{UV}^{\mathcal{A}}(u, v) < P_{UV}(u, v)\right\},$$

so that

$$\begin{aligned}\left\|P_{UV} - P_{UV}^{\mathcal{A}}\right\| &= P_{UV}(B) - P_{UV}^{\mathcal{A}}(B)\\ &= P_{UV}(B)\\ &\leqslant \eta.\end{aligned}$$

□

The previous proposition enables a lower bound for H^η in terms of $H_{\min}(P_{UV}|V)$. For $0 < \eta < 1$ and pmf Q_V on $\mathcal{V}$, consider the set $\mathcal{A} = \mathcal{A}(Q_V) \subseteq \mathcal{U} \times \mathcal{V}$, where

$$\mathcal{A} = \mathcal{U} \times \{v \in \mathcal{V}\colon P_V(v) \geqslant \eta Q_V(v)\}.$$

Then,

$$P_{UV}(\mathcal{A}) = 1 - P_{UV}(\mathcal{A}^c) \geqslant 1 - \eta. \tag{5.8}$$

Furthermore, for each $(u, v) \in \mathcal{A}$, we have

$$\begin{aligned} -\log \frac{P_{UV}^{\mathcal{A}}(u,v)}{P_V^{\mathcal{A}}(v)} &= -\log \frac{P_{UV}(u,v)}{P_V(v)} \\ &\geqslant -\log \frac{P_{UV}(u,v)}{\eta Q_V(v)}, \end{aligned}$$

whereby

$$H_{\min}\left(P_{UV}^{\mathcal{A}} | P_V^{\mathcal{A}}\right) \geqslant H_{\min}(P_{UV}|Q_V) - \log\frac{1}{\eta},$$

so that by Proposition 5.14,

$$H^{\eta} \geqslant H_{\min}(P_{UV}|Q_V) - \log\frac{1}{\eta}. \tag{5.9}$$

Since the previous bound holds for every Q_V, we get

$$H^{\eta} \geqslant H_{\min}(P_{UV}|V) - \log\frac{1}{\eta}. \tag{5.10}$$

Remark 5.15. The balanced coloring and leftover hash Lemmas 5.11 and 5.13 hold in weaker form with $H_{\min}(P_{UV}|V) - \log\frac{1}{\eta}$ replacing H^{η}, by (5.10).

Another popular form of extractor lemmas considers the case where $V = (V_1, V_2)$ is a $\mathcal{V}_1 \times \mathcal{V}_2$-valued rv where $\mathcal{V}_1$ and $\mathcal{V}_2$ are countable and finite sets, respectively. In a typical application, V_1 corresponds to side information available to an eavesdropper and V_2 to public communication involved in the underlying protocol. For this case, by choosing

$$Q_{V_1V_2}(v_1, v_2) = \frac{Q_{V_1}(v_1)}{|\mathcal{V}_2|}, \quad v_1 \in \mathcal{V}_1, v_2 \in \mathcal{V}_2,$$

we get

$$H_{\min}(P_{UV}|V) \geqslant H_{\min}(P_{UV}|Q_V) \geqslant H_{\min}(P_{UV_1}|V_1) - \log|\mathcal{V}_2|.$$

Remark 5.16. By the previous bound and (5.9), the balanced coloring and leftover hash Lemmas 5.11 and 5.13 hold with H^η replaced by $H_{\min}(P_{UV_1}|V_1) - \log|\mathcal{V}_2| - \log\frac{1}{\eta}$. In fact, the mentioned leftover hash lemma can be improved by directly modifying the proof of Lemma 5.4. This version, stated without proof, is

$$\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) \leqslant \frac{1}{2}\sqrt{2^{\log k - H_{\min}(P_{UV}|V)}}, \tag{5.11}$$

which for the case $V = (V_1, V_2)$ can be weakened using the specific choice of $Q_{V_1V_2}$ above to get

$$\sigma_{\mathtt{var}}(\Phi(U); V_1, V_2, \Phi) \leqslant \frac{1}{2}\sqrt{2^{\log k + \log|\mathcal{V}_2| - H_{\min}\left(P_{UV_1}|V_1\right)}}. \tag{5.12}$$

Finally, an improved extraction in (5.12) can be had by smoothing with respect to P_{UV}. Specifically, denote by $H^\eta_{\min}(P_{UV}|V)$ the smooth conditional minentropy of U given V defined by

$$H^\eta_{\min}(P_{UV}|V) \triangleq \sup_{Q_{UV}\colon \|P_{UV} - Q_{UV}\| \leqslant \eta} H_{\min}(Q_{UV}|V), \quad 0 < \eta < 1.$$

The following incarnation of the leftover hash lemma is a simple extension of (5.12) and is a generalization of all the versions stated above.

Lemma 5.17. Let U, V_1 and V_2 be $\mathcal{U}$-, $\mathcal{V}_1$- and $\mathcal{V}_2$-valued rvs with joint pmf $P_{UV_1V_2}$, where $\mathcal{U}$ and $\mathcal{V}_1$ are countable sets and $\mathcal{V}_2$ is a finite set. Let $\mathcal{F}$ be a UHF of length k consisting of mappings $\phi\colon \mathcal{U} \to \{1, ..., k\}$. Then, for $0 < \eta < 1$, for a random mapping Φ distributed uniformly on $\mathcal{F}$ and independent of U, V_1, V_2, it holds that

$$\sigma_{\mathtt{var}}(\Phi(U); V_1, V_2, \Phi) \leqslant 2\eta + \frac{1}{2}\sqrt{2^{\log k + \log|\mathcal{V}_2| - H^\eta_{\min}\left(P_{UV_1}|V_1\right)}}.$$

Lemma 5.17 suggests the following two-step procedure for SK generation for the multiterminal setup introduced in §3.1. The terminals in $\mathcal{M}$ first generate ϵ-CR $L = L(U_\mathcal{M}, X_\mathcal{M})$ by means of interactive communication $\mathbf{F} = \mathbf{F}(U_\mathcal{M}, X_\mathcal{M})$. Then the terminals extract from this CR an SK K, in accordance with the the balanced coloring or leftover hash lemmas, maintaining a low variational secrecy index. When the terminals observe i.i.d. data (say, in time), a refinement of this procedure is possible in which the next result plays a central role.

5.4.2 The case of i.i.d. $\{(U_i, V_i)\}_{i=1}^n$

We close this chapter with an extractor lemma for the case when U and V_1 correspond to n i.i.d. repetitions $(U_i, V_{1i})_{i=1}^n$ with pmf P_{UV_1}. Assume for simplicity that the sets $\mathcal{U}$ and $\mathcal{V}_1$ are finite. By a Chernoff bound, for each $\delta > 0$, there exists a constant $c = c(\delta) > 0$ such that for each $n \geqslant 1$, the set

$$\mathcal{A} = \left\{(u^n, v^n) \in \mathcal{U}^n \times {\mathcal{V}_1}^n \colon\ -\log P_{U^n|V_1^n}(u^n|v^n) > n\big[H(U|V_1) - \delta\big]\right\}$$

has probability

$$P_{U^nV^n}(\mathcal{A}) \geqslant 1 - 2^{-nc}.$$

Thus, by Proposition 5.14 with $\eta = \eta(n) = 2^{-nc}$,

$$\begin{aligned} H_{\min}^{\eta}\left(P_{U^nV_1^n}|V_1^n\right) &\geqslant H_{\min}\left(P^{\mathcal{A}}_{U^nV_1^n}|V_1^n\right) \\ &\geqslant H_{\min}\left(P^{\mathcal{A}}_{U^nV_1^n}|P_{V_1^n}\right) \\ &= \inf_{(u^n,v^n)\in\mathcal{A}} -\log \frac{P_{U^n|V_1^n}(u^n|v^n)}{P_{U^nV_1^n}(\mathcal{A})} \\ &\geqslant n\big[H(U|V_1) - \delta\big] + \log\left(1 - 2^{-nc}\right), \end{aligned} \tag{5.13}$$

where the last inequality uses the definition of $\mathcal{A}$ and $P_{U^nV_1^n}(\mathcal{A}) \geqslant 1 - 2^{-nc}$. Lemma 5.17, combined with the bound above, yields the following mainstay of several of the achievability results in the chapters to come.

Lemma 5.18. Consider rvs U^n, V_1^n, V_2 such that $\{(U_i, V_{1i})\}_i^n$ are i.i.d. repetitions of the rv (U, V_1), where U, V_1, V_2 are $\mathcal{U}$-, $\mathcal{V}_1$-, $\mathcal{V}_2$-valued rvs, $|\mathcal{U}| < \infty, |\mathcal{V}_1| < \infty, |\mathcal{V}_2| < \infty$. Then, for every $\delta > 0$ there is $c = c(\delta) > 0$ such that for every $n \geqslant 1$ and every

$$\log k < n\big[H(U|V_1) - 2\delta\big] - \log|\mathcal{V}_2| + \log\left(1 - 2^{-nc}\right) + 2, \tag{5.14}$$

a random mapping Φ distributed uniformly on a UHF $\mathcal{F}$ of length k consisting of mappings $\phi\colon \mathcal{U} \to \{1, ..., k\}$ and independent of U^n, V_1^n, V_2 satisfies

$$\sigma_{\mathtt{var}}(\Phi(U^n); V_1^n, V_2, \Phi) \leqslant 2^{-n\min\{c,\delta\}+\log 3}.$$

Remark 5.19. In this section, smoothing involved optimization with respect to the pmf Q_{UV} such that $\|P_{UV} - Q_{UV}\| \leqslant \eta$. However, the proofs of Lemmas 5.11 and 5.13 hold even when P_{UV} is a subdistribution, i.e., when

$$\sum_{u,v} P_{UV}(u,v) \leqslant 1.$$

Thus, all the results of this section can be extended with smoothing performed with respect to a subdistribution Q_{UV} such that

$$\|P_{UV} - Q_{UV}\| = \frac{1}{2}\sum_{u,v} |P_{UV}(u,v) - Q_{U,V}(u,v)| \leqslant \eta.$$

A useful choice of such a subdistribution is given by

$$Q_{UV}(u,v) = P_{UV}(u,v)\,\mathbb{1}((u,v) \in \mathcal{A}),$$

which satisfies $\|P_{UV} - Q_{UV}\| \leqslant \eta$ if $P_{UV}(\mathcal{A}^c) \leqslant \eta$. This notion of smoothing has many interesting applications and will be used in Chapter 8. Also, it leads to an improvement in Lemma 5.18 and allows the omission of the $\log(1-2^{-nc})$ term in (5.14) by replacing the distribution $P^{\mathcal{A}}_{U^nV^n}$ in (5.13) with a subdistribution $P_{U^nV^n}\mathbb{1}((U^n, V^n) \in \mathcal{A})$.

5.5 Story of results

The concept of randomness extraction was introduced by von Neumann in [90] who considered the generation of uniformly distributed random bits from biased coin tosses. The problems addressed in this chapter are of a slightly different hue as we do not insist on producing exactly uniform bits but seek random bits with pmf close to the uniform. This latter direction was considered first by Santha and Vazirani [70, 71]. In fact, the theoretical computer science community has contributed an extensive body of work on extractors; see [88, Chapter 6] for a survey. This work includes extractors for a family of sources that will render certain computational tasks feasible in a specific (randomized) computational complexity class, while often demonstrating the infeasibility of deterministic extractors; the work in [70, 71] provides one such example of an infeasibility result. In contrast, our focus is on a less-ambitious extraction of randomness from a fixed source. This, or an extension

to a family of i.i.d. sources, suffices for the applications considered in this monograph. In such cases, our seeded extractors can be derandomized easily. For fundamental limits of randomness extraction (of nearly uniform bits), see [89].

The definition of conditional minentropy with optimization with respect to Q_V was introduced in [66]. The operational form in (5.1) was given in [47]. The first instance of the leftover hash lemma, with a conditional entropy-based notion of leakage, appeared in [7]. Lemma 5.4 is from [41], and the term "leftover hash" appeared first in [42] where a strengthening of the result of [41] was given with Rényi entropy of order 2 in place of minentropy. The version here using UHFs is from [36], where side information was introduced as well. Variations of this result using conditional Rényi entropy of order 2 can be found in ([6, 37, 24]) A basic form of the balanced coloring lemma without side information was introduced in [2] (see also [16]), where it was noted that the superexponentially decaying form of the bound enables uniform extraction for any i.i.d. distribution in a family of exponential size with a uniformly bounded minentropy. Variants of the balanced coloring lemma admitting side information were obtained in [21]. The idea of smoothing in the context of the leftover hash lemma was introduced in [68, 66, 69]. While the original versions of balanced coloring were not stated in terms of minentropy and nor did they rely on smoothing, the modified forms stated above enable a unified treatment of the leftover hash and balanced coloring lemmas. The final form of the leftover hash Lemma 5.17 was given in [39], and followed readily from [66].

Part II

Applications

6

Secret Key Capacity for the Multiterminal Source Model

Secret key generation for the multiterminal source model, in which each terminal observes one component of a discrete memoryless multiple source, is investigated in this chapter. The model is given in §6.1. Its secret key capacity problem is studied in detail in §6.2 and a single-letter characterization is provided. The achievability proof brings out an inherent connection to a multiterminal data compression problem of omniscience, namely that of recovering at each of m terminals the source components observed by all the other terminals. The latter setting does not involve any secrecy constraints. A strong converse is shown based on the results of Chapter 4. The secret key capacity is seen to equal shared information, thereby giving operational meaning to the latter notion. Secret key generation for a special "pairwise independent network" model is considered in §6.3. The achievability proof of SK capacity shows a matching connection to a combinatorial problem of maximal packing of spanning trees in an associated multigraph.

6.1 Multiterminal source model

The multiterminal source model is a garnished version of the basic setup introduced in §3.1.

Let $X_1, \ldots, X_m,\ m \geqslant 2$, be rvs with finite alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m$, and joint pmf $P_{X_1 \cdots X_m}$. Consider a discrete memoryless multiple source (DMMS) with generic rv $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ and comprising n i.i.d. repetitions $X_{\mathcal{M}}^n = (X_1^n, \ldots, X_m^n)$ of $X_{\mathcal{M}},\ n \geqslant 1$. For a set of terminals $\mathcal{M} = \{1, \ldots, m\}$, assume that each terminal $i \in \mathcal{M}$ observes the ith component X_i^n of $X_{\mathcal{M}}^n$. Randomization is allowed at the terminals, with the finite-valued rv U_i denoting the local randomness at terminal $i \in \mathcal{M}$. We assume that the rvs $U_1, \ldots, U_m$ are mutually independent and that $U_{\mathcal{M}} = (U_1, \ldots, U_m)$ is independent of $X_{\mathcal{M}}^n$. All the terminals know $P_{U_{\mathcal{M}} X_{\mathcal{M}}^n}$.

The terminals in $\mathcal{M}$ engage in interactive communication $\mathbf{F}$ as in Definition 3.1, where now f_{ji} is allowed to yield any function of (U_i, X_i^n) and of the previous communication $\phi_{ji},\ 1 \leqslant j \leqslant r,\ 1 \leqslant i \leqslant m$. Note that $\mathbf{F} = \mathbf{F}^{(n)}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$. Of particular interest is *simple communication* which, recalling Definition 3.1, is $\mathbf{F} = (F_1, \ldots, F_m)$, where $F_i = f_i{}^{(n)}(U_i, X_i^n),\ i \in \mathcal{M}$. The terminals cooperate, using public interactive communication $\mathbf{F}$, to generate a SK which is concealed from an eavesdropper with access to $\mathbf{F}$.

6.2 Secret key capacity

The notion of a multiterminal SK has been described in §4.1. Now for a multiterminal source model, we define achievable SK rates and SK capacity.

Definition 6.1. $R \geqslant 0$ is an achievable SK rate for the terminals in $\mathcal{M}$ if there exist (ϵ_n, δ_n)-SKs $K^{(n)}$ for $\mathcal{M}$ with values in $\mathcal{K}^{(n)}$ using interactive communication $\mathbf{F}^{(n)}$, *i.e.*, there exist local estimates $K_i^{(n)} = K_i^{(n)}\left(U_i, X_i^n, \mathbf{F}^{(n)}\right),\ i \in \mathcal{M}$, of $K^{(n)}$ satisfying

$$\mathbb{P}\big(K_i^{(n)} = K^{(n)},\ i \in \mathcal{M}\big) \geqslant 1 - \epsilon_n, \quad \sigma_{\mathtt{var}}\big(K^{(n)}; \mathbf{F}^{(n)}\big) \leqslant \delta_n \qquad (6.1)$$

where $\epsilon_n \to 0$, $\delta_n \to 0$, and $\frac{1}{n} \log \left|\mathcal{K}^{(n)}\right| \to R$ as $n \to \infty$.

The largest achievable SK rate is the SK *capacity* C_S.

Thus, an SK $K^{(n)}$ is an ϵ_n-CR for $\mathcal{M}$ using $\mathbf{F}^{(n)}$ with the variational secrecy index for $K^{(n)}$ given $\mathbf{F}^{(n)}$ contained within δ_n.

Theorem 6.2 (SK capacity for $\mathcal{M}$)**.** The SK capacity for a multiterminal source model with generic rv $X_{\mathcal{M}} = \big(X_1, \ldots, X_m\big)$ is

$$C_S \;=\; SI(X_{\mathcal{M}}) \;=\; H(X_{\mathcal{M}}) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S | X_{S^c}), \qquad (6.2)$$

where the fractional partition λ and $\mathcal{S}(\mathcal{M})$ are as in Definition 3.4. Furthermore, SK capacity can be achieved without randomization at the terminals in $\mathcal{M}$ and with simple communication.

Corollary 6.3. It holds that

$$C_S \;=\; \min_{\pi} \frac{1}{|\pi| - 1} D\Big(P_{X_{\mathcal{M}}} \,||\, \prod_{k=1}^{|\pi|} P_{X_{\pi_k}}\Big), \qquad (6.3)$$

where the minimum is over all nontrivial partitions π of $\mathcal{M}$.

Theorem 6.2 says that the largest rate of ϵ_n-CR for $\mathcal{M}$ that is nearly independent of the public communication $\mathbf{F}^{(n)}$ used to generate it, is equal to the shared information $SI(X_{\mathcal{M}})$ for $X_1, \ldots, X_m$, defined in Remark 3.11.

Remark 6.4. By Lemma 2.4, SK capacity defined analogously as in Definition 6.1 but under the divergence secrecy index (see Definition 2.2), can be no larger. In fact, the capacity remains undiminished since the achievability proof below of Theorem 6.2 is with δ_n in (6.1) vanishing to zero exponentially in n, so that by Lemma 2.4 the divergence secrecy index, too, decays to zero.

The CR concept of omniscience, introduced in §3.2, will be of material significance in the proof of achievability of SK capacity in Theorem 6.2. This notion is discussed next, followed by the proof of Theorem 6.2.

6.2.1 The role of omniscience

For the multiterminal source model, omniscience for the terminals in $\mathcal{M}$ entails forming ϵ_n-CR $L^{(n)} = X_{\mathcal{M}}^n$ using interactive communication $\mathbf{F} = \mathbf{F}^{(n)}\big(U_{\mathcal{M}}, X_{\mathcal{M}}^n\big)$. Determining the smallest rate of communication that enables such omniscience is a problem of multiterminal source coding that is not constrained by any secrecy requirement. As shall be seen below, the concept of omniscience brings out an inherent connection between secrecy generation and source coding for the multiterminal source model.

Definition 6.5. The smallest achievable rate of communication for omniscience, termed *minimum* CO rate and denoted by R_{CO}, is the smallest number $R \geqslant 0$ such that for suitable interactive communication $\mathbf{F}^{(n)} = \mathbf{F}^{(n)}\big(U_{\mathcal{M}}, X_{\mathcal{M}}^n\big)$ of the terminals in $\mathcal{M}$ and ϵ_n with

$$\epsilon_n \to 0 \text{ and } \frac{1}{n} \log \big|\mathcal{F}^{(n)}\big| \to R \text{ as } n \to \infty,$$

$X_{\mathcal{M}}^n$ is ϵ_n-CR for $\mathcal{M}$, where $\mathcal{F}^{(n)}$ is the range of $\mathbf{F}^{(n)}$.

Theorem 6.6 (Minimum CO rate for $\mathcal{M}$)**.** The minimum CO rate for a multiterminal source model with generic rv $X_{\mathcal{M}} = \big(X_1, \ldots, X_m\big)$ is

$$R_{CO} \;=\; \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S | X_{S^c}), \tag{6.4}$$

and can be achieved without randomization at the terminals in $\mathcal{M}$ and with simple communication.

The achievability part of Theorem 6.6 is a particularization of a general result regarding a "normal source network without helpers." The latter is addressed next, followed by the proof of Theorem 6.6.

Normal source network without helpers

Let $X_1, \ldots, X_m, D_1, \ldots, D_m$, $m \geqslant 2$, be rvs with finite alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_m, \mathcal{D}_1, \ldots, \mathcal{D}_m$, respectively, and with known joint pmf $P_{X_1 \cdots X_m D_1 \cdots D_m}$. Consider n i.i.d. repetitions of $(X_{\mathcal{M}}, D_{\mathcal{M}}) = (X_1, \ldots, X_m, D_1, \ldots, D_m)$ denoted by $(X_{\mathcal{M}}^n, D_{\mathcal{M}}^n) = (X_1^n, \ldots, X_m^n, D_1^n, \ldots, D_m^n)$, $n \geqslant 1$.

For a DMMS with generic rv $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ as in §6.1, consider the following network source coding problem with decoder side information. Each component X_i^n of the DMMS is connected to exactly one encoder $f_i = f_i^{(n)} \colon \mathcal{X}_i^n \to \mathcal{M}_i = \{1, \ldots, M_i\}$ of rate $R_i = \frac{1}{n} \log M_i,\ i \in \mathcal{M}$. Each of m decoders $\psi_i = \psi_i^{(n)},\ i \in \mathcal{M}$, possesses side information D_i^n, and is required to decode, for a prescribed set $\mathcal{S}_i \subseteq \mathcal{M}$, the corresponding component sources $\{X_j^n, j \in \mathcal{S}_i\}$ of the DMMS. Each decoder $\psi_i,\ i \in \mathcal{M}$, is connected to only those encoders $\{f_j, j \in \mathcal{S}_i\}$ whose corresponding DMMS components it must decode (and so has no additional encoders as "helpers"), and thus is a mapping $\psi_i \colon \left(\times_{j \in \mathcal{S}_i} \mathcal{M}_j \right) \times \mathcal{D}_i^n \to \times_{j \in \mathcal{S}_i} \mathcal{X}_j^n$. These are the only connections in the network, and only simple communication in the form of encoder outputs is allowed. No randomization in encoding or decoding is assumed. The codes $\left\{ \left(f_i, \psi_i\right), i \in \mathcal{M} \right\}$ are required to satisfy

$$\mathbb{P}\left(\psi_i\Big(\left(f_j(X_j^n), j \in \mathcal{S}_i\right), D_i^n \Big) = \left(X_j^n, j \in \mathcal{S}_i\right),\ i \in \mathcal{M} \right) \ \geqslant\ 1 - \epsilon_n \tag{6.5}$$

where $\epsilon_n \to 0$ as $n \to \infty$.

Lemma 6.7 (Achievable rates for normal source network)**.** For the normal source network without helpers above, there exist codes $\left\{ \left(f_i^{(n)}, \psi_i^{(n)}\right),\ i \in \mathcal{M} \right\}$ satisfying (6.5) if the rates $R_i,\ i \in \mathcal{M}$, satisfy the "Slepian-Wolf" conditions

$$\sum_{j \in S} R_j \geqslant H\left(X_S \mid X_{\mathcal{S}_i \setminus S}, D_i\right), \quad S \subseteq \mathcal{S}_i,\ i \in \mathcal{M}. \tag{6.6}$$

Remark 6.8. In fact, randomly chosen $\{f_i^{(n)}, i \in \mathcal{M}\}$ with rates as in (6.6) (and suitably chosen decoders) will satisfy (6.5) with large probability.

Proof. Consider first the special case when the decoders possess no side information, *i.e.*, $D_i =$ constant, $i \in \mathcal{M}$. Then the claim follows by the Slepian-Wolf theorem applied to the decoder for each $i \in \mathcal{M}$. Specifically, there exist codes $\left\{ \left(f_i^{(n)}, \psi_i^{(n)}\right),\ i \in \mathcal{M} \right\}$ of rates $R_i,\ i \in \mathcal{M}$

satisfying

$$\sum_{j\in S} R_j \geqslant H\left(X_S \mid X_{\mathcal{S}_i\setminus S}\right), \quad S\subseteq \mathcal{S}_i,\ i\in\mathcal{M}. \tag{6.7}$$

that meet (6.5) (with $D_i^n = \text{constant},\ i\in\mathcal{M}$).

For the given general model, consider a corresponding modified setup without decoder side information obtained by the following simple artifice. Introduce dummy sources $\tilde{X}_i^n = D_i^n,\ i\in\mathcal{M}$, and connect each dummy source $\tilde{X}_i^n$ to a dummy encoder $\tilde{f}_i = \tilde{f}_i^{(n)}$ of rate $\tilde{R}_i = H(D_i)$. The output of each $\tilde{f}_i$ feeds only into a single decoder $\tilde{\psi}_i = \tilde{\phi}_i^{(n)}$ which is connected also to the encoders $\{f_j,\ j\in\mathcal{S}_i\}$. In this modified setup, each decoder $\tilde{\psi}_i,\ i\in\mathcal{M}$, is required to decode $\{X_j^n,\ j\in\mathcal{S}_i\}$ as well as $\tilde{X}_i^n$ from the outputs of $\{f_j,\ j\in\mathcal{S}_i\}$ and $\tilde{f}_i$, but without direct access to the side information D_i^n, under the requirement

$$\mathbb{P}\left(\tilde{\psi}_i\Big(\left(f_j(X_j^n), j\in\mathcal{S}_i\right); \tilde{f}_i\left(\tilde{X}_i^n\right)\Big) = \left(X_j^n, j\in\mathcal{S}_i; \tilde{X}_i^n\right),\ i\in\mathcal{M}\right) \geqslant 1-\epsilon_n$$

where $\epsilon_n\to 0$ as $n\to\infty$.

It is clear that if there exist codes $\left\{\left((f_i,\tilde{f}_i),\tilde{\psi}_i\right),\ i\in\mathcal{M}\right\}$ with encoders f_i of rates $R_i,\ i\in\mathcal{M}$ (and $\tilde{f}_i$ of rates $\tilde{R}_i = H(D_i),\ i\in\mathcal{M}$, as above), that satisfy the mentioned requirement for the modified setup, then the encoders $f_i,\ i\in\mathcal{M}$, together with suitable decoders $\psi_i,\ i\in\mathcal{M}$, will satisfy the less stringent (6.5) for the given model, too.

Thus, the proof is completed upon showing that numbers $R_i\geqslant 0,\ i\in\mathcal{M}$, that obey (6.6) will also meet conditions for the modified setup that are analogous to (6.7). The latter conditions are: for each $i\in\mathcal{M}$ and $S\subseteq\mathcal{S}_i$,

$$\sum_{j\in S} R_j \ \geqslant\ H\left(X_S \mid X_{\mathcal{S}_i\setminus S}, D_i\right),$$

and for S' such that $X_{S'} = \left(X_S, \tilde{X}_i\right) = (X_S, D_i)$,

$$\begin{aligned}\sum_{j\in S'} R_j \ &\geqslant H\left(X_{S'} \mid X_{\mathcal{S}_i\setminus S}\right)\\ &= H\left(D_i \mid X_{\mathcal{S}_i\setminus S}\right) + H\left(X_S \mid X_{\mathcal{S}_i\setminus S}, D_i\right).\end{aligned}$$

If (6.6) is satisfied, then the first condition coincides with it. The second condition is met, too, as its left-side equals

$$\sum_{j\in S'} R_j \;=\; H(D_i) + \sum_{j\in S} R_j$$

and is no smaller than the right-side, since $H(D_i) \geqslant H\left(D_i \mid X_{\mathcal{S}_i\setminus S}\right)$ and the first condition holds.

□

Proof of Theorem 6.6

Proof. Achievability: In Lemma 6.7, set $D_i = X_i$ and $\mathcal{S}_i = \mathcal{M}\setminus\{i\},\ i \in \mathcal{M}$. Then, there exist codes $\left\{\left(f_i^{(n)}, \psi_i^{(n)}\right),\ i \in \mathcal{M}\right\}$ using which $X_{\mathcal{M}}^n$ is achievable as ϵ_n-CR for $\mathcal{M}$, with $\epsilon_n \to 0$ as $n \to \infty$, if the rates $R_i,\ i \in \mathcal{M}$, satisfy the Slepian-Wolf conditions (6.6) which now are simply

$$\sum_{i\in S} R_i \geqslant H(X_S \mid X_{S^c}), \quad S \in \mathcal{S}(\mathcal{M}), \tag{6.8}$$

where $\mathcal{S}(\mathcal{M})$ is as in §3.1. The codes do not use randomization at the terminals in $\mathcal{M}$, and the encoder outputs constitute simple communication $\mathbf{F} = (F_1, \ldots, F_m)$, $F_i = {f_i}^{(n)}\left(X_i^n\right),\ i \in \mathcal{M}$, of rate

$$\frac{1}{n}\log\prod_{i=1}^m M_i \;=\; \sum_{i=1}^m \frac{1}{n}\log M_i \;=\; \sum_{i=1}^m R_i.$$

Clearly the minimum sum rate $\min_{(R_1,\ldots,R_m)} \sum_{i=1}^m R_i$ subject to (6.8) is an achievable CO rate for $\mathcal{M}$. By the duality theorem of linear programming, this minimum is equal to $\max_\lambda \sum_{S\in\mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c})$.

Converse: Suppose that $X_{\mathcal{M}}^n$ is achievable as ϵ_n-CR for $\mathcal{M}$ using interactive communication $\mathbf{F}^{(n)}$, where $\epsilon_n \to 0$ as $n \to \infty$. By Lemma 3.9, with X_S^n and $X_{S^c}^n$ in the roles of X_S and X_{S^c}, respectively, we have

$$H(\mathbf{F}^n) \;\geqslant \max_\lambda \sum_{S\in\mathcal{S}(\mathcal{M})} \lambda_S H\big(X_S^n \mid X_{S^c}^n\big) - m\big(n\epsilon_n \log|\mathcal{X}_{\mathcal{M}}| + h(\epsilon_n)\big),$$

so that

$$\liminf_n \frac{1}{n}\log|\mathcal{F}^{(n)}| \geqslant \liminf_n \frac{1}{n} H(\mathbf{F}^n) \geqslant \max_\lambda \sum_{S\in\mathcal{S}(\mathcal{M})} \lambda_S H\big(X_S \mid X_{S^c}\big).$$

□

6.2.2 Proof of Theorem 6.2

Proof. The achievability of $H(X_{\mathcal{M}}) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c})$ as an SK rate is shown in two steps: The terminals in $\mathcal{M}$ first achieve CR in the form of omniscience using $\mathbf{F}$; then, each terminal extracts an SK in the form of uniform randomness almost independent of $\mathbf{F}$ from this CR, as enabled by Lemma 5.18. It is followed by the converse part of the proof which is an immediate consequence of the common randomness entropy bound in §4.2.1.

Achievability: First, fixing an arbitrary $\nu > 0$, we get by Theorem 6.6 that the terminals in $\mathcal{M}$ can achieve omniscience, *i.e.*, $X_{\mathcal{M}}^n$ as ϵ_n-CR, using simple communication $\mathbf{F}^{(n)} = \left(F_1^{(n)}, \ldots, F_m^{(n)}\right)$ of range $\mathcal{F}^{(n)}$, say, and rate

$$R \;=\; \frac{1}{n} \log |\mathcal{F}^{(n)}| \;=\; \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c}) + \nu, \tag{6.9}$$

and without having to use local randomization, where $\epsilon_n \to 0$ as $n \to \infty$. We note thereby that any function of $X_{\mathcal{M}}^n$ also remains an ϵ_n-CR for $\mathcal{M}$ achievable with $\mathbf{F}^{(n)}$.

Next, Lemma 5.18, with

$$U^n = X_{\mathcal{M}}^n,\ V_1^n = \text{constant},\ V_2 = \mathbf{F}^{(n)} \text{ and } \mathcal{V}_2 = \mathcal{F}^{(n)},$$

guarantees, for any number $0 \leqslant H < H(X_{\mathcal{M}}) - R$, the existence of $\phi(X_{\mathcal{M}}^n)$ with values in $\left\{1, \ldots, \lfloor \exp(nH) \rfloor\right\}$ such that for $K^{(n)} = \phi(X_{\mathcal{M}}^n)$, the variational secrecy index $\sigma_{\mathtt{var}}\left(K^{(n)}; \mathbf{F}\right) \leqslant \delta_n$ where $\delta_n \to 0$ exponentially rapidly as $n \to \infty$. Since $\nu > 0$ in (6.9) can be chosen to be arbitrarily small, it follows that $H(X_{\mathcal{M}}) - \max_{\lambda} \sum_{S \in \mathcal{S}(\mathcal{M})} \lambda_S H(X_S \mid X_{S^c})$ is an achievable SK rate.

Converse: Suppose that $K^{(n)}$ represents an (ϵ_n, δ_n)-SK for $\mathcal{M}$ with values in $\mathcal{K}^{(n)}$ using interactive communication $\mathbf{F}^{(n)}$ and with $\sigma_{\mathtt{var}}\left(K^{(n)}; \mathbf{F}^{(n)}\right) \leqslant \delta_n$, where $\epsilon_n \to 0$ and $\delta_n \to 0$ as $n \to \infty$. Then, by Remark 4.7 with $X_{\mathcal{M}}^n, X_S^n$ and $X_{S^c}^n$ in the roles of $X_{\mathcal{M}}, X_S$ and X_{S^c},

respectively, and with $Z =$ constant, we get

$$\log\left|\mathcal{K}^{(n)}\right| \leqslant \frac{1}{1-m\epsilon_n-\delta_n}\left[H(X_{\mathcal{M}}^n) - \max_{\lambda}\sum_{S\in\mathcal{S}(\mathcal{M})}\lambda_S H(X_S^n \mid X_{S^c}^n)\right] + \frac{mh(\epsilon_n)+h(\delta_n)}{1-m\epsilon_n-\delta_n},$$

whereupon

$$\limsup_n \frac{1}{n}\log\left|\mathcal{K}^{(n)}\right| \leqslant H(X_{\mathcal{M}}) - \max_{\lambda}\sum_{S\in\mathcal{S}(\mathcal{M})}\lambda_S H(X_S \mid X_{S^c}).$$

The Corollary is immediate from (4.3) with $Z =$ constant.

□

6.2.3 Strong converse for SK capacity

An achievable SK rate $R \geqslant 0$ in Definition 6.1 involves (ϵ_n, δ_n)-SKs $K^{(n)}$ for the terminals in $\mathcal{M}$ with $\epsilon_n \to 0$ and $\delta_n \to 0$ in (6.1) as $n \to \infty$, and the converse proof of Theorem 6.2 makes explicit use of the limits. This converse can be strengthened to show that achievable SK rates cannot increase even if the requirements in (6.1) are eased so as to hold for fixed $\epsilon > 0,\ \delta > 0$ with $\epsilon + \delta < 1$.

Lemma 6.9 (Strong converse for SK capacity)**.** Fix $\epsilon > 0,\ \delta > 0$ with $\epsilon + \delta < 1$. Let $K^{(n)}$ represent an (ϵ, δ)-SK for $\mathcal{M}$ with values in $\mathcal{K}^{(n)}$ using interactive communication $\mathbf{F}^{(n)}$, and with $\sigma_{\mathtt{var}}\left(K^{(n)}; \mathbf{F}^{(n)}\right) \leqslant \delta$. Then,

$$\limsup_n \frac{1}{n}\log|\mathcal{K}^{(n)}| \leqslant SI\left(X_1, \ldots, X_m\right) = \min_{\pi}\frac{1}{|\pi|-1}D\big(P_{X_{\mathcal{M}}}\big\|\prod_{i=1}^{|\pi|}P_{X_{\pi_i}}\big).$$

Remark 6.10. The case $\epsilon+\delta \geqslant 1$ is uninteresting as it enables arbitrarily large achievable SK rates. For instance, Terminal 1 can generate a $(0,1)$-SK $\tilde{K}$, distributed uniformly on a set $\mathcal{K}$, which it communicates publicly to the remaining terminals; and each terminal $i \in \mathcal{M}$ generates a $(1,0)$-SK K_i distributed uniformly on $\mathcal{K}$ using U_i but requiring no communication whatsoever. Clearly, an rv K that equals $\tilde{K}$ with probability $1-\epsilon$ and K_1, say, with probability ϵ, constitutes an (ϵ, δ)-SK for $\delta = 1-\epsilon$, with arbitrarily large $\log|\mathcal{K}|$.

Proof. Apply Theorem 4.12 with

$$X_{\mathcal{M}} = X_{\mathcal{M}}^n,\ Z = \text{constant and } Q_{X_{\mathcal{M}}^n}^{\pi} = \prod_{i=1}^{|\pi|} P_{X_{\pi_i}}^n$$

to get that for every partition π of $\mathcal{M}$,

$$\log|\mathcal{K}^{(n)}| \leqslant \frac{1}{|\pi|-1}\Big[-\log\beta_{\epsilon+\delta+\eta}\big(P_{X_{\mathcal{M}}}^n, \prod_{i=1}^{|\pi|} P_{X_{\pi_i}}^n\big) + |\pi|\log(1/\eta)\Big] \tag{6.10}$$

where $0 < \eta < 1-\epsilon-\delta$. For each partition π of $\mathcal{M}$, by Stein's Lemma applied to the testing of the null hypothesis $P_{X_{\mathcal{M}}}^n$ versus the alternative hypothesis $\prod_{i=1}^{|\pi|} P_{X_{\pi_i}}^n$, the infimum $\beta_{\epsilon+\delta+\eta}\big(P_{X_{\mathcal{M}}}^n, \prod_{i=1}^{|\pi|} P_{X_{\pi_i}}^n\big)$ of the probability of error of type II when the probability of error of type I is constrained to not exceed $\epsilon+\delta+\eta$, satisfies

$$\lim_n\ -\frac{1}{n}\log\beta_{\epsilon+\delta+\eta}\big(P_{X_{\mathcal{M}}}^n, \prod_{i=1}^{|\pi|} P_{X_{\pi_i}}^n\big)\ =\ D\big(P_{X_{\mathcal{M}}}\big\|\prod_{i=1}^{|\pi|} P_{X_{\pi_i}}\big)$$

for all $0 < \epsilon+\delta+\eta < 1$. Hence, the assertion of the result follows from (6.10). □

6.3 Example: Pairwise Independent Network (PIN) model

The Pairwise Independent Network (PIN) model is a special form of the multiterminal source model of §6.1. It is a simplified description of a wireless communication network in which every pair of terminals share a "statistically-reciprocal link" in both directions, with all such pairwise links being mutually independent.

Specifically, the rvs $X_1, \ldots, X_m$ are of the form

$$X_i\ =\ (A_{ij},\ j \in \mathcal{M}\backslash\{i\}),\ i \in \mathcal{M},$$

where the "reciprocal pairs" of finite-valued rvs (A_{ij}, A_{ji}), $1 \leqslant i < j \leqslant m$, are mutually independent. Thus, every pair of terminals is associated with a corresponding pair of rvs, with such pairs of rvs being mutually independent.

The SK capacity for the PIN model, given by Theorem 6.2 and Corollary 6.3, simplifies as shown next.

Lemma 6.11 (SK capacity of the PIN model)**.** The SK capacity for the PIN model equals

$$C_S \;=\; \min_{\pi} \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} I(A_{ij} \wedge A_{ji})\Bigg], \tag{6.11}$$

where for a fixed nontrivial partition π of $\mathcal{M}$, a pair of indices (i,j) cross π if i and j are in different atoms of π.

Proof. For every partition π of $\mathcal{M}$ in (6.3),

$$D\Big(P_{X_{\mathcal{M}}} \,||\, \prod_{k=1}^{|\pi|} P_{X_{\pi_k}}\Big) \;=\; \sum_{k=1}^{|\pi|} H\left(X_{\pi_k}\right) - H\left(X_{\mathcal{M}}\right). \tag{6.12}$$

Specializing to the PIN model, for each $1 \leqslant k \leqslant |\pi|$,

$$H\left(X_{\pi_k}\right) \;=\; \sum_{\substack{1\leqslant i<j\leqslant m:\\ i\in\pi_k,\ j\in\pi_k}} H\left(A_{ij}, A_{ji}\right) + \sum_{i\in\pi_k,\ j\notin\pi_k} H\left(A_{ij}\right),$$

and

$$\begin{aligned} H\left(X_{\mathcal{M}}\right) &= \sum_{1\leqslant i<j\leqslant m} H\left(A_{ij}, A_{ji}\right) \\ &= \sum_{k=1}^{|\pi|} \Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ i\in\pi_k,\ j\in\pi_k}} H\left(A_{ij}, A_{ji}\right) + \sum_{\substack{1\leqslant i<j\leqslant m:\\ i\in\pi_k,\ j\notin\pi_k}} H\left(A_{ij}, A_{ji}\right) \Bigg]. \end{aligned}$$

Then in (6.12),

$$\begin{aligned} D\Big(P_{X_{\mathcal{M}}} \,||\, \prod_{k=1}^{|\pi|} P_{X_{\pi_k}}\Big) &= \sum_{k=1}^{|\pi|} \Bigg[\sum_{i\in\pi_k,\ j\notin\pi_k} H\left(A_{ij}\right) - \sum_{\substack{1\leqslant i<j\leqslant m:\\ i\in\pi_k,\ j\notin\pi_k}} H\left(A_{ij}, A_{ji}\right) \Bigg] \\ &= \sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} \Big[H\left(A_{ij}\right) + H\left(A_{ji}\right) - H\left(A_{ij}, A_{ji}\right) \Big] \\ &= \sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} I(A_{ij} \wedge A_{ji}), \end{aligned}$$

whence the claim. □

6.3.1 Achieving SK capacity for the PIN model by tree packing

The formula for SK capacity of the PIN model given above is in terms of linear combinations of mutual information terms that involve pairs of "reciprocal" rvs $\big\{ (A_{ij}, A_{ji}),\ 1 \leqslant i < j \leqslant m \big\}$. Each such mutual information connotes the maximum rate of a pairwise SK that can be generated by the corresponding pair of terminals in $\mathcal{M}$ from just their own local observations using public communication. The expression in (6.11) hints at the possibility of forming a groupwide SK for all the terminals in $\mathcal{M}$ by propagating mutually independent pairwise SKs of maximum rate, building on Examples 4.4 and 4.5. As will be seen below, a maximal packing of spanning trees in a multigraph associated with the PIN model yields that the corresponding rate of packing is always a lower bound for SK capacity, which is shown to be tight. This method thereby also affords an alternative achievability proof for the SK capacity of the PIN model. In fact, it shows directly the achievability of the expression in (6.3) for this special model.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a multigraph, *i.e.*, a connected undirected graph with no self-loops and with multiple edges possible between any vertex pair, whose vertex set $\mathcal{V} = \mathcal{M} = \{1, \ldots, m\}$ and edge set $\mathcal{E} = \{e_{ij} \geqslant 0,\ 1 \leqslant i < j \leqslant m\}$, where e_{ij} is the number of edges connecting the pair of vertices $i, j,\ 1 \leqslant i < j \leqslant m$.

Definition 6.12. A spanning tree of $\mathcal{G}$ is a subgraph of $\mathcal{G}$ that is a tree and whose vertex set is $\mathcal{M}$. A spanning tree packing of $\mathcal{G}$ is any collection of edge disjoint spanning trees of $\mathcal{G}$. Let $\mu(\mathcal{G})$ denote the maximum size of such a packing.

Next, assume without any loss of generality in the PIN model that all pairwise reciprocal mutual information values $I(A_{ij} \wedge A_{ji}),\ 1 \leqslant i < j \leqslant m$, are rational numbers. Let $\mathcal{N}$ denote the collection of positive integers n such that the number of edges between any pair of vertices i, j is equal to $nI(A_{ij} \wedge A_{ji})$ is integer-valued for all $1 \leqslant i < j \leqslant m$; the elements of $\mathcal{N}$ form an arithmetic progression. Consider a sequence of multigraphs $\{\mathcal{G}^{(n)} = \big(\mathcal{M}, \mathcal{E}^{(n)}\big),\ n \in \mathcal{N}\}$, where $\mathcal{E}^{(n)}$ is such that $e_{ij} = nI(A_{ij} \wedge A_{ji})$ for n in $\mathcal{N}$. Clearly, $\frac{1}{n}\mu(\mathcal{G}^{(n)})$ is nondecreasing for n in $\mathcal{N}$. We term $\sup_{n \in \mathcal{N}} \frac{1}{n}\mu(\mathcal{G}^{(n)})$ as the maximum rate of spanning

tree packing in the multigraph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$. The connection between SK generation for the PIN model and spanning tree packing is brought out below. Specifically, we present an alternative SK generation scheme using spanning tree packing which achieves capacity for the PIN model.

The SK is generated using a two-step procedure. The first is to generate *pairwise* SKs of maximal rate between pairs of terminals in $\mathcal{M}$. In the second step, bits from these pairwise SKs are propagated to form a groupwide SK for $\mathcal{M}$ of optimal rate; this step is associated with a maximal spanning tree packing of an appropriate multigraph.

(i) Fix $\nu > 0$ that is smaller than every positive $I\left(A_{ij} \wedge A_{ji}\right),\ 1 \leqslant i < j \leqslant m$. Each pair of Terminals i, j with $I\left(A_{ij} \wedge A_{ji}\right) > 0$ generate an (ϵ_n, δ_n)-SK $K_{ij} = K_{ij}^{(n)}\left(A_{ij}^n, A_{ji}^n\right)$ of rate $\frac{1}{n}\log|\mathcal{K}_{ij}^{(n)}| = I\left(A_{ij} \wedge A_{ji}\right) - \nu$ using public communication $F_{ij} = F_{ij}^{(n)}\left(A_{ij}^{(n)}, A_{ji}^{(n)}\right)$, where $\epsilon_n \to 0$ and

$$\sigma_{\mathtt{var}}\left(K_{ij}^{(n)}; F_{ij}^{(n)}\right) \leqslant \delta_n, \text{ with } \delta_n \to 0 \text{ exponentially as } n \to \infty. \tag{6.13}$$

The existence of such rvs $K_{ij},\ 1 \leqslant i < j \leqslant m$, comes from the proof of Theorem 6.2. We note that

$$\left(K_{ij}, F_{ij}\right),\ 1 \leqslant i < j \leqslant m, \quad \text{are mutually independent.} \tag{6.14}$$

(ii) Next, consider a sequence of multigraphs $\left\{\mathcal{G}_\nu^{(n)} = \left(\mathcal{M}, \widetilde{\mathcal{E}^{(n)}}\right),\ n \in \mathcal{N}\right\}$ with $\widetilde{\mathcal{E}^{(n)}}$ assigning $\left\lfloor n\big(I\left(A_{ij} \wedge A_{ji}\right) - \nu\big)\right\rfloor$ edges to each vertex pair i, j. We claim that every spanning tree in a spanning tree packing of $\mathcal{G}_\nu^{(n)}$ can be associated with one shared bit for $\mathcal{M}$ using public communication of which the former is independent. To see this, a common vertex i corresponding to edges (i, j) and $(i, j'),\ j \neq j'$, in a spanning tree, broadcasts the binary sum of two independent bits – one each from K_{ij} and $K_{ij'}$. This enables i, j, j' to share one of these two bits, with the shared bit being independent of the binary sum. Proceeding in this manner, all the vertices in $\mathcal{M}$ – being connected by the spanning tree – can share one bit that is independent jointly of all the publicly broadcast binary sums for the tree, whence the claim. This procedure, applied repeatedly to every edge disjoint spanning tree of $\mathcal{G}_\nu^{(n)}$, generates $\mu\left(\mathcal{G}_\nu^{(n)}\right)$ shared bits for $\mathcal{M}$.

Denote by $K = K^{(n)}\left(K_{ij}^{(n)},\ 1 \leqslant i < j \leqslant m\right)$ a $\mathcal{K}^{(n)}$-valued rv corresponding to these shared bits, and by $F = F\left(K_{ij}^{(n)},\ 1 \leqslant i < j \leqslant m\right)$ the $\mathcal{F}^{(n)}$-valued communication comprising the binary sum messages above. The next result shows that the SK so generated achieves the SK capacity.

Lemma 6.13. The rv K defined above constitutes an $\left(\epsilon_n, \tilde{\delta}_n\right)$-SK for $\mathcal{M}$, where

$$\tilde{\delta}_n = m(m-1)\delta_n \tag{6.15}$$

and its rate

$$\frac{1}{n}\log\left|\mathcal{K}^{(n)}\right| \ = \ \frac{1}{n}\mu\left(\mathcal{G}_\nu^{(n)}\right)$$

satisfies

$$\begin{aligned}
&\sup_{n\in\mathcal{N}} \frac{1}{n}\mu\left(\mathcal{G}_\nu^{(n)}\right) \\
&\geqslant \ \sup_{n\in\mathcal{N}} \frac{1}{n}\mu\left(\mathcal{G}^{(n)}\right) - \frac{m(m-1)\nu}{2} \\
&\geqslant \ \min_{\pi} \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} I(A_{ij} \wedge A_{ji})\Bigg] - \frac{m(m-1)\nu}{2}.
\end{aligned} \tag{6.16}$$

Corollary 6.14. The SK capacity of the PIN model is

$$\begin{aligned}
C_S \ &= \ \sup_{n\in\mathcal{N}} \frac{1}{n}\ \mu(\mathcal{G}^{(n)}) \\
&= \ \min_{\pi} \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} I(A_{ij} \wedge A_{ji})\Bigg],
\end{aligned} \tag{6.17}$$

where the minimum is over all nontrivial partitions π of $\mathcal{M}$.

Proof. Clearly, $K = K^{(n)}\left(K_{ij}^{(n)}, 1 \leqslant i < j \leqslant m\right)$ is an ϵ_n-CR for $\mathcal{M}$. By a classic graph-theoretic result of Nash-Williams and Tutte, given a multigraph $\mathcal{G} = (\mathcal{M}, \mathcal{E})$, the maximum number of edge disjoint spanning trees that can be packed in $\mathcal{G}$ is equal to

$$\mu\left(\mathcal{G}\right) \ = \ \min_{\pi}\left\lfloor \frac{1}{|\pi|-1}\left(\text{number of edges of } \mathcal{G} \text{ that cross } \pi\right)\right\rfloor, \tag{6.18}$$

where the minimum is over all nontrivial partitions of $\mathcal{M}$. Thus,

$$\begin{aligned}
&\mu\left(\mathcal{G}_\nu^{(n)}\right) \\
&= \min_\pi \left\lfloor \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} \left\lfloor n\Big(I(A_{ij}\wedge A_{ji})-\nu\Big)\right\rfloor\Bigg]\right\rfloor \\
&\geqslant \min_\pi \left\lfloor \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} nI(A_{ij}\wedge A_{ji})\Bigg]\right\rfloor \\
&\qquad\qquad - \frac{(n\nu+1)m(m-1)}{2} - 1.
\end{aligned}$$

Note that by (6.18) the first term on the right-side above equals $\mu(\mathcal{G}^{(n)})$. Thus, for $n \in \mathcal{N}$

$$\begin{aligned}
&\frac{1}{n}\mu\left(\mathcal{G}_\nu^{(n)}\right) \\
&\geqslant \frac{1}{n}\mu\left(\mathcal{G}^{(n)}\right) - \frac{m(m-1)\nu}{2} - \frac{m(m-1)-2}{2n} \\
&\geqslant \min_\pi \frac{1}{|\pi|-1}\Bigg[\sum_{\substack{1\leqslant i<j\leqslant m:\\ (i,j) \text{ crosses } \pi}} I(A_{ij}\wedge A_{ji})\Bigg] \\
&\qquad\qquad - \frac{m(m-1)\nu}{2} - \frac{m(m-1)-4}{2n},
\end{aligned}$$

which yields (6.16) upon taking the supremum over $n \in \mathcal{N}$ on both sides.

It remains to show (6.15). Denote $K_{(1)} = (K_{ij}, 1 \leqslant i < j \leqslant m)$ and $F_{(1)} = (F_{ij}, 1 \leqslant i < j \leqslant m)$ in step (i), and let $K_u = K_u^{(n)}(K_{(1)})$ represent the pairwise SK bits in $K_{(1)}$ that remain unused after the maximal spanning tree packing $\mathcal{G}_\nu^{(n)}$ yields K by means of F in step (ii). Let $\mathcal{K}_{(1)}^{(n)}$ and $\mathcal{K}_u^{(n)}$ represent the set of values of $K_{(1)}$ and K_u, respectively. Since there is a $1-1$ mapping between the bits corresponding to the edges in any spanning tree and the shared bit for $\mathcal{M}$ together with the binary sum messages that enabled such sharing for that tree, a consequent $1-1$ mapping exists between $K_{(1)}$ and (K, F, K_u). Furthermore

$$P_{\text{unif}}^{\mathcal{K}_{(1)}^{(n)}} = P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{\text{unif}}^{\mathcal{F}^{(n)}} \times P_{\text{unif}}^{\mathcal{K}_u^{(n)}}. \tag{6.19}$$

Then

$$\begin{aligned}
\sigma_{\mathtt{var}}\left(K; F_{(1)}, F\right) &= \left\| P_{KF_{(1)}F} - P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{F_{(1)}F} \right\| \\
&\leqslant \left\| P_{KFF_{(1)}} - P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{\text{unif}}^{\mathcal{F}^{(n)}} \times P_{F_{(1)}} \right\| \\
&\quad + \left\| P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{FF_{(1)}} - P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{\text{unif}}^{\mathcal{F}^{(n)}} \times P_{F_{(1)}} \right\| \\
&= \left\| P_{KFF_{(1)}} - P_{\text{unif}}^{\mathcal{K}^{(n)}} \times P_{\text{unif}}^{\mathcal{F}^{(n)}} \times P_{F_{(1)}} \right\| \\
&\quad + \left\| P_{FF_{(1)}} - P_{\text{unif}}^{\mathcal{F}^{(n)}} \times P_{F_{(1)}} \right\| \\
&\leqslant \left\| P_{K_{(1)}F_{(1)}} - P_{\text{unif}}^{\mathcal{K}_{(1)}^{(n)}} \times P_{F_{(1)}} \right\| \\
&\quad + \left\| P_{K_{(1)}F_{(1)}} - P_{\text{unif}}^{\mathcal{K}_{(1)}^{(n)}} \times P_{F_{(1)}} \right\| \\
&= 2\sigma_{\mathtt{var}}\left(K_{(1)}; F_{(1)}\right), \qquad (6.20)
\end{aligned}$$

where the second inequality above uses $K_{(1)} \equiv (K, F, K_u)$ and (6.19). Also,

$$\begin{aligned}
&\sigma_{\mathtt{var}}\left(K_{(1)}; F_{(1)}\right) \\
&= \sigma_{\mathtt{var}}\left((K_{ij}, 1 \leqslant i < j \leqslant m); F_{ij}, 1 \leqslant i < j \leqslant m)\right) \\
&= \left\| P_{(K_{ij}, 1\leqslant i<j\leqslant m)(F_{ij}, 1\leqslant i<j\leqslant m)} - P_{(K_{ij}, 1\leqslant i<j\leqslant m)} P_{(F_{ij}, 1\leqslant i<j\leqslant m)} \right\| \\
&= \left\| \prod_{1\leqslant i<j\leqslant m} P_{K_{ij}F_{ij}} - \prod_{1\leqslant i<j\leqslant m} P_{K_{ij}} P_{F_{ij}} \right\| \\
&\leqslant \sum_{1\leqslant i<j\leqslant m} \left\| P_{K^{ij}F^{ij}} - P_{K^{i-1,j-1}F^{i-1,j-1}} P_{K_{ij}} P_{F_{ij}} \right\| \\
&= \sum_{1\leqslant i<j\leqslant m} \left\| P_{K_{ij}F_{ij}} - P_{K_{ij}} P_{F_{ij}} \right\| \\
&= \sum_{1\leqslant i<j\leqslant m} \sigma_{\mathtt{var}}\left(K_{ij}; F_{ij}\right) \\
&\leqslant \frac{m(m-1)}{2}\delta_n, \qquad (6.21)
\end{aligned}$$

where the first inequality above uses Lemma 2.1 and (6.14), and the last inequality is by (6.13).

Gathering (6.20) and (6.21),

$$\sigma_{\mathtt{var}}\left(K; F_{(1)}, F\right) \;\leqslant\; \tilde{\delta}_n = m(m-1)\delta_n$$

which is (6.15)

The corollary holds by Lemma 6.11 since $\nu > 0$ was arbitrary. □

Remark 6.15. If the joint pmf of $P_{K_{(1)}F_{(1)}}$ equals

$$P_{K_{(1)}F_{(1)}} = P_{\mathrm{unif}}^{\mathcal{K}_{(1)}^{(n)}} \times P_{F_{(1)}} \tag{6.22}$$

with $P_{\mathrm{unif}}^{\mathcal{K}_{(1)}^{(n)}}$ as in (6.19), then (6.20) asserts that the SK generation scheme above converts perfect pairwise SKs $K_{(1)} = (K_{ij}, 1 \leqslant i < j \leqslant m)$ into a perfect SK K.

6.4 Story of results and open problems

SK generation for the two-terminal source model was introduced by Maurer in [54] and SK capacity characterized in [54, 1]. The multiterminal model with $m \geqslant 2$ terminals described above and the characterization of SK capacity in Theorem 6.2 are from [21]. The "shared information" form of SK capacity in Corollary 6.3 was derived as an upper bound for SK capacity in [21] and shown to be tight in [13]. The common randomness-based approach for analyzing SK rates pursued here was introduced by Ahlswede and Csiszár in [1, 2]. The specific connection to the data compression problem of omniscience generation was identified in [21]; our treatment closely follows [21]. Note that the omniscience setting is related to a general source network problem with no helpers considered in [18, 34]. However, since we allow interaction, the standard converse of [18, 34] does not apply here. Nevertheless, our achievability scheme entails simple communication and can be seen as a special case of general achievability in [18, 34]. A two-terminal version of the omniscience problem was considered first in [26]. A multiterminal version in a slightly different setting appeared in [99]. A strong converse for SK capacity with $\delta < (1 - \sqrt{\epsilon})^2$ was shown in [80], and the complete strong converse (for all $\epsilon + \delta < 1$) stated here appeared in [85, 87]. The PIN model of §6.3.1 was introduced in [103] and its SK

capacity characterized in this chapter is from [61, 59]. Variants of this model have been considered in [10, 11, 14].

Extensions and open problems. The multiterminal source model of this chapter is a reduction of a slightly broader "helper" model in [21] that entails SK generation by a subset of terminals in $\mathcal{A} \subseteq \mathcal{M} = \{1, \ldots, m\}$ with the cooperation of the remaining terminals in $\mathcal{A}^c = \mathcal{M} \backslash \mathcal{A}$ which participate in the public discussion. The setting in §6.1 has $\mathcal{A} = \mathcal{M}$. A further generalization is the "privacy" model in which the secret CR generated must be concealed simultaneously from an eavesdropper observing the public communication and from a specified subset $\mathcal{D} \subseteq \mathcal{A}^c$ of the helper terminals. The maximum rate of such a secret CR, termed the *private key* capacity, was characterized in [21] in terms of fractional partitions. However, even for the helper model with $\mathcal{A} \subset \mathcal{M}$, an analog of the partition-based expression in Corollary 6.3 (which now requires each feasible partition $\pi = \left(\pi_1, \ldots, \pi_{|\pi|}\right)$ of $\mathcal{M}$ to satisfy $\pi_k \cap \mathcal{A} \neq \varnothing,\ 1 \leqslant k \leqslant |\pi|,\ 2 \leqslant |\pi| \leqslant |\mathcal{A}|$) is an upper bound that is not tight, in general (see [10]). Since the strong converse proof given in §6.2.3 relies on such a partition-based upper bound, full strong converses for the helper model with $\mathcal{A} \subset \mathcal{M}$ and for the privacy model remain open. Nevertheless, the partial strong converse of [80], which is valid only for $\delta < (1 - \sqrt{\epsilon})^2$ with $\mathcal{A} = \mathcal{M}$, extends to the helper model.

A more comprehensive construct has the eavesdropper, in addition to observing the public discussion, also possessing side information Z^n that has a given joint pmf with $X_{\mathcal{M}}^n$. Termed the "wiretapper" model, the maximum rate of a SK for it is called *wiretap secret key* (WSK) capacity. A single-letter characterization of WSK capacity, in general, remains defiant even for the case $m = 2$ initiated in [54, 1]. Private key capacity forms an obvious upper bound that is tight only in special cases.

Furthermore, upper and lower bounds for WSK capacity that coincide with restricted interaction and side information were derived in [54, 1, 67] and for the multiterminal generalization in [21, 28]. In particular, an operational characterization of WSK capacity was given in [28], similar to that in [21], where it was shown to equal $H(X_{\mathcal{M}})$ minus the minimum communication rate for omniscience at a neutral observer

with side information Z^n.

Another direction involves the cooperative generation of groupwise SKs for different subsets of terminals in a multiterminal source model, with an SK for each group being concealed from other groups in addition to an eavesdropper observing the communication. An associated SK capacity region of achievable SK rate-tuples remains an open problem, in general; special cases have been considered in [101, 32].

SK generation for a multiterminal Gaussian source model was considered in [60] and SK capacity characterized. The strong converse in [87] extends to this model as well, albeit for communication protocols taking countably many values. On the other hand, the partial strong converse of [80] is valid for continuous-valued sources with general distribution and admitting general measurable interactive protocols. While the results of [60] hold for the helper model as well, these two strong converse results do not. Thus, even a partial strong converse is not available for the multiterminal helper model with continuous-valued sources and general measurable interactive protocols.

Finally, SK capacity for a two-terminal continuous-valued source model as well as the second order asymptotic term in SK-length for n i.i.d. observations, have been derived recently in [38, 39]. Interestingly, the achievability scheme therein uses interactive communication. In fact, examples exhibiting the necessity of interaction in such regimes are available. Complete analogs of these results for the wiretap model and the multiterminal model with $m \geqslant 2$ terminals remain open.

7

Minimum Communication for Secret Key Capacity

Considering a two-terminal source model, the minimum rate of interactive communication needed to generate a secret key of maximum rate is examined, and is shown to be related to a new interactive variant of Wyner's common information. Interpreting the latter for a pair of random variables as the minimum rate of a function of their i.i.d. repetitions that makes them conditionally independent, interactive common information is defined in §7.1 as its restriction when said function is a common randomness together with the interactive communication that generated it. Interactive common information which, unlike Wyner's common information, does not have a known single-letter formula, is shown to correspond to common randomness generated when a secret key of optimal rate is achieved using interactive communication of minimal rate. The consequent nonsingle-letter characterization of the desired minimum communication rate, stated in §7.2, is proved in §7.3. The necessity of interaction in communication for minimality is illustrated also in §7.2.

7.1 Communication and common randomness for secret keys

Consider SK agreement in the source model of §6.1 for the special case of two terminals, *i.e.*, when $m = 2$. Terminals 1 and 2 observe, respectively, the rvs X_1^n and X_2^n where (X_1^n, X_2^n) are n i.i.d. repetitions of (X_1, X_2). Consider an r-round interactive communication $\mathbf{F}$ as in Definition 3.1 of §3.1. We restrict ourselves to deterministic communication, *i.e.*, with the local randomization $U_1 = U_2 = \varnothing$. The rate of this communication is defined to be the maximum number of bits per observation transmitted in implementation of the corresponding communication protocol. Formally, denoting by $\|\mathbf{F}\|$ the cardinality of the range of the rv $\mathbf{F}$, the rate of communication $\mathbf{F}$ is given by $\frac{1}{n}\log\|\mathbf{F}\|$.

We seek to determine the minimum (asymptotic) rate of interactive communication required for generating an SK of maximum rate. Our approach entails characterizing the CR (see §3.2) generated when the terminals agree on such an SK, and relating the rate of communication to the rate of CR for a maximum rate SK. In particular, the minimum rate of CR generated in maximum rate SK agreement constitutes a *common information* (CI) quantity which is related closely to the Wyner's CI.

We interpret Wyner's CI for a pair of rvs (X_1, X_2) as the minimum rate of a function of their i.i.d. repetitions (X_1^n, X_2^n) that renders X_1^n and X_2^n conditionally independent.

Definition 7.1. A number $R \geqslant 0$ is an achievable CI rate if for every $0 < \epsilon < 1$ there exists an $n \geqslant 1$ and a finite-valued rv $L = L(X_1^n, X_2^n)$ of rate $(1/n)H(L) \leqslant R + \epsilon$ that satisfies the property

$$\frac{1}{n} I(X_1^n \wedge X_2^n \mid L) \leqslant \epsilon. \tag{7.1}$$

Obvious examples of such an rv are $L = (X_1^n, X_2^n)$ or X_1^n or X_2^n. The infimum of all achievable CI rates, denoted $CI_W(X_1 \wedge X_2)$, is called the CI of X_1 and X_2. The following theorem, stated without proof, characterizes $CI_W(X_1 \wedge X_2)$.

Theorem 7.2. The CI of rvs X_1, X_2 is

$$CI_W(X_1 \wedge X_2) = \min_W I(X_1, X_2 \wedge W), \tag{7.2}$$

where the rv W takes values in a finite set $\mathcal{W}$ with $|\mathcal{W}| \leqslant |\mathcal{X}_1||\mathcal{X}_2|$ and satisfies the Markov condition $X_1 \multimap W \multimap X_2$.

Note that for every W such that $X_1 \multimap W \multimap X_2$, we have

$$\begin{aligned} I(X_1, X_2 \wedge W) &= H(X_1, X_2) - H(X_1|W) - H(X_2|W) \\ &\geqslant H(X_1, X_2) - H(X_1|X_2) - H(X_2|X_1) \\ &= I(X_1 \wedge X_2), \end{aligned}$$

so that $CI_W(X_1 \wedge X_2) \geqslant I(X_1 \wedge X_2)$.

We shall see that the minimum rate of an interactive communication required to generate a maximum rate SK will be related closely to an interactive variant of Wyner's CI, defined next.

Definition 7.3. An achievable *r-interactive* CI rate is defined in a manner analogous to achievable CI rate, but with the restriction that the rvs L in (7.1) be ϵ-CR, *i.e.*, $L = (J, \mathbf{F})$, where $\mathbf{F}$ is an r-interactive communication and J is ϵ-recoverable from[1] $\mathbf{F}$. The infimum of all achievable r-interactive CI rates, denoted $CI_i^r(X_1; X_2)$, is called the r-interactive CI of the rvs X_1 and X_2. By definition, the nonnegative sequence $\{CI_i^r(X_1; X_2)\}_{r=1}^{\infty}$ is nonincreasing in r and is bounded below by $CI_W(X_1 \wedge X_2)$. Define interactive CI of the rvs X_1, X_2 as

$$CI_i(X_1 \wedge X_2) = \lim_{r\to\infty} CI_i^r(X_1; X_2).$$

Then, $CI_i(X_1 \wedge X_2) \geqslant CI_W(X_1 \wedge X_2) \geqslant 0$. Note that $CI_i^r(X_1; X_2)$ may not be symmetric in X_1 and X_2 since the communication is initiated at terminal $\mathcal{X}_1$. However, since

$$CI_i^{r+1}(X_1; X_2) \leqslant CI_i^r(X_2; X_1) \leqslant CI_i^{r-1}(X_1; X_2),$$

clearly,

$$\begin{aligned} CI_i(X_1 \wedge X_2) &= \lim_{r\to\infty} CI_i^r(X_1; X_2) \\ &= \lim_{r\to\infty} CI_i^r(X_2; X_1) \\ &= CI_i(X_2 \wedge X_1). \end{aligned} \tag{7.3}$$

[1]Since the set $\mathcal{M} = \{1, 2\}$ remains fixed in this chapter, we simply use the phrase "CR using $\mathbf{F}$" in place of "CR for $\mathcal{M}$ using $\mathbf{F}$"; see Definition 3.7.

Furthermore, for all $0 < \epsilon < 1$, $J = {X_1}^n$ is ϵ-recoverable from ${X_2}^n$ and a communication (of a Slepian-Wolf codeword) $F = F\,(X_1^n)$, and $L = (J, F)$ satisfies (7.1). Hence, $CI_i(X_1 \wedge X_2) \leqslant H(X_1)$; similarly, $CI_i(X_1 \wedge X_2) \leqslant H(X_2)$. To summarize, we have

$$0 \leqslant CI_W(X_1 \wedge X_2) \leqslant CI_i(X_1 \wedge X_2) \leqslant \min\{H(X_1), H(X_2)\}, \quad (7.4)$$

where the first and the last inequalities can be strict. In §7.2 we show an instance where the second inequality is strict.

The notion of r-interactive CI plays a pivotal role in maximum rate SK generation and is related closely to the minimum rate of an r-round interactive communication needed for generating such an SK. Loosely speaking, the results of the next section assert the following: A CR that satisfies (7.1) can be used to generate a maximum rate SK and conversely, a maximum rate SK yields a CR satisfying (7.1). In fact, such a CR of rate R can be recovered from an interactive communication of rate $R - C_S$, where C_S is the SK capacity for a two-terminal source model with rvs X_1, X_2 described in Chapter 6. Therefore, to find the minimum rate of interactive communication needed to generate a maximum rate SK, it is sufficient to characterize $CI_i(X_1 \wedge X_2)$.

To state these results, we need a formal definition of aforementioned optimal rates of communication.

Definition 7.4. A rate $R' \geqslant 0$ is an achievable *r-interactive communication rate for CI_i^r* if, for all $0 < \epsilon < 1$, there exists for some $n \geqslant 1$, an r-round interactive communication $\mathbf{F}$ of rate $(1/n) \log \|\mathbf{F}\| \leqslant R' + \epsilon$, and an ϵ-CR J using $\mathbf{F}$, with $L = (J, \mathbf{F})$ satisfying (7.1). Let R_{CI}^r denote the infimum of all achievable r-interactive communication rates for CI_i^r. Similarly, $R'' \geqslant 0$ is an achievable *r-interactive communication rate for SK capacity* if, for all $0 < \epsilon < 1$, there exists, for some $n \geqslant 1$, an r-round interactive communication $\mathbf{F}$ of rate $(1/n) \log \|\mathbf{F}\| \leqslant R'' + \epsilon$, and an[2] (ϵ, ϵ)-SK K, recoverable from $\mathbf{F}$, of rate

$$(1/n)H(K) \geqslant I(X_1 \wedge X_2) - \epsilon.$$

Let R_{SK}^r denotes the infimum of all achievable r-interactive communication rates for SK capacity. Note that by their definitions, both R_{CI}^r

[2]For convenience, epsilontics in this chapter involve a single vanishing ϵ.

and R^r_{SK} are nonincreasing with increasing r, and are bounded below by zero. Define

$$R_{CI} = \lim_{r\to\infty} R^r_{CI}, \quad R_{SK} = \lim_{r\to\infty} R^r_{SK}.$$

Although $R^r_{CI}(X_1; X_2)$ and $R^r_{SK}(X_1; X_2)$ are not equal to $R^r_{CI}(X_2; X_1)$ and $R^r_{SK}(X_2; X_1)$, respectively, the quantities R_{CI} and R_{SK} are symmetric in X_1 and X_2 using an argument similar to that leading to (7.3).

7.2 Communication rate for secret key capacity

The result below characterizes the minimum rate of interactive communication required for generating a maximum rate SK. Specially, it provides a structural characterization of R^r_{SK} and shows that it equals R^r_{CI}. Furthermore, it provides a characterization of the latter quantity in terms of $CI^r_i(X_1, X_2)$

Theorem 7.5. For every $r \geqslant 1$,

$$R^r_{SK} = R^r_{CI} = CI^r_i(X_1; X_1) - I(X_1 \wedge X_2). \tag{7.5}$$

Corollary 7.6 (Communication requirements for attaining SK capacity)**.** It holds that

$$R_{SK} = R_{CI} = CI_i(X_1 \wedge X_2) - I(X_1 \wedge X_2). \tag{7.6}$$

Remark. The relation (7.6) can be interpreted as follows. Any CR J recoverable from interactive communication $\mathbf{F}$, with $L = (J, \mathbf{F})$ satisfying (7.1), can be decomposed into two mutually independent parts: An SK K of maximum rate and the interactive communication $\mathbf{F}$. It follows upon rewriting (7.6) as $CI_i(X_1 \wedge X_2) = I(X_1 \wedge X_2) + R_{CI}$ that the communication $\mathbf{F}$ is (approximately) of rate R_{CI}. Furthermore, R_{CI} is the same as R_{SK}.

A computable characterization of the operational term $CI_i(X_1 \wedge X_2)$ is not known. However, the next result, stated without proof, gives a single-letter characterization of $CI^r_i(X_1; X_2)$.

Theorem 7.7. Given rvs X_1, X_2 and $r \geqslant 1$, we have

$$CI_i^r(X_1;X_2) = \min_{U_1,...,U_r} I(X_1, X_2 \wedge U_1, ..., U_r), \tag{7.7}$$

where the minimum is taken over rvs $U_1, ..., U_r$ taking values in finite sets $\mathcal{U}_1, ..., \mathcal{U}_r$, respectively, that satisfy the following conditions

$$\begin{aligned}
&(P1)\ U_{2i+1} \multimap X_1, U^{2i} \multimap X_2, \quad 0 \leqslant i \leqslant \lfloor (r-1)/2 \rfloor,\\
&\qquad U_{2i} \multimap X_2, U^{2i-1} \multimap X_1, \quad 1 \leqslant i \leqslant \lfloor r/2 \rfloor,\\
&(P2)\ X_1 \multimap U^r \multimap X_2,\\
&(P3)\ |\mathcal{U}_{2i+1}| \leqslant |\mathcal{X}_1| \prod_{j=1}^{2i} |\mathcal{U}_j| + 1, \quad 0 \leqslant i \leqslant \lfloor (r-1)/2 \rfloor,\\
&\qquad |\mathcal{U}_{2i}| \leqslant |\mathcal{X}_2| \prod_{j=1}^{2i-1} |\mathcal{U}_j| + 1, \quad 1 \leqslant i \leqslant \lfloor r/2 \rfloor,
\end{aligned}$$

with $\mathcal{U}_0 = \varnothing$ and $U_0 =$ constant.

Remark 7.8. Note that (7.7) has the same form as the expression for $CI_W(X_1 \wedge X_2)$ in (7.2) with W replaced by $(U_1, ..., U_r)$ satisfying the conditions above.

Remark 7.9. Theorem 7.5 complies with a central tenet of this monograph: SK generation is linked intrinsically to the efficient generation of CR. It illustrates this connection formally for the extreme case of a maximum rate SK. In general, it is of interest to study the connection between SK agreement and CR generation using an interactive communication of rate R. For $\rho \geqslant 0$, a rate $R \geqslant 0$ is an achievable CR rate for ρ if for every $0 < \epsilon < 1$ there exists for some $n \geqslant 1$ an ϵ-CR L using an r-round interactive communication $\mathbf{F}$, with

$$\frac{1}{n} H(L) \geqslant R - \epsilon,$$

of rate

$$\frac{1}{n} H(\mathbf{F}) \leqslant \rho + \epsilon.$$

Denote by $CR(\rho)$ the maximum achievable CR rate for ρ. Similarly, denote by $C(\rho)$ the maximum rate of an SK that can be generated

using a communication as above. It can be shown in a straightforward manner that

$$C(\rho) = CR(\rho) - \rho. \tag{7.8}$$

The graph of CR as a function of ρ is plotted in Fig. 7.1. We note that

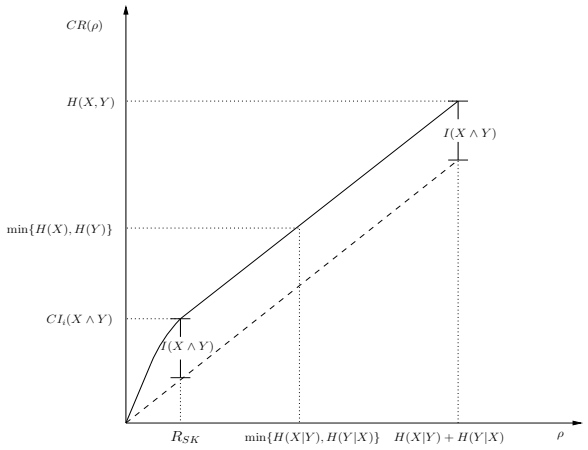

Figure 7.1: Minimum rate of communication R_{SK} for maximum rate SK generation

$CR(\rho)$ is an increasing and a concave function of ρ, as can be seen from a simple time-sharing argument. Since R_{SK} is the minimum rate of communication required to generate a maximum rate SK, $CR(\rho) - \rho = I(X \wedge Y)$ for $\rho \geqslant R_{SK}$. Thus, our results characterize the graph of $CR(\rho)$ for all $\rho \geqslant R_{SK}$. The quantity R_{SK} is the minimum value of ρ for which the slope of $CR(\rho)$ is 1; $CR\left(R_{SK}\right)$ is equal to the interactive common information $CI_i(X \wedge Y)$.

Remark 7.10 (Can interaction reduce the communication rate?)**.** In Theorem 6.2, we saw that it suffices to use simple communication to achieve SK capacity, *i.e.*, interaction is not helpful in improving the (asymptotic) maximum rate of SK generation. But can it help in reducing the rate of communication required for achieving SK capacity? Theorem 7.7 provides a tool for answering this question. By Theorem 7.5,

interaction will help in reducing the rate of communication required for attaining SK capacity iff $\min\{CI_i^r(X_1;X_2), CI_i^r(X_2;X_1)\}$ strictly decreases for some $r \geqslant 1$. Theorem 7.7 provides an expression for $\min\{CI_i^r(X_1;X_2), CI_i^r(X_2;X_1)\}$ which can analyzed for specific examples. In fact, a manipulation of this expression can be used to show that R_{SK} can be attained using simple communication for the case of binary symmetric rvs X_1, X_2 with

$$P_{X_1}(1) = 0.5 \text{ and } P_{X_2|X_1}(1|1) = P_{X_2|X_1}(0|0).$$

On the other hand, it can be shown that

$$\min\{CI_i^1(X_1;X_2), CI_i^1(X_2;X_1)\} > CI_i(X_1 \wedge X_2)$$

for ternary X_1, X_2 with joint pmf given by

$$\begin{bmatrix} a & a & a \\ b & a & a \\ a & c & a \end{bmatrix},$$

where a, b, c are nonnegative and satisfy

$$7a + b + c = 1, \quad c \neq a, \quad \text{and } 2a > b > a.$$

Thus, remarkably, interaction does help in reducing the rate of communication needed for generating a maximum rate SK. In fact, as will be clear from the proof of Theorem 7.5, the example above illustrates that it is not always optimal (from a viewpoint of communication rate) to extract a maximum rate SK from a CR $L = X_1$ or $L = X_2$.

7.3 Proof by randomness decomposition

We turn to the proof of Theorem 7.5. The main idea behind the proof is characterizing the CR generated when the terminals agree on a maximum rate SK. Heuristically, we show that the CR generated yields a maximum rate SK iff the observations of the two terminals are conditionally independent given the CR. Thus, the overall randomness in the system *decomposes* into two independent parts: first, the CR generated by the terminals, and second, residual independent local randomness

available at the terminals. This approach of decomposing the overall randomness into various independent parts, with operational significance, was used in our derivation of SK capacity in Chapter 6 as well, and is a common theme underlying many proofs in this monograph.

To prove Theorem 7.5, we begin by making a few simple observations. The first is a basic property of interactive communication for two terminals seen already in (3.2).

Lemma 7.11. Let $m = 2$. For an interactive communication $\mathbf{F}$, we have

$$H(\mathbf{F}) \geqslant H\left(\mathbf{F}|X_1^n\right) + H\left(\mathbf{F}|X_2^n\right).$$

In fact, an interactive communication can be compressed to a rate approximately equal to the right-side above.

Lemma 7.12. For an r-interactive communication $\mathbf{F}$, define

$$\mathbf{F}_i = \mathbf{F}\left(X_{1(n(i-1)+1)}^{ni}, X_{(2n(i-1)+1)}^{ni}\right), \quad 1 \leqslant i \leqslant k.$$

Then, for all $k \geqslant k_0(n, \epsilon, |\mathcal{X}_1|, |\mathcal{X}_2|)$ there exists an r-interactive communication $\mathbf{F}' = \mathbf{F}'\left(X_1^{nk}, X_2^{nk}\right)$ of rate

$$\frac{1}{nk}\log\|\mathbf{F}'\| \leqslant \frac{1}{n}\left[H\left(\mathbf{F}|X_1^n\right) + H\left(\mathbf{F}|X_2^n\right)\right] + \epsilon, \tag{7.9}$$

such that $\mathbf{F}^k = (\mathbf{F}_1, ..., \mathbf{F}_k)$ is an ϵ-CR using $\mathbf{F}'$.

Proof. Using a special case of Proposition 6.7 applied to the case of two terminals, there exist mappings $f_1, ..., f_r$ of $F_1^k, ..., F_r^k$, respectively, of rates[3]

$$\frac{1}{k}\log\|f_{2i+1}\| \leqslant H(F_{2i+1} \mid X_2^n, F_1, ..., F_{2i}) + \frac{n\epsilon}{2r}, \quad 0 \leqslant i \leqslant \lfloor (r-1)/2 \rfloor,$$
$$\frac{1}{k}\log\|f_{2i}\| \leqslant H(F_{2i} \mid X_1^n, F_1, ..., F_{2i-1}) + \frac{n\epsilon}{2r}, \quad 1 \leqslant i \leqslant \lfloor r/2 \rfloor,$$

such that

$$F_{2i+1}^k \text{ is } \frac{\epsilon}{2r}\text{-recoverable from } \left(f_{2i+1}(F_{2i+1}^k), X_2^N, F_1^k, ..., F_{2i}^k\right),\ 0 \leqslant i \leqslant \lfloor (r-1)/2 \rfloor,$$
$$F_{2i}^k \text{ is } \frac{\epsilon}{2r}\text{-recoverable from } \left(f_{2i}(F_{2i}^k), X_1^N, F_1^k, ..., F_{2i-1}^k\right),\ 1 \leqslant i \leqslant \lfloor r/2 \rfloor,$$

[3]Such optimal rate codes obtained by the random binning argument in the proof of Proposition 6.7 will be referred to as "Slepian-Wolf" codes.

for all k sufficiently large. Thus, the communication $\mathbf{F}'$ given by $F_i' = f_i\left(F_i^k\right)$, $1 \leqslant i \leqslant r$, constitutes the required communication of rate

$$\frac{1}{nk}\log\|\mathbf{F}'\| \leqslant \frac{1}{n}\left[H\left(\mathbf{F}|X_1^n\right) + H\left(\mathbf{F}|X_2^n\right)\right] + \epsilon.$$

□

The following simple observation lies at the heart of the proof of Theorem 7.5 and leads to the aforementioned randomness decomposition.

Lemma 7.13. (*A General Decomposition*) For a CR J using an interactive communication $\mathbf{F}$, we have

$$\begin{aligned} nI(X_1 \wedge X_2) = I\left(X_1^n \wedge X_2^n \mid J, \mathbf{F}\right) + H(J, \mathbf{F}) - H\left(\mathbf{F} \mid X_1^n\right) \\ - H\left(\mathbf{F} \mid X_2^n\right) - H\left(J \mid X_1^n, \mathbf{F}\right) - H\left(J \mid X_2^n, \mathbf{F}\right). \end{aligned} \tag{7.10}$$

Proof. For $T = T\left(X_1^n, X_2^n\right)$, we have

$$\begin{aligned} & nI(X_1 \wedge X_2) \\ &= H\left(X_1^n, X_2^n\right) - H\left(X_1^n \mid X_2^n\right) - H\left(X_2^n \mid X_1^n\right) \\ &= H\left(X_1^n, X_2^n \mid T\right) - H\left(X_1^n \mid X_2^n, T\right) - H\left(X_2^n \mid X_1^n, T\right) \\ &\quad + H(T) - H\left(T \mid X_1^n\right) - H\left(T \mid X_2^n\right) \\ &= I\left(X_1^n \wedge X_2^n \mid T\right) + H(T) - H\left(T \mid X_1^n\right) - H\left(T \mid X_2^n\right). \end{aligned}$$

Lemma 7.13 follows upon choosing $T = (J, \mathbf{F})$. □

Note that a simplification of (7.10) gives

$$\begin{aligned} I(X_1 \wedge X_2) \leqslant \frac{1}{n}\Bigg[I\left(X_1^n \wedge X_2^n \mid J, \mathbf{F}\right) + H(J, \mathbf{F}) \\ - H\left(\mathbf{F} \mid X_1^n\right) - H\left(\mathbf{F} \mid X_2^n\right) \Bigg]. \end{aligned} \tag{7.11}$$

If J is an ϵ-CR using $\mathbf{F}$, Fano's inequality implies

$$\frac{1}{n}\left[H\left(J \mid X_1^n, \mathbf{F}\right) + H\left(J \mid X_2^n, \mathbf{F}\right)\right] \leqslant 2\epsilon \log|\mathcal{X}_1||\mathcal{X}_2| + 2h(\epsilon), \tag{7.12}$$

where $\delta(\epsilon) \triangleq 2\epsilon \log |\mathcal{X}_1||\mathcal{X}_2| + 2h(\epsilon) \to 0$ as $\epsilon \to 0$. Combining (7.10), (7.12) we get

$$I(X_1 \wedge X_2) \geqslant \frac{1}{n}\Big[I\left(X_1^n \wedge X_2^n \mid J, \mathbf{F}\right) + H(J, \mathbf{F}) - H\left(\mathbf{F} \mid X_1^n\right) - H\left(\mathbf{F} \mid X_2^n\right) \Big] - \delta(\epsilon), \quad (7.13)$$

and further, by Lemma 7.11,

$$I(X_1 \wedge X_2) \geqslant \frac{1}{n}\left[I\left(X_1^n \wedge X_2^n \mid J, \mathbf{F}\right) + H(J, \mathbf{F}) - H(\mathbf{F}) \right] - \delta(\epsilon). \quad (7.14)$$

Proof of Theorem 7.5.

We now prove (7.5). The proof of (7.6) then follows upon taking the limit $r \to \infty$ on both sides of (7.5). The proof of (7.5) follows from claims 1–3 below. In particular, the proofs of these claims establish a structural equivalence between a maximum rate SK and an SK of rate $\approx \frac{1}{n} H(J \mid \mathbf{F})$ extracted from a CR J using $\mathbf{F}$ such that $L = (J, \mathbf{F})$ satisfies (7.1).

Claim 1: $R_{CI}^r \geqslant CI_i^r(X_1; X_2) - I(X_1 \wedge X_2)$.

Proof. By the definition of R_{CI}^r, for every $0 < \epsilon < 1$ there exists, for some $n \geqslant 1$, an r-interactive communication $\mathbf{F}$ of rate

$$\frac{1}{n} \log \|\mathbf{F}\| \leqslant R_{CI}^r + \epsilon, \quad (7.15)$$

and J, an ϵ-CR J using $\mathbf{F}$, such that $L = (J, \mathbf{F})$ satisfies (7.1). It follows from (7.14) that

$$\frac{1}{n} H(J, \mathbf{F}) \leqslant I(X_1 \wedge X_2) + \frac{1}{n} H(\mathbf{F}) + \delta(\epsilon),$$

which with (7.15) gives

$$\frac{1}{n} H(J, \mathbf{F}) \leqslant I(X_1 \wedge X_2) + R_{CI}^r + \epsilon + \delta(\epsilon). \quad (7.16)$$

Since $(J, \mathbf{F})$ satisfies

$$\frac{1}{n} I\left(X_1^n \wedge X_2^n \mid J, \mathbf{F}\right) \leqslant \epsilon \leqslant \epsilon + \delta(\epsilon),$$

the inequality (7.16), along with the fact that $\epsilon + \delta(\epsilon) \to 0$ as $\epsilon \to 0$, implies that $I(X_1 \wedge X_2) + R_{CI}^r$ is an achievable r-interactive CI rate; hence, $CI_i^r(X_1; X_2) \leqslant I(X_1 \wedge X_2) + R_{CI}^r$.

Claim 2: $R_{SK}^r \geqslant R_{CI}^r$.

Proof. Using the definition of R_{SK}^r, for $0 < \epsilon < 1$ there exists, for some $n \geqslant 1$, an r-interactive communication $\mathbf{F}$ of rate $\frac{1}{n} \log \|\mathbf{F}\| \leqslant R_{SK}^r + \epsilon$ and an ϵ-SK K recoverable from $\mathbf{F}$ of rate

$$\frac{1}{n} H(K) \geqslant I(X_1 \wedge X_2) - \epsilon. \tag{7.17}$$

By choosing $J = K$ in (7.14) and rearranging the terms we get,

$$\frac{1}{n} I\left(X_1^n \wedge X_2^n \mid K, \mathbf{F}\right) \leqslant I(X_1 \wedge X_2) - \frac{1}{n} H(K \mid \mathbf{F}) + \delta(\epsilon).$$

Next, from $(1/n) I(K \wedge \mathbf{F}) < \epsilon$, we have

$$\begin{aligned} \frac{1}{n} I\left(X_1^n \wedge X_2^n \mid K, \mathbf{F}\right) &\leqslant I(X_1 \wedge X_2) - \frac{1}{n} H(K) + \epsilon + \delta(\epsilon) \\ &\leqslant 2\epsilon + \delta(\epsilon), \end{aligned}$$

where the last inequality follows from (7.17). Since $2\epsilon + \delta(\epsilon) \to 0$ as $\epsilon \to 0$, R_{SK}^r is an achievable r-interactive communication rate for CI_i^r, and thus, $R_{SK}^r \geqslant R_{CI}^r$.

Claim 3: $R_{SK}^r \leqslant CI_i^r(X_1; X_2) - I(X_1 \wedge X_2)$.

Proof. For $0 < \epsilon < 1$, let J be an ϵ-CR recoverable from an r-interactive communication $\mathbf{F}$, with

$$\frac{1}{n} H(J, \mathbf{F}) \leqslant CI_i^r(X_1; X_2) + \epsilon, \tag{7.18}$$

such that $L = (J, \mathbf{F})$ satisfies (7.1). So by (7.11),

$$\begin{aligned} \frac{1}{n} \left[H(\mathbf{F} \mid X_1^n) + H(\mathbf{F} \mid X_2^n)\right] &\leqslant \frac{1}{n} H(J, \mathbf{F}) - I(X_1 \wedge X_2) + \epsilon \\ &\leqslant CI_i^r(X_1; X_2) - I(X_1 \wedge X_2) + 2\epsilon. \end{aligned} \tag{7.19}$$

To prove the assertion in claim 3, we show that for some $N \geqslant 1$ there exists a $\Delta(\epsilon)$-SK $K = K(X_1^N, X_2^N)$ of rate

$$\frac{1}{n} \log \|K\| \geqslant I(X_1 \wedge X_2) - \Delta(\epsilon)$$

recoverable from an r-interactive communication $\mathbf{F}'' = \mathbf{F}''(X_1^N, X_2^N)$ of rate

$$\frac{1}{N}\log\|\mathbf{F}''\| \leqslant \frac{1}{n}\left[H(\mathbf{F} \mid X_1^n) + H(\mathbf{F} \mid X_2^n)\right] + \Delta(\epsilon) - 2\epsilon, \tag{7.20}$$

where $\Delta(\epsilon) \to 0$ as $\epsilon \to 0$. Then (7.20), along with (7.19), would yield

$$\frac{1}{N}\log\|\mathbf{F}''\| \leqslant CI_i^r(X_1; X_2) - I(X_1 \wedge X_2) + \Delta(\epsilon), \tag{7.21}$$

so that $CI_i^r(X_1; X_2) - I(X_1 \wedge X_2)$ is an achievable r-interactive communication rate for SK capacity, thereby establishing the claim.

It remains to find K and $\mathbf{F}''$ as above. To that end, let J be recovered as $J_1 = J_1(X_1^n, \mathbf{F})$ and $J_2 = J_2(X_2^n, \mathbf{F})$ by terminals 1 and 2, respectively, *i.e.*,

$$\mathbb{P}\left(J = J_1 = J_2\right) \geqslant 1 - \epsilon.$$

Further, for $k \geqslant 1$, let

$$J_{1i} = J_1\left(X_{1(n(i-1)+1)}^{ni}, \mathbf{F}_i\right), \qquad J_{2i} = J_2\left(X_{(2n(i-1)+1)}^{ni}, \mathbf{F}_i\right), \quad 1 \leqslant i \leqslant k,$$

where $\mathbf{F}_i = \mathbf{F}\left(X_{1(n(i-1)+1)}^{ni}, X_{(2n(i-1)+1)}^{ni}\right)$. For odd r, we find an r-interactive communication $\mathbf{F}''$ such that $\left(J_1^k, \mathbf{F}^k\right)$ is an ϵ-CR recoverable from $\mathbf{F}''$, for all k sufficiently large; then the SK K will be chosen to be a function of $\left(J_1^k, \mathbf{F}^k\right)$ of appropriate rate. The proof for even r is similar and is obtained by interchanging the roles of J_1 and J_2. In particular, by Lemma 7.12, for all k sufficiently large there exists an r-interactive communication $\mathbf{F}'$ such that $\mathbf{F}^k$ is ϵ-CR recoverable from $\mathbf{F}'$ of rate given by (7.9). Next, from Fano's inequality,

$$\frac{1}{n}\max\{H(J \mid J_1); H(J_1 \mid J_2)\} \leqslant \epsilon \log|\mathcal{X}_1||\mathcal{X}_2| + h(\epsilon). \tag{7.22}$$

By Proposition 6.7, there exists a mapping f of J_1^k of rate

$$\frac{1}{k}\log\|f\| \leqslant H(J_1 \mid J_2) + n\epsilon \tag{7.23}$$

such that

$$J_1^k \text{ is } \epsilon\text{-recoverable from } \left(f\left(J_1^k\right), J_2^k\right), \tag{7.24}$$

for all k sufficiently large. It follows from (7.22), (7.23) that

$$\frac{1}{nk}\log\|f\| \leqslant \epsilon + \epsilon\log|\mathcal{X}_1||\mathcal{X}_2| + h(\epsilon). \tag{7.25}$$

For $N = nk$, we define the r-interactive communication $\mathbf{F}'' = \mathbf{F}''\left(X_1^N, X_2^N\right)$ as

$$\begin{aligned} F_i'' &= F_i', \quad 1 \leqslant i \leqslant r-1, \\ F_k'' &= F_r', f(J_1^k), \quad i = r, \end{aligned}$$

Thus, $\left(J_1^k, \mathbf{F}^k\right)$ is 2ϵ-CR recoverable from $\mathbf{F}''$, where by (7.9) and (7.25) the rate of communication $\mathbf{F}''$ is bounded as

$$\begin{aligned} &\frac{1}{nk}\log\|\mathbf{F}''\| \\ &\leqslant \frac{1}{n}\left[H\left(\mathbf{F}|X_1^n\right) + H\left(\mathbf{F}|X_2^n\right)\right] + 2\epsilon + \epsilon\log|\mathcal{X}_1||\mathcal{X}_2| + h(\epsilon). \end{aligned} \tag{7.26}$$

Finally, to construct the SK $K = K\left(J_1^k, \mathbf{F}^k\right)$, using Lemma 5.18 with

$$U = (J_1, \mathbf{F}),\ V_1 = \phi,\ n = k,\ V_2 = \mathbf{F}',$$

we get from (7.26) that there exists a function K of $J_1^k, \mathbf{F}^k$ such that

$$\begin{aligned} \frac{1}{k}\log\|K\| &\geqslant H(U) - \frac{1}{k}\log\|\mathbf{F}''\| \\ &\geqslant H(J_1, \mathbf{F}) - H(\mathbf{F} \mid X_1^n) - H(\mathbf{F} \mid X_2^n) \\ &\qquad - n(2\epsilon + \epsilon\log|\mathcal{X}_1||\mathcal{X}_2| + h(\epsilon)), \end{aligned} \tag{7.27}$$

and

$$I(K \wedge \mathbf{F}') \leqslant \exp(-ck)$$

where $c > 0$, for all sufficiently large k. We get from (7.27) and (7.11) that the rate of K is bounded below as follows:

$$\begin{aligned} \frac{1}{nk}\log\|K\| \geqslant I(X_1 \wedge X_2) - \frac{1}{n}I\left(X_1^n \wedge X_2^n \mid J_1, \mathbf{F}\right) \\ - 2\epsilon - \epsilon\log|\mathcal{X}_1||\mathcal{X}_2| - h(\epsilon). \end{aligned} \tag{7.28}$$

Observe that

$$\begin{aligned} I(X_1^n \wedge X_2^n \mid J, \mathbf{F}) &= I(J_1, X_1^n \wedge X_2^n \mid J, \mathbf{F}) \\ &\geqslant I(X_1^n \wedge X_2^n \mid J, J_1, \mathbf{F}) \\ &\geqslant I(X_1^n \wedge X_2^n \mid J_1, \mathbf{F}) - H(J \mid J_1), \end{aligned}$$

which, along with (7.22) and the fact that $L = (J, \mathbf{F})$ satisfies (7.1), yields

$$\frac{1}{n} I(X_1^n \wedge X_2^n \mid J_1, \mathbf{F}) \leqslant \epsilon + \epsilon \log |\mathcal{X}_1||\mathcal{X}_2| + h(\epsilon). \tag{7.29}$$

Upon combining (7.28) and (7.29) we get,

$$\frac{1}{nk} \log \|K\| \geqslant I(X_1 \wedge X_2) - 3\epsilon - 2\epsilon \log |\mathcal{X}_1||\mathcal{X}_2| - 2h(\epsilon).$$

Thus, for $\Delta(\epsilon) = 4\epsilon + 2\epsilon \log |\mathcal{X}_1||\mathcal{X}_2| + 2h(\epsilon)$, K is a $\Delta(\epsilon)$-SK of rate $(1/nk) \log \|K\| \geqslant I(X_1 \wedge X_2) - \Delta(\epsilon)$, recoverable from r-interactive communication $\mathbf{F}''$ which, with (7.26), completes the proof. □

7.4 Story of results and open problems

SK generation with rate-limited communication was considered first in [20], preceded by a related study in [2] of CR generation by two terminals under similar restrictions. The problem is especially interesting for the two-terminal Gaussian source model where the quantization-based scheme used in [60] requires communication of unbounded rate. For this model, SK capacity using rate-limited one-way communication was characterized in [92]. This chapter closely follows [75] where the minimum rate of interactive communication for attaining SK capacity for two terminals was characterized. Lemma 7.12 is an instance of compression of interactive protocols using communication of rate equal to "intrinsic information," a topic of considerable current interest.

Extensions and open problems. Extensions of the results of this chapter to $m \geqslant 2$ terminals constitute an interesting open direction; partial results were reported in [56, 57]. In particular, a sufficient condition is provided in [57] for equality between R_{SK} and R_{CI} to hold in the PIN model and examples are given that satisfy this condition. In

another direction, the minimum communication rate is characterized in [14] for attaining SK capacity in a hypergraph PIN model using linear communication schemes which, by example, is shown to exceed the rate required by general schemes. In a similar vein, but without any restriction on communication protocols, recent work in [55] provides an upper bound for R_{SK} for a hypergraph PIN model; the bound, however, is shown to be not tight in general.

Remark 7.10 provides an instance where interaction helps reduce the communication rate for achieving SK capacity. The interesting question of characterizing conditions under which interaction is indeed helpful in achieving SK capacity remains unresolved even for two terminals. A related question concerning conditions which necessitate all the terminals to communicate in a multiterminal setting was addressed in [57].

It is also of interest to examine if the standard SK generation method (in the computer science literature) of recovering the observations of at least one of the terminals is suboptimal with regard to communication rate. In [75], it is shown that for two terminals observing the components of a binary symmetric source, *i.e.*, X_1 and X_2 such that $P_{X_1}(1) = 0.5$ and

$$P_{X_2|X_1}(1|0) = P_{X_2|X_1}(0|1) = \delta, \tag{7.30}$$

recovering the observations of one of the terminals is unavoidable in achieving SK capacity. It was conjectured that same must be true for all two-terminal models with binary sources. Interestingly, a recent work [50] claims to have resolved this conjecture in the negative, showing that it is only true when (7.30) holds.

Lastly, referring to §7.2, a characterization of $CR(\rho)$ for $\rho < R_{SK}$ is central to the characterization of $C(\rho)$, and this, along with a single-letter characterization of R_{SK}, remain open.

8

Secure Function Computation with Trusted Parties

Our formulation of secure function computation for the multiterminal source model of Chapter 6 requires the terminals to compute a given function of their collective DMMS observations using interactive communication that does not give away the function value. This setting differs materially from the celebrated classical problem in which each terminal, using its local data together with the computed function of the collective data, can glean no additional information regarding the data observed by other terminals. The question of finding a necessary and sufficient condition for a function to be computed securely by multiple terminals is posed formally in §8.1. It follows readily that the entropy of the function must not exceed the SK capacity of the multiterminal model. That this condition is also sufficient is shown in §8.2 by an achievability proof that relies on the techniques of Chapter 5. A single-shot general necessary condition for secure computability is derived in §8.3 and illustrated by examples.

8.1 Secure function computation

We consider secure function computation with trusted parties for the multiterminal source model of §6.1. A set of terminals $\mathcal{M} = \{1, \ldots, m\}$, $m \geqslant 2$, observe the respective components of a DMMS $X_{\mathcal{M}}^n = (X_1^n, \ldots, X_m^n)$ consisting of n i.i.d. repetitions of the rv $X_{\mathcal{M}}$, with Terminal i observing the component X_i^n. Let $g\colon \mathcal{X}_{\mathcal{M}} \to \mathcal{Y}$ be a given mapping, where $\mathcal{Y}$ is a finite alphabet. For $n \geqslant 1$, the mapping $g^n\colon \mathcal{X}_{\mathcal{M}}^n \to \mathcal{Y}^n$ is defined by

$$g^n(x_{\mathcal{M}}^n) = (g(x_{11}, \ldots, x_{m1}), \ldots, g(x_{1n}, \ldots, x_{mn})), \qquad x_{\mathcal{M}}^n = (x_1^n, ..., x_m^n) \in \mathcal{X}_{\mathcal{M}}^n.$$

For convenience, we shall denote the rv $g^n(X_{\mathcal{M}}^n)$ by $G^n, n \geqslant 1$, and, in particular, $G^1 = g(X_{\mathcal{M}})$ simply by G. The terminals in $\mathcal{M}$ wish to "compute securely" the function $g^n(x_{\mathcal{M}}^n)$ for $x_{\mathcal{M}}^n$ in $\mathcal{X}_{\mathcal{M}}^n$. To this end, they engage in interactive communication $\mathbf{F}$ as in Definition 3.1. Randomization is allowed at the terminals, with the finite-valued rv U_i denoting the local randomness[1] at Terminal $i \in \mathcal{M}$. Specifically, the interactive communication $\mathbf{F}$ consists of f_{ji} which is allowed to yield any function of (U_i, X_i^n) and of the previous communication ϕ_{ji}, $1 \leqslant j \leqslant r$, $1 \leqslant i \leqslant m$. Note that $\mathbf{F} = \mathbf{F}^{(n)}(U_{\mathcal{M}}, X_{\mathcal{M}}^n)$. Recall that a communication $\mathbf{F}$ constitutes a simple communication if the message of Terminal i depends only on its local observations, *i.e.*, $F_i = f_i^{(n)}(U_i, X_i^n)$, $i \in \mathcal{M}$.

Definition 8.1. For $\epsilon_n, \delta_n > 0, n \geqslant 1$, we say that g is (ϵ_n, δ_n)-*securely computable* ((ϵ_n, δ_n)- SC) from observations of length n, randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F} = \mathbf{F}^{(n)}$, if

(i) g^n is ϵ_n- recoverable from $\mathbf{F}$, *i.e.*, there exists $\widehat{g}_i^{(n)}$ satisfying

$$\mathbb{P}\left(\widehat{g}_i^{(n)}(U_i, X_i^n, \mathbf{F}) = G^n, i \in \mathcal{M}\right) \geqslant 1 - \epsilon_n, \tag{8.1}$$

and

[1] As before, we assume that the rvs $U_1, \ldots, U_m$ are mutually independent and that $U_{\mathcal{M}} = (U_1, \ldots, U_m)$ is independent of $X_{\mathcal{M}}^n$. All the terminals know $P_{U_{\mathcal{M}} X_{\mathcal{M}}^n} = P_{U_1} \times \cdots \times P_{U_m} P_{X_{\mathcal{M}}^n}$, $n \geqslant 1$.

(ii) g^n satisfies the secrecy condition

$$\|P_{G^n\mathbf{F}} - P_{G^n} \times P_{\mathbf{F}}\| \leqslant \delta_n. \tag{8.2}$$

By definition, an (ϵ_n, δ_n)-SC function g is recoverable (as g^n) at the terminals in $\mathcal{M}$ and is effectively concealed from an eavesdropper with access to the public communication $\mathbf{F}$.

Definition 8.2. We say that g is *securely computable* if g is (ϵ_n, δ_n)- SC from observations of length n, suitable randomization $U_{\mathcal{M}}$ and public communication $\mathbf{F}$, such that $\lim\limits_{n\to\infty} \epsilon_n + \delta_n = 0$.

Remark 8.3. Note that if secrecy in (8.2) is defined instead using Kullback-Leibler divergence, in view of Pinsker's inequality the class of securely computable functions can only decrease. In fact, we shall see that if a function satisfies our sufficiency condition for secure computability, it can be computed securely even under a divergence secrecy criterion; see Remark 8.7 below.

We seek to answer the following elemental question: When is a given g securely computable by the terminals in $\mathcal{M}$? A simple necessary condition for a function g to be securely computable follows upon comparing Definition 8.2 and Definition 6.1 of an achievable SK rate. Specifically, if g is an (ϵ_n, δ_n)-securely computable, then there exists a communication $\mathbf{F}$ such that G^n is ϵ_n-recoverable from $\mathbf{F}$ and (8.2) holds. Furthermore, applying Lemma 5.18 with G_i in the role of U_i, $1 \leqslant i \leqslant n$, and $V_1 = V_2 = \varnothing$, we get that for every $\eta > 0$ and a sufficiently large n, there exists a $\mathcal{K}$-valued rv $K = K(G^n)$ and $c > 0$ such that

$$\frac{1}{n} \log |\mathcal{K}| \geqslant H(G) - \eta,$$

and

$$\left\|P_K - P^{\mathcal{K}}_{\mathtt{unif}}\right\| \leqslant 2^{-n \min\{c,\delta\} + \log 3}. \tag{8.3}$$

On combining (8.2) and (8.3), we obtain by the triangle inequality that

$$\begin{aligned}
\sigma_{\mathtt{var}}(K; \mathbf{F}) &\leqslant \|P_{K\mathbf{F}} - P_K \times P_{\mathbf{F}}\| + \|P_K - P_{\mathtt{unif}}\| \\
&\leqslant \|P_{G^n\mathbf{F}} - P_{G^n} \times P_{\mathbf{F}}\| + \|P_K - P_{\mathtt{unif}}\| \\
&\leqslant \delta_n + 2^{-n \min\{c,\delta\} + \log 3}.
\end{aligned}$$

Thus, since K is a function of G^n and G^n is ϵ_n-recoverable from $\mathbf{F}$, the rv K constitutes an (ϵ_n, δ'_n)-SK, with $\delta'_n = \delta_n + 2^{-n \min\{c,\delta\}+\log 3}$, of rate larger than $H(G) - \eta$. It follows by Theorem 6.2 that

$$H(G) - \eta \leqslant C_S,$$

where C_S is the SK capacity of the multiterminal source model with rv $X_{\mathcal{M}}$. Therefore, since $\eta > 0$ was arbitrary, if a function g is securely computable, then it is necessary that

$$H(G) \leqslant C_S. \tag{8.4}$$

Remarkably, it transpires that $H(G) < C_S$ is a sufficient condition for g to be securely computable. Before showing this general result, we illustrate it in a specific example with a scheme for securely computing a particular function g under the assumption $H(G) < C_S$.

Example 8.4. Let $m = 2$, and let X_1, X_2 be $\{0,1\}$-valued rvs with

$$P_{X_1}(1) = p = 1 - P_{X_1}(0), \quad 0 < p < 1,$$
$$P_{X_2|X_1}(1 \mid 0) = P_{X_2|X_1}(0 \mid 1) = \delta, \quad 0 < \delta < \frac{1}{2}.$$

Let $g(x_1, x_2) = x_1 + x_2 \mod 2$.

By Theorem 6.2, the SK capacity for this source model is

$$C_S = I(X_1 \wedge X_2) = h(p * \delta) - h(\delta),$$

where $p * \delta = (1-p)\delta + p(1-\delta)$ and $h(\cdot)$ is binary entropy.

Since $H(G) = h(\delta)$, the condition $H(G) < C_S$ is the same as

$$2h(\delta) < h(p * \delta). \tag{8.5}$$

Under this condition, we give a simple scheme for the secure computation of g when $p = 1/2$. This scheme, in turn, builds upon on an SK agreement scheme for this example, which we address first. When $p = 1/2$, we can write

$$X_1^n = X_2^n + G^n \mod 2 \tag{8.6}$$

with G^n being independent separately of X_2^n and X_1^n. While Lemma 6.7 shows that X_1^n can be recovered from X_2^n and a randomly chosen function f of X_1^n of rate $H(X_1|X_2) = H(G) = h(\delta)$, for this specific example

it is seen that the role of f can be played by a linear function of X_1^n. Specifically, there exists a binary linear code, of rate $\cong 1 - h(\delta)$, with parity check matrix $\mathbf{P}$ such that X_1^n is ϵ_n-recoverable from (F_1, X_2^n) at Terminal 2, where the Slepian-Wolf codeword $F_1 = \mathbf{P}X_1^n$ constitutes public communication from Terminal 1, and where ϵ_n decays to 0 exponentially rapidly in n. Further, let $K = K(X_1^n)$ be the location of X_1^n in the coset of the standard array corresponding to $\mathbf{P}$. By the previous observation, K too is ϵ_n-recoverable from (F_1, X_2^n) at Terminal 2. It is easy to see that K constitutes a "perfect" SK for Terminals 1 and 2 from F_1, of rate $\cong I(X_1 \wedge X_2) = 1 - h(\delta)$, and satisfies

$$I(K \wedge F_1) = 0. \tag{8.7}$$

The SK agreement scheme above can be extended to deliver a scheme for securely computing g. In a first round of communication, the terminals execute the SK agreement scheme using F_1 above, thereby recovering K at both terminals. Furthermore, Terminal 2 recovers X_1^n and thereby forms an estimate $\widehat{G^n}$ of G^n; we assume without loss of generality that $\widehat{G^n}$ has been compressed losslessly with a small probability of error using a block code of rate $H(G)$.

Observe from (8.6) that $K = K(X_1^n) = K(X_2^n + G^n)$ and $F_1 = F_1(X_1^n) = F_1(X_2^n + G^n)$. Since G^n is independent of X_2^n, it follows that conditioned on each fixed value $G^n = \mathbf{g}$, the (common) argument of K and F_1, namely $X_2^n + G^n$, has a conditional pmf that equals the pmf of $X_2^n + \mathbf{g}$ which, in turn, coincides with the pmf of $X_1^n + \mathbf{g}$, *i.e.*, a permutation of the pmf of X_1^n. Hence by (8.7),

$$I(K \wedge F_1, G^n) = I(K \wedge F_1 \mid G^n) = 0, \tag{8.8}$$

since $I(K \wedge G^n) \leqslant I(X_1^n \wedge G^n) = 0$. This independence of K and (F_1, G^n) enables the secure communication of the estimate $\widehat{G^n}$ of G^n by Terminal 2 to Terminal 1. Specifically, Terminal 2 sends $\widehat{G^n}$ in encrypted form as

$$F_2 = \widehat{G^n} + K \mod 2$$

(all represented in bits), with encryption feasible since

$$H(G) = h(\delta) < 1 - h(\delta) \cong \frac{1}{n} H(K),$$

by the sufficient condition (8.5). Terminal 1 then decrypts F_2 using K to recover $\widehat{G^n}$, completing the computation of g^n at both the terminals.

The computation of g^n is secure since

$$I(G^n \wedge F_1, F_2) = I(G^n \wedge F_1) + I(G^n \wedge F_2 \mid F_1)$$

is small; specifically, the first term equals 0 since

$$I(G^n \wedge F_1) \leqslant I(G^n \wedge X_1^n) = 0,$$

while the second term is bounded according to

$$\begin{aligned} I(G^n \wedge F_2 \mid F_1) &= H(\widehat{G^n} + K \mid F_1) - H(\widehat{G^n} + K \mid F_1, G^n) \\ &\leqslant H(K) - H(G^n + K \mid F_1, G^n) + \delta_n, \\ &\qquad \text{with } \delta_n(\epsilon_n) \to 0 \\ &= I(K \wedge F_1, G^n) + \delta_n = \delta_n, \end{aligned}$$

where the intermediate step uses Fano's inequality and the exponential decay of ϵ_n to 0, and the last equality is by (8.8). This establishes secrecy under Kullback-Leibler divergence; secrecy under variational distance follows by the Pinsker's inequality. □

8.2 Characterization of secure computability

While the example above shows the sufficiency of the necessary condition (8.4) for secure computability, the scheme used is asymmetric with respect to the terminals whose extension to $m > 2$ terminals is not clear. Nevertheless, the main result of this section given below shows that the necessary condition (8.4) satisfied by a securely computable g is (almost) sufficient as well.

Theorem 8.5. A function g is securely computable by $\mathcal{M}$ if

$$H(G) < C_S. \tag{8.9}$$

Furthermore, under this condition the function g is securely computable using a simple communication $\mathbf{F}$.

Conversely, if g is securely computable by $\mathcal{M}$, then $H(G) \leqslant C_S$.

8.2.1 Proof of Theorem 8.5

The proof of the necessary part has been seen already in §8.1. Now, we show that a function g is securely computable for $\mathcal{M}$ using a simple communication $\mathbf{F}$ if

$$H(G) < C_S,$$

which, by Theorem 6.2 and Theorem 6.6, is the same as

$$R_{CO} < H(X_{\mathcal{M}}|G).$$

For $0 < \gamma < H(X_{\mathcal{M}}|G) - R_{CO}$, let the rvs Φ_i, $i \in \mathcal{M}$, be distributed uniformly on the family of all mappings $\{\phi\colon \mathcal{X}^n \to \{1, ..., \lceil 2^{nR_i} \rceil\}\}$, respectively, whose rates R_i are chosen to satisfy (6.8) and

$$\sum_{i\in\mathcal{M}} R_i < R_{CO} + \frac{\gamma}{2}. \tag{8.10}$$

Denote $\Phi = (\Phi_1, ..., \Phi_m)$, $\Phi^i = (\Phi_1, ..., \Phi_i)$, and $\Phi_{-i} = \Phi_{\mathcal{M}\setminus\{i\}}$. By Remark 6.8, there exist decoders ψ_i, $i \in \mathcal{M}$, and $c > 0$ such that

$$\mathbb{P}\left(\psi_i\left(X_i^n, \Phi_{-i}(X_{-i}^n)\right) \neq X_{\mathcal{M}}^n\right) \leqslant \epsilon_n, \quad i \in \mathcal{M}, \tag{8.11}$$

thereby guaranteeing ϵ_n-recovery of $X_{\mathcal{M}}^n$, and consequently also of the function g with $\epsilon_n = 2^{-nc}$, at all the terminals in $\mathcal{M}$. Note that these probabilities are computed with respect to the pmf of $(X_{\mathcal{M}}^n, \Phi)$.

It remains to establish secrecy of the computation of g. To this end, we shall show that there exists $c' > 0$ such that

$$\mathbb{E}_\Phi\left[\left\| P_{G^n\Phi(X_{\mathcal{M}}^n)} - P_{G^n} \times P_{\Phi(X_{\mathcal{M}}^n)} \right\|\right] \leqslant 2^{-nc'}, \tag{8.12}$$

This is done in three steps.

Step 1. Noting that for rvs U, V with joint pmf P_{UV} and any pmf Q_V of V,

$$\begin{aligned} \|P_{UV} - P_U \times P_V\| &\leqslant \|P_{UV} - P_U \times Q_V\| + \|P_V - Q_V\| \\ &\leqslant 2\,\|P_{UV} - P_U \times Q_V\|, \end{aligned}$$

we have

$$\begin{aligned}
&\mathbb{E}_\Phi\left[\left\|P_{G^n\Phi(X^n_{\mathcal{M}})} - P_{G^n}\times P_{\Phi(X^n_{\mathcal{M}})}\right\|\right]\\
&\leqslant 2\mathbb{E}_\Phi\left[\left\|P_{G^n\Phi(X^n_{\mathcal{M}})} - P_{G^n}\times P^m_{\mathtt{unif}}\right\|\right]\\
&\leqslant 2\mathbb{E}_\Phi\left[\sum_{\mathbf{g}} P_{G^n}(\mathbf{g})\left\|P_{\Phi(X^n_{\mathcal{M}})|G^n=\mathbf{g}} - P^m_{\mathtt{unif}}\right\|\right],
\end{aligned}$$

where $P^m_{\mathtt{unif}} = P_{\mathtt{unif},1}\times ...\times P_{\mathtt{unif},m}$ and $P_{\mathtt{unif},i}$ denotes the uniform pmf on $\{1,...,\lceil 2^{nR_i}\rceil\}$. Denoting $F_i = \Phi_i(X^n_i)$ and $\mathbf{F} = (F_1,...,F_m)$ for brevity, it follows upon applying Lemma 2.1 to the summands in the right-side above with $P_{U^m} = P_{F_1...F_m|G^n=\mathbf{g}}$ and $Q_{U^m} = P^m_{\mathtt{unif}}$ that

$$\begin{aligned}
&\mathbb{E}_\Phi\left[\left\|P_{G^n\Phi(X^n_{\mathcal{M}})} - P_{G^n}\times P_{\Phi(X^n_{\mathcal{M}})}\right\|\right]\\
&\leqslant 2\mathbb{E}_\Phi\left[\sum_{\mathbf{g}} P_{G^n}(\mathbf{g})\sum_{i=1}^m\left\|P_{F^i|G^n=\mathbf{g}} - P_{F^{i-1}|G^n=\mathbf{g}}\times P_{\mathtt{unif},i}\right\|\right]\\
&\leqslant 2\sum_{i=1}^m \mathbb{E}_\Phi\left[\left\|P_{G^n\mathbf{F}} - P_{G^nF_{-i}}\times P_{\mathtt{unif},i}\right\|\right] \qquad (8.13)
\end{aligned}$$

where the previous right-side is by the triangle inequality. Therefore, in order to establish that the secrecy requirement is met, it suffices to show that $\left(G^n, \Phi_{-i}(X^n_{-i})\right)$ is secure from $\Phi_i(X^n_i)$ for every $i\in\mathcal{M}$.

Step 2. Next, we focus on the summand on the right-side of (8.13). Denote by $\mathcal{F}_{-i}$ the set of all mappings ϕ_{-i} such that for ϵ_n as in (8.11)

$$\mathbb{P}\left(\psi_i(X^n_i,\phi_{-i}(X^n_{-i}))\neq X^n_{\mathcal{M}}\right)\leqslant\sqrt{\epsilon_n}. \qquad (8.14)$$

Therefore, by (8.11),

$$P_{\Phi_{-i}}\left(\mathcal{F}^c_{-i}\right)\leqslant\sqrt{\epsilon_n} = 2^{-nc/2},$$

which further yields, using the independence of $\Phi_1, ..., \Phi_m$, that

$$\begin{aligned}
&\mathbb{E}_{\Phi}\left[\left\|P_{G^n\mathbf{F}} - P_{G^nF_{-i}} \times P_{\mathtt{unif},i}\right\|\right] \\
=&\mathbb{E}_{\Phi}\left[\left\|P_{G^n\Phi(X^n_{\mathcal{M}})} - G^nP_{\Phi_{-i}(X^n_{-i})} \times P_{\mathtt{unif},i}\right\|\right] \\
&\leqslant 2^{-nc/2} + \sum_{\phi_{-i}\in\mathcal{F}_{-i}} P_{\Phi_{-i}}(\phi_{-i}) \times \\
&\qquad\qquad \mathbb{E}_{\Phi_i}\left[\left\|P_{\Phi_i(X^n_i)\phi_{-i}(X^n_{-i})G^n} - P_{\mathtt{unif},i} \times P_{\phi_{-i}(X^n_{-i})G^n}\right\|\right] \\
&= 2^{-nc/2} + \sum_{\phi_{-i}\in\mathcal{F}_{-i}} P_{\Phi_{-i}}(\phi_{-i})\, \sigma_{\mathtt{var}}\left(\Phi_i(X^n_i); G^n, \phi_{-i}(X^n_{-i}), \Phi_i\right).
\end{aligned} \tag{8.15}$$

Therefore, the overall secrecy proof will be complete upon showing that for every $\phi_{-i} \in \mathcal{F}_{-i}$, the rv $\Phi_i(X^n_i)$ remains secure from $(G^n, \phi_{-i}(X^n_{-i}))$, which is done next.

Step 3. To bound $\sigma_{\mathtt{var}}\left(\Phi_i(X^n_i); G^n, \phi_{-i}(X^n_{-i}), \Phi_i\right)$, we shall take recourse to the form of leftover hash given in Lemma 5.17 with $U = X^n_i$, $V_1 = G^n$, and $V_2 = \phi_{-i}(X^n_{-i})$. Our proof relies on the following key observation: Since the mappings $\phi_{-i} \in \mathcal{F}_{-i}$ facilitate the recovery of $X^n_{\mathcal{M}}$ by X^n_i, we can use $P_{X^n_{\mathcal{M}}|G^n}$ in place of $P_{X^n_i|G^n}$ in estimating minentropy.

To formalize this heuristic observation, we rely on smoothing using a subdistribution as described in Remark 5.19. Specifically, we use the following extension of Lemma 5.17.

Lemma 8.6. Consider rvs U', U and V with joint pmf $P_{U'UV}$ such that U' is η-recoverable from (U, V), *i.e.*, there exists a mapping ψ satisfying

$$\mathbb{P}\left(\psi(U, V) = U'\right) \geqslant 1 - \eta. \tag{8.16}$$

Let Φ be chosen uniformly over a UHF of length k. Then,

$$\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) \leqslant 2\eta + \frac{1}{2}\sqrt{2^{\log k - H^{\eta/2}_{\min}(P_{U'V}|V)}},$$

where the smoothing of $H_{\min}$ is over all subdistributions within variational distance $\eta/2$ of $P_{U'V}$. In particular, for $V = (V_1, V_2)$,

$$\sigma_{\mathtt{var}}(\Phi(U); V_1, V_2, \Phi) \leqslant 2\eta + \frac{1}{2}\sqrt{2^{\log k + \log|\mathcal{V}_2| - H^{\eta/2}_{\min}\left(P_{U'V_1}|V_1\right)}}.$$

Proof. Using a smoothing argument with (5.11), we get

$$\sigma_{\mathtt{var}}(\Phi(U); V, \Phi) \leqslant 2\eta + \frac{1}{2}\sqrt{2^{\log k - H^{\eta}_{\min}(P_{UV}|V)}}.$$

Thus, to prove the first claim, it suffices to show that

$$H^{\eta}_{\min}\left(P_{UV}|V\right) \geqslant H^{\eta/2}_{\min}\left(P_{U'V}|V\right); \tag{8.17}$$

the second claim follows from the first upon noting that any $\eta/2$-smoothing of P_{UV_1} can be obtained as a marginal of an $\eta/2$-smoothing of $P_{UV_1V_2}$ and restricting the optimization over $Q_{V_1V_2}$ in the definition of conditional minentropy to $Q_{V_1V_2}\left(v_1, v_2\right) = Q_{V_1}\left(v_1\right)/|\mathcal{V}_2|$.

It remains to establish (8.17). To this end, consider a subdistribution $Q_{U'V}$ such that

$$\left\|P_{U'V} - Q_{U'V}\right\| \leqslant \frac{\eta}{2}.$$

Then, for $Q_{U'UV} \triangleq Q_{U'V}P_{U|U'V}$.

$$\left\|P_{U'UV} - Q_{U'UV}\right\| \leqslant \frac{\eta}{2}.$$

Thus, the subdistribution $\tilde{Q}_{UV}$ defined by

$$\tilde{Q}_{UV}(u, v) = Q_{U'UV}\left(\psi(u, v), u, v\right)$$

satisfies

$$\begin{aligned}
\left\|P_{UV} - \tilde{Q}_{UV}\right\| &\leqslant \frac{1}{2}\sum_{u,v}\left|P_{UV}\left(u, v\right) - P_{U'UV}\left(\psi(u, v), u, v\right)\right| \\
&\quad + \left\|P_{U'UV} - Q_{U'UV}\right\| \\
&\leqslant \frac{1}{2}\sum_{u,v}\left|P_{UV}\left(u, v\right) - P_{U'UV}\left(\psi(u, v), u, v\right)\right| + \frac{\eta}{2} \\
&= \frac{1}{2}\sum_{u',u,v} P_{U'UV}\left(u', u, v\right)\mathbb{1}(u' \neq \psi(u, v)) + \frac{\eta}{2} \\
&\leqslant \eta.
\end{aligned}$$

Therefore,

$$H^{\eta}_{\min}(P_{UV}|V) \geqslant H_{\min}(\tilde{Q}_{UV}|V).$$

The proof is completed upon noting that for every subdistribution $\tilde{P}_V$ such that $\mathtt{supp}(\tilde{P}_V) \supseteq \mathtt{supp}(Q_V)$,

$$\begin{aligned} H_{\min}(\tilde{Q}_{UV}|V) &\geqslant H_{\min}(\tilde{Q}_{UV}|\tilde{P}_V) \\ &= -\log \max_{u,v} \frac{\tilde{Q}_{UV}(u,v)}{\tilde{P}_V(v)} \\ &= -\log \max_{u,v} \frac{Q_{U'UV}(\psi(u,v),u,v)}{\tilde{P}_V(v)} \\ &\geqslant -\log \max_{u',u,v} \frac{Q_{U'UV}(u',u,v)}{\tilde{P}_V(v)} \\ &\geqslant -\log \max_{u',v} \frac{Q_{U'V}(u',v)}{\tilde{P}_V(v)} \\ &= H_{\min}(Q_{U'V}|\tilde{P}_V). \end{aligned}$$

Since $Q_{U'V}$ and $\tilde{P}_V$ were arbitrary subdistributions such that $\|P_{U'V} - Q_{U'V}\| \leqslant \eta/2$ and $\mathtt{supp}(\tilde{P}_V) \supseteq \mathtt{supp}(Q_V)$, it follows upon combining the inequalities above that

$$H^{\eta}_{\min}(P_{UV}|V) \geqslant H_{\min}(\tilde{Q}_{UV}|V) \geqslant H^{\eta/2}_{\min}(P_{U'V}|V),$$

which completes the proof. □

We apply Lemma 8.6 with $U = X_i^n$, $U' = X_{\mathcal{M}}^n$, $V_1 = G^n$, and $V_2 = \phi_{-i}(X_{-i}^n)$. It follows from (8.14) that for every $\eta \geqslant 2^{-nc/2}$

$$\begin{aligned} &\sigma_{\mathtt{var}}(\Phi_i(X_i^n); G^n, \psi_{-i}(X_{-i}^n), \Phi) \\ &\leqslant 2\eta + \frac{1}{2}\sqrt{2^{nR_i + n\sum_{j\in\mathcal{M}, j\neq i} R_j - H^{\eta/2}_{\min}\left(P_{X^n_{\mathcal{M}}G^n}|G^n\right)}}. \end{aligned}$$

Next, proceeding as in §5.4.2, given $\delta > 0$, the minentropy $H^{\eta/2}_{\min}\left(P_{X^n_{\mathcal{M}}G^n}|G^n\right)$ is bounded below as

$$H^{\eta/2}_{\min}\left(P_{X^n_{\mathcal{M}}G^n}|G^n\right) \geqslant n[H(X_{\mathcal{M}}|G) - \gamma/4],$$

for all n sufficiently large, provided $\eta \geqslant 2^{-nc_0}$ for a sufficiently large c_0; the arguments above will hold as long as $\eta \geqslant 2^{-n\min\{c/2, c_0\}}$. Note that the $\log(1 - 2^{-nc_0})$ term in (5.13) does not appear here since, following Remark 5.19, we are smoothing over subdistributions. Thus, for n

sufficiently large, with $c_1 = \min\{c/2, c_0\}$,

$$\begin{aligned}
&\sigma_{\mathtt{var}}(\Phi_i(X_i^n); G^n, \psi_{-i}(X_{-i}^n), \Phi) \\
&\leqslant 2 \cdot 2^{-nc_1} + \frac{1}{2}\sqrt{2^{n\sum_{j\in\mathcal{M}} R_i - nH(X_{\mathcal{M}}|G) + n\gamma/4}} \\
&\leqslant 2 \cdot 2^{-nc_1} + \frac{1}{2}\sqrt{2^{nR_{CO} - nH(X_{\mathcal{M}}|G) + n3\gamma/4}}, \qquad (8.18)
\end{aligned}$$

where the previous inequality is by (8.10).

Finally, upon combining (8.13), (8.15), and (8.18), and using the sufficient condition $H(X_{\mathcal{M}}) - R_{CO} > \gamma$, we get

$$\mathbb{E}_\Phi\left[\left\| P_{G^n \Phi(X_{\mathcal{M}}^n)} - P_{G^n} \times P_{\Phi(X_{\mathcal{M}}^n)} \right\|\right] \leqslant 7m2^{-n\min\{c/2, c_1, \gamma/8\}},$$

which establishes (8.12), thereby completing the proof of Theorem 8.5. □

Remark 8.7. We have shown that, under the sufficiency condition $H(G) < C_s$, there exists a communication $\mathbf{F}$ which attains omniscience and satisfies for every $i \in \mathcal{M}$,

$$\sigma_{\mathtt{var}}(F_i; G^n, F_{-i}) \leqslant 2^{-nc}$$

for a constant $c > 0$. Using Lemma 2.4, this in turn implies

$$\sigma_{\mathtt{div}}(F_i; G^n, F_{-i}) = nR_i - H(F_i) + I(F_i \wedge G^n, F_{-i}) \to 0 \text{ as } n \to \infty.$$

Note that by the chain rule for mutual information

$$I(G^n \wedge \mathbf{F}) \leqslant \sum_{i=1}^{m} I(F_i \wedge G^n, F_{-i}),$$

which along with the observation above yields

$$\lim_{n\to\infty} I(G^n \wedge \mathbf{F}) = 0,$$

namely the secrecy of communication $\mathbf{F}$ under the Kullback-Leibler divergence-based secrecy criterion.

8.3 General necessary condition for secure computability

We close this chapter with a necessary condition for secure computability of a function in a "single-shot" setting. Specifically, given $\epsilon, \delta > 0$

such that $\epsilon + \delta < 1$, we seek to characterize functions $g\colon \mathcal{X}_{\mathcal{M}} \to \mathcal{Y}$ that are (ϵ, δ)-SC in the sense of Definition 8.1 for $n = 1$. Throughout this section, we refer to such a function g as an (ϵ, δ)-SC function.

Recall that the necessary condition in Theorem 8.5 for the asymptotic formulation (*i.e.*, for n large) of the characterization sought above follows upon observing that if the terminals can compute securely the function g, then they can extract a SK of rate $H(G)$ from G^n. Therefore, $H(G)$ must be necessarily less than the maximum rate of a SK that can be generated, namely the SK capacity C_S. We extend this principle to the single-shot setting by using the conditional independence testing upper bound for SK length given in Theorem 4.12.

Corollary 8.8. For $0 \leqslant \epsilon, \delta < 1$ with $\epsilon + \delta < 1$, if a function g is (ϵ, δ)-SC, then

$$\begin{aligned} &H_{\min}^{\xi}(P_G) \\ &\leqslant \frac{1}{|\pi| - 1}\Big[- \log \beta_{\mu}\big(P_{X_{\mathcal{M}}}, \mathrm{Q}_{X_{\mathcal{M}}}^{\pi}\big) + |\pi| \log(1/\eta)\Big] \\ &\qquad\qquad + 2\log(1/2\zeta) + 1, \quad \mathrm{Q}_{X_{\mathcal{M}}}^{\pi} \in \mathcal{Q}(\pi), \end{aligned} \tag{8.19}$$

for every $\mu = \epsilon + \delta + 2\xi + \zeta + \eta$ with $\xi, \zeta, \eta > 0$ such that $\mu < 1$, and for every partition π of $\mathcal{M}$, where $\mathrm{Q}_{X_{\mathcal{M}}}^{\pi}$ is as in (4.8) with $Z =$ constant.

Proof. The proof is based on extracting an $(\epsilon, \delta + 2\xi + \zeta)$-SK from the securely computed function G. Specifically, suppose that g is (ϵ, δ)-SC from public communication $\mathbf{F}$. Using Lemma 5.17 with $U = G$ and $V_1 = V_2 =$ constant, we get that there exists a $\mathcal{K}$-valued rv $K = K(G)$ with $\log|\mathcal{K}| = \lfloor H_{\min}^{\xi}(P_G) - 2\log(1/2\zeta) \rfloor$ satisfying

$$\big\| \mathrm{P}_{K(G)} - P_{\mathtt{unif}}^{\mathcal{K}} \big\| \leqslant 2\xi + \zeta.$$

Therefore,

$$\begin{aligned} &\big\| \mathrm{P}_{K(G)\mathbf{F}} - \mathrm{P}_{\mathrm{unif}} \times \mathrm{P}_{\mathbf{F}} \big\| \\ &\leqslant \big\| \mathrm{P}_{K(G)\mathbf{F}} - \mathrm{P}_{K(G)} \times \mathrm{P}_{\mathbf{F}} \big\| \\ &\qquad + \big\| \mathrm{P}_{K(G)} \times \mathrm{P}_{\mathbf{F}} - \mathrm{P}_{\mathrm{unif}} \times \mathrm{P}_{\mathbf{F}} \big\| \\ &\leqslant \| \mathrm{P}_{G\mathbf{F}} - \mathrm{P}_G \times \mathrm{P}_{\mathbf{F}} \| + \big\| \mathrm{P}_{K(G)} - \mathrm{P}_{\mathrm{unif}} \big\| \\ &\leqslant \delta + 2\xi + \zeta, \end{aligned}$$

where the final inequality uses the fact that g is (ϵ, δ)-SC from $\mathbf{F}$. Furthermore, since G is ϵ-recoverable from $\mathbf{F}$, so is $K = K(G)$. Therefore, K constitutes an (ϵ, δ)-SK of length $\lfloor H_{\min}^{\xi}(P_G) - 2\log(1/2\zeta) \rfloor$ and the claimed necessary condition follows from the bound for the maximum length of an (ϵ, δ)-SK given in Theorem 4.12. □

We conclude this section with two illustrative examples.

Example 8.9. (Computing functions of independent observations using a perfect SK). Suppose that each terminal i in $\mathcal{M}$ observes U_i, where the rvs $U_1, ..., U_m$ are mutually independent. Furthermore, all the terminals share a k-bit perfect SK K which is independent of $U_{\mathcal{M}}$. How many bits k are required to render the function $g(U_1, ..., U_m)$ an (ϵ, δ)-SC function?

Note that the observation of Terminal i is $X_i = (U_i, K)$. For a partition π of $\mathcal{M}$, in order to bound $-\log \beta_\epsilon \left(P_{X_{\mathcal{M}}}, \prod_{i=1}^{|\pi|} P_{X_{\pi_i}} \right)$, we take recourse to Lemma 4.15. Specifically, noting that

$$\frac{P_{X_{\mathcal{M}}}}{\prod_{i=1}^{|\pi|} P_{X_{\pi_i}}} = {P_K}^{(1-|\pi|)},$$

it follows from Lemma 4.15 with $\gamma = (|\pi| - 1)k$ that

$$-\log \beta_\mu \left(P_{X_{\mathcal{M}}}, \prod_{i=1}^{|\pi|} P_{X_{\pi_i}} \right) \leqslant (|\pi| - 1)k + \log(1 - \mu).$$

Therefore, by Corollary 8.8 a necessary condition for g to be (ϵ, δ)-SC is

$$H_{\min}^{\xi}(P_G) \leqslant k + \frac{1}{|\pi| - 1} \left(|\pi| \log \frac{1}{\eta} + \log \frac{1}{1 - \mu} \right) + 2\log \frac{1}{2\zeta} + 1, \tag{8.20}$$

for every $\xi, \zeta, \eta > 0$ satisfying $\mu = \epsilon + \delta + 2\xi + \zeta + \eta < 1$. Note that the finest partition, *i.e.*, $|\pi| = m$, gives the best lower bound on k in (8.20).

For the special case when $U_i = B_i^n$, a sequence of independent, unbiased bits, and

$$g(B_1^n, ..., B_m^n) = B_{11} \oplus ... \oplus B_{m1}, ..., B_{1n} \oplus ... \oplus B_{mn},$$

i.e., the terminals seek to compute the (element-wise) parities of the bit sequences, it holds that $H_{\min}^{\xi}(P_G) \geqslant n$. Therefore, g is (ϵ, δ)-SC only if $n \leqslant k + O(1)$. We remark that this necessary condition is also (almost) sufficient. Indeed, if $n \leqslant k$, all the terminals but Terminal m can reveal all their bits $B_1^n, \ldots, B_{m-1}^n$ and Terminal m can communicate $B_1^n \oplus \ldots \oplus B_m^n \oplus K_n$, where K_n denotes any n out of k bits of K. Clearly, this results in a secure computation of g.

Example 8.10. (Secure transmission). Terminal 1 seeks to transmit to Terminal 2 a message rv M of known pmf P_M, with the terminals sharing a k-bit perfect SK K. To this end, they communicate interactively using a communication $\mathbf{F}$, enabling Terminal 2 to form an estimate $\hat{M}$ of M. This protocol accomplishes (ϵ, δ)-secure transmission if $\mathbb{P}\left(M = \hat{M}\right) \geqslant 1 - \epsilon$ and

$$\|P_{M\mathbf{F}} - P_M \times P_{\mathbf{F}}\| \leqslant \delta.$$

Shannon's classic result implies that $(0, 0)$-secure transmission is feasible only if k is at least $\log |\mathcal{M}|$, where $|\mathcal{M}|$ denotes the size of the message set.[2] This prompts the question: Can we relax this constraint for $\epsilon, \delta > 0$? We give a necessary condition below for the feasibility of (ϵ, δ)-secure transmission by relating it to the previous example.

Specifically, let the observations of Terminals 1 and 2 be $X_1 = (M, K)$ and $X_2 = K$, respectively. Then, (ϵ, δ)-secure transmission of M is equivalent to the function $g(X_1, X_2) = M$ being (ϵ, δ)-SC. Therefore, using (8.20), (ϵ, δ)-secure transmission of M is feasible only if

$$H_{\min}^{\xi}(P_M) \leqslant k + 2\log\frac{1}{\eta} + \log\frac{1}{1-\mu} + 2\log\frac{1}{2\zeta} + 1, \tag{8.21}$$

for every $\xi, \zeta, \eta > 0$ satisfying $\mu = \epsilon + \delta + 2\xi + \zeta + \eta < 1$.

Condition (8.21) brings out a tradeoff between k and $\epsilon + \delta$. As an illustration, consider a message M as a rv Y taking values in a set $\mathcal{Y} = \{0,1\}^n \cup \{0,1\}^{2n}$ with the following pmf:

$$P_Y(y) = \begin{cases} \frac{1}{2} \cdot \frac{1}{2^n} & y \in \{0,1\}^n \\ \frac{1}{2} \cdot \frac{1}{2^{2n}} & y \in \{0,1\}^{2n} \end{cases}.$$

[2] This is a slight generalization of Shannon's original result.

For $\epsilon+\delta = 0$, we know that secure transmission will require k to exceed the *worst-case message length* $2n$. By allowing $\epsilon + \delta > 0$, can we make do with fewer SK bits, for instance, with $k = H(M) = (3/2)n + 1$ (noting that the *average message length* equals $(3/2)n$)? The necessary condition above says that this is not possible if $\epsilon + \delta < 1/2$. Indeed, since $H_{\min}^{\xi}(P_Y) \geqslant 2n$ for $\xi = 1/4$, we get from (8.21) that the message $M = Y$ can be (ϵ, δ)-securely transmitted only if $2n \leqslant k + O(1)$, where the constant depends on ϵ and δ.

8.4 Story of results and open problems

Shannon's information theoretic analysis of the one-time pad in his seminal paper [72] is indeed the first instance of the secure computing problem considered here, as pointed out in Example 8.10. A generalization that entails secure computing of a function, in the sense considered in this chapter, by two terminals that share perfect SKs was introduced in [63]. Both these formulations differ from the standard notion of secure computing considered in the cryptography literature, following Yao's seminal work [100], where the communication channel is trusted but both terminals strive to reveal the least amount of information about themselves to each other. Also, as for the SK generation problem in Chapter 6, we have restricted ourselves to the *honest-but-curious* setting where the adversary can only eavesdrop without tampering with the protocol. The contents of this chapter are drawn largely from [81, 82, 83] where our secure computing problem was defined, and Theorem 8.5 proved. The linear Slepian-Wolf compression scheme of Example 8.4 was given in [97], and the corresponding SK generation scheme in [102]. The general necessary condition and the single-shot examples of §8.3 are from [85, 87]. Note that several extensions of Shannon's original one-time pad result allowing nonzero leakage and probability of error are available; see, for instance, [45, Problems 2.12 and 2.13]. Example 8.10 strengthens such results by allowing interactive communication.

Extensions and open problems. In [82, 83], securely computable functions were characterized in a larger setting of the helper model

(see §6.4) where only a subset $\mathcal{A}$ of the terminals in $\mathcal{M}$ seek to securely compute the function g (with terminals in $\mathcal{A}^c$ serving as helpers for communication). An even further generalization was considered in [74, 77] where each terminal seeks to compute a different function without revealing the value of a (yet possibly different) function g_0. A general conjecture regarding the characterization of securely computable functions was made in [77], with justification only in special cases (see [76] for a discussion). A characterization of securely computable functions in this context, and its extension where the eavesdropper has access to correlated side information Z^n, constitute an open avenue.

A different view of secure computing, and a sketch of an alternative proof of characterization of a securely computable function, was given in [12]. In another direction, secure sampling and secure channel simulation were considered in [91] and [27], respectively. Secure distributed source coding (see, for instance, [64, 31, 73]), which addresses the tradeoff between communication rate and equivocation, albeit not for interactive communication, is also related closely to secure computing in our sense. All these directions have seen progress only in fits and starts, and interesting open problems beckon.

9

Secret Key Capacity for the Multiterminal Channel Model

In the multiterminal channel model, a subset of k terminals, $1 \leqslant k \leqslant m-1$, govern the inputs of a noisy but secure discrete memoryless multiaccess channel with the remaining $m-k$ terminals receiving the channel outputs. In between transmissions over the channel, all the terminals additionally can communicate among themselves publicly as in the source model of Chapter 6. The secret key capacity problem is formulated in §9.1 with secrecy required from an eavesdropper observing the communication. A general single-letter characterization of secret key capacity remains defiant. General lower bounds for secret key capacity are obtained in §9.2 through achievability proofs using source emulation techniques that are redolent of Chapter 6. Also, a general upper bound is presented that is built upon a precursor bound from Chapter 4; this upper bound can be improved for special channels. A full characterization of secret key capacity is presented in §9.3 when the channel has a single input terminal. For the mirroring channel model with a single output terminal, a general characterization of secret key capacity is open. However, interesting connections exist between it and the transmission capacity region of the multiple access channel with and without feedback.

9.1 Multiterminal channel model

The multiterminal channel model with $m \geqslant 2$ terminals is based on an underlying discrete memoryless channel (DMC) with multiple inputs and multiple outputs. Consider a DMC

$$W\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_k \to \mathcal{X}_{k+1} \times \cdots \times \mathcal{X}_m$$

with finite input and output alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_k$ and $\mathcal{X}_{k+1}, \ldots, \mathcal{X}_m$, respectively, $1 \leqslant k \leqslant m-1$. Terminals $1, \ldots, k$ control the inputs of the DMC W over which they transmit n-length sequences, while terminals $k+1, \ldots, m$ receive the corresponding n-length output sequences. Transmission and reception over the DMC W are *secure*. After every transmission over the DMC and corresponding reception, and prior to any next transmission, the terminals in $\mathcal{M}$ can engage in interactive communication over a public noiseless channel of unlimited capacity. Hereafter, we note a distinction between secure "transmission" over the DMC from the input terminals $1, \ldots, k$ to the output terminals $k+1, \ldots, m$, and public "communication" among the terminals in $\mathcal{M}$, and accordingly term them simply as transmission or communication. The communication is observed by all the terminals in $\mathcal{M}$ as also by an eavesdropper. Randomization is permitted at the terminals for transmission or communication, with the finite-valued rv U_i representing the local randomness at terminal $i \in \mathcal{M}$. The rvs $U_1, \ldots, U_m$ are taken to be mutually independent.

It will be convenient to use the following shorthand notation. For integers $1 \leqslant a \leqslant b \leqslant m$, we write $[a, b] = (a, \ldots, b)$. Given rvs $Z_i,\ i \in \mathcal{M}$, we denote $Z_{[a,b]} = (Z_a, \ldots, Z_b)$.

The protocol for transmission and communication is as follows. At each time instant $t = 1, \ldots, n$, the input terminals in $[1, k]$ transmit the symbols $X_{1t}, \ldots, X_{kt}$ over the DMC and the output terminals in $[k+1, m]$ observe (instantaneously) the corresponding output symbols $X_{k+1,t}, \ldots, X_{mt}$. Additionally, during each time interval $(t, t+1)$ between the tth and $(t+1)$th transmissions — for $t = 1, \ldots, n-1$ and after the nth transmission for $t = n$ — the input and output terminals in $\mathcal{M}$ participate in interactive communication as in Definition 3.1. In each interval $(t, t+1)$, hereafter simply called interval t, the commu-

nication of the terminals in $\mathcal{M}$ will be represented collectively by $F_{(t)}$, and furthermore, we denote $F^{(t)} = \big(F_{(1)}, \ldots, F_{(t)}\big)$, $t = 2, \ldots, n-1$, and $\mathbf{F} = \big(F_{(1)}, \ldots, F_{(n)}\big)$.

Each input terminal $i \in [1, k]$, determines its tth transmission symbol X_{it} as a function of U_i for $t = 1$ and of $\big(U_i, F_{(1)}, \ldots, F_{(t-1)}\big)$ for $t = 2, \ldots, n$. Furthermore, in interval t, the communication of terminal $i \in \mathcal{M}$ in any round is allowed to be a function of U_i, the symbols $(X_{i1}, \ldots, X_{it})$ earlier generated or observed by terminal i, and all earlier communication $\big(F_{(1)}, \ldots, F_{(t-1)}\big)$ in previous intervals along with communication in prior rounds in the current interval. The terminals in $\mathcal{M}$ cooperate, through transmission and communication, to generate a SK which is concealed from an eavesdropper observing the communication $\mathbf{F}$.

9.2 Secret key capacity: General lower and upper bounds

Achievable SK rates and SK capacity for a mutiterminal channel model with DMC W, hereafter termed simply channel model W, are defined analogously as in Definition 6.1.

Definition 9.1. $R \geqslant 0$ is an achievable SK rate for the terminals in $\mathcal{M}$ if there exist (ϵ_n, δ_n)-SKs $K^{(n)}$ for $\mathcal{M}$ with values in $\mathcal{K}^{(n)}$, obtained by means of n uses of the DMC W together with interactive communication $\mathbf{F}^{(n)}$, i.e., there exist local estimates $K_i^{(n)} = K_i^{(n)}\big(U_i, X_i^n, \mathbf{F}^{(n)}\big)$, $i \in \mathcal{M}$, of $K^{(n)}$ satisfying

$$\mathbb{P}\big(K_i^{(n)} = K^{(n)},\ i \in \mathcal{M}\big) \geqslant 1 - \epsilon_n, \quad \sigma_{\mathtt{var}}\big(K^{(n)}; \mathbf{F}^{(n)}\big) \leqslant \delta_n \tag{9.1}$$

where

$$\epsilon_n \to 0, \quad \delta_n \to 0 \text{ and } \frac{1}{n} \log \big|\mathcal{K}^{(n)}\big| \to R \text{ as } n \to \infty.$$

The largest achievable SK rate is the SK *capacity* C_S.

A general single-letter characterization of SK capacity for the channel model W, unlike for its source model counterpart in Chapter 6, remains unresolved. General single-letter lower and upper bounds for SK capacity are described below. An achievability technique of "source

emulation" gives rise to a variant of the source model of Chapter 6; achievable SK rates for the latter yield, in turn, a general lower bound for the SK capacity of the channel model. A general upper bound is obtained by means of a converse proof based on methods developed in Chapter 4 combined with suitable entropy inequalities. The lower and upper bounds agree only in special cases. Both admit improvements for a particular class of channel models.

9.2.1 General lower bound for SK capacity by source emulation

Simple source emulation affords an obvious means for generating a SK for a channel model W. Consider finite-valued independent rvs $X_1, \ldots, X_k$ with a given joint pmf

$$P_{X_{[1,k]}} \;=\; \prod_{i=1}^{k} P_{X_i}. \tag{9.2}$$

The input terminals in $[1,k]$ generate – using local randomness $U_1, \ldots, U_k$ – n i.i.d. repetitions of $X_1, \ldots, X_k$, respectively, and transmit them over the DMC W, thereby creating i.i.d. repetitions of $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ with joint pmf

$$P_{X_{\mathcal{M}}}\left(x_{[1,m]}\right) \;=\; \left(\prod_{i=1}^{k} P_{X_i}\left(x_i\right)\right) W\left(x_{[k+1,m]} \mid x_{[1,k]}\right), \quad x_{[1,m]} \in \bigtimes_{i=1}^{m} \mathcal{X}_i, \tag{9.3}$$

with each output terminal $i \in [k+1, m]$ observing n i.i.d. repetitions of X_i. By this operational strategy, the channel model W emulates a multiterminal source model with generic rv $X_{\mathcal{M}}$ where $P_{X_{\mathcal{M}}}$ is specified by (9.3), and ipso facto the SK capacity of the latter model, given by Theorem 6.2, is a lower bound for the SK capacity of the former. Furthermore, this bound can be improved by maximization over all input pmfs $P_{X_{[1,k]}}$ of the form (9.2). Thus, recalling Theorem 6.2,

$$C_S \;\geqslant\; \max_{P_{X_{[1,k]}}} \; \min_{\lambda} \quad H(X_{\mathcal{M}}) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}). \tag{9.4}$$

General source emulation is an improved operational strategy for generating a SK of enhanced rate for a channel model by using, in

effect, an auxiliary source. For an auxiliary rv V with values in a finite set $\mathcal{V}$, consider rvs $V, X_1, \ldots, X_k$ such that $X_1, \ldots, X_k$ are conditionally independent given V. If the rvs $X_1, \ldots, X_k$ are inputs to the DMC W, the corresponding output rvs satisfy the Markov condition

$$V \multimap X_{[1,k]} \multimap X_{[k+1,m]}. \tag{9.5}$$

Then the pmf of $(V, X_{\mathcal{M}})$ is given by

$$P_{VX_{\mathcal{M}}}\left(v, x_{[1,m]}\right) = P_V\left(v\right)\left(\prod_{i=1}^{k} P_{X_i|V}\left(x_i \mid v\right)\right) W\left(x_{[k+1,m]} \mid x_{[1,k]}\right), \quad v \in \mathcal{V},\ x_{[1,m]} \in \mathcal{X}_{\mathcal{M}}. \tag{9.6}$$

Terminal 1, say, using local randomness U_1 generates $V^n = (V_1, \ldots, V_n)$ comprising n i.i.d. repetitions of V and communicates V^n publicly to all the remaining terminals in $\mathcal{M}$; the eavesdropper, too, has access to V^n. Knowing $V^n = v^n = (v_1, \ldots, v_n), v^n \in \mathcal{V}^n$, the input terminals in $[1, k]$ – using their local randomness $U_1, \ldots, U_k$ – generate and transmit i.i.d. rvs $X_{[1,k]t}, t = 1, \ldots, n$, over the DMC W, with conditional pmf given $V^n = v^n$ as

$$P_{X^n_{[1,k]}|V^n}\left(x^n_{[1,k]} \mid v^n\right) = \prod_{t=1}^{n}\left(\prod_{i=1}^{k} P_{X_i|V}\left(x_{it} \mid v_t\right)\right), \quad x^n_{[1,k]} \in \mathcal{X}^n_{[1,k]},\ v^n \in \mathcal{V}^n.$$

The corresponding channel outputs $X^n_{[k+1,m]}$ satisfy

$$V^n \multimap X^n_{[1,k]} \multimap X^n_{[k+1,m]}, \qquad P_{X^n_{[k+1,m]}|X^n_{[1,k]}} = W^n,$$

with $X_{[k+1,m]t}, t = 1, \ldots, n$, being i.i.d.

The channel model thereby emulates a multiterminal source model with generic rv $X_{\mathcal{M}}$, initiating it with communication V^n where $P_{VX_{\mathcal{M}}}$ is given by (9.6). This initial public communication, to which the eavesdropper is also privy, makes the latter model vary somewhat from the standard source model of §6.1. The SK capacity of this emulated source model makes for a lower bound for the SK capacity of the channel model. Note that for $V =$ constant, simple source emulation obtains.

Theorem 9.2 (Lower bound for SK capacity)**.** The SK capacity of a channel model W is bounded below as

$$C_S \geqslant \max_{P_{VX_{[1,k]}}} \min_{\lambda} H(X_{\mathcal{M}} \mid V) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}, V) \qquad (9.7)$$

where $P_{VX_{\mathcal{M}}}$ is as in (9.6) (with W given). The right-side represents the largest SK rate achievable by general source emulation. This SK rate can be achieved also by a protocol in which the input terminals do not communicate publicly and transmit mutually independent sequences over the DMC W that are not necessarily i.i.d.

Remark 9.3. The maximum in (9.7) with respect to $P_{VX_{\mathcal{M}}}$ is achieved as the cardinality of $\mathcal{V}$ can be bounded by standard techniques involving the "support lemma."

Proof. The proof of the first two assertions of the theorem entails a minor modification of the proof of Theorem 6.2, and is sketched. Let $P_{VX_{\mathcal{M}}}$ attain the maximum in (9.7). Terminal 1 generates $V^n = (V_1, \ldots, V_n)$ as above and communicates it publicly to all the (remaining) terminals in $\mathcal{M}$.

By general source emulation as above, a multiterminal source model results with generic rv $X_{\mathcal{M}}$ and with all the terminals in $\mathcal{M}$ having additional access to V^n. Considering this model, a straightforward modification of the achievability proof of Theorem 6.6 with each decoder $\psi_i,\ i \in \mathcal{M}$, now possessing additional side information V^n, yields that the terminals in $\mathcal{M}$ can form $X_{\mathcal{M}}^n$ as ϵ_n-CR using communication $\mathbf{F} = \mathbf{F}^{(n)}$ of range $\mathcal{F}^{(n)}$ and rate

$$R \;=\; \frac{1}{n} \log |\mathcal{F}^{(n)}| \;=\; \max_{\lambda} \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}, V) + \nu,$$

where $\epsilon_n \to 0$ as $n \to \infty$ and $\nu > 0$ is arbitrary.

Then, Lemma 5.18 with

$$U^n = X_{\mathcal{M}}^n,\ V_1^n = V^n,\ V_2 = \mathbf{F}^{(n)} \text{ and } \mathcal{V}_2 = \mathcal{F}^{(n)}$$

yields, for any number $0 \leqslant H \leqslant H(X_{\mathcal{M}} \mid V) - R$, the existence of $K^{(n)} = \phi(X_{\mathcal{M}}^n)$ as ϵ_n-CR with values in $\mathcal{K}^{(n)} = \{1, \ldots, \lfloor \exp(nH) \rfloor\}$

and variational secrecy index $\sigma_{\mathtt{var}}\left(K^{(n)}; V^n, \mathbf{F}^{(n)}\right) \leqslant \delta_n$, where $\epsilon_n \to 0$ and $\delta_n \to 0$ exponentially as $n \to \infty$. Since

$$H\left(X_{\mathcal{M}} \mid V\right) - R = H\left(X_{\mathcal{M}} \mid V\right) - \max_{\lambda} \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}, V) - \nu$$

with arbitrarily small $\nu > 0$, it follows that the right-side of (9.7) is an achievable SK rate for the channel model W. Furthermore, a natural and straightforward extension of Theorem 6.2 to a multiterminal source model with generic rv $X_{\mathcal{M}}$ and with all the terminals in $\mathcal{M}$ as well as the eavesdropper having additional access to V^n, shows that the right-side of (9.7) is the largest SK rate achievable by general source emulation.

In the general source emulation protocol above, first Terminal 1 communicates V^n publicly. Then the DMC W is used n times for the transmission and reception, respectively, of $X_{[1,k]}^n$ and $X_{[k+1,m]}^n$, with the input terminals in $[1, k]$ using their local randomness $U_1, \ldots, U_k$ to generate $X_{[1,k]}^n$. Upon completion of the use of the DMC, the ensuing public communication $\mathbf{F}^{(n)} = (F_1, \ldots, F_m)$ entails each terminal $i \in \mathcal{M}$ sending at most one public message $F_i = f_i\left(X_i^n\right)$ without using U_i, by Theorem 6.2, whereupon the terminals in $\mathcal{M}$ form a SK $K^{(n)}$ of rate approaching the right-side of (9.7). We show next that the input terminals in $[1, k]$ need not communicate publicly, by altering the pmfs of the input rvs $X_i^n,\ i \in [1, k]$. Consider an (ϵ_n, δ_n)-SK $K^{(n)}$ as above. The recoverability and variational secrecy index of $K^{(n)}$ imply that

$$\frac{1}{\epsilon_n} \sum_{i \in \mathcal{M}} \mathbb{P}\left(K_i^{(n)}(X_i^n, V^n, \mathbf{F}^{(n)}) \neq K^{(n)}\right) + \frac{1}{\delta_n} \sigma_{\mathtt{var}}\big(K^{(n)}; V^n, \mathbf{F}^{(n)}\big) \leqslant m + 1.$$

Since

$$\begin{aligned}
&\mathbb{P}\left(K_i^{(n)}(X_i^n, V^n, \mathbf{F}^{(n)}) \neq K^{(n)}\right) \\
&= \sum_{v^n, j_1, \ldots, j_k} \mathbb{P}\left(V^n = v^n, F_1 = j_1, \ldots, F_k = j_k\right) \\
&\times \mathbb{P}\left(K_i^{(n)}(X_i^n, V^n, \mathbf{F}^{(n)}) \neq K^{(n)} \mid V^n = v^n, F_1 = j_1, \ldots, F_k = j_k\right)
\end{aligned}$$

and

$$\begin{aligned}
&\sigma_{\mathtt{var}}\big(K^{(n)}; V^n, \mathbf{F}^{(n)}\big) \\
&= \sum_{v^n, j_1, \ldots, j_k} \mathbb{P}\left(V^n = v^n, F_1 = j_1, \ldots, F_k = j_k\right) \\
&\times \sigma_{\mathtt{var}}\big(K^{(n)}; F_{k+1}, \ldots F_m \mid V^n = v^n, F_1 = j_1, \ldots, F_k = j_k\big)
\end{aligned}$$

where $\sigma_{\mathtt{var}}\left(\cdot \mid \cdot\right)$ in the right-side above equals

$$\begin{aligned}
\Big\| & P_{K^{(n)} F_{k+1} \cdots F_m | V^n F_1 \cdots F_k}\left(\cdot \mid v^n, j_1, \ldots, j_k\right) \\
& - P_{\mathrm{unif}}^{(n)} \times P_{F_{k+1} \cdots F_m | V^n F_1 \cdots F_k}\left(\cdot \mid v^n, j_1, \ldots, j_k\right) \Big\|,
\end{aligned}$$

it follows that for some $v^{*n} = (v_1^*, \ldots, v_n^*)$ and $j_1^*, \ldots, j_k^*$,

$$\begin{aligned}
&\frac{1}{\epsilon_n} \sum_{i \in \mathcal{M}} \mathbb{P}\left(K_i^{(n)}(X_i^n, V^n, \mathbf{F}^{(n)}) \neq K^{(n)} \mid V^n = v^{*n}, F_1 = j_1^*, \ldots, F_k = j_k^*\right) \\
&+ \frac{1}{\delta_n} \sigma_{\mathtt{var}}\big(K^{(n)}; F_{k+1}, \ldots, F_m \mid V^n = v^{*n}, F_1 = j_1^*, \ldots, F_k = j_k^*\big) \leqslant m + 1.
\end{aligned}$$

This shows that if the joint pmf of $V^n, X_1^n, \ldots X_k^n$ is changed to its conditional pmf given

$$V^n = v^{*n},\ F_1 = f_1\left(X_1^n\right) = j_1^*,\ \ldots,\ F_k = f_k\left(X_k^n\right) = j_k^*,$$

the same protocol as above renders $K^{(n)}$ an $((m+1)\epsilon_n, (m+1)\delta_n)$-SK. Under this conditional pmf, with probability 1 the i.i.d. sequence V^{*n} becomes a deterministic sequence v^{*n} and the public message $F_i = f_i\left(X_i^n\right)$ of each input terminal $i \in [1, k]$ equals a constant j_i^*, implying in effect no public communication by the input terminals. Furthermore, the input terminals in $[1, k]$ can transmit over the DMC W at each $t = 1, \ldots, n$, mutually independent rvs $X_{1t}, \ldots, X_{kt}$ with joint pmf

$$P_{X_{[1,k]t}} \ = \ P_{X_{[1,k]t} | V = v_t^*} \ = \ \prod_{i=1}^{k} P_{X_{it} | V = v_t^*},$$

noting that the second equality invokes the conditional independence of $X_1, \ldots, X_k$ given V in (9.6) that underlies general source emulation. Finally, for each $i \in [1, k]$, changing the pmf of X_i^n to the conditional

pmf of X_i^n given $V^n = v^{*n}, F_1 = j_1^*, \ldots, F_k = j_k^*$, which now becomes the conditional pmf of X_i^n given $f_i(X_i^n) = j_i^*$, makes each channel input sequence X_i^n possibly non-i.i.d.

□

9.2.2 General upper bound for SK capacity

In order to state our general upper bound for SK capacity, for rvs $V, X_{\mathcal{M}}$ satisfying (9.5) and

$$P_{VX_{\mathcal{M}}}\left(v, x_{[1,m]}\right) = P_V(v)\, P_{X_{[1,k]}|V}\left(x_{[1,k]} \mid v\right) W\left(x_{[k+1,m]} \mid x_{[1,k]}\right), \quad v \in \mathcal{V},\ x_{[1,m]} \in \mathcal{X}_{\mathcal{M}} \tag{9.8}$$

(but not necessarily the conditional independence of $X_1, \ldots, X_k$ given V in (9.6)) and any fractional partition λ of $\mathcal{M}$, denote

$$G(P_{VX_{\mathcal{M}}}, \lambda) \;=\; H(X_{\mathcal{M}} \mid V) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S \mid X_{S^c}, V) \tag{9.9}$$

and

$$G\left(P_{VX_{[1,k]}}, \lambda\right) \;=\; H\left(X_{[1,k]} \mid V\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_{[1,k]\cap S} \mid X_{[1,k]\cap S^c}, V\right). \tag{9.10}$$

By Lemma 3.5, observe that

$$G(P_{VX_{\mathcal{M}}}, \lambda) \;\geqslant\; 0$$

upon setting $\mathbf{F} = X_{\mathcal{M}}$ and $Y = V$ therein. In the same vein, we have

$$G\left(P_{VX_{[1,k]}}, \lambda\right) \;=\; H\left(X_{[1,k]} \mid V\right) - \sum_{S' \in \mathcal{S}_{[1,k]}} \lambda'_{S'} H\left(X_{S'} \mid X_{S'^c}, V\right) \geqslant 0,$$

where $\lambda' = \left\{\lambda'_{S'}, S' \in \mathcal{S}_{[1,k]}\right\}$ with

$$\lambda'_{S'} \;=\; \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ [1,k]\cap S = S'} \lambda_S$$

is a fractional partition of $[1,k]$ since

$$\sum_{S' \in \mathcal{S}_{[1,k]}:\ S' \ni i} \lambda'_{S'} = \sum_{S' \in \mathcal{S}_{[1,k]}:\ S' \ni i} \ \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ [1,k] \cap S = S'} \lambda_S$$
$$= \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \ni i} \lambda_S = 1 \quad \text{for all } i \in [1,k].$$

Suppose that $K^{(n)}$ represents an (ϵ_n, δ_n)-SK for $\mathcal{M}$ with values in $\mathcal{K}^{(n)}$, achievable with randomization $U_{\mathcal{M}}$, n uses of the DMC W and interactive communication $\mathbf{F}^{(n)}$, and with $\sigma_{\mathtt{var}}\left(K^{(n)}; \mathbf{F}^{(n)}\right) \leqslant \delta_n$ where $\epsilon_n \to 0$ and $\delta_n \to 0$ as $n \to \infty$. Then, by Remark 4.7 with (U_i, X_i^n) in the role of $X_i,\ i \in \mathcal{M}$, and $\mathbf{F}^{(n)}$ in the role of Z, we get

$$\frac{1}{n} \log \left|\mathcal{K}^{(n)}\right| \leqslant \frac{\alpha_n}{n} \Bigg[H\left(U_{\mathcal{M}}, X_{\mathcal{M}}^n \mid \mathbf{F}^{(n)}\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(U_S, X_S^n \mid U_{S^c}, X_{S^c}^n, \mathbf{F}^{(n)}\right) \Bigg] + \beta_n \tag{9.11}$$

with

$$\alpha_n = \frac{1}{1 - m\epsilon_n - \delta_n} \to 1 \text{ and } \beta_n = \frac{mh(\epsilon_n) + h(\delta_n)}{1 - m\epsilon_n - \delta_n} \to 0 \text{ as } n \to \infty.$$

A main step in obtaining our upper bound for C_S is to show that the expression within $\Big[\cdots \Big]$ in the right-side is bounded above by

$$\sum_{t=1}^{n} \Bigg[\left(H\left(X_{\mathcal{M}t}\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_{St} \mid X_{S^c t}\right) \right) - \left(H\left(X_{[1,k]t}\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_{([1,k] \cap S)t} \mid X_{([1,k] \cap S^c)t}\right) \right) \Bigg], \tag{9.12}$$

where the pmf of $X_{\mathcal{M}t},\ t = 1, \ldots, n$, is

$$P_{X_{\mathcal{M}t}}(x_{\mathcal{M}}) \ = \ P_{X_{[1,k]t}}(x_{[1,k]}) W\left(x_{[k+1,m]} \mid x_{[1,k]}\right), \quad x_{\mathcal{M}} \in \mathcal{X}_{\mathcal{M}}.$$

This step, involving a rather tortuous manipulation of information quantities, is omitted.

Finally, we simplify (9.12) by the following standard technique of "single-letterization." Let V be an auxiliary rv distributed uniformly

on $\{1,\dots,n\}$ and independent of $X_{\mathcal{M}}^n$, and set $\tilde{X}_i = X_{iV},\ i \in \mathcal{M}$. Then

$$\sum_{t=1}^{n} H\left(X_{\mathcal{M}t}\right) = nH\left(\tilde{X}_{\mathcal{M}} \mid V\right),$$
$$\sum_{t=1}^{n} \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_{St} \mid X_{S^c t}\right) = n \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(\tilde{X}_S \mid \tilde{X}_{S^c}, V\right), \text{ etc.},$$

and it holds that

$$V \multimap \tilde{X}_{[1,k]} \multimap \tilde{X}_{[k+1,m]} \quad \text{and} \quad P_{\tilde{X}_{[k+1,m]}|\tilde{X}_{[1,k]}} = W.$$

Shedding the tildes we obtain for the new $X_{\mathcal{M}}$, from (9.11), (9.12) and recalling (9.9), (9.10), that

$$\limsup_n \frac{1}{n} \log \left|\mathcal{K}^{(n)}\right| \leqslant G\left(P_{VX_{\mathcal{M}}}, \lambda\right) - G\left(P_{VX_{[1,k]}}, \lambda\right)$$

for every fractional partition λ of $\mathcal{M}$. This leads to the following general upper bound for C_S.

Theorem 9.4 (Upper bound for SK capacity)**.** The SK capacity of a channel model W is bounded above as

$$C_S \leqslant \sup_{P_{VX_{[1,k]}}} \inf_{\lambda} G\left(P_{VX_{\mathcal{M}}}, \lambda\right) - G\left(P_{VX_{[1,k]}}, \lambda\right), \tag{9.13}$$

where $P_{VX_{\mathcal{M}}}$ is as in (9.8) (with W given).

Remark 9.5. The upper bound in (9.13) can be weakened as

$$\begin{aligned} C_S &\leqslant \max_{P_{VX_{\mathcal{M}}}} \min_{\lambda} G\left(P_{VX_{\mathcal{M}}}, \lambda\right) \\ &= \max_{P_{VX_{\mathcal{M}}}} \min_{\lambda} H\left(X_{\mathcal{M}} \mid V\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_S \mid X_{S^c}, V\right), \end{aligned} \tag{9.14}$$

where the maximum is over all $P_{VX_{\mathcal{M}}}$ in (9.8). This weakening differs from the lower bound in Theorem 9.2 by the lack of the conditional independence of $X_1,\dots,X_k$ given V. In fact, even the upper bound in (9.13) can be improved for a special class of channel models.

9.3 Special cases

9.3.1 Channel model with single input

In the special case of a channel model with a sole input terminal, the SK capacity is characterized fully.

Theorem 9.6 (DMC with a single input terminal)**.** The SK capacity of a channel model W with a single input terminal, i.e., $k = 1$, is

$$C_S \;=\; \min_{\lambda} \max_{P_{X_{\mathcal{M}}}} \; H(X_{\mathcal{M}}) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}) \tag{9.15}$$

where the maximum is over all $P_{X_{\mathcal{M}}}$ in (9.3), and can be achieved by simple source emulation.

Corollary 9.7 (Single-input single-output channel model)**.** The SK capacity of a channel model W with $m = 2,\ k = 1$, is

$$C_S \;=\; \max_{P_{X_1}\,:\, P_{X_2|X_1}=W} I\left(X_1 \wedge X_2\right)$$

and can be achieved without any public communication.

Proof. Achievability: Denoting

$$G\left(P_{X_1}, \lambda\right) = H(X_{\mathcal{M}}) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S | X_{S^c}), \tag{9.16}$$

we see by (9.4) that $\max_{P_{X_1}} \min_{\lambda} G\left(P_{X_1}, \lambda\right)$ is an achievable SK rate by simple source emulation, noting that the maximum in (9.4) with respect to $P_{X_{\mathcal{M}}}$ when $k = 1$ is, in effect, over P_{X_1} by (9.3). The achievability of the right-side of (9.15) is established by showing that

$$\max_{P_{X_1}} \min_{\lambda} G\left(P_{X_1}, \lambda\right) = \min_{\lambda} \max_{P_{X_1}} G\left(P_{X_1}, \lambda\right). \tag{9.17}$$

Observe that $G\left(P_{X_1}, \lambda\right)$ is a continuous function of P_{X_1} and λ, defined over convex compact sets. Moreover, $G\left(P_{X_1}, \lambda\right)$ is affine in λ and, as shown below, concave in P_{X_1}. Hence, (9.17) holds by the minimax theorem. For the remaining check of the mentioned concavity, since λ is a fractional partition of $\mathcal{M}$ (see §3.1), we have $\sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \ni 1} \lambda_S = 1$

and so (9.16) can be written as

$$\begin{aligned} G\left(P_{X_1}, \lambda\right) = & \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \ni 1} \lambda_S \big[H(X_{\mathcal{M}}) - H(X_S | X_{S^c}) \big] \\ & - \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \not\ni 1} \lambda_S H(X_S | X_{S^c}) \\ = & \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \ni 1} \lambda_S H(X_{S^c}) \\ & - \sum_{S \in \mathcal{S}_{\mathcal{M}}:\ S \not\ni 1} \lambda_S \big[H\left(X_{\mathcal{M}} \mid X_1\right) - H\left(X_{S^c} \mid X_1\right) \big]. \end{aligned}$$

Since $P_{X_{S^c}}$ is affine in P_{X_1}, $H(X_{S^c})$ is concave in P_{X_1}. Also, $H\left(X_{\mathcal{M}} \mid X_1\right)$ and $H\left(X_{S^c} \mid X_1\right)$ are affine in P_{X_1}. Hence, $G\left(P_{X_1}, \lambda\right)$ is concave in P_{X_1}.

Converse: By the weakened form (9.14) of the general upper bound in Theorem 9.4,

$$\begin{aligned} C_S \ &\leqslant \ \min_{\lambda} \max_{P_{V X_{\mathcal{M}}}} H\left(X_{\mathcal{M}} \mid V\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_S \mid X_{S^c}, V\right) \\ &= \ \min_{\lambda} \max_{P_{X_{\mathcal{M}}}} H\left(X_{\mathcal{M}}\right) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H\left(X_S \mid X_{S^c}\right) \\ &= \ \min_{\lambda} \max_{P_{X_1}} G\left(P_{X_1}, \lambda\right) \end{aligned}$$

where the first equality is seen readily from the optimality of a point mass pmf for V.

The corollary is immediate since the claimed maximum SK rate, a simplification of (9.15) for $m = 2,\ k = 1$, can be achieved by Terminal 1 transmitting a message as SK to Terminal 2 using a capacity achieving code for the DMC W.

□

9.3.2 Channel model with single output

A general single-letter characterization of SK capacity for a channel model W with a sole output terminal, is not known. However, there are intriguing connections between achievable SK rates and the transmission capacity region of the multiple access channel (MAC) defined

by W, without and with feedback from the output terminal. Some of these links are described below.

Let $W\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_{m-1} \to \mathcal{X}_m$ specify a DMC with $k = m-1$ input terminals and a single output terminal. Let $\mathcal{C}$ denote the (transmission) capacity region without feedback of the discrete memoryless MAC W under the average probability of decoding error criterion. Namely, $\mathcal{C}$ is the set of all achievable encoding rate tuples $(R_1, \ldots, R_{m-1})$ for which there exist encoders

$$e_i\colon \mathcal{K}_i \to \mathcal{X}_i^n \quad \text{with} \quad |\mathcal{K}_i| = \lceil \exp\left(nR_i'\right) \rceil \tag{9.18}$$

where R_i' is arbitrarily close to $R_i,\ i \in \mathcal{M}$, with the following attribute: if the MAC inputs are $X_i^n = e_i\left(K_i\right)$, where K_i is distributed uniformly on $\mathcal{K}_i,\ i \in [1, m-1]$, and the rvs $K_i,\ i \in [1, m-1]$ are mutually independent, then $K_i,\ i \in [1, m-1]$ are recoverable simultaneously by a decoder from the MAC output X_m^n with probability approaching 1 as $n \to \infty$. By definition, $\mathcal{C}$ is a closed convex set.

The capacity region $\mathcal{C}$ of the MAC W has the following classic characterization.

Theorem 9.8 (Capacity region of the MAC W)**.** The capacity region $\mathcal{C}$ of the MAC W is the set of rate tuples $(R_1, \ldots, R_{m-1})$ such that

$$0 \leqslant \sum_{i \in S} R_i \leqslant I\left(X_S \wedge X_m \mid X_{S^c \setminus \{m\}}, V\right), \quad S \subseteq [1, m-1], \tag{9.19}$$

for some rvs $V, X_{\mathcal{M}}$ with values in $\mathcal{V}, \mathcal{X}_{\mathcal{M}}$, respectively, and with pmf satisfying (9.5), (9.6), where $|\mathcal{V}| \leqslant k - 1$.

When noiseless causal feedback additionally is available from the MAC output terminal m to the input terminals in $[1, m-1]$, the encoder mappings and MAC inputs in (9.18) are modified as

$$\begin{aligned} &e_{it}\colon \mathcal{K}_i \times \mathcal{X}_m^{t-1} \to \mathcal{X}_i,\ t = 1, \ldots, n, \\ &\quad X_i^n = \left(e_{i1}\left(K_i\right), e_{i2}\left(K_i, X_{m1}\right), \ldots, e_{im}\left(K_i, X_m^{n-1}\right)\right),\ i \in [1, m-1]. \end{aligned} \tag{9.20}$$

Achievable rate tuples $(R_1, \ldots, R_{m-1})$ are defined analogously as for a MAC without feedback, and the capacity region $\mathcal{C}_f$ of the MAC W

with feedback is the set of all achievable rate tuples. Unlike for $\mathcal{C}$, a single-letter characterization of $\mathcal{C}_f$ is known only in special cases but not in general. It is known that $\mathcal{C}_f$ can properly contain $\mathcal{C}$.

The links between the SK capacity C_S for the channel model W and the capacity regions $\mathcal{C}$ and $\mathcal{C}_f$ for the MAC W, involve the largest achievable equal-rate tuples in $\mathcal{C}$ and $\mathcal{C}_f$:

$$R^* \triangleq \max\{R\colon\ (R,\ldots,R) \in \mathcal{C}\}\,, \quad R_f^* \triangleq \max\{R\colon\ (R,\ldots,R) \in \mathcal{C}_f\}\,. \tag{9.21}$$

Whereas a single-letter expression for R^* obtains from that of $\mathcal{C}$ above, an analogous expression for R_f^* is known only in special cases.

Theorem 9.9 (Lower bound for, and positivity of, SK capacity)**.** (i) The SK capacity of a channel model $W\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_{m-1} \to \mathcal{X}_m$ is bounded below as

$$C_S \ \geqslant\ R^*, \tag{9.22}$$

and R^* can be achieved as an SK rate by a protocol in which the input terminals do not communicate, and transmit mutually independent sequences over the DMC W that are not necessarily i.i.d., and the output terminal communicates after completion of transmissions over the DMC.
(ii) $C_S > 0$ iff there exists $(R_1,\ldots,R_{m-1}) \in \mathcal{C}$ such that $R_i > 0$ for each $i \in [1, m-1]$.

Proof. (i) Suppose that $(R,\ldots,R) \in \mathcal{C}$, with $R > 0$. Then, alluding to (9.18), there exist mutually independent rvs $K_i,\ i \in [1, m-1]$, each distributed uniformly on $\mathcal{K} \triangleq \{1,\ldots,\lceil \exp(nR')\rceil\}$ with R' arbitrarily close to R, and codewords $X_i^n = e_i\,(K_i)\,,\ i \in [1, m-1]$ as inputs to the DMC W, such that $K_i,\ i \in [1, m-1]$ are ϵ_n-recoverable simultaneously from the DMC output X_m^n, where $\epsilon_n \to 0$ as $n \to \infty$. Now, fixing an arbitrary $i_1 \in [1, m-1]$, terminal m communicates

$$\mathbf{F} \ =\ (K_{i_1} + K_i \mod |\mathcal{K}|,\ i \in [1, m-1]\backslash\{i_1\})\,.$$

Clearly, $\sigma_{\mathtt{var}}\,(K_{i_1}; \mathbf{F}) = 0$ and K_{i_1} is an $(\epsilon_n, 0)$-SK for $\mathcal{M}$ with $\epsilon_n \to 0$ as $n \to \infty$. Hence, $C_S \geqslant R^*$.

(ii) The sufficiency of the condition for $C_S > 0$ is obvious from (i). For necessity, suppose that no $(R_1, \ldots, R_{m-1})$ as in the condition exists. Then, by the convexity of $\mathcal{C}$ it must hold for some $i_1 \in [1, m-1]$ that $R_{i_1} = 0$ for every $(R_1, \ldots, R_{m-1}) \in \mathcal{C}$ which, in turn, implies that $W(x_m \mid x_1, \ldots, x_{m-1})$ does not depend on x_{i_1}. Consolidating the terminals in $\mathcal{M}\backslash\{i_1\}$ to form a coalition A, any use of the DMC W is tantamount to a randomization performed by A (since the DMC input x_{i_1} does not affect the output x_m). However, the two parties $\{i_1\}$ and A, with only local randomization and communication as resources, cannot generate any secret CR; hence, neither can the terminals in $\mathcal{M}$ in the original model. Thus, $C_S = 0$. □

The proof of Theorem 9.9(i) shows a simple protocol for achieving R^* as an SK rate. In fact, the general source emulation protocol of §9.2.1, too, achieves an SK rate of R^* but cannot exceed it.

Theorem 9.10 (Maximum SK rate by general source emulation)**.** For a channel model $W\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_{m-1} \to \mathcal{X}_m$, a necessary and sufficient condition for R to be an achievable SK rate by general source emulation is $(R, \ldots, R) \in \mathcal{C}$.

Proof. For necessity, consider general source emulation using $V, X_{\mathcal{M}}$ whose pmf $P_{VX_{\mathcal{M}}}$ satisfies (9.5), (9.6) with $k = m-1$. By Theorem 9.2, this protocol achieves an SK rate

$$R = \min_{\lambda} \widetilde{G}(V, X_{\mathcal{M}}, \lambda) = \min_{\lambda} H(X_{\mathcal{M}} \mid V) - \sum_{S \in \mathcal{S}_{\mathcal{M}}} \lambda_S H(X_S \mid X_{S^c}, V). \tag{9.23}$$

Using the conditional independence of $X_1, \ldots, X_{m-1}$ given V and the Markov condition $V \multimap X_{[1,m-1]} \multimap X_m$, we have in (9.23) that

$$\begin{aligned} H(X_{\mathcal{M}} \mid V) &= H\left(X_{[1,m-1]} \mid V\right) + H\left(X_m \mid X_{[1,m-1]}, V\right) \\ &= \sum_{i=1}^{m-1} H(X_i \mid V) + H\left(X_m \mid X_{[1,m-1]}\right); \end{aligned}$$

for $S \ni m$,

$$\begin{aligned} H\left(X_S \mid X_{S^c}, V\right) &= H\left(X_{S\backslash\{m\}}, X_m \mid X_{S^c}, V\right) \\ &= H\left(X_{S\backslash\{m\}} \mid X_{S^c}, V\right) + H\left(X_m \mid X_{[1,m-1]}, V\right) \\ &= \sum_{i \in S\backslash\{m\}} H\left(X_i \mid V\right) + H\left(X_m \mid X_{[1,m-1]}\right); \end{aligned}$$

and for $S \not\ni m$,

$$\begin{aligned} H\left(X_S \mid X_{S^c}, V\right) &= H\left(X_S \mid X_{S^c\backslash\{m\}}, V\right) - I\left(X_S \wedge X_m \mid X_{S^c\backslash\{m\}}, V\right) \\ &= \sum_{i \in S} H\left(X_i \mid V\right) - I\left(X_S \wedge X_m \mid X_{S^c\backslash\{m\}}, V\right). \end{aligned}$$

These, combined with $\sum_{S \in \mathcal{S}_\mathcal{M}:\, S \ni i} \lambda_S = 1,\ i \in \mathcal{M}$, yield the simplification

$$\widetilde{G}\left(V, X_\mathcal{M}, \lambda\right) \ = \sum_{S \in \mathcal{S}_\mathcal{M}:\, S \not\ni m} \lambda_S I\left(X_S \wedge X_m \mid X_{S^c\backslash\{m\}}, V\right). \tag{9.24}$$

For each fixed $\widetilde{S} \subseteq [1, m-1]$, choose $\lambda = \{\lambda_S, S \in \mathcal{S}_\mathcal{M}\}$ as

$\lambda_S = \dfrac{1}{|\widetilde{S}|}$ if $S = \widetilde{S}$ or $S = \mathcal{M}\backslash\{i\}$ for some $i \in \widetilde{S}$, and $\lambda_S = 0$ otherwise.

For this choice of λ,

$$\widetilde{G}\left(V, X_\mathcal{M}, \lambda\right) \ = \frac{1}{|\widetilde{S}|} I\left(X_{\widetilde{S}} \wedge X_m \mid X_{\widetilde{S}^c\backslash\{m\}}, V\right),$$

so that R in (9.23) satisfies

$$R|\widetilde{S}| \ \leqslant \ I\left(X_{\widetilde{S}} \wedge X_m \mid X_{\widetilde{S}^c\backslash\{m\}}, V\right) \tag{9.25}$$

for every $\widetilde{S} \subseteq [1, m-1]$. By Theorem 9.8, we get that $(R, \ldots, R) \in \mathcal{C}$.

Turning to sufficiency, suppose that $(R, \ldots, R) \in \mathcal{C}$. By Theorem 9.8, for some $V, X_\mathcal{M}$ with $X_1, \ldots, X_{m-1}$ conditionally independent given V, $V \multimap X_{[1,m-1]} \multimap X_m$ and $P_{X_m|X_{[1,m-1]}} = W$, the inequalities in (9.25) are met. By (9.24), (9.25),

$$\begin{aligned} \widetilde{G}\left(V, X_\mathcal{M}, \lambda\right) \ &\geqslant \sum_{S \in \mathcal{S}_\mathcal{M}:\, S \not\ni m} \lambda_S R|S| \\ &\geqslant \ R\left(\sum_{S \in \mathcal{S}_\mathcal{M}} \lambda_S |S| - \sum_{S \in \mathcal{S}_\mathcal{M}:\, S \ni m} \lambda_S (m-1)\right). \end{aligned}$$

Since

$$\sum_{S\in\mathcal{S}_{\mathcal{M}}} \lambda_S|S| = \sum_{i=1}^{m} \sum_{S\in\mathcal{S}_{\mathcal{M}}:\ S\ni i} \lambda_S = m \ \text{ and } \sum_{S\in\mathcal{S}_{\mathcal{M}}:\ S\ni m} \lambda_S = 1,$$

we get that $\min_\lambda \widetilde{G}\,(V, X_{\mathcal{M}}, \lambda) \ \geqslant\ R$, so that by Theorem 9.2 R is an SK rate achievable by general source emulation. □

Remark 9.11. We remark that general source emulation can strictly outperform simple source emulation by achieving higher SK rates. In the same vein as Theorem 9.10, it can be seen that R is an achievable SK rate by simple source emulation iff $(R, \ldots, R)$ is in the polyhedron

$$\left\{ (R_1, \ldots, R_{m-1}):\ R_i \geqslant 0, \sum_{i\in S} R_i \leqslant I\left(X_S \wedge X_m \mid X_{S^c\setminus\{m\}}\right), \right.$$
$$\left. S \subseteq [1, m-1] \right\}$$

for some i.i.d. rvs $X_1, \ldots, X_{m-1}$ and with $P_{X_m|X_{[1,m-1]}} = W$. The capacity region $\mathcal{C}$ of the MAC W is the convex closure of the union of all such polyhedra whereas the union is known to be nonconvex in general, explaining the remark.

The simple SK generation protocol of Theorem 9.9(i) as well as general source emulation achieve an SK rate of at most R^*. This SK rate of R^* cannot be bettered even if general source emulation were relaxed so as to free each channel input sequence $X_{i1}, \ldots, X_{in}$ from being i.i.d., $i \in [1, m-1]$. Such a relaxation is realized as follows. First, the terminals in $\mathcal{M}$ engage in an initial round of communication, depicted collectively by $F_{(0)} = F_{(0)}\,(U_{\mathcal{M}})$. Next, channel transmissions $X_{i1}, \ldots, X_{in}$, not necessarily i.i.d., $i \in [1, m-1]$, occur with no intervening communication, i.e., $F_{(t)} =$ constant in intervals $t = 1, \ldots, n-1$. Then the terminals in $\mathcal{M}$ enter into a concluding round of communication $F_{(n)} = F_{(n)}\left(U_{\mathcal{M}}, F_{(0)}, X_{\mathcal{M}}^n\right)$. Finally, an SK for $\mathcal{M}$ is generated in the form of local estimates

$$K_i^{(n)} \ = \ K_i^{(n)}\left(U_i, X_i^n, F_{(0)}, \mathbf{F}\right), \quad i \in \mathcal{M}, \tag{9.26}$$

where $\mathbf{F} = \left(F_{(1)}, \ldots, F_{(n)}\right)$.

We note that in all the three protocols above, the inputs to the DMC W are chosen without knowledge of its previous outputs. This raises the question: Can complex protocols that select channel inputs based on feedback from the output terminal, attain SK rates exceeding R^*? While the secrecy requirement precludes full public feedback by terminal m, a coding scheme with judicious partial feedback which affords a gain in (secure) transmission rates that exceed the information leaked by feedback, might enhance SK rates beyond R^* and perhaps nearing $R_f^* \geqslant R^*$.

Consider the following augmented SK generation protocol entailing feedback-dependent channel transmissions and limited input communication, of which general source emulation and its relaxation above are special cases. After an initial round of communication by the terminals in $\mathcal{M}$, described collectively by $F_{(0)} = F_{(0)}\left(U_{\mathcal{M}}\right)$, all the communication in intervals $t = 1, \ldots, n-1$ are only by the DMC output terminal m with $F_{(t)} = F_{(t)}\left(U_m, X_m^t, F_{(0)}, F^{(t-1)}\right)$. The channel transmissions $X_{it} = X_{it}\left(U_i, X_i^{t-1}, F_{(0)}, F^{(t-1)}\right),\ t = 1, \ldots, n$, from each input terminal $i \in [1, m-1]$ depend causally on all the information available to it. Upon completion of the channel transmissions, a last round of communication $F_{(n)} = F_{(n)}\left(U_{\mathcal{M}}, X_{\mathcal{M}}^n, F_{(0)}, F^{(n-1)}\right)$ occurs by the terminals in $\mathcal{M}$. Thereupon, an SK for $\mathcal{M}$ is generated as in (9.26) with $\mathbf{F} = \left(F_{(1)}, \ldots, F_{(n)}\right)$ as described here. Such a protocol can be shown to achieve an SK rate that is bounded above by R_f^*, and achieves R_f^* for a symmetric DMC with

$$W\left(x_m \mid x_1, \ldots, x_{m-1}\right) \;=\; W\left(x_m \mid x_{\sigma(1)}, \ldots, x_{\sigma(m-1)}\right) \tag{9.27}$$

for every permutation σ of $\{1, \ldots, m-1\}$.

Owing to the lack of a general single-letter characterization of $\mathcal{C}_f$ and R_f^*, proof techniques for the previous protocol circumvent this difficulty by converting transmission schemes for a MAC with feedback directly into SK generation schemes, all involving multi letter expressions. The proofs of R_f^* as an upper bound for SK rates achieved by the augmented protocol and the achievability of R_f^* for a symmetric MAC, are omitted.

It remains open whether a sufficiently complex protocol can attain, in general, an SK rate of R_f^* or more. We close with an example in-

volving a MAC W for which a single-letter characterization of $\mathcal{C}_f \supset \mathcal{C}$ is known.

Example 9.12 (Arithmetic adder MAC)**.** Consider a DMC $W\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_3$ with $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$, $\mathcal{X}_3 = \{0,1,2\}$, and

$$W(x_3 \mid x_1, x_2) \;=\; \mathbb{1}(x_3 = x_1 + x_2).$$

The capacity region of the MAC W is

$$\mathcal{C} = \{(R_1, R_2) :\; 0 \leqslant R_1, R_2 \leqslant 1,\; R_1 + R_2 \leqslant 1.5\}$$

and the capacity region with feedback is

$$\mathcal{C}_f = \{(R_1, R_2) :\; 0 \leqslant R_1, R_2 \leqslant 1,\; R_1 + R_2 \leqslant 1.58226\}.$$

For a class of channel models that include the present W, a refinement of the proof of Theorem 9.4 yields an improved upper bound $C_S \leqslant R_f^*$. Since W is a symmetric DMC (9.27), R_f^* is an achievable SK rate by the previous augmented SK generation protocol. Hence, $C_S = R_f^* = 0.79113$.

9.4 Story of results and open problems

The realization that secure transmission, from a legitimate transmitter to a legitimate receiver over a noisy channel when an eavesdropper has access to wiretap side information, can be enhanced by public communication was illustrated first in [48] and shown comprehensively in [53]. SK capacity for a single-input single-output predecessor of the multiterminal channel model of this chapter, stated as Corollary 9.7, was characterized in [54, 1]. Our treatment follows [22, 23, 84].

The SK capacity for a single-input multiple-output channel model in Theorem 9.6 was obtained in [22]. Simple source emulation sufficed to achieve SK capacity, and a general upper bound for achievable SK rates was shown to be tight. Developed further for a channel model with multiple input and output terminals, the lower bound for SK capacity based on general source emulation in Theorem 9.2 and the upper bound in Theorem 9.4, are from [23]. Links between achievable SK rates for a channel model with a single output terminal and the transmission

capacity region of the underlying MAC in §9.3.2, were obtained in [23]. The refinement for a special class of channel models of the upper bound of Theorem 9.4 in terms of the farthest equal-rate tuple in the MAC feedback capacity region, and described in the concluding segment of §9.3.2, was shown to be tight in [84].

Extensions and open problems. As for its source counterpart in Chapter 6, the multiterminal channel model of this chapter is a specialization of helper and privacy models of broader scope described in [22, 23]. The corresponding secrecy capacities for a single-input model were resolved fully in [22]. These capacities are shown to be achievable by simple source emulation, and by even simpler means that involve no public communication by the channel input terminal. When the eavesdropper additionally possesses wiretap side information Z^n, known secrecy capacities for associated privacy models yield upper bounds for wiretap secrecy capacities that are tight only under special circumstances [22]. More general results for this latter model can be found in [29].

For the expanded helper and privacy settings in [23] of the general multiterminal channel model of §9.1, only partial results are available. The SK capacity is unknown even in the simplest setting of a SK for all the terminals in $\mathcal{M}$ that is concealed from an eavesdropper with access to (only) the public communication. The discussion in §9.3.2 exposes the inadequacies of both the lower bound in Theorem 9.2 emerging from general source emulation and the upper bound in Theorem 9.4. Evidence of the difficulty in creating suitable communication-based correlations among channel input terminals is manifest already in the special case of a multiaccess channel model with a sole output terminal described in that section. However, offering potential insights for an approach are beneficial links to the capacity region (with and without feedback) of the underlying MAC, identified in [23, 84]. The crux of the challenge in this SK capacity characterization is a precise understanding of the tradeoffs between gains in channel transmission rates using coding schemes that are bolstered by interactive communication protocols, and the information leaked by the latter. Extensions to wiretap secret key capacity remain unvanquished.

Acknowledgements

This monograph grew out of collaborative works with Imre Csiszár, Sirin Nitinawarat and Shun Watanabe, to whom we are deeply indebted. We are grateful to Shun for saving us from many a subtle flaw with his critique, and for permitting us to include unpublished matter; to Madhu Sudan and Alex Vardy for pointed and helpful guidance; and to the anonymous referees for their observations. We thank Sergio Verdú and Mike Casey for inviting us to write; and Sergio for safeguarding technical and linguistic clarity.

We gratefully acknowledge the support of this research and writing project by the U.S. National Science Foundation under Grants CCF1117546 and CCF1527354, and the Start-Up Grant of the Indian Institute of Science, Bangalore.

References

[1] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–Part I: Secret sharing. *IEEE Trans. Inf. Theory*, 39(4):1121–1132, July 1993.

[2] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography–Part II: CR capacity. *IEEE Trans. Inf. Theory*, 44(1):225–240, January 1998.

[3] R. Ahlswede and I. Csiszár. On oblivious transfer capacity. *Information Theory, Combinatorics, and Search Theory*, pages 145–166, 2013.

[4] K. Audenaert. A sharp Fannes-type inequality for the von Neumann entropy. *Physical Review A*, 40:8127–8136, 2007. arXiv:quant-ph/0610146.

[5] M. Bellare, S. Tessaro, and A. Vardy. *Semantic Security for the Wiretap Channel*, pages 294–311. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[6] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. Generalized privacy amplification. *IEEE Trans. Inf. Theory*, 41(6):1915–1923, November 1995.

[7] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17(2):210–229, 1988.

[8] M. Braverman. Coding for interactive computation: progress and challenges. *Proc. Conference on Communication, Control, and Computing (Allerton)*, pages 1914–1921, October 2012.

[9] N.J. Cerf, S. Massar, and S. Schneider. Multipartite classical and quantum secrecy monotones. *Physical Review A*, 66(4):042309, October 2002.

[10] C. Chan. Generating secret in a network. *Ph. D. Dissertation, Massachussetts Institute of Technology*, 2010.

[11] C. Chan. Linear perfect secret key agreement. In *Proc. Information Theory Workshop (ITW)*, pages 723–726, October 2011.

[12] C. Chan. Agreement of a restricted secret key. *Proc. IEEE International Symposium on Information Theory*, pages 1782–1786, July 2012.

[13] C. Chan and L. Zheng. Mutual dependence for secret key agreement. *Proc. Annual Conference on Information Sciences and Systems (CISS)*, March 2010.

[14] T.A. Courtade and T.R. Halford. Coded cooperative data exchange for a secret key. *IEEE Trans. Inf. Theory*, 62(7):3785–3795, July 2016.

[15] C. Crépeau and J. Kilian. *Weakening Security Assumptions and Oblivious Transfer*, pages 2–7. Springer New York, New York, NY, 1990.

[16] I. Csiszár. Almost independence and secrecy capacity. *Prob. Pered. Inform.*, 32(1):48–57, 1996.

[17] I Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[18] I. Csiszár and J. Körner. Towards a general theory of source networks. *IEEE Trans. Inf. Theory*, 26(2):155–165, March 1980.

[19] I. Csiszár and J. Körner. *Information theory: Coding theorems for discrete memoryless channels. 2nd edition.* Cambridge University Press, 2011.

[20] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Trans. Inf. Theory*, 46(2):344–366, March 2000.

[21] I. Csiszár and P. Narayan. Secrecy capacities for multiple terminals. *IEEE Trans. Inf. Theory*, 50(12):3047–3061, December 2004.

[22] I. Csiszár and P. Narayan. Secrecy capacities for multiterminal channel models. *IEEE Trans. Inf. Theory*, 54(6):2437–2452, June 2008.

[23] I. Csiszár and P. Narayan. Secrecy generation for multiaccess channel models. *IEEE Trans. Inf. Theory*, 59(1):17–31, January 2013.

[24] S. Fehr and S. Berens. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory*, 60(11):6801–6810, November 2014.

[25] M. Fitzi, S. Wolf, and J. Wullschleger. *Pseudo-signatures, Broadcast, and Multi-party Computation from Correlated Randomness*, pages 562–578. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[26] A. El Gamal and A. Orlitsky. Interactive data compression. In *Proc. Annual Symposium on Foundations of Computer Science*, pages 100–108, October 1984.

[27] A. Gohari, M.H. Yassaee, and M.R. Aref. Secure channel simulation. In *Information Theory Workshop (ITW)*, pages 406–410, September 2012.

[28] A. A. Gohari and V. Anantharam. Information-theoretic key agreement of multiple terminals - Part I. *IEEE Trans. Inf. Theory*, 56(8):3973 – 3996, August 2010.

[29] A. A. Gohari and V. Anantharam. Information-theoretic key agreement of multiple terminals - Part II: Channel model. *IEEE Trans. Inf. Theory*, 56(8):3997 – 4010, August 2010.

[30] S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270 – 299, 1984.

[31] D. Gunduz, E. Erkip, and H.V. Poor. Lossless compression with security constraints. *Proc. IEEE International Symposium on Information Theory*, pages 111–115, July 2008.

[32] Y. Liang H. Zhang, L. lai and H. Wang. The capacity region of the source-type model for secret key and private key generation. *IEEE Trans. Inf. Theory*, 60(10):6389–6398, October 2014.

[33] T. S. Han. *Information-Spectrum Methods in Information Theory [English Translation]*. Series: Stochastic Modelling and Applied Probability, Vol. 50, Springer, 2003.

[34] T. S. Han and K. Kobayashi. A unified achievable rate region for a general class of multiterminal source coding systems. *IEEE Trans. Inf. Theory*, 26(3):277–288, May 1980.

[35] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993.

[36] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28:12–24, 1999.

[37] M. Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Trans. Inf. Theory*, 57(6):3989–4001, June 2011.

[38] M. Hayashi, H. Tyagi, and S. Watanabe. Secret key agreement: General capacity and second-order asymptotics. *Proc. IEEE International Symposium on Information Theory*, pages 1136–1140, July 2014.

[39] M. Hayashi, H. Tyagi, and S. Watanabe. Secret key agreement: General capacity and second-order asymptotics. *IEEE Trans. Inf. Theory*, 62(7), May 2016.

[40] H. Imai, K. Morozov, A. C. Nascimento, and A. Winter. Efficient protocols achieving the commitment capacity of noisy correlations. *Proc. IEEE International Symposium on Information Theory*, pages 1432–1436, 2006.

[41] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proc. ACM Symposium on Theory of Computing*, pages 12–24, 1989.

[42] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *Proc. Annual Symposium on Foundations of Computer Science*, pages 248–253, November 1989.

[43] M. Iwamoto and K. Ohta. Security notions for information theoretically secure encryptions. *Proc. IEEE International Symposium on Information Theory*, pages 1777–1781, July 2011.

[44] A. H. Kaspi. Two-way source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, 31(6):735–740, November 1985.

[45] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman & Hall/CRC, 2007.

[46] J. Kilian. Founding crpytography on oblivious transfer. In *Proc. Symposium on Theory of Computing*, pages 20–31, 1988.

[47] R. Konig, R. Renner, and C. Schaffner. The operational meaning of min- and max-entropy. *IEEE Trans. Inf. Theory*, 55(9):4337 – 4347, September 2009.

[48] S. K. Leung-Yan-Cheong. Multi-user and wiretap channels including feedback. *Ph. D. Dissertation, Stanford University*, 1976.

[49] Y. Liang, H. V. Poor, and S. Shamai (Shitz). Information theoretic security. *Found. Trends Commun. Inf. Theory*, 5(4-5):355–580, April 2009.

[50] J. Liu, P. Cuff, and S. Verdú. Common randomness and key generation with limited interaction. *CoRR*, abs/1601.00899v2, 2016.

[51] M. Madiman and P. Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inf. Theory*, 56(6):2699–2713, June 2010.

[52] J. L. Massey. An introduction to contemporary cryptology. *Proc. the IEEE*, 76(5):533–549, 1988.

[53] U. M. Maurer. Provably secure key distribution based on independent channels. *IEEE Information Theory Workshop (ITW)*, pages 49–71, June 1990.

[54] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory*, 39(3):733–742, May 1993.

[55] M. Mukherjee, C. Chan, N. Kashyap, and Q. Zhou. Bounds on the communication rate needed to achieve SK capacity in the hypergraphical source model. *CoRR*, arXiv:1601.05377v2, 2016.

[56] M. Mukherjee and N. Kashyap. On the communication complexity of secret key generation in the multiterminal source model. *Proc. IEEE Symposium on Information Theory*, June 2014.

[57] M. Mukherjee, N. Kashyap, and Y. Sankarasubramaniam. On the public communication needed to achieve SK capacity in the multiterminal source model. *CoRR*, abs/1507.02874, 2015.

[58] A.C.A. Nascimento and A. Winter. On the oblivious transfer capacity of noisy correlations. *Proc. IEEE International Symposium on Information Theory*, pages 1871–1875, July 2009.

[59] S. Nitinawarat and P. Narayan. Perfect omniscience, perfect secrecy, and Steiner tree packing. *IEEE Trans. Inform. Theory*, 56(12):6490 – 6500, December 2010.

[60] S. Nitinawarat and P. Narayan. Secret key generation for correlated Gaussian sources. *IEEE Trans. Inf. Theory*, 58(6):3373–3391, June 2012.

[61] S. Nitinawarat, C. Ye, A.Barg, P. Narayan, and A. Reznik. Secret key generation for a pairwise independent network model. *IEEE Trans. Inform. Theory*, 56(12):6482 – 6489, December 2010.

[62] A. Orlitsky. Worst-case interactive communication I: Two messages are almost optimal. *IEEE Trans. Inf. Theory*, 36(5):1111–1126, September 1990.

[63] A. Orlitsky and A. El Gamal. Communication with secrecy constraints. In *Proc. ACM Symposium on Theory of Computing*, pages 217–224, October 1984.

[64] V. Prabhakaran and K. Ramchandran. On secure distributed source coding. In *IEEE Information Theory Workshop (ITW)*, pages 442–447, September 2007.

[65] P. A. Regalia, A. Khisti, Y. Liang, and S. Tomasin (editors). Secure communications via physical-layer and information-theoretic techniques. *Proc. the IEEE*, 103(10), October 2015.

[66] R. Renner. Security of quantum key distribution. *Ph. D. Dissertation, ETH Zurich*, 2005.

[67] R. Renner and S. Wolf. *New Bounds in Secret-Key Agreement: The Gap between Formation and Secrecy Extraction*, pages 562–577. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[68] R. Renner and S. Wolf. Smooth Rényi entropy and applications. *Proc. IEEE International Symposium on Information Theory*, page 233, June 2004.

[69] R. Renner and S. Wolf. *Simple and Tight Bounds for Information Reconciliation and Privacy Amplification*, pages 199–216. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[70] M. Santha and U. V. Vazirani. Generating quasi-random sequences from slightly-random sources. In *Proc. Annual Symposium on Foundations of Computer Science*, pages 434–440, October 1984.

[71] M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75 – 87, 1986.

[72] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, October 1949.

[73] R. Tandon, S. Ulukus, and K. Ramchandran. Secure source coding with a helper. *Proc. Conference on Communication, Control, and Computing (Allerton)*, pages 1061–1068, October 2009.

[74] H. Tyagi. Distributed computing with privacy. *Proc. IEEE International Symposium on Information Theory*, pages 1157–1161, July 2012.

[75] H. Tyagi. Common information and secret key capacity. *IEEE Trans. Inf. Theory*, 59(9):5627–5640, September 2013.

[76] H. Tyagi. Common randomness principles of secrecy. *Ph. D. Dissertation, Univeristy of Maryland, College Park*, 2013.

[77] H. Tyagi. Distributed function computation with confidentiality. *IEEE Journal on Selected Areas in Communications*, 31(4):691–701, April 2013.

[78] H. Tyagi, N. Kashyap, Y. Sankarasubramaniam, and K. Viswanathan. Fault-tolerant secret key generation. *Proc. IEEE International Symposium on Information Theory*, pages 1787–1791, July 2012.

[79] H. Tyagi and P. Narayan. How many queries will resolve common randomness? *IEEE Trans. Inf. Theory*, 59(9):5363–5378, September 2013.

[80] H. Tyagi and P. Narayan. How many queries will resolve common randomness? *Proc. IEEE International Symposium on Information Theory*, pages 3165–3169, July 2013.

[81] H. Tyagi, P. Narayan, and P. Gupta. Secure computing. *Proc. IEEE International Symposium on Information Theory*, pages 2612–2616, June 2010.

[82] H. Tyagi, P. Narayan, and P. Gupta. When is a function securely computable? *Proc. IEEE International Symposium on Information Theory*, pages 2876–2880, August 2011.

[83] H. Tyagi, P. Narayan, and P. Gupta. When is a function securely computable? *IEEE Trans. Inf. Theory*, 57(10):6337–6350, October 2011.

[84] H. Tyagi and S. Watanabe. Secret key capacity for multiple access channel with public feedback. *Proc. Conference on Communication, Control, and Computing (Allerton)*, pages 1–7, October 2013.

[85] H. Tyagi and S. Watanabe. *A Bound for Multiparty Secret Key Agreement and Implications for a Problem of Secure Computing*, pages 369–386. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[86] H. Tyagi and S. Watanabe. Unpublished notes. 2014.

[87] H. Tyagi and S. Watanabe. Converses for secret key agreement and secure computing. *IEEE Trans. Inf. Theory*, 61:4809–4827, September 2015.

[88] S. P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7:1–336, 2012.

[89] S. Vembu and S. Verdú. Generating random bits from an arbitrary source: Fundamental limits. *IEEE Trans. Inf. Theory*, 41(5):1322–1332, Sep. 1995.

[90] J. von Neumann. Various techniques used in connection with random digits. 1963.

[91] Y. Wang and P. Ishwar. On unconditionally secure multi-party sampling from scratch. *Proc. IEEE Symposium on Information Theory (ISIT)*, pages 1782–1786, July 2011.

[92] S. Watanabe and Y. Oohama. Secret key agreement from vector Gaussian sources by rate limited public communication. *Information Forensics and Security, IEEE Transactions on*, 6(3):541–550, September 2011.

[93] S. Winkler and J. Wullschleger. *On the Efficiency of Classical and Quantum Oblivious Transfer Reductions*, pages 707–723. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[94] S. Winkler and J. Wullschleger. On the efficiency of classical and quantum secure function evaluation. *IEEE Trans. Inf. Theory*, 60(6):3123–3143, June 2014.

[95] A. Winter, A.C.A. Nascimento, and H. Imai. *Commitment Capacity of Discrete Memoryless Channels*, pages 35–51. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[96] S. Wolf and J. Wullschleger. New monotones and lower bounds in unconditional two-party computation. *IEEE Trans. Inf. Theory*, 54(6):2792–2797, June 2008.

[97] A. D. Wyner. Recent results in the Shannon theory. *IEEE Trans. Inf. Theory*, 20(1):2–10, January 1974.

[98] A. D. Wyner. The wiretap channel. *Bell System Technical Journal*, 54(8):1355–1367, October 1975.

[99] A. D. Wyner, J. K. Wolf, and F. M. J. Willems. Communicating via a processing broadcast satellite. *IEEE Trans. Inf. Theory*, 48(6):1243–1249, June 2002.

[100] A. C. Yao. Some complexity questions related to distributive computing. *Proc. Annual Symposium on Theory of Computing*, pages 209–213, 1979.

[101] C. Ye and P. Narayan. The secret key-private key capacity region for three terminals. *Proc. IEEE International Symposium on Information Theory*, pages 2142–2146, September 2005.

[102] C. Ye and P. Narayan. Secret key and private key constructions for simple multiterminal source models. *IEEE Trans. Inf. Theory*, 58(2):639–651, February 2012.

[103] C. Ye and A. Reznik. Group secret key generation algorithms. *Proc. IEEE International Symposium on Information Theory*, pages 2596–2600, June 2007.

[104] Z. Zhang. Estimating mutual information via Kolmogorov distance. *IEEE Trans. Inf. Theory*, 53(9):3280–3282, September 2007.