Generating Secret in a Network

by

Chung Chan

Bachelor of Science, Massachusetts Institute of Technology (2004)
Master of Engineering, Massachusetts Institute of Technology (2005)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author ....
Department of Electrical Engineering and Computer Science
August 25th, 2010

Certified by..
Lizhong Zheng
Associate Professor
Thesis Supervisor

Accepted by..
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Generating Secret in a Network

by

Chung Chan

Submitted to the Department of Electrical Engineering and Computer Science
on August 25th, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

This monograph studies the theory of information through the multiuser secret key agreement problem. A general notion of mutual dependence is established for the secrecy capacity, as a natural generalization of Shannon's mutual information to the multivariate case. Under linear-type source models, this capacity can be achieved practically by linear network codes. In addition to being an unusual application of the network coding solution to a secrecy problem, it gives secrecy capacity an interpretation of network information flow and partition connectivity, further confirming the intuitive meaning of secrecy capacity as mutual dependence. New identities in submodular function optimization and matroid theory are discovered in proving these results. A framework is also developed to view matroids as graphs, allowing certain theory on graphs to generalize to matroids.

In order to study cooperation schemes in a network, a general channel model with multiple inputs is formulated. Single-letter secrecy capacity upper bounds are derived using the Shearer-type lemma. Lower bounds are obtained with a new cooperation scheme called the mixed source emulation. In the same way that mixed strategies may surpass pure strategies in zero-sum games, mixed source emulation outperforms the conventional pure source emulation approach in terms of the achievable key rate. Necessary and sufficient conditions are derived for tightness of these secrecy bounds, which shows that secrecy capacity can be characterized for a larger class of channels than the broadcast-type channels considered in previous work. The mixed source emulation scheme is also shown to be unnecessary for some channels while insufficient for others. The possibility of a better cooperative scheme becomes apparent, but a general scheme remains to be found.

Thesis Supervisor: Lizhong Zheng
Title: Associate Professor

Acknowledgments

I would like to acknowledge the support of the Research Laboratory of Electronics at Massachusetts Institute of Technology (MIT), Laboratory of Information and Decision Systems at MIT, and the Shun Hing Institute of Advanced Engineering at the Chinese University of Hong Kong (CUHK). In particular, I would like to thank my supervisor, Professor Lizhong Zheng, and my thesis committee members, Professor Vivek K. Goyal and Professor Gregory W. Wornell, for their help and valuable comments. I would like to thank Professor Angela Y. Zhang, Professor Anthony M. So, and my advisor for giving me an opportunity to work at CUHK, where I had great freedom to work on problems of my interest that eventually find their ways to this thesis. I would also like to thank Professor Raymond W. Yeung for the postdoctoral research opportunity on network coding in the Information Engineering Department of CUHK.

In addition to the generous support of my education, I would like to thank MIT and the alumni for providing many recreational facilities, such as the ice-skating rink, the sailing pavillion, the swimming pool and the heavy sandbags, which I often enjoy much more than my research. To show my gratitude, I am comitting myself to sailing every non-rainy day starting from my 10-year anniversary at MIT on Aug 17 until I leave on Aug 28. The ten years of experience here has been transformational. I have gained many precious memories being with my colleagues, teachers, students, roommates, host family, classmates and friends. Thank you all for supporting me while I was lost, putting up with or pointing out my mistakes. I am especially thankful to my host family for showing me around Boston and New York City, my friends in San Diego for helping me during my internship at Qualcomm, and my previous roommate in Ashdown who spended tremedous time and effort helping my dad fight with liver cancer. Above all, I am most grateful for my family, my parents and my brother. Your love and care has given me great support.

Chung Chan

Aug 17, 2010

Massachusetts Institute of Technology

Contents

Chapter 1

Secret Key Agreement

This monograph focuses on a problem called *secret key agreement* [1, 12, 13, 42]. It is a study of how a network of users can generate and share a common secret using some private and public resources. The emphasis here is *not* on the communication of a given secret from one user to the others, but rather on the generation of a common secret that is based on correlated private observations, and agreed upon after public discussion. Hence, the title of this dissertation is "*generating secret in a network*".

1.1 A secret agreement game

Let us introduce the topic informally with the following game that is played by a group of people. One player is selected to be a wiretapper while others are users. Each player writes something on a piece of paper that they cannot reveal to others.

Rule: The users win iff everyone except the wiretapper writes down the same thing.

In other words, the wiretapper wins if the users cannot agree on a common secret. To help the users settle on a common secret, they are allowed to discuss publicly with each other as long as they are clearly heard by everyone including the wiretapper. Is there a winning strategy for the users?

Although the users are allowed to discuss, they cannot communicate any secret in public. So, is it even possible to obtain a common secret? Suppose the users discuss and learn that they are all interested in basketball but the wiretapper is not; they all know the winning team of a recent match but the wiretapper does not. Then, the users can simply put down the winning team as their common secret. If some users missed the match, others can help them by naming a few players in the winning team. Even though the users cannot communicate any secret to the others in public, they can extract it from their correlated private observations.

If this game is played repeatedly with the same users and wiretapper, the users will eventually lose by running out of ideas for their common secret. Intuitively, the closer the users are, the more the secrets they can share. Can we equate the amount of secrets to the correlation of private observations of the users? If so, is there a systematic and practical way to consolidate such correlation into secrets? If the users are given some time to discuss privately without the wiretapper, how can they cooperate to enhance their correlation? More importantly, what is the significance of solving this specific problem in cryptography? We will address these questions in the information-theoretic framework by Shannon [51].

1. Can we equate secrets to the correlation of the users?
2. If so, how to consolidate such correlation into secrets?
3. How can users cooperate to enhance their correlation?
4. What is the significance of secret generation?

1.2 Information-theoretic cryptography

Historically, the idea of secret key agreement by public discussion is a creative solution to some of the challenges in the development of information-theoretic cryptography. On the one hand, it is an application of information theory to the study of provably secure systems; on the other hand, it provides a nice framework to study the nature and properties of information. In the following, we highlight some of the key advances and challenges in the theory of information developed under this secrecy framework.

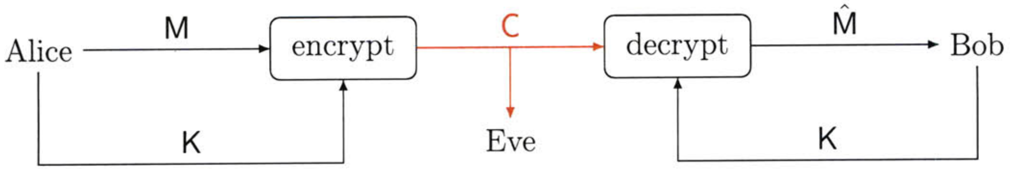


Figure 1-1: Symmetric key cipher

1.2.1 Shannon cipher system

In [51], Shannon applied the information theory he invented in [52] to study the *symmetric-key cipher* in Figure 1-1. Alice and Bob share a secret key K unknown to Eve. Alice wants to communicate a message M for Bob that is independent of K. She uses K to encrypt M into a cryptogram C, which is sent to Bob in public. Bob uses K to decrypt C into the message estimate $\hat{\mathsf{M}}$, while Eve tries to learn M from C without the key. The goal is to have Bob recover the message, i.e. $\hat{\mathsf{M}} = \mathsf{M}$, but Eve learn nothing about it, i.e. M independent of C.

Shannon showed that this *perfect secrecy* is possible if and only if the length of the key is no shorter than the length of the message. Necessity follows easily from the properties of Shannon's measure: the entropy H and mutual information I.[1]

$$H(\mathsf{M}) \overset{\text{(a)}}{=} H(\mathsf{M}|\mathsf{C}) \overset{\text{(b)}}{=} H(\mathsf{M}|\mathsf{CK}) + I(\mathsf{M} \wedge \mathsf{K}|\mathsf{C}) \overset{\text{(c)}}{\leq} H(\mathsf{K})$$

The length of the message after compression is the entropy $H(\mathsf{M})$. Since perfect secrecy requires that C is independent of M, the entropy is not decreased by the knowledge of C, which gives the equality (a). The randomness $H(\mathsf{M}|\mathsf{C})$ conditioned on C can be decomposed into two parts: $H(\mathsf{M}|\mathsf{CK})$, which measures the randomness independent of K; and $I(\mathsf{M} \wedge \mathsf{K}|\mathsf{C})$, which measures the randomness correlated with K. This gives the identity (b). Since Bob has to recover M perfectly, the first part is 0; the second part measures only part of the randomness of K and is therefore upper bounded by $H(\mathsf{K})$ in (c). Hence, the key K has to be as random as the message M if it completely resolves the randomness of the message M given the cryptogram C that is independent of M.

[1] See [8, 11] or Section A.1 for definitions.

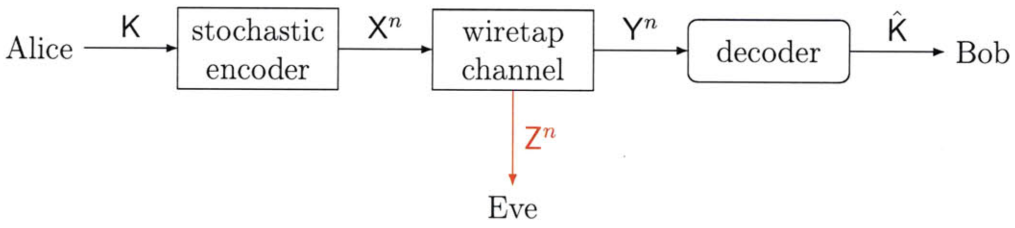


Figure 1-2: Wiretap channel

The last inequality (c) can be achieved with equality by the *one-time pad*,

$$\mathsf{C} := \mathsf{M} \oplus \mathsf{K} \qquad \text{(encryption)}$$

$$\hat{\mathsf{M}} := \mathsf{C} \oplus \mathsf{K} \qquad \text{(decryption)}$$

where M and K are uniformly random bits and $\oplus$ is elementwise modulo-2 addition. It is easy to show that C remains uniformly random given M, and $\hat{\mathsf{M}} = \mathsf{M}$ as desired.

The optimistic view of this result is that secrecy can be stored in the form of a key. If two users can generate one bit of common secret, whether it corresponds to meaningful data or not, any one of them can use it to encrypt one bit of private message for the other. The pessimistic view, however, is that the users have to find a way to share a secret key long enough to encrypt their private message perfectly securely. The original problem of communicating a private message turns into the new problem of generating a common secret key.

Lesson: Secrecy can be stored in a key for secure communication.
Challenge: Secure communication of a long message requires a long key.

1.2.2 Wiretap channel

A remedy to the problem of generating a long common secret key was given by Wyner in [61]. He showed that a secret key is not needed for secure communication when the wiretapper's observation is degraded by additional channel noise. This result was extended to the general broadcast-type wiretap channel in Figure 1-2 by Csiszár [10]. Alice encodes uniformly random bits K into a channel input sequence X^n, and sends it

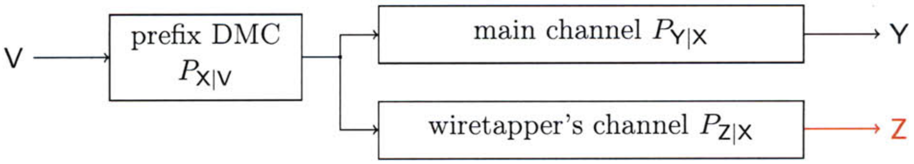

V
prefix DMC
$P_{\mathsf{X}|\mathsf{V}}$
main channel $P_{\mathsf{Y}|\mathsf{X}}$
wiretapper's channel $P_{\mathsf{Z}|\mathsf{X}}$
Y
Z

Figure 1-3: Prefix DMC

over a discrete memoryless broadcast channel with transition probability $P_{\mathsf{YZ}|\mathsf{X}}$. Bob decodes his observation Y^n into the estimate $\hat{\mathsf{K}}$. The wiretapper, Eve, attempts to learn K from her own observation Z^n. The objective is to have

$$\begin{aligned} &\Pr\{\mathsf{K} \neq \hat{\mathsf{K}}\} \to 0 && \text{(recoverability)} \\ &\frac{1}{n} I(\mathsf{K} \wedge \mathsf{Z}^n) \to 0 && \text{(secrecy)} \end{aligned}$$

as n increases to infinity. This ensures that K can be obtained by Bob but not Eve asymptotically. The notion of asymptotic secrecy can also be strengthened for free by the technique of privacy amplification [7, 40] such that the leakage of information $I(\mathsf{K} \wedge \mathsf{Z}^n)$ does not grow with the length of the message. The maximum rate of K, called the *secrecy capacity*, has the single-letter characterization

$$C_{\mathrm{w}} = \max_{\mathsf{V} \leftrightarrow \mathsf{X} \leftrightarrow \mathsf{YZ}} [I(\mathsf{V} \wedge \mathsf{Y}) - I(\mathsf{V} \wedge \mathsf{Z})] \tag{1.1}$$

where the maximization is over the choice of the virtual input distribution P_{V} and a prefix discrete memoryless channel (DMC) $P_{\mathsf{X}|\mathsf{V}}$ that Alice artificially introduces to corrupt the wiretapper's channel more than the main channel as shown in Figure 1-3.

We can think of $I(\mathsf{V} \wedge \mathsf{Y})$ and $I(\mathsf{V} \wedge \mathsf{Z})$ in (1.1) roughly as the strength of the main channel and the wiretapper's channel respectively. Thus, if the main channel is (strictly) more capable [36] than the wiretapper's channel, Alice can communicate private messages without a key at constant rate $C_{\mathrm{s}} > 0$. By choosing a uniformly random private message, she can use the channel to share a common secret with Bob, and store this secrecy in the form of a key for later use as described in the previous section. Unfortunately, this approach fails if the wiretapper's channel is less noisy [36] than the main channel, i.e. $C_{\mathrm{s}} = 0$. This can happen, for instance, in wireless communication when the wiretapper intercepts the signal at a point in between the sender and the intended receiver. The wiretapper has a less noisy channel because he is closer to the source.

Lesson:	A more capable main channel supports secure communication.
Challenge:	The approach fails when the wiretapper's channel is less noisy.

1.2.3 Public discussion

Suppose in addition to the use of a wiretap channel, Alice and Bob can publicly discuss with each other in front of Eve noiselessly. Maurer [41] showed by the following example that secure communication is possible even if the the wiretapper's channel is less noisy than the main channel.

Example 1.1 (Binary symmetric wiretap channel) Consider in Figure 1-2

$$\mathsf{Y} := \mathsf{X} \oplus \mathsf{N}_1 \qquad \text{(main channel)}$$

$$\mathsf{Z} := \mathsf{X} \oplus \mathsf{N}_2 \qquad \text{(wiretapper's channel)}$$

where the additive noise $\mathsf{N}_1 \sim \mathrm{Bern}_{0.18}$ and $\mathsf{N}_2 \sim \mathrm{Bern}_{0.1}$ are independent Bernoulli distributed random variables, equal to 1 with probabilities 0.18 and 0.1 respectively. If we cascade two wiretapper's channel in series, we have a binary symmetric channel with crossover probability $2 \times (0.1) \times (1 - 0.1) = 0.18$, which is equivalent to that of the main channel. Thus, the wiretapper's channel is less noisy than the main channel, or more precisely, the main channel is a *stochastically degraded* version of the wiretapper's channel. Without public discussion, the secrecy capacity (1.1) is 0 by the data processing theorem [8].

Suppose Alice and Bob can discuss at any time in public. Alice first chooses the input sequence to the wiretap channel as the discrete memoryless source (DMS) $\mathsf{X} \sim \mathrm{Bern}_{0.5}$. This effectively turns the channel into a discrete multiple memoryless source (DMMS) $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ [11]. Bob then encodes a secret key K into a binary n-

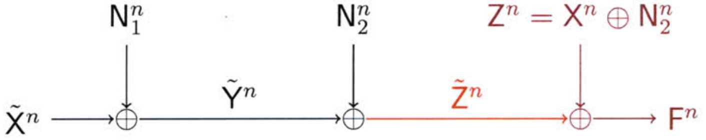


Figure 1-4: Effective wiretap channel $P_{\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}|\tilde{\mathsf{X}}}$

sequence $\tilde{\mathsf{X}}^n$ and publicly reveals

$$\mathsf{F}^n := \tilde{\mathsf{X}}^n \oplus \mathsf{Y}^n = \tilde{\mathsf{X}}^n \oplus \mathsf{X}^n \oplus \mathsf{N}_1^n$$

after observing the channel n times. Then, Alice attempts to recover K from

$$\tilde{\mathsf{Y}}^n := \mathsf{F}^n \oplus \mathsf{X}^n = \tilde{\mathsf{X}}^n \oplus \mathsf{N}_1^n$$

Since Eve also observes the public message F^n, she can generate

$$\tilde{\mathsf{Z}}^n := \mathsf{F}^n \oplus \mathsf{Z}^n = \tilde{\mathsf{X}}^n \oplus \mathsf{N}_1^n \oplus \mathsf{N}_2^n$$

This is indeed a *sufficient statistics* [8] for Eve in decoding K. i.e.

$$\begin{aligned} 0 &= I(\tilde{\mathsf{X}}^n \wedge \mathsf{F}^n\mathsf{Z}^n|\tilde{\mathsf{Z}}^n) \\ &\overset{\text{(a)}}{=} I(\tilde{\mathsf{X}}^n \wedge \mathsf{F}^n|\mathsf{Z}^n\tilde{\mathsf{Z}}^n) + I(\tilde{\mathsf{X}}^n \wedge \mathsf{Z}^n|\tilde{\mathsf{Z}}^n) \end{aligned}$$

We can see this easily from Figure 1-4. Given $\tilde{\mathsf{Z}}^n$ and Z^n, the public message F^n is a redundant observation for Eve since she can recover it by the sum $\tilde{\mathsf{Z}}^n \oplus \mathsf{Z}^n$. Thus, the first mutual information in (a) is 0. Since X^n is uniformly distributed and independent of $(\mathsf{N}_1^n, \mathsf{N}_2^n)$, we have $\mathsf{Z}^n = \mathsf{X}^n \oplus \mathsf{N}_2^n$ uniformly distributed regardless of the realization of $\tilde{\mathsf{Z}}^n$ and $\tilde{\mathsf{X}}^n$, and so the second mutual information in (a) is 0.

Effectively, there is a wiretap channel $P_{\tilde{\mathsf{Y}}\tilde{\mathsf{Z}}|\tilde{\mathsf{X}}}$ from Bob to Alice and Eve but the wiretapper's channel $P_{\tilde{\mathsf{Z}}|\tilde{\mathsf{X}}}$ now becomes a physically degraded version of the main channel $P_{\tilde{\mathsf{Y}}|\tilde{\mathsf{X}}}$. *The secrecy capacity is therefore strictly positive.* □

In the above scheme, the users turn the wiretap channel into a DMMS, with which they have some correlation not observable by the wiretapper. Then, public discussion allows the users to turn this correlation into a secret key, by reorienting the DMMS into a more favorable wiretap channel. What makes this possible? Suppose instead of having N_1 independent of N_2, we have $\mathsf{N}_1 = \mathsf{N}_2 \oplus \mathsf{N}_3$ for some N_3 independent of N_2. Then, the above scheme fails because the main channel is always a *physically degraded* version of the wiretapper's channel before or after the reorientation. Any correlation generated between Alice and Bob can be observed by Eve and so the secrecy capacity is 0 even with public discussion.

Lesson: Public discussion helps reorient the wiretap channel into a better form.

1.2.4 Secret key agreement

Secret key agreement is basically the idea of using public discussion to extract a common secret key from the correlation among the users, whether it comes from a wiretap channel or a DMMS. Maurer [42] and independently Ahlswede and Csiszár [1] formulated this problem for the source and channel models involving a wiretapper and two authorized users, one of which is a sender for the channel model. Example 1.1 discussed in the previous section is an example of the channel model. An example of the source model is as follows.

Example 1.2 Let B_0, B_1 and J be three independent uniformly random bits. Alice observes $\mathsf{X} := (\mathsf{B}_0, \mathsf{B}_1)$; Bob observes $\mathsf{Y} := (\mathsf{B}_\mathsf{J}, \mathsf{J})$; Eve does not have any private observation, i.e. $\mathsf{Z} = \emptyset$. Alice and Bob cannot choose B_0 nor B_1 as the key since Bob may not observe it. They cannot choose J as the key either because Alice may not observe it. It can be proved more formally [22] that the users cannot agree on any secret without public discussion. With public discussion, however, Bob can reveal J in public. Then, Alice and Bob can choose B_J as the key because the public message J allows Alice to recover B_J with her private observation but does not leak any information to Eve as $I(\mathsf{B}_\mathsf{J} \wedge \mathsf{J}) = 0$. □

The above scheme is indeed optimal since the secrecy capacity without any wiretapper's private information is $I(\mathsf{X}\wedge\mathsf{Y})$, which is 1 bit in this case. For the general source model where Alice, Bob and Eve observes the DMS X, Y and Z respectively, [1] showed that the following secret key rates are achievable.

$$\max_{\mathsf{U}\leftrightarrow\mathsf{V}\leftrightarrow\mathsf{X}\leftrightarrow\mathsf{YZ}}[I(\mathsf{V}\wedge\mathsf{Y}|\mathsf{U})-I(\mathsf{V}\wedge\mathsf{Z}|\mathsf{U})] \qquad \text{(forward secrecy capacity)}$$

$$\max_{\mathsf{U}\leftrightarrow\mathsf{V}\leftrightarrow\mathsf{Y}\leftrightarrow\mathsf{XZ}}[I(\mathsf{V}\wedge\mathsf{X}|\mathsf{U})-I(\mathsf{V}\wedge\mathsf{Z}|\mathsf{U})] \qquad \text{(backward secrecy capacity)}$$

The maximization is over P_{U} and $P_{\mathsf{V}|\mathsf{U}}$ in accordance with the stated Markov chain. The *forward secrecy capacity* is the maximum achievable key rate when only Alice can speak publicly, while the *backward secrecy capacity* is the maximum key rate when only Bob can speak. Since a wiretap channel can be turned into a DMMS, the above expressions also lower bound the secrecy capacity for the channel model.

Unfortunately, there is no known matching upper bound on the secrecy capacity for general source and channel model with arbitrary public discussion. [42] and [1] gave the conditional information upper bounds $I(\mathsf{X}\wedge\mathsf{Y}|\mathsf{Z})$ and $\max_{P_{\mathsf{X}}} I(\mathsf{X}\wedge\mathsf{Y}|\mathsf{Z})$ for the source and channel model respectively. They are tight in the special case when X, Y, and Z forms a Markov chain in any order. The bound for the source model was improved in [43] to the intrinsic conditional mutual information

$$I(\mathsf{X}\wedge\mathsf{Y}\downarrow\mathsf{Z}):=\min_{\mathsf{XY}\leftrightarrow\mathsf{Z}\leftrightarrow\mathsf{U}} I(\mathsf{X}\wedge\mathsf{Y}|\mathsf{U})$$

Skripsky [54], Renner and Wolf [48] showed that the bound is loose and further improved the bound to the double intrinsic information

$$I(\mathsf{X}\wedge\mathsf{Y}\downarrow\downarrow\mathsf{Z}):=\min_{\mathsf{U}}[I(\mathsf{X}\wedge\mathsf{Y}\downarrow\mathsf{ZU})+H(\mathsf{U})]$$

Gohari and Anantharam [23] strictly improves this bound to

$$\min_{\mathsf{U}}[I(\mathsf{X}\wedge\mathsf{Y}|\mathsf{U})+I(\mathsf{XY}\wedge\mathsf{U}|\mathsf{Z})]$$

and the more complicated bounds in [24] for the source and channel models.

Lesson: Public discussion consolidates correlation into a secret key.

Challenge: Secrecy capacity is unknown for the general source or channel model.

1.3 Contributions

We study the secret key agreement problem in a general multiuser setting. The focus is on how a large group of users can cooperate to generate a common secret key when the wiretapper has no private observation. This element is missing in the two-user one-wiretapper model [1, 42] described in the previous section. In the following, we give an overview of some recent results and our contributions in this direction.

Csiszár and Narayan [12] initiated the study of secret key agreement in the multiuser setting. Under the source model where every user $i \in V$ observes a correlated discrete memoryless source Z_i, a subset $A \subseteq V$ of the users, called the *active users*, want to share a secret key. The secrecy capacity for the special case without wiretapper's side information can be characterized by the linear program (LP),

$$C_{\mathrm{s}} = \min_{\lambda} \left[H(\mathsf{Z}_V) - \sum_{B} \lambda_B H(\mathsf{Z}_B | \mathsf{Z}_{B^c}) \right] \tag{1.2}$$

where Z_B denotes the random vector $(\mathsf{Z}_i : i \in B)$, and $\lambda = (\lambda_B \geq 0 : \emptyset \neq B \not\supseteq A)$ is required to satisfy the linear constraint $\sum_{B \ni i} \lambda_B = 1$ for all $i \in V$. The proposed capacity-achieving scheme in [12] reveals a meaningful duality relationship between the secret key agreement problem and the source coding problem of *communication for omniscience*, i.e. the users publicly discuss until every active user learns the entire source Z_V. It becomes evident that the secrecy capacity captures a fairly general notion of mutual dependence among the private sources, which appears to be a natural extension of Shannon's mutual information to the multivariate case.

However, such a conclusion is premature because the LP characterization of the secrecy capacity is rather unintuitive. Although λ in (1.2) carries the meaning of fractional partition in combinatorics [49], and Lagrangian multipliers in optimization theory [5], no clear information-theoretic interpretation is known, except for the following mutual dependence upper bound derived in [12],

$$C_{\mathrm{s}} \leq \min_{\mathcal{P}} \frac{1}{|\mathcal{P}|-1} D\left(P_{\mathsf{Z}_V} \,\middle\|\, \prod_{C\in\mathcal{P}} P_{\mathsf{Z}_C} \right) \tag{1.3}$$

where $\mathcal{P}$ is a set partition of V into at least two parts such that every part overlaps with A, and $D(\cdot\|\cdot)$ denotes the information divergence.[2] Same as Shannon's mutual information, it is a divergence expression from the joint distribution P_{Z_V} to the product of the marginal distributions P_{Z_C}'s. However, this bound is proved to be tight only up to the three-user case in [12]. There is no obvious reason why it should be tight in general, nor is there a known counter-example that proves otherwise.

In Chapter 2, we show that the bound is tight when all users are active, i.e. $A = V$ [6]. Not only does this establish the desired notion of mutual dependence for secret key agreement, the tightness result generalizes to a new identity for the submodular function optimization in Section B.1. In section 2.4, we extend the capacity-achieving scheme in [12] to give a strengthened duality relationship with the source coding problem when the total public discussion cannot exceed a given rate. This allows us to see in mathematical terms that a shared secret is the consensus that needs not be publicly discussed. To put it simply, if the users can agree on 2 bits of information after 1 bit of public discussion, then the remaining 1 bit is a shared secret. Conversely, if the users can agree on 1 secret bit after 1 bit of public discussion, they can agree on 2 bits of information since the secret is (nearly) independent of the public messages.

Similar to Shannon's channel coding theorem [52], the capacity-achieving scheme in [12] for secret key agreement uses a random coding argument that does not guarantee any practical code. In Chapter 3, we describe a practical linear network coding [28] approach to secret key agreement when the source has a general linear dependence structure. In particular, single-source network coding solution is optimal for the case when all users are active. This optimality is captured by a new and surpris-

[2] See [8] or Section A.1.

ingly general identity for matroids [50] in Section B.2, extending the work of Frank et al. [20], and Bang-Jensen and Thomassé [3]. In addition to being a practical solution, the network coding approach gives a theoretically interesting interpretation of secrecy capacity as information flow. It also has the combinatorial interpretation as partition connectivity in Section 3.2 in the special case when the dependence structure of the source can be captured by a dependency hypergraph. This extends the work of Ye and Reznik [62] on pairwise independent network, and Li et al. [37] on undirected networks.

In the subsequent paper [13], Csiszár and Narayan studied the secret key agreement problem under a multiterminal broadcast-type channel model. The secrecy capacity is characterized as a minimax problem

$$C_{\mathrm{s}} = \max_{P_{\mathsf{Z}_1}} \min_{\lambda} \left[H(\mathsf{Z}_V) - \sum_{B} \lambda_B H(\mathsf{Z}_B | \mathsf{Z}_{B^c}) \right] \tag{1.4}$$

where $P_{\mathsf{Z}_1} \in \mathscr{P}(Z_1)$ is an input distribution on support set Z_1 for the broadcast channel $P_{\mathsf{Z}_{V\setminus\{1\}}|\mathsf{Z}_1}$, and λ is a fractional partition as in (1.2). The expression looks almost identical to (1.2) except for the maximization over P_{Z_1}. Indeed, the capacity is attained by the *pure source emulation* approach that uses the channel as a source after sending independent input symbols.[3] Compared to the source model, however, this is less passive in the sense that the correlation among the users is not just given as it is, but created by sending an input sequence optimized over the choice of the distribution.

For a large network, however, it is more realistic to expect multiple channels relating subsets of the users, rather than a single broadcast channel covering all users. Allowing only one sender in the model is also quite restrictive, for it does not allow one to study how users can cooperate to enhance their correlation. To capture this interesting component in the study of networks, we extend the broadcast-type channel model to a general multiterminal network in Chapter 4. Each user can simultaneously be a sender and a receiver. The model is formulated in Chapter 5.

[3] See Example 1.1.

Minimax characterizations of the secrecy capacity, its upper and lower bounds are derived in Chapter 6. The lower bound is achieved by a new cooperative scheme called the *mixed source emulation*. It employs the well-known idea of mixed strategy [56] by considering the minimax characterization as the payoff of a virtual zero-sum game. In section 6.3, we give a simple example, called the *coupling channel*, for which mixed source emulation strictly outperforms the conventional pure source emulation approach.

Chapter 7 analyzes the tightness of the secrecy bounds. Although secrecy capacity is unknown for the multiterminal network, we find that the secrecy bounds match under some general conditions in Section 7.2. This includes the broadcast-type channels in [13] and some classes of channels that are not broadcast-type such as the interference-free channels in Section 7.2.1. In particular, we consider a class of finite homomorphic channels in Section 7.2.2. Due to the group structure of the channels, it is optimal to simply have each sender transmit uniformly independent input symbols. Cooperation turns out to be unnecessary in this case. In Section 7.3, however, we show that even mixed source emulation is strictly suboptimal for an example referred to as the *consensus* channel. It is also clear from the example that there are adaptive approach to cooperate better than mixed source emulation, and so the proposed secrecy lower bound can be loose.

We have divide the subject into two parts. In Part I, we focus on the source model, and establish the notion of mutual dependence for secret key agreement in Chapter 2, and interpret it as information flow and partition connectivity using the linear network coding scheme in Chapter 3. The combinatorial framework for the proofs is presented in Appendix B. In Part II, we introduce the multiterminal network model in Chapter 4 with an concise overview of the main results detailed in the subsequent chapters. We strongly recommend skipping the technical details of Part II for the first reading. Part II can also be read before Part I for a more detailed description of the secret key agreement protocol and the derivation of the secrecy capacity.

Part I

Mutual Dependence of Random Sources

Chapter 2

Multivariate Correlation

In this part, we focus on the multiterminal source model for secret key agreement by Csiszár and Narayan [12]. An informal introduction was given in Chapter 1 and a more general multiterminal channel model will be formulated in Part II.

The source model involves a set V of users/terminals and a wiretapper. As illustrated in Figure 2-1, each user $i \in V$, denoted as T_i,

1. randomizes independently by generating a continuous random variable U_i with the probability density function chosen as $P_{\mathsf{U}_V} = \prod_{i \in V} P_{\mathsf{U}_i}$,
2. observes n samples of a finitely-valued DMS Z_i that is distributed according to some given joint probability mass function P_{Z_V},
3. and then discuss publicly and interactively by broadcasting messages that are functions of his accumulated knowledge, which includes his randomization, private source, and the previous public messages. e.g. if user i broadcasts the j-th public message F_j, he can set it as a deterministic function of U_i, Z_i^n and the previous messages $\mathsf{F}_{[j-1]} := \{\mathsf{F}_1, \dots, \mathsf{F}_{j-1}\}$.

A subset $A \subseteq V$ of the users, called the *active users*, want to share a secret key. Each active user $i \in A$ generates an individual key K_i as a function of his accumulated knowledge so that the keys are almost surely the same, i.e.

$$\Pr\{\exists i \in A,\ \mathsf{K}_i \neq \mathsf{K}\} \to 0 \qquad \text{(recoverability)}$$

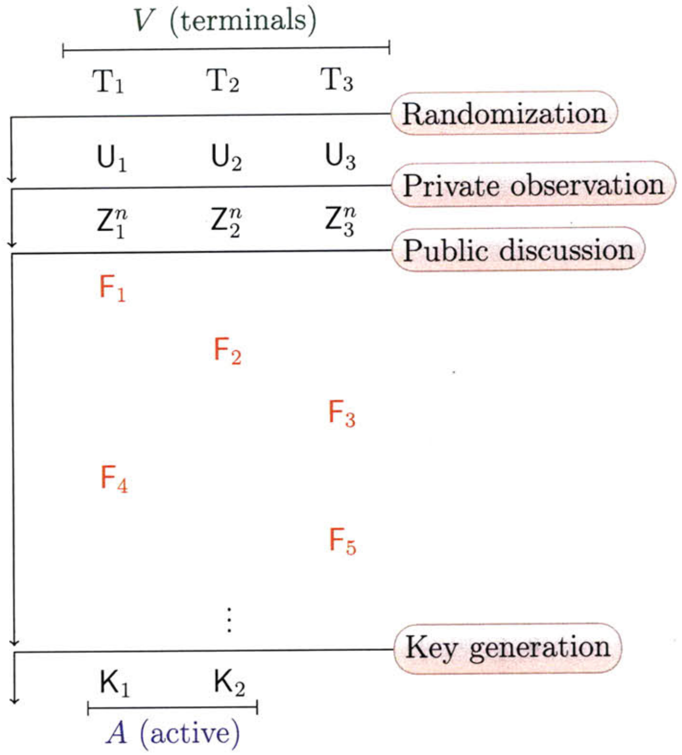


Figure 2-1: Simple source model for secret key agreement

as $n \to \infty$ for some finitely-valued random variable K taking values from a finite set K, the cardinality of which is growing exponentially in n. The common secret key must remain nearly uniformly distributed to the wiretapper who observes the entire public discussion F, i.e.

$$\frac{1}{n}\left[\log|K| - H(\mathsf{K}|\mathsf{F})\right] \to 0 \qquad \text{(secrecy)}$$

The users who are not active, called the *helpers*, are allowed but not required to learn the secret key. To understand this secrecy condition, recall that

$$\log|K| \overset{\text{(a)}}{\geq} H(\mathsf{K}) \overset{\text{(b)}}{\geq} H(\mathsf{K}|\mathsf{F})$$

where (a) is because uniform distribution maximizes entropy [8], and (b) is because conditioning reduces entropy. The secrecy condition roughly requires both inequalities to be satisfied with equality in an asymptotical sense so that the key appears to be uniformly distributed even conditioned on the public messages.

Secrecy capacity is the achievable key rate

$$C_{\mathrm{s}} := \sup_{\mathsf{U}_V,\mathsf{F},\mathsf{K}_A} \liminf_{n\to\infty} \frac{1}{n}\log|K|$$

maximized over the choice of the randomization, public discussion, and key generation functions. It is characterized by [12] as the following linear program

$$C_{\mathrm{s}} = \min_{\lambda}\left[H(\mathsf{Z}_V) - \sum_{B}\lambda_B H(\mathsf{Z}_B|\mathsf{Z}_{B^c})\right] \tag{2.1}$$

where λ is a vector of non-negative weights λ_B with $B \subseteq V : \emptyset \neq B \not\supseteq A$ and

$$\sum_{B\ni i}\lambda_B = 1 \qquad \text{for all } i \in V$$

Note that this expression is even simpler to compute than the channel capacity, which involves maximization of a non-linear function over the input distribution. However, λ does not have a clear information-theoretic meaning. Although secrecy capacity intuitively reflects how correlated the privates sources Z_i's are as argued informally in Chapter 1, we cannot see this directly from the LP characterization except for small networks. In the two-user case where $V = A = [2]$, the expression is simply Shannon's mutual information between the private sources Z_1 and Z_2, which can be regarded roughly as the distance from the joint distribution $P_{\mathsf{Z}_1\mathsf{Z}_2}$ to the product distribution $P_{\mathsf{Z}_1}P_{\mathsf{Z}_2}$. In the three-user case, it also simplifies to a minimum divergence expression from the joint distribution to the product of marginal distributions.[1] If this can be done in general when $A = V$, we have a heuristically meaningful expression for the secrecy capacity as the minimum distance from the joint distribution to the product of marginal distributions, which is a natural generalization of Shannon's mutual information to the multivariate case. However, it appears to be quite difficult even in the four-user case, and was posed as an open question in [12].

In Section 2.1, we establish this minimum divergence expression for secret key

[1] See [12] or the expression in Example 2.1 of the following section.

agreement when all users are active [6]. Section 2.2 considers the general case when some users are not active. Some alternative definitions of correlations are given in Section 2.3 as a comparison. In Section 2.4, we strengthen the duality with source coding in [12] which gives a meaningful interpretation of mutual dependence.

2.1 Mutual Dependence

In information theory, the dependence between any two random variables is captured by Shannon's mutual information

$$\begin{aligned} I(\mathsf{Z}_1 \wedge \mathsf{Z}_2) &:= H(\mathsf{Z}_1) + H(\mathsf{Z}_2) - H(\mathsf{Z}_1\mathsf{Z}_2) \\ &= \mathrm{E}\left[\frac{P_{\mathsf{Z}_1\mathsf{Z}_2}(\mathsf{Z}_1,\mathsf{Z}_2)}{P_{\mathsf{Z}_1}(\mathsf{Z}_1)P_{\mathsf{Z}_2}(\mathsf{Z}_2)}\right] \\ &=: D(P_{\mathsf{Z}_1\mathsf{Z}_2}\|P_{\mathsf{Z}_1}P_{\mathsf{Z}_2}) \end{aligned} \tag{2.2}$$

where $P_{\mathsf{Z}_1\mathsf{Z}_2}$ denotes the distribution of Z_1 and Z_2, $D(\cdot\|\cdot)$ is the information divergence, and $H(\cdot)$ is the entropy measure.[2] It has various operational meanings spanning over the source and channel coding theories. A heuristically appealing extension [12] to the multivariate case with more than two random variables is the following mutual dependence expression.

Definition 2.1 (Mutual Dependence) For any finitely-valued random vector $\mathsf{Z}_V := (\mathsf{Z}_i : i \in V)$ with $|V| \geq 2$, the *mutual dependence* of Z_V is defined as

$$\begin{aligned} I(\mathsf{Z}_V) &:= \min_{\mathcal{P}\in\Pi} \frac{1}{|\mathcal{P}|-1}\left[\sum_{C\in\mathcal{P}} H(\mathsf{Z}_C) - H(\mathsf{Z}_V)\right] \\ &= \min_{\mathcal{P}\in\Pi} \frac{1}{|\mathcal{P}|-1} \mathrm{E}\left[\frac{P_{\mathsf{Z}_V}(\mathsf{Z}_V)}{\prod_{C\in\mathcal{P}} P_{\mathsf{Z}_C}(\mathsf{Z}_C)}\right] \\ &= \min_{\mathcal{P}\in\Pi} \frac{1}{|\mathcal{P}|-1} D\left(P_{\mathsf{Z}_V} \middle\| \prod_{C\in\mathcal{P}} P_{\mathsf{Z}_C}\right) \end{aligned} \tag{2.3}$$

where Π is the collection of set partitions $\mathcal{P}$ of V into at least 2 non-empty sets. □

[2] See Section A.1 for an introduction.

(2.3) is a natural generalization of (2.2) since both are expressed in terms of the divergence from the joint distribution to the product distribution of certain marginals. When there are more than two random variables, there is more than one way to partition them into groups. We can view (2.3) informally as the minimum distance from the joint distribution to the product distributions according to the different partitions. To explain the normalization factor $(|\mathcal{P}|-1)$, consider the simple case when $\mathsf{Z}_i = \mathsf{Z}$ for all $i \in V$. The divergence $D(P_{\mathsf{Z}_V} \| \prod_{C\in\mathcal{P}} P_{\mathsf{Z}_C})$ equals $(|\mathcal{P}|-1)H(\mathsf{Z})$, which is the amount $H(\mathsf{Z})$ of common randomness in Z_C's overcounted by a factor of $(|\mathcal{P}|-1)$ times. Thus, (2.3) roughly measures the minimum randomness shared among the parts in a partition of Z_V. It has a clear operational meaning when $|V| \leq 3$.

Example 2.1 Mutual dependence (2.3) reduces to the usual mutual information when $|V| = 2$. i.e. $I(\mathsf{Z}_{\{1,2\}}) = I(\mathsf{Z}_1 \wedge \mathsf{Z}_2)$. With $V := [3] := \{1,2,3\}$, we have

$$I(\mathsf{Z}_{[3]}) = \min\left\{ I(\mathsf{Z}_1 \wedge \mathsf{Z}_2\mathsf{Z}_3), I(\mathsf{Z}_2 \wedge \mathsf{Z}_1\mathsf{Z}_3), I(\mathsf{Z}_3 \wedge \mathsf{Z}_1\mathsf{Z}_2), \frac{1}{2}\left[\sum_{i\in[3]} H(\mathsf{Z}_i) - H(\mathsf{Z}_{[3]})\right]\right\}$$

It was proved in [12] that this equals the secrecy capacity of the secret key agreement problem involving three users, i.e. $V = A = [3]$. □

Indeed, this operational meaning of secrecy capacity can be established more generally for any V and DMMS P_{Z_V}.

Theorem 2.1 *Given a finite ground set $V : |V| \geq 2$, the mutual dependence in* (2.3) *satisfies*

$$I(\mathsf{Z}_V) = H(\mathsf{Z}_V) - \max_{\lambda\in\Lambda} \sum_{B\in\mathscr{F}} \lambda_B H(\mathsf{Z}_B|\mathsf{Z}_{B^c}) \tag{2.4}$$

where $\mathscr{F} := 2^V \setminus \{V\}$, Λ is defined as the collection of fractional partitions $\lambda := (\lambda_B : B \in \mathscr{F})$ of V, i.e. $\lambda_B \geq 0$ for all $B \in \mathscr{F}$ and $\sum_{B\in\mathscr{F}:i\in B} \lambda_B = 1$ for all $i \in V$. □

This establishes the desired operational meaning for the mutual dependence since the expression on the R.H.S. of (2.4) is the secrecy capacity when $A = V$. The proof relies on a simple property of information: mutual information is non-negative. More precisely, for any subsets $B_1, B_2 \subseteq V$, we have

$$I(\mathsf{Z}_{B_1} \wedge \mathsf{Z}_{B_2} | \mathsf{Z}_{B_1 \cap B_2}) \geq 0$$

or equivalently that entropy is *submodular* [21], i.e.

$$H(\mathsf{Z}_{B_1}) + H(\mathsf{Z}_{B_2}) \geq H(\mathsf{Z}_{B_1 \cap B_2}) + H(\mathsf{Z}_{B_1 \cup B_2})$$

If we replace entropy by an arbitrary submodular function, (2.4) can be extended to the more general identity (B.7) in Section B.1, which equates a combinatorial optimization problem over set partitions to an LP problem, which is easy to compute.

PROOF (THEOREM 2.1) Define $h : \mathscr{F} \mapsto \mathbb{R}$ as

$$h(B) := H(\mathsf{Z}_B | \mathsf{Z}_{B^c}) \qquad \text{for all } B \in \mathscr{F} \tag{2.5}$$

with the convention $h(\emptyset) = 0$. The submodularity of entropy [21] implies the *supermodularity* of h as follows.

Subclaim 2.1A *h is* supermodular. *i.e. for all $B_1, B_2 \in \mathscr{F} : B_1 \cap B_2, B_1 \cup B_2 \in \mathscr{F}$,*

$$h(B_1) + h(B_2) \leq h(B_1 \cap B_2) + h(B_1 \cup B_2) \tag{2.6}$$

Equivalently, $-h$ is submodular. ◁

PROOF Consider proving the non-trivial case where B_1 and B_2 are non-empty. By the positivity of mutual information $I(\mathsf{Z}_{B_1^c} \wedge \mathsf{Z}_{B_2^c} | \mathsf{Z}_{B_1^c \cap B_2^c}) \geq 0$, we have

$$H(\mathsf{Z}_{B_1^c} | \mathsf{Z}_{B_1^c \cap B_2^c}) \geq H(\mathsf{Z}_{B_1^c} | \mathsf{Z}_{B_2^c})$$
$$H(\mathsf{Z}_{B_1^c}) + H(\mathsf{Z}_{B_2^c}) \geq H(\mathsf{Z}_{B_1^c \cup B_2^c}) + H(\mathsf{Z}_{B_1^c \cap B_2^c})$$

(2.6) follows since $h(B) = H(\mathsf{Z}_V) - H(\mathsf{Z}_{B^c})$. ◀

By the strong duality theorem [15], the *primal* LP (2.4) equals its LP *dual*,

$$\begin{aligned} &\text{minimize} && \sum_{i\in V} r_i && && \textbf{(2.7a)} \\ &\text{subject to} && \sum_{i\in B} r_i \geq h(B) && \text{for all } B \in \mathscr{F} && \textbf{(2.7b)} \end{aligned}$$

The supermodularity of h translates to the following property on the tight constraints of the dual problem.

Subclaim 2.1B *For any feasible solution r to the dual linear program* (2.7)*, and $B_1, B_2 \in \mathscr{F} : B_1 \cap B_2, B_1 \cup B_2 \in \mathscr{F}$, if B_1 and B_2 are tight constraints, i.e.*

$$\sum_{i\in B_j} r = h(B_j) \qquad \textit{for } j = 1,2 \tag{2.8a}$$

then $B_1 \cup B_2$ is also a tight constraint, i.e.

$$\sum_{i\in B_1\cup B_2} r_i = h(B_1 \cup B_2) \tag{2.8b}$$

n.b. $B_1 \cap B_2$ is also tight but we do not need it for the proof of Theorem 2.1. ◁

Proof Since $B_1 \cup B_2 \in \mathscr{F}$, we immediately have $\sum_{i\in B_1\cup B_2} r_i \geq h(B_1 \cup B_2)$ by (2.7b). The reverse inequality can be proved as follows,

$$\begin{aligned} \sum_{i\in B_1\cup B_2} r_i &= \sum_{i\in B_1} r_i + \sum_{i\in B_2} r_i - \sum_{i\in B_1\cap B_2} r_i \\ &\overset{\text{(a)}}{\leq} h(B_1) + h(B_2) - h(B_1 \cap B_2) \\ &\overset{\text{(b)}}{\leq} h(B_1 \cup B_2) \end{aligned}$$

(a) is by (2.8a) on B_1 and B_2, and (2.7b) on $B_1 \cap B_2 \in \mathscr{F}$; (b) is by Subclaim 2.1A. With a similar argument, we also have $\sum_{i\in B_1\cap B_2} r_i = h(B_1 \cap B_2)$. ◀

Let λ^* be an optimal solution to the primal LP (2.4).[3] Define its support set as

$$\mathcal{B} := \{B \in \mathscr{F} : \lambda^*_B > 0\} \tag{2.9}$$

and the corresponding partition of V as

$$\mathcal{P}^* := \left\{ \left(\bigcup \{B \in \mathcal{B} : B \not\ni i\} \right)^c : i \in V \right\} \tag{2.10}$$

Subclaim 2.1C *$\mathcal{P}^*$ belongs to Π. (See Definition 2.1.)* ◁

PROOF Define the relation R on V as

$$i \sim_R j \iff i \in C_j \qquad \text{for } i, j \in V$$

where $C_i := (\bigcup\{B \in \mathcal{B} : B \not\ni i\})^c$. By definition (2.10), $\mathcal{P}^* = \{C_i : i \in V\}$. To show that $\mathcal{P}^*$ is a partition of V, it suffices to show that $\sim_R$ is an equivalence relation on V as follows,

$$\begin{aligned} i \sim_R j &\iff i \notin \bigcup\{B \in \mathcal{B} : B \not\ni j\} \\ &\iff \{B \in \mathcal{B} : B \not\ni i\} \supseteq \{B \in \mathcal{B} : B \not\ni j\} \\ &\iff \{B \in \mathcal{B} : i \in B\} \subseteq \{B \in \mathcal{B} : j \in B\} \end{aligned}$$

i.e. any set in $\mathcal{B}$ that contains i also contains j. Using this simplification, it is easy to see that $\sim_R$ satisfies the defining properties of an equivalence relation:

- *Reflexivity:* R is reflexive since $i \in C_i$ trivially for $i \in V$.
- *Transitivity:* Suppose $i \sim_R j$ and $j \sim_R k$ for some $i, j, k \in V$. Then,

$$\{B \in \mathcal{B} : i \in B\} \subseteq \{B \in \mathcal{B} : j \in B\} \subseteq \{B \in \mathcal{B} : k \in B\}$$

which implies $i \sim_R k$ as desired.

[3] λ^* exists or equivalently Λ is non-empty. For example, $\lambda_{\{i\}} = 1$ for $i \in V$ is a fractional partition in Λ. For the more general case considered in Theorem B.1, Λ may be empty.

- *Symmetry:* Suppose to the contrary that $i \sim_R j$ but $j \not\sim_R i$. Then,

  $$\{B \in \mathcal{B} : i \in B\} \subsetneq \{B \in \mathcal{B} : j \in B\}$$

  This implies, by definition (2.9) of $\mathcal{B}$ that

  $$\sum_{B \ni i} \lambda_B^* < \sum_{B \ni j} \lambda_B^*$$

  which is the desired contradiction since both sides equal 1 by the definition of Λ in Theorem (2.1).

Finally, to argue that $|\mathcal{P}^*| \geq 2$, note that $\mathcal{B} \neq \emptyset$ since $\sum_{B \in \mathscr{F}} \lambda_B^* > 0$. Since any $B \in \mathscr{F}$ satisfies $B \neq V$, we have $C_i \neq V$ for all $i \in V$ as desired. ◂

Not only is $\mathcal{P}^*$ a feasible solution in Π for (2.3), but it is also optimal, as a consequence of the supermodularity of h and the duality theorem.

Subclaim 2.1D *For any optimal r^* to the dual problem* (2.7),

$$\sum_{i \in C^c} r_i^* = h(C^c) \qquad \textit{for all } C \in \mathcal{P}^* \tag{2.11}$$

i.e. every part of $\mathcal{P}^*$ corresponds to a tight constraint. ◃

Proof By the complementary slackness theorem [15, Theorem 5.4], $\sum_{i \in B} r_i^* = h(B)$ for all $B \in \mathcal{B}$. By Subclaim 2.1B, we have

$$\sum_{i \in \bigcup\{B \in \mathcal{B} : B \not\ni i\}} r_i^* = h\left(\bigcup\{B \in \mathcal{B} : B \not\ni i\}\right) \qquad \text{for all } i \in V$$

which gives the desired equality (2.11) under (2.10). ◂

The primal/dual optimality criteria [15, Theorem 5.5] implies that the fractional partition

$$\left(\frac{\mathbb{1}\{B^c \in \mathcal{P}^*\}}{|\mathcal{P}^*| - 1} : B \in \mathscr{F}\right) \in \Lambda$$

is optimal to (2.4). More precisely, for all feasible r to the dual (2.7) and $\mathcal{P} \in \Pi$,

$$\begin{aligned} H(\mathsf{Z}_V) - \max_{\lambda\in\Lambda} \sum_{B\in\mathscr{F}} \lambda_B H(\mathsf{Z}_B|\mathsf{Z}_{B^c}) &\overset{\text{(a)}}{\leq} H(\mathsf{Z}_V) - \sum_{i\in V} r_i && \text{by duality} \\ &= H(\mathsf{Z}_V) - \frac{1}{|\mathcal{P}|-1} \sum_{C\in\mathcal{P}} \sum_{i\in C^c} r_i \\ &\overset{\text{(b)}}{\leq} H(\mathsf{Z}_V) - \frac{1}{|\mathcal{P}|-1} \sum_{C\in\mathcal{P}} H(\mathsf{Z}_{C^c}|\mathsf{Z}_C) && \text{by (2.7b)} \\ &= \frac{1}{|\mathcal{P}|-1} \left[\sum_{C\in\mathcal{P}} H(\mathsf{Z}_C) - H(\mathsf{Z}_V) \right] \end{aligned}$$

When we set r to an optimal solution r^*, (a) is satisfied with equality by the strong duality theorem. When we set $\mathcal{P}$ to $\mathcal{P}^*$, which is valid by Subclaim 2.1C, (b) is also satisfied with equality by Subclaim 2.1D. This gives the desired equality (2.4) and completes the *proof of Theorem 2.1.* ■

2.2 Slackness of mutual dependence bound

For the more general case when $A \subseteq V$, [12] gives the following mutual dependence upper bound on the secrecy capacity C_{s} with unlimited public discussion

$$C_{\mathrm{s}} = \min_{\lambda\in\Lambda(\mathscr{F}(A),V)} \left[H(\mathsf{Z}_V) - \sum_{B\in\mathscr{F}(A)} \lambda_B H(\mathsf{Z}_B|\mathsf{Z}_{B^c}) \right] \tag{2.12}$$

$$\leq \min_{\mathcal{P}\in\Pi(\mathscr{F}(A),V)} \frac{1}{|\mathcal{P}|-1} D\left(P_{\mathsf{Z}_V} \middle\| \prod_{C\in\mathcal{P}} P_{\mathsf{Z}_C} \right) \tag{2.13}$$

where $\mathscr{F}(A)$, $\Lambda(\mathscr{F}(A), V)$ and $\Pi(\mathscr{F}(A), V)$ are defined in (B.1a), (B.5), and (B.3) respectively in Appendix B.1. $\Pi(\mathscr{F}(A), V)$, in particular, refers to the set of partitions of V into at least two parts such that every part intersects A.

In the previous section, we show that this bound is tight when $A = V$. The bound was also shown to be tight for $|V| \leq 3$ in [12]. However, tightness of this bound was not known even in the case when $|V| = 4$. There was no apparent reasons why it should be tight while no counter-example was found that could suggest otherwise. In

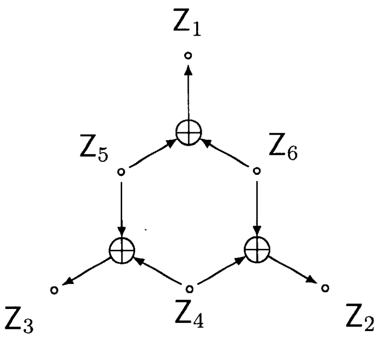


Figure 2-2: Finite linear source for Example 2.2

the following, We will show that it is loose by a minimal example with $|V| = 6$ and $|A| = 3$. Thus, the bound is tight only for small networks. It does not carry the operational meaning of secrecy capacity for bigger networks with helpers.

Example 2.2 Let $V = [6]$, $A = [3]$, and Z_V be a finite linear source with

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} Z_4 \\ Z_5 \\ Z_6 \end{bmatrix}$$

where Z_4, Z_5 and Z_6 are independent uniformly random bits. The matrix multiplication is over the binary field $\mathbb{F}_2$. This is illustrated in Figure 2-2. □

Proposition 2.1 *The secrecy capacity* (2.12) *and mutual dependence upper bound* (2.13) *are* $\frac{3}{4}$ *and* 1 *bit respectively for Example 2.2. The bound is therefore slack.* □

PROOF It is easy to see that $H(Z_C) = |C|$ for any $C \subseteq V$ such that $|C| \leq 2$. When $|C| = 3$, $H(Z_C)$ can be 2 or 3 depending on the linear dependence of the elements in Z_C. It can be verified that $H(Z_C) = 2$ if $C \in \mathcal{S} := \{\{1,5,6\},\{2,4,6\},\{3,4,5\},\{4,5,6\}\}$. For instance, $Z_1 = Z_5 \oplus Z_6$ and so $H(Z_{\{1,5,6\}}) = 2$. When $|C| > 3$, we have $H(Z_C) = 3$ since there can be at most three mutually independent bits. In summary, we have

$$H(Z_C) = \min\{|C|, 3\} - \mathbb{1}_{\mathcal{S}}(C) \qquad \textbf{(2.14)}$$

where $\mathbb{1}_{\mathcal{S}}(C)$ is the indicator function equal to 1 if $C \in \mathcal{S}$ and 0 otherwise.

Solving the LP in (2.12) for the secrecy capacity, we obtain an optimal λ with

$$\lambda_B = \begin{cases} \frac{1}{4} & , B = \{2,3,4\}, \{1,3,5\}, \{1,2,6\}, \{1\}^c, \{2\}^c, \{3\}^c \\ 0 & \text{otherwise} \end{cases}$$

The LP dual [15] of (2.12) is

$$C_\text{s} = H(\mathsf{Z}_V) - \min_{\boldsymbol{r}} \sum_{i \in V} r_i \tag{2.15}$$

where $\boldsymbol{r} := (r_i : i \in V)$ is subject to the constraints that

$$\sum_{i \in B} r_i \geq H(\mathsf{Z}_B|\mathsf{Z}_{B^c}) \qquad \text{for all } B \in \mathscr{F}(A)$$

The optimal solution to the dual problem is $r_i = \frac{1}{4}$ for $i \in A$ and $r_i = \frac{1}{2}$ for $i \in A^c$. It can be shown that both the primal (2.12) and the dual (2.15) LP's evaluate to the same desired value $\frac{3}{4}$, which verifies the optimality of the solutions by duality.

To show that (2.13) is 1, first note that $|\mathcal{P}| \geq 3$ because $|A| = 3$ and every $C \in \mathcal{P}$ must contain a different element in A by the definition (B.3) of $\Pi(\mathscr{F}(A), V)$. It suffices to show that the normalized divergence in (2.13), i.e.

$$\circledast := \frac{1}{|\mathcal{P}| - 1} \left[\sum_{C \in \mathcal{P}} H(\mathsf{Z}_C) - H(\mathsf{Z}_V) \right]$$

is at least 1 for each of the following cases of $\mathcal{P}$:

1. $|\mathcal{P}| = 2$, i.e. $\mathcal{P} = (C_1, C_2)$.
   (a) $|C| > 3$ for some $C \in \mathcal{P}$. Assume $|C_1| > 3$ for definiteness. $H(\mathsf{Z}_{C_1}) = H(\mathsf{Z}_V) = 3$ and $H(\mathsf{Z}_{C_2}) \geq 1$ by (2.14). Then, $\circledast = H(\mathsf{Z}_{C_1}) + H(\mathsf{Z}_{C_2}) - H(\mathsf{Z}_V) \geq 3 + 1 - 3 = 1$ as desired.
   (b) $|C_1| = |C_2| = 3$. Then, $H(\mathsf{Z}_{C_i}) \geq 2$ by (2.14) for $i \in [2]$. $\circledast \geq 2 + 2 - 3 = 1$ as desired.
2. $|\mathcal{P}| = 3$, i.e. $\mathcal{P} = (C_1, C_2, C_3)$.

(a) $|C| > 3$ for some $C \in \mathcal{P}$. By (2.14), $H(\mathsf{Z}_C)$ is at least 3 for some $C \in \mathcal{P}$ and 1 for others. $\circledast \geq \frac{3+1+1-3}{2} = 1$.

(b) $|C| = 3$ for some $C \in \mathcal{P}$. Assume $|C_1| = 3$, $|C_2| = 2$ and $|C_3| = 1$ for definiteness. By (2.14), $H(\mathsf{Z}_{C_i})$ is at least 2 for $i \in [2]$ and 1 for $i = 3$. Thus, $\circledast \geq \frac{2+2+1-3}{2} = 1$.

(c) $|C| = 2$ for all $C \in \mathcal{P}$. Then, $H(\mathsf{Z}_C) = 2$ by (2.14) for all $C \in \mathcal{P}$. $\circledast \geq \frac{2+2+2-3}{2} = 1.5$.

Finally, the minimum value of 1 for (2.13) is attained by the optimal partition $\mathcal{P} = (\{1\}, \{2\}, \{1,2\}^c)$ since $H(\mathsf{Z}_1) = H(\mathsf{Z}_2) = 1$, $H(\mathsf{Z}_{[2]^c}) = 3$ and so $\circledast = \frac{1+1+3-3}{2} = 1$. ■

With a simple trick, we can also generate similar examples for different values of $|V|$ and $|A|$ as follows.

Theorem 2.2 *There are examples for which the mutual dependence upper bound in* (2.13) *is loose when* $|V| \geq 6$ *and* $|A| \in \{3, \dots, |V| - 3\}$. □

PROOF By Proposition 2.1, Example 2.2 is the desired example with $|V| = 6$ and $|A| = 3$. Other examples can be constructed from this by duplicating active users and/or helpers. For examples, let $1'$ be a new active user who observes the same component source Z_1 as user 1. This is the desired example for $|V| = 7$ and $|A| = 4$ because both the secrecy capacity and mutual dependence upper bound remain unchanged under the same DMMS. Alternatively, let $6'$ be a new helper who observes the component source Z_6. This gives the desired example for $|V| = 7$ and $|A| = 3$ by the same argument. ■

Example 2.2 is indeed minimal in terms of the number of users and active users. In other words, (2.13) is tight for $2 \leq |A| \leq| V| \leq 5$ or $|V| = 6 > 2|A|$, regardless of the choice of Z_V. In Appendix C.3, we give in Theorem C.1 a generalization of the mutual dependence upper bound (2.13) and a tightness test which involves only a finite number of test cases. A computer program is written to automate the test and show that the bound is tight for small networks. This result again relies only on the simple fact that mutual information is non-negative.

2.3 Other measures of correlation

Needless to say, there have been many attempts in generalizing Shannon's mutual information or understanding it in different settings. We will give a survey of various measures of correlations, as comparisons to the notion of mutual dependence we consider here.

The structure of correlation is rather complex in the multivariate case. There are, for instance, different notions of independence: mutual, pairwise, conditional and semi-independence [27]. For the set $\mathsf{Z}_V := (\mathsf{Z}_1, \dots, \mathsf{Z}_m)$ of m random variables, Watanabe's total correlation [57] is defined as

$$\sum_{i=1}^{m} H(\mathsf{Z}_i) - H(\mathsf{Z}_V) = D(P_{\mathsf{Z}_V} \| \prod_{i \in V} P_{\mathsf{Z}_i})$$

It has a simple interpretation as the redundancy in expected length when entropy encoding Z_V with respect to the product of the marginal distributions instead of the joint distribution [8]. McGill's interaction information [44] for three random variables is defined as

$$I(\mathsf{Z}_i \wedge \mathsf{Z}_j | \mathsf{Z}_k) - I(\mathsf{Z}_i \wedge \mathsf{Z}_j)$$

which is symmetric for every permutation (i, j, k) of $(1, 2, 3)$. It is interpreted as the gain (or loss if negative) in correlation with an additional random variable. Jakulin and Bratko [30] gave a general formula for the m-way interaction information

$$-\sum_{A \subseteq V} (-1)^{|V|-|A|} H(\mathsf{Z}_A)$$

which is equal in magnitude to the following co-information [4]

$$-\sum_{A \subseteq V} (-1)^{|A|} H(\mathsf{Z}_A)$$

first derived by Yeung [63] by treating shannon's information measure as a measure over a σ-field. Treating each random variable as a set, it has the mathematical interpretation as a quantity commonly possessed by every set. However, it can be negative unintuitively. To fix this, Han [25] investigated the mathematical structure connecting the various notions of multivariate correlation. He defined the *linear correlative entropy space* as the set of linear functions of entropies equal to zero when the random variables are mutually independent. Total correlation, interaction information and co-information are all special cases. He then derived in [26] the conditions for the correlative functions in the space to be non-negative and symmetric, just as the total correlation.

The notion of mutual dependence we consider here in (2.3) can also be expressed as functions of the entropies, and indeed, functions of the partial correlations defined by Watanabe [57] with an additional normalization factor. It is nonnegative, symmetric and correlative but not linear in the entropies, and therefore does not fall into the linear correlative entropy space. Nonetheless, it can be computed easily as an LP by Theorem 2.1 and carries the operational meaning from secret key agreement. In comparison, total/partial correlations are more specific to a particular product of the marginal distributions. Interaction information focuses more on the change rather than the total amount of correlation. Co-information assumes a priori that information behaves like objects we can normally measure in volume. This rather idealistic assumption might have lead to the peculiarity that the measure can be negative.

For the simplest case of two random variables, there are also some alternative correlative measures other than Shannon's mutual information. Wyner's common information [60]

$$C(\mathsf{X} \wedge \mathsf{Y}) := \inf_{\mathsf{X} \leftrightarrow \mathsf{U} \leftrightarrow \mathsf{Y}} I(\mathsf{U} \wedge \mathsf{XY})$$

is the minimum rate of common randomness needed for two otherwise independent simulators to generate outputs that approximate the statistics of a DMMS (X, Y) arbitrarily closely. Common information is more difficult to compute [59] than mutual information. It is no smaller than $I(\mathsf{X} \wedge \mathsf{Y})$ because the common randomness may contain a component independent of one of the source it simulates. The remaining component has rate equal to the mutual information, as pointed out in [14], via the alternative setting of channel simulation with common randomness.

Gác and Körner's common information [22, 58] is $J(\mathsf{X} \wedge \mathsf{Y}) := H(\mathsf{V})$ where V is the *maximum common function* of X and Y. It is easy to compute using the ergodic decomposition [22] and is no larger than mutual information because it is indeed the maximum amount of secret key we can generate without public discussion. With public discussion, the secret key rate increases to the mutual information.

2.4 Duality with source coding

If the mutual dependence in (2.3) is a natural generalization of Shannon's mutual information, we should expect to see confirmations from other related problems in source coding or channel coding. Indeed, Csiszár and Narayan [12] already discovered a duality between the secret key agreement problem and the source coding problem of communication for omniscience (CO). To explain this, suppose the users want to independently compress their observations into public messages such that every active user can recover the entire source Z_V. What is the smallest sum rate of the messages required? If the active users do not observe any correlated source, then $H(\mathsf{Z}_V)$ is needed by the Slepian-Wolf source coding theorem [8]. If the active users observe some correlated sources, however, they can use them as side information to recover the entire source. The smallest sum rate R_{co}, called the smallest rate of communication for omniscience, can be smaller than $H(\mathsf{Z}_V)$. The maximum savings in the source coding rate below $H(\mathsf{Z}_V)$ turns out to be the secrecy capacity [12].

$$C_{\text{s}} = H(\mathsf{Z}_V) - R_{\text{co}} \tag{2.16}$$

In this section, we extend the result to a general duality between the problem of secret key agreement and the problem of maximum common randomness (MCR) under a rate constraint on the public discussion. It has the following information-theoretic appeal:

1. Maximum common randomness can be attained *strongly* by first agreeing on a maximum secret key and vice versa.

2. Secrecy capacity has the interpretation of mutual dependence as the maximum common randomness that needs not be publicly discussed.

The duality holds in the asymptotic sense. There may not be any polynomial reductions between the two problems. In other words, a practical solution to one problem does not guarantee a practical solution to the other. This is unlike the duality between secret key agreement and network coding to be described in Chapter 3, where a practical network code gives rise to a practical secret key agreement scheme.

We now define the problems of rate-constrained secret key agreement and maximum common randomness as follows. For a slightly more general result, we incorporate a set $D \subsetneq V$ of untrusted users in the model. They are *not* active by definition, i.e. $A \subseteq D^c$, and their observations are known to the wiretapper. This means that the secret key should be asymptotically independent of both the public messages and the private knowledge of the untrusted users. A more detailed description can be found in Chapter 5 or [12].

Definition 2.2 (Rate-constrained SKA) Consider the source model where terminal $i \in V$ observes finitely-valued memoryless source Z_i, and

$$A \subseteq D^c \subseteq V : |A| \geq 2$$

where A and D are the sets of active and respectively untrusted users. The secrecy capacity under public discussion rate R is

$$C_{\mathrm{s}}(R) := \sup \liminf_{n \to \infty} \frac{1}{n} \log |K| \tag{2.17}$$

by choosing a sequence in n of $(\mathsf{K}, \mathsf{U}_{D^c}, \mathsf{F}, \mathsf{K}_A)$ where

- K is a finitely-valued random variable with support set K,
- U_i for $i \in D^c$ are continuous-valued random variables independent of Z_V with density function $\prod_{i \in D^c} P_{\mathsf{U}_i}$,
- for some positive integer r, $\mathsf{F} := \mathsf{F}_{[r]}$ is a vector of F_j for $j \in [r]$ defined as some function $F_j(\mathsf{U}_{i_j}, \mathsf{Z}_j^n, \mathsf{Z}_D^n)$ for some $i_j \in D^c$ with finite range $|F_j| \in [2, \infty)$, and

- K_i for $i \in A$ is some function $K_i(\mathsf{U}_i, \mathsf{Z}_i^n, \mathsf{Z}_D^n, \mathsf{F})$ taking values from K,

subject to the constraints that

$$\Pr\{\exists i \in A, \mathsf{K}_i \neq \mathsf{K}\} \leq \epsilon_n \to 0 \quad \text{(recoverability)} \tag{2.18a}$$

$$\frac{1}{n}\left[\log|K| - H(\mathsf{K}|\mathsf{Z}_D^n\mathsf{F})\right] \leq \delta_n \to 0 \quad \text{(secrecy)} \tag{2.18b}$$

$$\limsup_{n\to\infty}\left[\frac{1}{n}\log|F| - R\right] \leq 0 \quad \text{(discussion rate)} \tag{2.18c}$$

$C_{\mathrm{s}}(R)$ is said to be *strongly achievable* if ϵ_n and δ_n decays exponentially to 0. □

In words, $C_{\mathrm{s}}(R)$ is the maximum rate of secret key K that can be recovered by terminal $i \in A$ as K_i after private randomization U_i, observation Z_i^n by terminal $i \in D^c$, public observation Z_D^n and discussion F at a rate below R. Note that randomization and public discussion are carried out before and respectively after observing the entire source for n time units. Furthermore, Z_D^n is revealed in public a priori without the need for additional public discussion.

The problem of maximum common randomness is defined in a similar way except that the secrecy constraint is replaced by the uniformity constraint.

Definition 2.3 (MCR) Consider the same source as in Definition 2.2 where terminal $i \in V$ observes finitely-valued memoryless source Z_i, and $A \subseteq D^c \subseteq V : |A| \geq 2$. The maximum common randomness capacity under public discussion rate R is

$$C(R) := \sup \liminf_{n\to\infty} \frac{1}{n}\log|L| \tag{2.19}$$

subject to the constraints that

$$\Pr\{\exists i \in A, \mathsf{L}_i \neq \mathsf{L}\} \leq \epsilon_n \to 0 \quad \text{(recoverability)} \tag{2.20a}$$

$$\frac{1}{n}\left[\log|L| - H(\mathsf{L}|\mathsf{Z}_D^n)\right] \leq \delta_n \to 0 \quad \text{(uniformity)} \tag{2.20b}$$

$$\limsup_{n\to\infty}\left[\frac{1}{n}\log|F| - R\right] \leq 0 \quad \text{(discussion rate)} \tag{2.20c}$$

where the common randomness L is a finitely-valued random variable, the public discussion F and the private randomizations U_{D^c} are as defined before in Definition 2.2, and $\mathsf{L}_i := L_i(\mathsf{U}_i, \mathsf{Z}_i^n, \mathsf{Z}_D^n, \mathsf{F})$ is the estimate of L by each active user $i \in A$. $C(R)$ is said to be *strongly achievable* if ϵ_n and δ_n decays exponentially to 0. □

The common randomness L defined above is the same as the secret key K in Definition 2.2 except that L needs not be asymptotically independent of F. Both of them need to be almost surely recoverable by the active users, and nearly uniformly distributed over their support sets. The component sources Z_D of the untrusted users are also assumed to be publicly known but do not count towards the discussion rate.

Theorem 2.3 (Duality) *The rate-constrained secrecy capacity $C_{\mathrm{s}}(R)$ and the common randomness capacity $C(R)$ defined in Definition 2.2 and 2.3 are both strongly achievable and satisfy the following duality,*

$$C(R) = C_{\mathrm{s}}(R) + R \qquad \textit{for all } R \geq 0 \tag{2.21}$$

(2.16) *is a special case when $R = R_{\mathrm{co}}$, where $C(R)$ becomes $H(\mathsf{Z}_V)$.* □

This is illustrated in Figure 2-3. C_0 is the maximum amount of common randomness with a negligible amount of public discussion. This is also the secrecy capacity since the public discussion reveals only a negligible amount of the common randomness. At $R = r$, $C(r) = C_{\mathrm{s}}(r) + r$ increases in r at a rate larger than 1 by the monotonicity of $C_{\mathrm{s}}(r)$. Eventually, this rate decreases to 1 at some point $R = R_{\mathrm{s}}$, where further increasing R does not increase $C_{\mathrm{s}}(R)$. This gives the secrecy capacity $C_{\mathrm{s}} = \lim_{R\to\infty} C_{\mathrm{s}}(R)$ with unlimited discussion rate. The smallest rate R_{co} of communication for omniscience is when the common randomness is the entire DMMS Z_V. By the result of [12], $(R_{\mathrm{s}}, H(\mathsf{Z}_V))$ is a point on the curve to the right of R_{s} because the secrecy capacity C_{s} can be attained through communication for omniscience.

It is *not* known whether there is a single-letter characterization for the curve over $R \in (0, R_{\mathrm{co}})$. The smallest rate R_{s} to attain secrecy capacity C_{s} is also unknown in general as mentioned in [12]. Nonetheless, we have an intuitive duality between the rate-constrained SKA and MCR problems, which is evident from the duality between

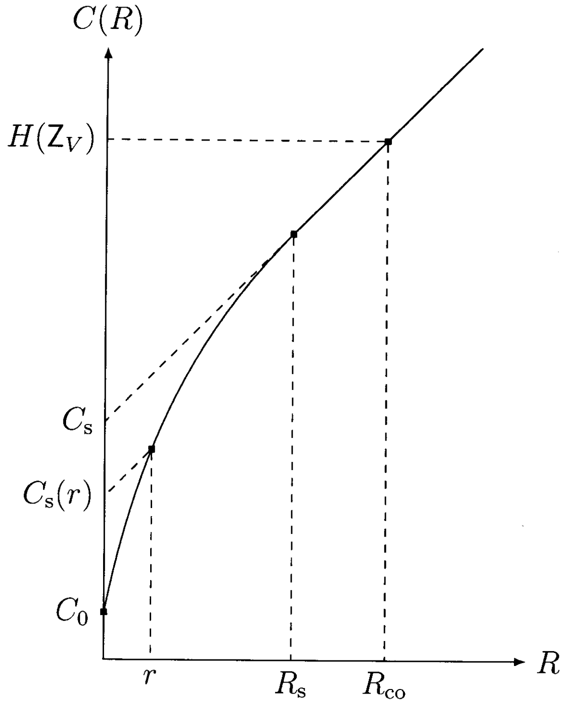


Figure 2-3: Duality between rate-constrained SKA and MCR

the SKA and CO problems in [12]. On the one hand, if the active users can agree on $C(R)$ bits of common randomness using R bits of public discussion, the wiretapper can learn at most R bits of the common randomness from the public messages. The remaining $C(R) - R$ bits should remain secret. On the other hand, since the active users can agree on $C_\mathrm{s}(R)$ bits of secret using R bits of public discussion, they can use both the secret key and the public messages as common randomness. Since the secret key is nearly uniformly distributed given the public messages, which can also be compressed to nearly uniformly distributed random variables, the amount of common randomness is at least $C_\mathrm{s}(R) + R$.

The additional fact that both capacities can be strongly achievable follows from the technical lemmas below:

1. source coding can be done efficiently with exponentially decaying error probability by Proposition A.6, which is a result from [11], and
2. the uniformity/secrecy condition can be satisfied in the strong sense by Lemma A.4, which is a straight-forward extension of [12, Lemma B.2].

The detailed proof of Theorem 2.3 is given in Section A.5.

Chapter 3

Linear Network Coding

To prove that the LP characterization (1.2) of the secrecy capacity is achievable, Csiszár and Narayan [12] applied the source coding solution of communication for omniscience for secrecy key agreement. This duality between the source coding and secrecy problems was described and extended in Section 2.4. The idea is to have the users publicly discuss at the smallest rate R_{co} until every active user recovers the entire source Z_V almost surely, attaining omniscience. Then, by a random coding argument, the active users can extract the desired secret key as a function of Z_V at rate $H(\mathsf{Z}_V) - R_{\text{co}}$. The LP characterization of R_{co} gives the desired secrecy capacity.

Although the duality with source coding gives a systematic solution to secret key agreement, it does not guarantee a practical solution. Much like the channel coding theorem, the random coding argument does not impose any structure to the code. In the worst case, the public discussion and key functions may be exponentially complex in the constraint length n, which may have to be very large for the asymptotic recoverability and secrecy conditions to be meaningful. There is no polynomial reduction between the source coding and the secrecy problems, and so a practical solution for one problem does not necessarily entail a practical solution for the other.

In this chapter, we will prove a duality between the channel coding and the secrecy problems by giving a practical linear network coding solution to secret key agreement. Network coding [28] is normally a solution to the channel coding problem over a network with noise-free links but it turns out to apply here as well to the secrecy problem.

We will start with a source model that has a general linear dependence structure and then strengthen the results for a relatively more specific source, whose dependence structure can be captured by a hypergraph. Recoverability and secrecy conditions can be satisfied perfectly for some finite n, and the linear operations involved in the public discussion and key generation have only polynomial complexity in n.

3.1 Finite linear source

We will illustrate the main ideas with the following example of a private source.

Example 3.1 Let $V = [3]$ and $\mathsf{Z}_3 = \mathsf{Z}_1 \oplus \mathsf{Z}_2$ where Z_1 and Z_2 are uniformly distributed independent bits, and $\oplus$ is the XOR operation (or modulo two sum). □

This is called a finite linear source because the observations are linearly related.

Definition 3.1 (Finite linear source) Z_V is a *finite linear source* if the component source Z_i for user $i \in V$ are vectors of random variables that can be expressed as linear combinations of a common set of independent random variables that are uniformly distributed over a common finite field. In matrix notation,

$$\mathbf{z} = \boldsymbol{H}\mathbf{x} = \begin{bmatrix} \boldsymbol{I} \\ \bar{\boldsymbol{H}} \end{bmatrix} \mathbf{x} \tag{3.1}$$

where $\mathbf{x}$ is the vector of independent random variables uniformly distributed over a finite field $\mathbb{F}_q$ of order q, $\boldsymbol{H}$ is a matrix of elements in $\mathbb{F}_q$ consisting of the identity matrix $\boldsymbol{I}$ and submatrix $\bar{\boldsymbol{H}}$, and $\mathbf{z}$ is the vector of all random variables partitioned into Z_V. Without loss of generality, we assume that the elements in $\mathbf{x}$ can be partitioned into X_V where $\mathsf{X}_i \subseteq \mathsf{Z}_i$ for $i \in V$.[1] X_V is called a *base* of Z_V, while $\boldsymbol{H}$ and $\bar{\boldsymbol{H}}$ are called a *representation* and a *basic representation* of Z_V respectively. □

[1] To argue this, we can first assume $\boldsymbol{H}$ has full column rank without loss of generality. Then, there exists an invertible submatrix $\bar{\boldsymbol{H}}$ of rows from $\mathbf{H}$. Rewriting (3.1) as $\mathbf{z} = (\boldsymbol{H}\bar{\boldsymbol{H}}^{-1})(\bar{\boldsymbol{H}}\mathbf{x})$ gives the desired structure.

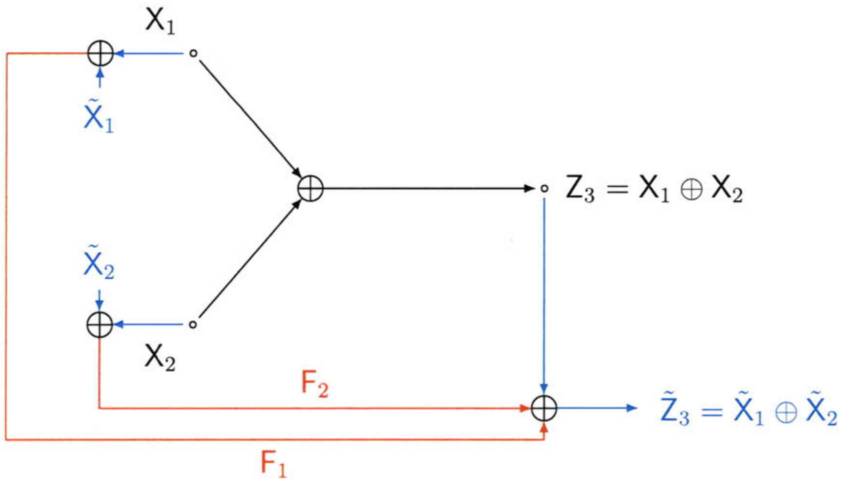

X_1
$\tilde{\mathsf{X}}_1$
$\mathsf{Z}_3 = \mathsf{X}_1 \oplus \mathsf{X}_2$
$\tilde{\mathsf{X}}_2$
X_2
F_2
$\tilde{\mathsf{Z}}_3 = \tilde{\mathsf{X}}_1 \oplus \tilde{\mathsf{X}}_2$
F_1

Figure 3-1: Turning private source to private channel

For Example 3.1, $(\mathsf{Z}_1, \mathsf{Z}_2)$ is a base since we can write

$$\overbrace{\begin{bmatrix}\mathsf{Z}_1\\ \mathsf{Z}_2\\ \mathsf{Z}_3\end{bmatrix}}^{\mathsf{z}} = \overbrace{\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 1\end{bmatrix}}^{H}\overbrace{\begin{bmatrix}\mathsf{Z}_1\\ \mathsf{Z}_2\end{bmatrix}}^{\mathsf{x}}\begin{matrix}\}\mathsf{X}_1\\ \}\mathsf{X}_2\end{matrix}$$

with $\bar{\boldsymbol{H}} = \begin{bmatrix}1 & 1\end{bmatrix}$, and the base $(\mathsf{X}_1, \mathsf{X}_2)$ uniformly distributed.

Indeed, with public discussion, we can convert the private source into a private channel characterized by the transfer matrix $\boldsymbol{H}$ (or simply $\bar{\boldsymbol{H}}$). This is illustrated in Figure 3-1 for Example 3.1. Let $\tilde{\mathsf{X}}_1$ and $\tilde{\mathsf{X}}_2$ be arbitrary secrets of user 1 and 2 respectively. The users publicly reveal the cryptograms $\tilde{\mathsf{X}}_1 \oplus \mathsf{X}_1$ and $\tilde{\mathsf{X}}_2 \oplus \mathsf{X}_2$, which are independent of the secrets by the uniformity of the base.[2] User 3 adds the cryptograms to Z_3 and observes effectively the following sum,

$$\tilde{\mathsf{Z}}_3 := \mathsf{Z}_3 \oplus (\tilde{\mathsf{X}}_1 \oplus \mathsf{X}_1) \oplus (\tilde{\mathsf{X}}_2 \oplus \mathsf{X}_2) = \tilde{\mathsf{X}}_1 \oplus \tilde{\mathsf{X}}_2$$

since $\mathsf{Z}_3 = \mathsf{X}_1 \oplus \mathsf{X}_2$ by definition. This is the desired channel characterized by $\bar{\boldsymbol{H}}$. As user 1 and 2 observe $\tilde{\mathsf{X}}_1$ and $\tilde{\mathsf{X}}_2$ trivially, we have the effective private channel $\boldsymbol{H}$.

[2] This is the one-time pad [51], which is perfectly secure.

Proposition 3.1 (Source to channel) *With public discussion, the private finite linear source Z_V in Definition 3.1 can be used as a deterministic linear private channel characterized by the transfer matrix $\boldsymbol{H}$ in* (3.1) *with inputs and outputs partitioned by the users as in X_V and Z_V respectively.* □

PROOF Let $\tilde{\mathbf{x}}$ be a random vector in $\mathbb{F}_q$ (not necessarily uniformly distributed) independent of $\mathbf{x}$ in (3.1) but with the same dimension. Define

$$\begin{aligned}\tilde{\mathbf{z}} &:= \boldsymbol{H}(\tilde{\mathbf{x}} + \mathbf{x}) - \mathbf{z} \\ &= \boldsymbol{H}(\tilde{\mathbf{x}} + \mathbf{x}) - \boldsymbol{H}\mathbf{x} = \boldsymbol{H}\tilde{\mathbf{x}}\end{aligned}$$

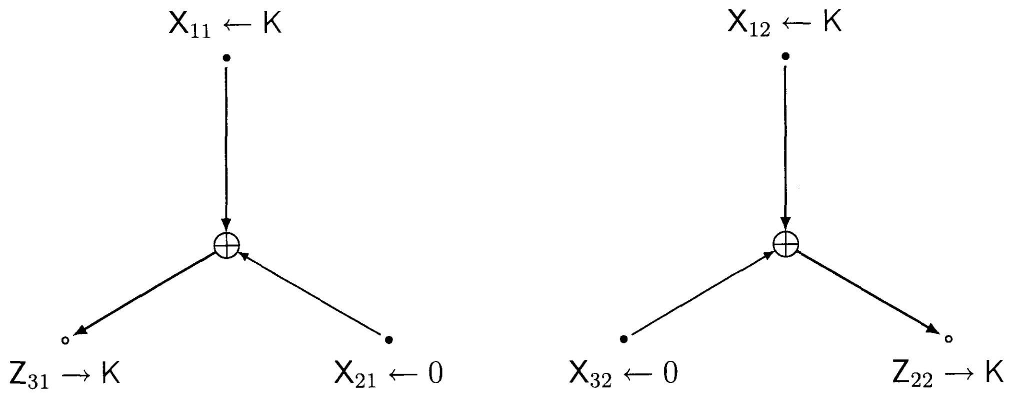
By definition, each element in $\tilde{\mathbf{z}}$ can be generated by the user who observes the corresponding element in $\mathbf{z}$ and the vector $\tilde{\mathbf{x}}+\mathbf{x}$ of cryptograms. Since the cryptograms are independent of the secrets $\tilde{\mathbf{x}}$ by the uniformity of $\mathbf{x}$, revealing them in public effectively gives a private channel $\boldsymbol{H}$ from inputs in $\tilde{\mathbf{x}}$ to outputs in $\tilde{\mathbf{z}}$. ■

In essence of the above proposition, we can treat X_V and Z_V as the inputs and outputs of a private channel. Users can share a secret key simply by generating it at a source node and multicasting it to the others through the private channel. In Example 3.1, user 1 and 3 can share a secret bit K by setting the inputs as $\mathsf{X}_1 \leftarrow \mathsf{K}$ and $\mathsf{X}_2 \leftarrow 0$. This gives the output $\mathsf{Z}_3 \rightarrow \mathsf{K}$ as desired.

Suppose user 2 also wants to share the key. We can extend the source model to two time units and let Z_{it} be the observation of user $i \in [3]$ at time $t \in [2]$. Let

$$(\mathsf{X}_{11}, \mathsf{X}_{21}, \mathsf{X}_{12}, \mathsf{X}_{32}) := (\mathsf{Z}_{11}, \mathsf{Z}_{21}, \mathsf{Z}_{12}, \mathsf{Z}_{32}) \tag{3.2}$$

be the inputs. Then, setting $\mathsf{X}_{11} \leftarrow \mathsf{K}$ and $\mathsf{X}_{21} \leftarrow 0$ spread the key bit from user 1 to 3 at time 1, while setting $\mathsf{X}_{12} \leftarrow \mathsf{K}$ and $\mathsf{X}_{32} \leftarrow 0$ spread the key bit from user 1 to 2 at time 2. The key rate achieved is 0.5 bits. This network coding approach is summarized in Figure 3-2 and can be generalized as follows for any finite linear source.



(a) Time 1

(b) Time 2

Figure 3-2: Network code for Example 3.1: X_{it} and Z_{it} denote the input • and respectively output ∘ of user i at time t. The routes of information flow are highlighted.

Definition 3.2 (Single source) Given a finite linear source Z_V in Definition 3.1, the active users in A can share a secret key as follows.

1. Extend the source Z_V over $n \in \mathbb{P}$ time units to Z_V^n.
2. Pick a source node $s \in A$.
3. Pick a base X_V^n of Z_V^n as the inputs of the effective private channel.[3]
4. Have the source s generate a secret key K and multicast it to all active users in A through the private channel using a linear network code [28].

K is chosen to be uniformly distributed and required to be perfectly recoverable by all users in A. No additional public discussion is performed other than that required to convert the source to a private channel in Proposition 3.1.[4] □

For the linear network code in Figure 3-2 for Example 3.1, we have chosen $n = 2$, $s = 1$, $X_{V1} = (Z_{11}, Z_{21}, \emptyset)$ and $X_{V2} = (Z_{12}, \emptyset, Z_{32})$. This network coding approach is indeed optimal because the mutual dependence in (2.3) evaluates to 0.5 bits as follows,

$$\begin{aligned} C_s &\leq \frac{1}{2} D(P_{Z_1Z_2Z_3} \| P_{Z_1}P_{Z_2}P_{Z_3}) \\ &= \frac{1}{2}\left[H(Z_1) + H(Z_2) + H(Z_3) - H(Z_1Z_2Z_3)\right] \\ &= \frac{1}{2}\left[1 + 1 + 1 - 2\right] = 0.5 \end{aligned}$$

[3] See the conversion from source to channel in Proposition 3.1.

[4] Additional channel uses beyond n times may be needed for the linear network code.

In general, when given a finite linear source, what is the key rate achievable by the single-source linear network coding scheme in Definition 3.2? Does it reach the secrecy capacity? To answer these questions, we need only characterize the maximum network throughput and compare it to the secrecy capacity. Intuitively, the larger the rank of the transfer matrix $\boldsymbol{H}$ is, the larger the correlation between the channel inputs and outputs, and so the larger the achievable key rate by network coding. We can characterize the achievable key rate using such rank function in the language of matroid theory [50].

Definition 3.3 (Linear matroid) Given a finite linear source Z_V in Definition 3.1, let $Z_V = \bigcup_{i \in V} Z_i$ be a set of elements that indexes the corresponding random element in the source $\mathsf{Z}_V = (\mathsf{Z}_i : i \in V)$, or the rows of $\boldsymbol{H}$ in correspondence to the way Z_V partitions $\mathbf{z}$.[5] Define the *rank function* $r : Z_V \mapsto \mathbb{N}$ with $r(T)$ for $T \subseteq Z_V$ being the rank of the submatrix consisting of the rows of $\boldsymbol{H}$ indexed by the elements in T. The pair (Z_V, r) is called the *(linear) matroid* for Z_V, and we have

$$r(Z_C) = \frac{H(\mathsf{Z}_C)}{\log q} \qquad \text{for all } C \subseteq V$$

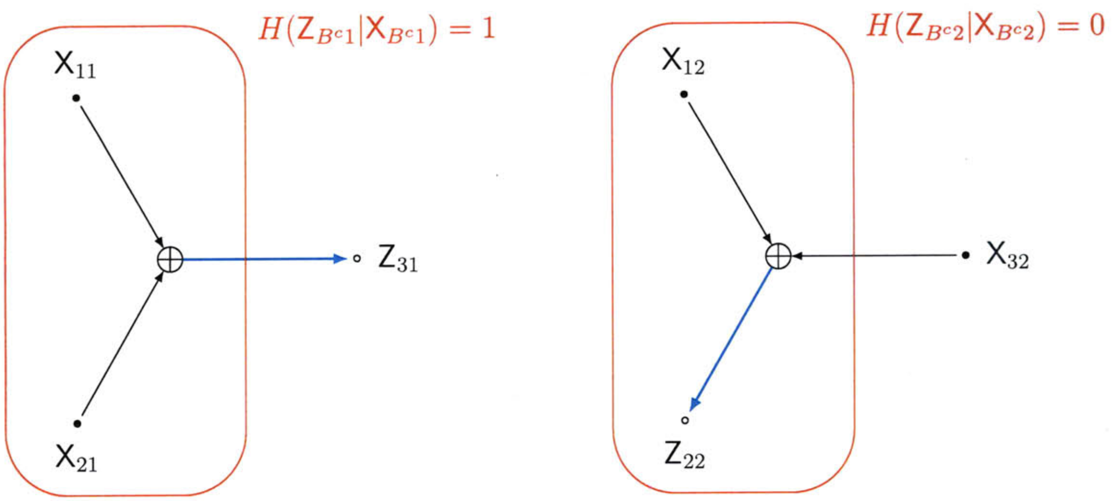
$\boldsymbol{H}$ is called a *representation* of the linear maroid.[6] $\bar{\boldsymbol{H}}$ is referred to as a *basic representation* of the linear matroid.

We call $r(T|U) := r(T \cup U) - r(U)$ for $T, U \subseteq Z_V$ the *conditional rank* of T given U. $\mathcal{X}$ is defined as the set of *bases* X_V with $X_i \subseteq Z_i$ disjoint and $r(X_V) = r(Z_V)$. It is easy to see that the elements in a base $X_V \in \mathcal{X}$ indexes the elements in a base X_V of the finite linear source Z_V. □

In Example 3.1, the rank of $\boldsymbol{H}$ is $r(Z_V) = H(\mathsf{Z}_V) = 2$. The set $\mathcal{X}$ of bases are (Z_1, Z_2), (Z_2, Z_3) and (Z_1, Z_3). The achievable key rate can be characterized using r (and $\mathcal{X}$) as follows. Given the time extension n, channel inputs X_V^n and the source

[5] For example, if Z_1 is a vector of two random bits $(\boldsymbol{h}_1^T \mathbf{x}, \boldsymbol{h}_2^T \mathbf{x})$, then $Z_1 = (1, 2)$ is a vector of the corresponding row indices of $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots \end{bmatrix}^T$.

[6] The convention we use here is that elements of the linear matroid are represented by rows of $\boldsymbol{H}$ instead of the columns.



(a) Time 1

(b) Time 2

Figure 3-3: Network throughput for Example 3.1: cut values for $B = [2]$

node $s \in A$, the network throughput has the following min-cut characterization [2],

$$\frac{1}{n} \min_{B \subseteq V: s \in B \not\supseteq A} r(Z^n_{B^c} | X^n_{B^c}) \qquad \text{(min-cut)}$$

Figure 3-3 illustrates the computation of the above expression for the linear network code in Figure 3-2. Consider $B = [2]$, which satisfies the constraint $s \in B \not\supseteq A$ for $s = 1$ and $A = V = [3]$. At time 1, $r(Z_{B^c1}|X_{B^c1})$ equals $H(\mathsf{Z}_{B^c1}|\mathsf{X}_{B^c1}) = 1$ because there is one bit of channel output outside B, namely Z_{31}, that is controlled by some channel inputs inside B, e.g. X_{11}. This quantity, called the cut value, is the maximum possible information flow from B to B^c. At time 2, however, $r(Z_{B^c2}|X_{B^c2})$ equals $H(\mathsf{Z}_{B^c2}|\mathsf{X}_{B^c2}) = 0$ because there is no output outside B that is controlled by any input inside B. i.e. there is no information flow from B to B^c. The average information flow is 0.5 bits, which turns out to be the minimum cut value among all valid choices of B, namely $\{1\}$, $\{1,2\}$ and $\{1,3\}$.

Maximizing the min-cut value over all possible choices of the time extension n of the channel inputs X_V give the maximum achievable secret key rate for the single-source network coding scheme. Using the memorylessness of the channel, we have the following single-letter characterization, which can be computed numerically for simple networks.

Theorem 3.1 (Single source) *Given a finite linear source* Z_V *in Definition 3.1, the secret key rate* $C_{\text{sn}}^A(\log q)$ *bits achievable by the single-source network coding scheme in Definition 3.2 can be characterized by the matroid* (Z_V, r) *in Definition 3.3 as*

$$C_{\text{sn}}^A := \max_{P_{\hat{\mathsf{X}}_V} \in \mathscr{P}(\mathcal{X})} \min_{B \subseteq V: s \in B \not\supseteq A} \mathrm{E}[r(Z_{B^c}|\hat{\mathsf{X}}_{B^c})] \tag{3.3a}$$

$$= \max_{P_{\hat{\mathsf{X}}_V} \in \mathscr{P}(\mathcal{X})} \min_{\mathcal{P} \in \Pi(A)} \frac{\sum_{C \in \mathcal{P}} \mathrm{E}[r(Z_C|\hat{\mathsf{X}}_C)]}{|\mathcal{P}| - 1} \tag{3.3b}$$

where $\hat{\mathsf{X}}_V$ *is a random variable distributed as* $P_{\hat{\mathsf{X}}_V}$ *over the set* $\mathcal{X}$ *of bases of* (Z_V, r) *and* $\Pi(A)$ *is defined in* (B.4a) *as the collection of* $\mathcal{P} = \{C_1, \dots, C_k\} \subseteq 2^V \setminus \{\emptyset\}$ *such that* $k \geq 2$*, every* C_i *contains an element in* A*, and every element in* A *is contained in exactly one* C_i*.*[7] *n.b.* (3.3b) *is independent of the choice of* s*, and so as* (3.3a).

Furthermore, the secret key can be perfectly secret and recoverable in the sense that it can be independent of the public messages and recoverable by the active users with zero error probability. The rate can approach $C_{\text{sn}}^A(\log q)$ *bits with a gap in the order of* $|V|/n$*.* □

The single-source network coding approach is optimal if the throughput $C_{\text{sn}}^A(\log q)$ equals the secrecy capacity. This is indeed the case when all users are active, i.e. $A = V$. As we have proven earlier in Theorem 2.1, the secrecy capacity equals the mutual dependence (2.3) when all users are active. For the finite linear source, it is easy to simplify the divergence expression to the following form

$$C_{\text{s}} = \min_{\mathcal{P} \in \Pi} \frac{1}{|\mathcal{P}| - 1} \sum_{C \in \mathcal{P}} r(Z_C|X_C) \qquad \text{for any } X_V \in \mathscr{X}$$

which has the combinatorial interpretation of *partition connectivity* [3, 20]. It is invariant to the choice of $X_V \in \mathscr{X}$, and turns out to equal the min-cut characterization of the network throughput. Figure 3-4 illustrates the computation of this expression for Example 3.1. Consider $\mathcal{P} = (\{1\}, \{2\}, \{3\})$ and $X_V = (Z_1, Z_2, \emptyset)$. Each $r(Z_C|X_C)$ is the cut value that gives the maximum amount of information that can possibly flow

[7] Elements in A^c can be in any number (or none) of the parts of $\mathcal{P}$.

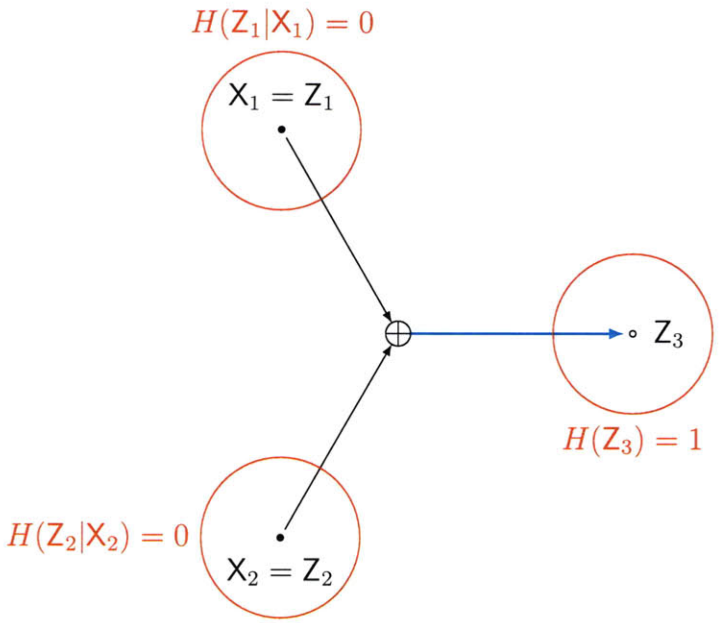


Figure 3-4: Secrecy capacity for Example 3.1: $\mathcal{P} = (\{1\}, \{2\}, \{3\})$

from C^c into $C \in \mathcal{P}$. Since there is only one output bit, namely Z_3, that is controlled by some input from other users, only one part in $\mathcal{P}$ has a cut value of 1 bit. This turns out to achieve the minimum in the partition connectivity expression, which is 0.5 bits as desired.

Corollary 3.1 C^A_{sn} *is independent of* $s \in A$ *by* (3.3b). *When all users are active, i.e.* $A = V$, (3.3b) *becomes the secrecy capacity* $C^V_{\text{s}}(\log q)$

$$C^V_{\text{sn}} = C^V_{\text{s}} = \min_{\mathcal{P} \in \Pi} \frac{\sum_{C \in \mathcal{P}} r(Z_C) - r(Z_V)}{|\mathcal{P}| - 1} \tag{3.4}$$

where $\Pi = \Pi(V)$. *In other words, the single-source network coding approach to secret key agreement is optimal when all users are active.* □

Proof By the random coding argument in [2], the asymptotic throughput of a multicast session from source $s \in A$ to other active users in A with input X_V chosen for the input is

$$\min_{B \subseteq V : s \in B \not\supseteq A} r(Z_{B^c} | X_{B^c})$$

This can be approached in the order of $|V|/n$ with perfect recoverability and secrecy.[8]

[8] Perfect secrecy is immediate since no additional public discussion is needed. To argue perfect

Maximizing the throughput over the choice of the base $X_V^n \in \mathcal{X}^n$ of the n-extension $Z_V^n := (Z_{Vt} : t \in [n])$ gives the rate

$$\begin{aligned}\frac{1}{n} \max_{X_V^n \in \mathcal{X}^n} \min_{B \subseteq V : s \in B \not\supseteq A} r(Z_{B^c}^n | X_{B^c}^n) &= \frac{1}{n} \max_{X_V^n \in \mathcal{X}^n} \min_{B \subseteq V : s \in B \not\supseteq A} \sum_{t \in [n]} r(Z_{B^c t} | X_{B^c t}) \\ &= \max_{X_V^n \in \mathcal{X}^n} \min_{B \subseteq V : s \in B \not\supseteq A} \sum_{X_V \in \mathcal{X}} \frac{N(X_V | X_V^n)}{n} r(Z_{B^c} | X_{B^c})\end{aligned}$$

where $N(X_V|X_V^n)$ is the number of occurrences of X_V in the sequence X_V^n. As $n \to \infty$, the above rate approaches (3.3a) as desired in the order of $1/n$. (3.3b) follows from a surprisingly general identity for matroids in Theorem B.2. (3.3b) is trivially independent of $s \in A$, so as (3.3a). When $A = V$, the expression in (3.3b) inside the maximization is independent of P_{X_V} because

$$\begin{aligned}\sum_{C \in \mathcal{P}} r(Z_C | X_C) &\overset{\text{(a)}}{=} \sum_{C \in \mathcal{P}} \left(r(Z_C) - \sum_{i \in C} |X_i| \right) \\ &\overset{\text{(b)}}{=} \sum_{C \in \mathcal{P}} r(Z_C) - r(Z_V)\end{aligned}$$

where (a) is because $X_C \subseteq Z_C$ and $r(X_i) = |X_i|$, and (b) is because $\sum_{C \ni i} 1 = 1$ and $\sum_{i \in V} |X_i| = |X_V| = r(X_V) = r(Z_V)$ by the definition of a base.

(3.4) for the corollary also follows from the last expression. Since $H(\mathsf{Z}_C) = r(Z_C) \log q$, the R.H.S. of (b) is the divergence $D(P_{\mathsf{Z}_V} \| \prod_{C \in \mathcal{P}} P_{\mathsf{Z}_C}) / \log q$, and so (3.3b) becomes the secrecy capacity in $\log q$ bits as desired from (2.4). ■

In the other case when $A \subsetneq V$, the single-source network coding approach still applies but its optimality is unknown. For the simple networks we have randomly generated so far, it is optimal. No counter-example has been found.

recoverability, note that the average error probability in [2] of a random linear code at rate R under a linear deterministic network is at most $2^{nR - (n - |V|) \min_{B \subseteq V : s \in B \not\supseteq A} r(Z_{B^c} | X_{B^c})}$. A code makes an error only if at least one message gives the same observation at a destination node in A as another message. For linear codes, existence of one such message implies by linearity that all other messages have this problem, and so the error probability must be at least $1/2$. If the average error probability is less than $1/2$, i.e. the error exponent bigger than 1, there must be a linear code that attains zero error probability as desired.

3.2 Source with dependency hypergraph

In the previous section, we described a single-source network coding approach for the finite linear source model where the dependency of the private observations can be captured by a (linear) matroid. By viewing matroids partitioned by vertices as edges in graphs, we discovered a general identity (B.33) in matroid theory that proves the optimality of the network coding approach in the case when all users are active.

In this section, we will derive some stronger results in the special case when the dependency of the source can be captured by a hypergraph. For example, we can derive a better delay guarantee for communicating the secret key bits by network coding, and prove that the network coding approach is also optimal when there are only two active users but an arbitrary number of helpers. When all users are active, the secrecy capacity corresponds to a more concrete notion of partition connectivity [3] of the dependency hypergraph. This gives a theoretically appealing confirmation of secrecy capacity as a measure of mutual dependence described in Chapter 2.

Example 3.1 considered in the previous section is indeed an example of a source with dependency hypergraph. The precise definition is as follows.

Definition 3.4 (Source with dependency hypergraph) A *hypergraph* is a tuple $H = (V, E, \phi)$ of vertex set V, edge set E, and edge function $\phi : E \mapsto 2^V \setminus \{\emptyset\}$ with $|\phi(e)| \geq 2$ for all $e \in E$.[9] Given two hypergraphs $H_1 = (V, E_1, \phi_1)$ and $H_2 = (V, E_2, \phi_2)$ on the same vertex set, the *disjoint union* $H_1 \sqcup H_2$ is a hypergraph $H = (V, E, \phi)$ with $E = \{(e, i) : e \in E_i\}$ and

$$\phi((e,i)) = \phi_i(e) \qquad \text{for all } (e,i) \in E$$

i.e. we distinguish between the edges from H_1 and H_2 by the additional index i.

Given the hypergraph $H = H_1 \sqcup H_2$ and finite group $\mathbb{G}$ of order $|\mathbb{G}| = q$, define $\mathsf{z}_i^{(e,1)} = \mathbf{z}^{(e,1)}$ for every $e \in E_1$ and $i \in \phi_1(e)$ as an independent random variable uniformly distributed over $\mathbb{G}$, and $\mathbf{z}^{(e,2)} := (\mathsf{z}_i^{(e,2)} : i \in \phi_2(e))$ as an independent

[9] It is possible but not of interest to consider singleton edges here.

random vector uniformly distributed over a subset of $\mathbb{G}^{|\phi_2(e)|}$ with zero sum,

$$\sum_{i \in \phi_2(e)} \mathsf{z}_i^{(e,2)} = 0 \qquad \text{for all } e \in E_2 \tag{3.5}$$

Define the source Z_V such that Z_i is the (ordered) set of $\mathsf{z}_i^{(e,j)}$ for $e \in E_j : i \in \phi_j(e)$ and $j \in [2]$. H is referred to as the *dependency hypergraph* of Z_V. □

For Example 3.1, the dependency hypergraph H consists of an empty hypergraph H_1 with no edge, and a hypergraph H_2 having one edge, say e, with $\phi_2(e) = [3]$. The random vector $\mathbf{z}^{(e,2)}$ is simply $(\mathsf{Z}_1, \mathsf{Z}_2, -\mathsf{Z}_3)$, which satisfies the constraint (3.5) trivially. Another straightforward example is to have $\mathsf{Z}_1 = \mathsf{Z}_2 = \mathsf{Z}_3$ be a uniformly random bit. In this case, the dependency hypergraph consists of an empty hypergraph H_2 with no edges, and a hypergraph H_1 having one edge e with $\phi_1(e) = [3]$. The random variable $\mathsf{z}^{(e,1)}$ observed by every node in $\phi_1(e)$ is simply Z_1.

We can also combine the previous two sources into one, with $\mathsf{Z}_1 = (\mathsf{z}^{(e,1)}, \mathsf{z}_1^{(e',2)})$, $\mathsf{Z}_2 = (\mathsf{z}^{(e,1)}, \mathsf{z}_2^{(e',2)})$ and $\mathsf{Z}_3 = (\mathsf{z}^{(e,1)}, \mathsf{z}_3^{(e',2)})$ such that $\mathsf{z}^{(e,1)}$, $\mathsf{z}_1^{(e',2)}$ and $\mathsf{z}_2^{(e',2)}$ are independent uniformly random bits while $\mathsf{z}_3^{(e',2)} = -(\mathsf{z}_1^{(e',2)} + \mathsf{z}_2^{(e',2)})$. Then, the dependency hypergraph H consists of a hypergraph H_1 with the edge e and a hypergraph H_2 with the edge e' where $\phi_1(e) = \phi_2(e') = [3]$.

A source with dependency hypergraph is a special case of the finite linear source in Definition 3.1 when $\mathbb{G}$ is a field. To see this, define an arbitrary root function $\rho_2 : E_2 \mapsto V$ with $\rho_2(e) \in \phi_2(e)$ for all $e \in E_2$. Then,

$$\mathbf{x} := (\mathsf{z}^{(e_1,1)}, \mathsf{z}_i^{(e_2,2)} : e_1 \in E_1, e_2 \in E_2, i \in \phi_2(e_2) \setminus \{\rho_2(e_2)\})$$

is a vector of independent random variables uniformly distributed over $\mathbb{G}$. By (3.5), the remaining random variables $\mathsf{z}_{\rho_2(e)}^{(e,2)}$ for $e \in E_2$ are simply negative sums of subsets of elements in $\mathbf{x}$. This satisfies the defining property (3.1) of a finite linear source for some representation $\boldsymbol{H}$ as desired. In general, regardless of whether $\mathbb{G}$ is a field, the matrix form (3.1) still applies for some matrix $\bar{\boldsymbol{H}}$ consisting of ±1 or 0. Indeed, the choice of a base X_V for Z_V corresponds to the choice of a root for every edge in H.

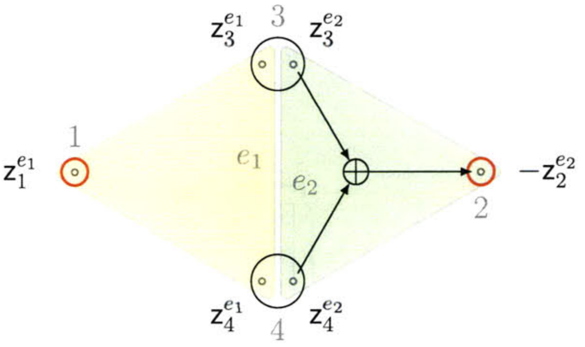


Figure 3-5: Dependency hypergraph for Example 3.2

Definition 3.5 (Root function) A star hypergraph [3] $H^* = (V, E, \phi, \rho)$ consists of a hypergraph $H = (V, E, \phi)$ and a root function $\rho : E \mapsto V$ with $\rho(e) \in \phi(e)$. Let $H^* = H_1^* \sqcup H_2^*$ be the star hypergraph of the dependency hypergraph in Definition 3.4 with the root functions $\rho : E \mapsto V$ and $\rho_i : E_i \mapsto V$ for $i \in [2]$ such that

$$\rho_i(e) = \rho((e,i)) \qquad \text{for all } (e,i) \in E$$

Define X_V as follows,

$$\mathsf{X}_i := (\mathsf{z}_i^{(e_1,1)}, \mathsf{z}_j^{(e_2,2)} : e_1 \in E_1, e_2 \in E_2, i = \rho_1(e_1), j \in \phi_2(e_2) \setminus \{\rho_2(e_2)\}) \qquad \textbf{(3.6)}$$

X_V is a base of Z_V, and every base can be defined this way for some ρ. □

As in Proposition 3.1, the source model can be converted to a private channel $\boldsymbol{H}$ or simply $\bar{\boldsymbol{H}}$ with the help of public discussion. In terms of the dependency hypergraph, we can view each edge as an independent transmission link as illustrated below.

Example 3.2 Let $V = [4]$, $A = [2]$, $\mathbb{G} = \mathbb{F}_2$, and Z_V be a source with the dependency hypergraph $H = H_1 \sqcup H_2$ in Definition 3.4 where H_1 and H_2 contain the 3-edges e_1 and e_2 respectively with

$$\phi_1(e_1) = \{1, 3, 4\} \quad \text{and} \quad \phi_2(e_2) = \{2, 3, 4\}$$

but no other edges. This is shown in Figure 3-6. □

Suppose we choose node 1 and 2 as the roots of e_1 and e_2 respectively, i.e. $\rho_1(e_1) = 1$ and $\rho_2(e_2) = 2$. Then, (3.6) gives $\mathsf{X}_1 = \mathsf{z}_1^{e_1}$, $\mathsf{X}_3 = \mathsf{z}_3^{e_2}$, $\mathsf{X}_4 = \mathsf{z}_4^{e_2}$ and $\mathsf{X}_2 = \emptyset$. In matrix form,

$$\begin{bmatrix} \mathsf{z}_1^{e_1} \\ \mathsf{z}_3^{e_2} \\ \mathsf{z}_4^{e_2} \\ \hdashline \mathsf{z}_3^{e_1} \\ \mathsf{z}_4^{e_1} \\ \mathsf{z}_2^{e_2} \end{bmatrix} = \begin{array}{r} \boldsymbol{I}\left\{\vphantom{\begin{matrix}1\\0\\0\end{matrix}}\right. \\ \bar{\boldsymbol{H}}\left\{\vphantom{\begin{matrix}1\\1\\0\end{matrix}}\right. \end{array} \overbrace{\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hdashline 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right]}^{\boldsymbol{H}} \begin{bmatrix} \mathsf{X}_1 \\ \mathsf{X}_3 \\ \mathsf{X}_4 \end{bmatrix}$$

Viewing $\bar{\boldsymbol{H}}$ as a channel, the first column corresponds to a broadcast link from sender $\rho_1(e_1)$ to receivers in $\phi_1(e_1) \setminus \{\rho_1(e_1)\}$. The last row corresponds to an interference link from senders in $\phi_2(e_2) \setminus \{\rho_2(e_2)\}$ to the receiver $\rho_2(e_2)$. Other rows and columns must be unit vectors by independence of the links.

Definition 3.6 (Hyperedges as links) We say that an edge e in a hypergraph $H = (V, E, \phi)$ is used as

- an *undirected broadcast link* in the sense that a sender $\rho(e) \in \phi(e)$ can be chosen at each time to send a unit of information noiselessly to all receivers in $\phi(e) \setminus \{\rho(e)\}$.
- an *undirected interference link* in the sense that a receiver $\rho(e) \in \phi(e)$ can be chosen to observe the sum of inputs from the senders in $\phi(e) \setminus \{\rho(e)\}$.
- a *selectable link* in the sense that a sender and a receiver can be chosen from $\phi(e)$ for a point-to-point link.

After converting the edges to independent directed links, the resulting composite channel is representable by a transfer matrix $\bar{\boldsymbol{H}}$ of which every non-zero entry equals 1 and is contained in a unit column or row vector.

The usual notion of a network flow can be built upon these various notions of links. A *unit flow* of length l is a sequence $u_1, e_1, u_2, e_2, \dots, u_{l+1}$ with u_i's being distinct vertices in V and e_i being an edge with sender selected as u_i and receiver selected as u_{i+1}. An *outbranching* from s to a set of nodes $A \setminus \{s\}$ is a collection of edges with the choice of a sender and one or more receivers for each edge, such that there is a unique unit flow from s to every node in $A\setminus\{s\}$. The *outcut* $\delta^+_{H^*}: 2^V \mapsto 2^E$ and *incut* $\delta^-_{H^*}$ are defined as

$$\delta^+_{H^*}(B) := \delta^-_{H^*}(B^c) := \{e \in E : \rho(e) \in B \not\supseteq \phi(e)\} \tag{3.7a}$$

for $B \subsetneq V$. The cut $\delta_H : 2^V \mapsto 2^E$ is defined as

$$\delta_H(B) := \{e \in E : B \not\supseteq \phi(e) \not\subseteq B^c\} \tag{3.7b}$$

The values of the cuts are simply their cardinalities. □

It follows from Proposition 3.1 that every edge in H_1 can be used as an undirected broadcast link or selectable link, while every edge in H_2 can be used as an undirected interference link or a selectable link. In Example 3.2, e_1 can be used as a broadcast link from sender 1 to receivers in $\{2,3\}$, which can also be viewed trivially as a point-to-point link from sender 1 to receiver 3. e_2 can be used as an interference link from senders in $\{3,4\}$ to receiver 2. If sender 4 transmits 0 over the link, we effectively have a point-to-point link from sender 3 to receiver 2. Using e_1 and e_2 as selectable links this way, there is a unit flow of information from user 1 to 3, namely the sequence $1, e_1, 3, e_2, 2$. This is shown in Figure 3-6. In general, when there are only 2 active users, we can use both the edges from H_1 and H_2 as selectable links and route the secret key from one active user to the other optimally as follows.

Definition 3.7 (Unicast by routing) Given a source Z_V with a dependency hypergraph $H = H_1 \sqcup H_2$ in Definition 3.4, two active users, i.e. $|A| = 2$, can share a secret key as follows.

1. Pick a source node $s \in A$ and a destination $t \in A\setminus\{s\}$.
2. Use the edges in H as selectable link in Definition 3.6 by choosing a sender and a receiver for every edge.
3. Decompose the links into edge-disjoint unit flows from s to t. Have the source s generate and route independent parts of a secret key K through each unit flow.

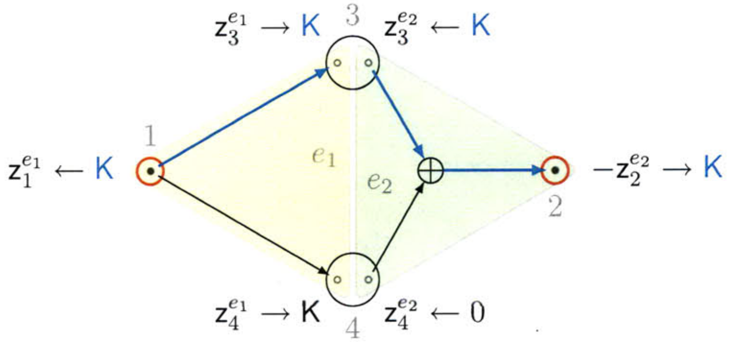


Figure 3-6: Network code for Example 3.2

This is a specialization of the scheme in Definition 3.2 without extension in time nor coding at any nodes. □

Theorem 3.2 *Given a source Z_V with a dependency hypergraph $H = H_1 \sqcup H_2$ in Definition 3.4, the routing solution in Definition 3.7 achieves the secret key rate $C_{\text{sn}}^A(\log q)$ bits in (3.3), which equals the secrecy capacity $C_{\text{s}}^A(\log q)$,*

$$C_{\text{sn}}^A = C_{\text{s}}^A = \min_{B \subseteq V : s \in B \not\ni t} |\delta_H(B)| \tag{3.8}$$

where $A = \{s, t\}$, and δ_H is defined in (3.7b). This can be attained non-asymptotically with no delay, and is independent of the choice of s. □

Proof The fact that (3.8) is the maximum number of edge-disjoint unit flows in H from s to t follows from Theorem B.3, which is a generalization of Menger's theorem to hypergraphs. It remains to show that (3.8) is the secrecy capacity. By [12],

$$C_{\text{s}}^A(\log q) \leq \min_{B \subseteq V : s \in B \not\ni t} D(P_{\mathsf{Z}_V} \| P_{\mathsf{Z}_B} P_{\mathsf{Z}_{B^c}})$$

Consider a star hypergraph $H^* = H_1^* \sqcup H_2^*$ and the corresponding base X_V defined in Definition 3.5.

$$\begin{aligned} D(P_{\mathsf{Z}_V} \| P_{\mathsf{Z}_B} P_{\mathsf{Z}_{B^c}}) &= H(\mathsf{Z}_B) + H(\mathsf{Z}_{B^c}) - [H(\mathsf{X}_B) + H(\mathsf{X}_{B^c})] \\ &= H(\mathsf{Z}_B | \mathsf{X}_B) + H(\mathsf{Z}_{B^c} | \mathsf{X}_{B^c}) \end{aligned}$$

because $H(\mathsf{Z}_V) = H(\mathsf{X}_V) = H(\mathsf{X}_B) + H(\mathsf{X}_{B^c})$ by the definition of a base. The last two entropy terms evaluate to the following cut values by (3.6),

$$\frac{H(\mathsf{Z}_B|\mathsf{X}_B)}{\log q} = |\delta^-_{H_1^*}(B)| + |\delta^+_{H_2^*}(B)|$$
$$\frac{H(\mathsf{Z}_{B^c}|\mathsf{X}_{B^c})}{\log q} = |\delta^+_{H_1^*}(B)| + |\delta^-_{H_2^*}(B)|$$

By (3.7), the sum of the above gives (3.8) as desired. ■

$|A| = 2$ is essential for the proof of optimality. For instance, Example 3.1 with $A = [3]$ cannot attain the secrecy capacity of 0.5 bits without an extension of $n \geq 2$. Furthermore, coding may sometimes be necessary as shown by the following example.

Example 3.3 Let $A = V = [4]$, $\mathbb{G} = \mathbb{F}_2$, and Z_V be a source with dependency hypergraph $H = H_1$, i.e. H_2 being empty, in Definition 3.4 where $E := \{123, 134, 124\}$ and $\phi(ijk) := \{i, j, k\} \in V$. This is illustrated in Figure 3-7. □

It is easy to argue that at least 2 edges are needed to give an outbranching that supports 1 bit of information flow from a source node to all other nodes. Since H has only 3 edges, there are at most 3 edge-disjoint outbranchings for every 2 time units. Thus, routing independent secret key over edge-disjoint outbranchings can attain a maximum rate of $\frac{3}{2}$ bits. With linear network coding, however, the secrecy capacity of 2 bits is achievable as follows:

1. Choose user 1 as the source.
2. Select user 1 as the sender for all edges in H, and all other users as receivers.
3. Have user 1 generate two independent secret key bits K_1 and K_2 uniformly distributed over $\mathbb{G}$, and then send K_1, K_2 and $\mathsf{K}_1 \oplus \mathsf{K}_2$ respectively over the broadcast links 123, 134 and 124.

Since every user has access to at least two links, they can recover the key bits perfectly. This is shown in Figure 3-7.

In general, we can specialize the scheme in Definition 3.2 as follows with convolutional network code, which has a better delay guarantee without diminishing the achievable secret key rate.

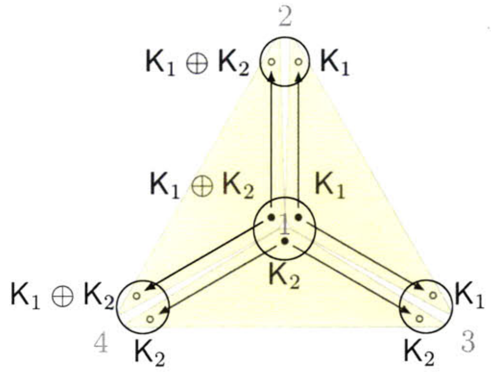


Figure 3-7: Network code for Example 3.3

Definition 3.8 (Convolutional code) Given a source Z_V with a dependency hypergraph $H = H_1 \sqcup H_2$ in Definition 3.4, the active users in A with $2 < |A| \leq| V|$ can share a secret key as follows.

1. Extend the source over $n \in \mathbb{P}$ time units.
2. Pick a source node $s \in A$.
3. Let $\tilde{H} = (V, \tilde{E}, \tilde{\phi}) = \bigsqcup_{i \in [n]} H$ be the n-extended hypergraph of H. Use the edges in $\tilde{H}_1$ as undirected broadcast links and the edges in $\tilde{H}_2$ as selectable links defined in Definition 3.6. i.e. choose a sender for each edge in $\tilde{H}_1$ and a receiver for each edge in $\tilde{H}_2$.
4. Have the source s generate a secret key K and multicast it to all active users in A using a convolutional network code [34].

If $\mathbb{G}$ is not a field, we use the field with maximum order less than q. □

Instead of encoding a long message over a large block of time, convolution code encodes a data stream continuously in time, allowing mixing of the data in the transmitted and received signals. Although it may not work well with the more general finite linear source in the previous section, it applies naturally here for the special source model with dependency hypergraph. The continuous encoding mode has a better delay guarantee and less demand for buffering. The achievable key rate and other details of the convolutional network code are given in the following theorem.[10]

[10] See [28, 34] for more details on convolutional network codes.

Theorem 3.3 *Given a source Z_V with a dependency hypergraph $H = H_1 \sqcup H_2$ in Definition 3.4, the convolutional network coding scheme in Definition 3.8 achieves asymptotically the secret key rate $C^A_{\text{sn}}(\log q)$ bits in* (3.3)*, which simplifies to*

$$C^A_{\text{sn}} = \max_{P_{\mathsf{H}_1^*}, P_{\mathsf{H}_2^*}} \min_{B \subseteq V : s \in B \not\supseteq A} \mathrm{E}\left[|\delta^{+}_{\mathsf{H}_1^*}(B)| + |\delta^{-}_{\mathsf{H}_2^*}(B)|\right] \tag{3.9a}$$

$$= \max_{P_{\mathsf{H}_1^*}, P_{\mathsf{H}_2^*}} \min_{\mathcal{P} \in \Pi(A)} \frac{\sum_{C \in \mathcal{P}} \mathrm{E}\left[|\delta^{-}_{\mathsf{H}_1^*}(C)| + |\delta^{+}_{\mathsf{H}_2^*}(C)|\right]}{|\mathcal{P}| - 1} \tag{3.9b}$$

where H_1^* and H_2^* are random star hypergraphs of H_1 and H_2 respectively with random root functions. This is independent of the choice of $s \in A$. The maximum delay μ is upper bounded by

$$\mu \leq n^3 C^A_{\text{sn}} |E| \log_q(|A| n C^A_{\text{sn}} q) \tag{3.10}$$

where n is the time-extension in step 1 *of Definition 3.8. When all users are active, i.e. $A = V$, the secrecy capacity in* (3.4) *is attained as a consequence of Corollary 3.1, which is*

$$C^V_{\text{s}} = \min_{\mathcal{P} \in \Pi} \frac{\sum_{C \in \mathcal{P}} \left[|\delta^{-}_{H_1^*}(C)| + |\delta^{+}_{H_2^*}(C)|\right]}{|\mathcal{P}| - 1}$$

where H_1^* and H_2^* are arbitrary star hypergraphs of H_1 and H_2 respectively. Furthermore, $n \leq |V| - 1$ by (B.33)*, which can be substituted in* (3.10) *to give a bound on delay.* □

C^V_{s} equals p_H defined in (B.51), which has the combinatorial interpretation of partition connectivity for hypergraphs [3]. The optimal partition $\mathcal{P}$ also has the intuitive meaning of highly connected/dependent nodes as described in Proposition B.6.

Proof The fact that edges in H_2 can be used as selectable links instead of undirected interference links follows from Corollary B.2 that edges in H_2 can be shrunk to 2-edges without diminishing the min-cut value in (3.9). The fact that (3.3) evaluates to (3.9) under (3.6) follows from the fact that

$$r(Z_{B^c}|X_{B^c}) = \frac{H(\mathsf{Z}_{B^c}|\mathsf{X}_{B^c})}{\log q} = |\delta^{+}_{H_1^*}(B)| + |\delta^{-}_{H_2^*}(B)| \qquad \text{for any } B \subseteq V$$

It remains to show that there is a convolutional code that attains the throughput in (3.9) with a maximum delay of (3.10). After step 3 in Definition 3.9, every edge in $\tilde{H}_1$ is a directed broadcast link while every edge in $\tilde{H}_2$ is a point-to-point link, which is just a special kind of broadcast link. Thus, step 4 is essentially a network coding problem with directed broadcast links. Without loss of generality, we assume H_2 is empty and construct the desired convolutional code for $H = H_1$ that attains the rate

$$d_{\tilde{H}^*}(A, s) := \min_{B\subseteq V: s\in B\not\supseteq A} |\delta^+_{\tilde{H}^*}(B)| \tag{3.11}$$

Let t be the time index, D be an indeterminate for time delay, u be the total number of input processes, and $\mathbb{F}_{q^k}$ for some positive integer $k \in \mathbb{P}$ be the support set for each sample. Let the generating function of the input process $j \in [u]$ be

$$\mathsf{X}_j(D) := \sum_{t\geq 0} \mathsf{X}_{jt} D^t$$

where X_{jt} are independent and uniformly random over $\mathbb{F}_{q^k}$. $(\mathsf{X}_{jt} : t \in \mathbb{P})$ is the j-th uniformly random data stream originated from the source s and to be communicated to every node in $A \setminus \{s\}$. The generating function of the edge process at $e \in \tilde{E}$ is defined as

$$\mathsf{Y}_e(D) = \sum_{\substack{e'\in\tilde{E}:\\ \tilde{\rho}(e)\in\tilde{\phi}(e')\setminus\tilde{\rho}(e')}} a_{ee'} D \mathsf{Y}_{e'}(D) + \sum_{\substack{j\in[u]:\\ \tilde{\rho}(e)=s}} b_{ej} \mathsf{X}_j(D)$$

where $a_{ee'}, b_{ej} \in \mathbb{F}_{q^k}$. This is generated by node $\tilde{\rho}(e)$ and received by nodes in $\tilde{\phi}(e)$. The additional unit delay on the incoming edge processes is sufficient (but not necessary[11]) to avoid cyclic dependency in the information flow when the line graph of $\tilde{H}^*$ contains a cycle. Define the generating function of the output process $j \in [u]$ at node $i \in A \setminus \{s\}$ as

$$\mathsf{Z}_{ij}(D) = \sum_{\substack{e\in\tilde{E}:\\ i\in\tilde{\phi}(e)\setminus\tilde{\rho}(e)}} c_{ije} \mathsf{Y}_e(D)$$

where $c_{ije} \in \mathbb{F}_{q^k}$. This can be regarded as a summary of the incoming edge processes

[11] As mentioned in [28], it is sufficient to have one unit delay associated with every cycle in the line digraph of $\tilde{H}^*$.

at node i for decoding the input processes. To define the final decoding step, consider the following matrix notation. Let

$$\begin{aligned}
\boldsymbol{a} &:= (a_{ee'} : e, e' \in \tilde{E}, \tilde{\rho}(e) \in \tilde{\phi}(e') \setminus \tilde{\rho}(e')) \\
\boldsymbol{b} &:= (b_{ej} : e \in \tilde{E}, j \in [u], \tilde{\rho}(e) = s) \\
\boldsymbol{c}_i &:= (c_{ije} : j \in [u], e \in \tilde{E}, \tilde{\phi}(e) \setminus \tilde{\rho}(e) \ni i)
\end{aligned}$$

and $\boldsymbol{A} = [A_{ee'}]$, $\boldsymbol{B} = [B_{ej}]$ and $\boldsymbol{C}_i = [C_{ije}]$ be the matrices with row index first such that the entries $A_{ee'}$, B_{ej} and C_{ije} equal $a_{ee'}$, b_{ej}, and c_{ije} respectively if defined and 0 otherwise. Then, with $\mathbf{X}(D) = (\mathsf{X}_j(D))$, $\mathbf{Y}(D) = (\mathsf{Y}_e(D))$ and $\mathbf{Z}_i(D) = (\mathsf{Z}_{ij}(D))$ defined as the vectors of the specified input, edge and output processes, we have the matrix form of the convolutional code

$$\begin{aligned}
\mathbf{Y}(D) &= D\boldsymbol{A}\mathbf{Y}(D) + \boldsymbol{B}\mathbf{X}(D) \\
\mathbf{Z}_i(D) &= \boldsymbol{C}_i\mathbf{Y}(D) && \text{for all } i \in A \setminus \{s\}
\end{aligned}$$

Combining these equations, we have

$$\mathbf{Z}_i(D) = \overbrace{\boldsymbol{C}_i(\boldsymbol{I} - D\boldsymbol{A})^{-1}\boldsymbol{B}}^{\boldsymbol{H}_i}\mathbf{X}(D) \qquad \text{for all } i \in A \setminus \{s\} \tag{3.12}$$

The matrix inverse $(\boldsymbol{I} - D\boldsymbol{A})^{-1}$ is well-defined because the system $\boldsymbol{H}_i$ is realizable in a distributed fashion by construction.[12] The determinant $|\boldsymbol{I} - D\boldsymbol{A}|$ is a non-zero polynomial of D with constant term equal to $|\boldsymbol{I} - 0\boldsymbol{A}| = 1$ and so

$$(\boldsymbol{I} - D\boldsymbol{A})^{-1} = \frac{1}{|\boldsymbol{I} - D\boldsymbol{A}|}\operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A}) \tag{3.13}$$

where $\operatorname{adj}(\boldsymbol{M})$ is the *adjugate matrix*, the entry at row r and column c of which is $(-1)^{r+c}$ times the determinant of $\boldsymbol{M}$ with row c and column r removed.

Finally, define the decoding at node $i \in A \setminus \{s\}$ using the system $\hat{\mathbf{C}}_i$ as

$$\mathbf{X}(D)D^{\mu_i} = \hat{\mathbf{C}}_i\mathbf{Z}_i(D)$$

[12] Every processing step satisfies the causality requirement under the assumption that the nodes process information independently.

which returns the input processes with delay μ_i. This is feasible if

$$\hat{\mathsf{C}}_i := \mathsf{C}_i \boldsymbol{H}_i^{-1} D^{\mu_i}$$

is realizable, i.e. $\boldsymbol{H}_i^{-1}$ is well-defined and every entry viewed as a rational function of D has at most μ_i poles at 0.

We first show that $\boldsymbol{H}_i^{-1}$ is well-defined for all $i \in A \setminus \{s\}$ if $d_{\tilde{H}^*}(A, s) \geq u$ in (3.11) and k is chosen sufficiently large. Then, we upper bound k and the delay $\max_{i \in A \setminus \{s\}} \mu_i$. Since $\boldsymbol{H}_i^{-1} := \operatorname{adj}(\boldsymbol{H}_i)/|\boldsymbol{H}_i|$, it is well defined iff $|\boldsymbol{H}_i|$ is a non-zero rational function of D. By (3.12) and (3.13),[13]

$$|\boldsymbol{H}_i| = \frac{\overbrace{|\boldsymbol{C}_i \operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A})\boldsymbol{B}|}^{\xi_i}}{|\boldsymbol{I} - D\boldsymbol{A}|^u} \tag{3.14}$$

Thus, it suffices to show that

$$\xi := \prod_{i \in A \setminus \{s\}} \xi_i$$

is a non-zero polynomial of D for some choice of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{F}_{q^k}$ and $k \in \mathbb{P}$.

By (3.11), $d_{\tilde{H}^*}(A, s) \geq u$ implies that $\min_{B \subseteq V : s \in B \not\ni i} |\delta_{\tilde{H}^*}(B)| \geq u$ for all $i \in A \setminus \{s\}$, which implies under Theorem B.3 or the extension [3, Theorem 4.1] of the Menger's theorem that there exists u edge-disjoint unit flows from s to i. This, in turn, implies that $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ can be chosen for each i such that ξ_i is a non-zero polynomial $\xi_i(D)$ of D. Thus, ξ is a non-zero polynomial $\xi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, D)$. Choose k such that q^k is larger than the maximum degree $\deg(\xi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}))$ of ξ by viewing ξ as a polynomial of each of the variables in $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ but D as some constant in $\mathbb{F}_{q^k}$. i.e.

$$k = \lfloor \log_q(\deg \xi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})) \rfloor + 1 \tag{3.15}$$

It follows by an inductive argument as in [28, Lemma 2.1] that ξ is a non-zero polynomial of D for some choice of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. More precisely, arrange the variables $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, D)$

[13] We have also used the simple fact that $|\alpha \boldsymbol{M}| = \alpha^u |\boldsymbol{M}|$ for any u-by-u matrix $\boldsymbol{M}$.

in a sequence $(x_1, \ldots, x_l = D)$. The polynomial $\xi(x_2, \ldots, x_l)$ is non-zero by some choice of $x_1 \in \mathbb{F}_{q^k}$ since each coefficient is a polynomial of x_1 with degree strictly smaller than q^k, and therefore cannot have all elements in $\mathbb{F}_{q^k}$ as roots. Similarly, given $\xi(x_i, \ldots, x_l)$ is a non-zero polynomial for some choice of $x_{[i-1]}$, we have a choice of x_i such that $\xi(x_{i+1}, \ldots, x_l)$ is a non-zero polynomial. By induction, there is a choice of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ such that $\xi(D)$ is a non-zero polynomial, and so $\boldsymbol{H}_i^{-1}$ is well-defined for all $i \in A \setminus \{s\}$ as desired.

To upper bound k, we have

$$\begin{aligned} \deg(\xi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})) &\le (|A| - 1) \max_{i \in V \setminus \{s\}} \deg(\xi_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})) \\ &= (|A| - 1)u \end{aligned}$$

where the last equality is because the entries of $\boldsymbol{C}_i \operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A})\boldsymbol{B}$ is a degree-1 polynomial in $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$, and so its determinant, namely $\xi_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, has degree at most u. It follows from (3.15) that

$$k \le \log_q |A| uq \tag{3.16}$$

To upper bound μ_i, denote $n_{\mathrm{Z},0}(\boldsymbol{M})$ and $n_{\mathrm{P},0}(\boldsymbol{M})$ as the maximum numbers of zeros and respectively poles at 0 over the entries of matrix $\boldsymbol{M}$, viewed as rational functions of D. Then,

$$\begin{aligned} \mu_i := n_{\mathrm{P},0}(\boldsymbol{H}_i^{-1}) &\le n_{\mathrm{Z},0}(|\boldsymbol{H}_i|) + \overbrace{n_{\mathrm{P},0}(\operatorname{adj}(\boldsymbol{H}_i))}^{\overset{(a)}{=} 0} \\ &\overset{(b)}{=} n_{\mathrm{Z},0}(\overbrace{|\boldsymbol{C}_i \operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A})\boldsymbol{B}|}^{\xi_i}) \\ &\le \deg(\xi_i(D)) \overset{(c)}{=} u(|\tilde{E}| - 1) \end{aligned}$$

(a) is because $n_{\mathrm{P},0}(\operatorname{adj}(\boldsymbol{H}_i)) \le (u-1) n_{\mathrm{P},0}(\boldsymbol{H}_i) = (u-1) n_{\mathrm{Z},0}(|\boldsymbol{I} - D\boldsymbol{A}|) = 0$ since $|\boldsymbol{I} - D\boldsymbol{A}|$ is a polynomial of D with no zero;[14] (b) is by (3.14); (c) is because every entry of $\operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A})$ is a polynomial of D with degree at most $|\tilde{E}| - 1$, so as the

[14] The factor $(u-1)$ comes from the definition of $\operatorname{adj}(\boldsymbol{H}_i)$ that every entry is a determinant of a matrix with dimension $(u-1)$.

entries of the linear combination $\boldsymbol{C}_i \operatorname{adj}(\boldsymbol{I} - D\boldsymbol{A})\boldsymbol{B}$.[15] Since $|\tilde{E}| = n|E|$, we have

$$\max_{i \in A \setminus \{s\}} \mu_i \leq u(n|E| - 1) \tag{3.17}$$

This completes the proof since we have a convolutional code that achieves (3.11) asymptotically with a finite field extension (3.16) and delay (3.17). The overall delay is upper bounded as follows,

$$\max_{i \in A \setminus \{s\}} nk(\mu_i + 1) \leq n^3 C_{\text{sn}}^A |E| \log_q |A| n C_{\text{sn}}^A q$$

since n extension is needed to convert H to $\tilde{H}$, an additional factor of k is needed for the field extension, and a delay of $\mu_i + 1$ frames of nk time units can guarantee that the input process can be decoded by every node $i \in A \setminus \{s\}$. The R.H.S. is obtained by (3.16), (3.17) and the inequalities $nC_{\text{sn}}^A \geq d_{\tilde{H}^*}(A, s) \geq u$. ■

When all users are active, and H_1 is empty or consists of only 2-edges, we can have an even better delay guarantee with the following routing solution.

Definition 3.9 (Routing) Given a source Z_V with a dependency hypergraph $H = H_2$ in Definition 3.4, i.e. H_1 being empty, all users in $A = V$ can share a secret key as follows.

1. Extend the source over $n \in \mathbb{P}$ time units.
2. Pick a source node $s \in A$.
3. Let $\tilde{H} = (V, \tilde{E}, \tilde{\phi}) = \bigsqcup_{i \in [n]} H$ be the n-extended hypergraph of H. Use the edges in $\tilde{H}$ as selectable links.
4. Decompose the links into edge-disjoint outbranchings from s to $V \setminus \{s\}$. Have the source s generate and route independent parts of a secret key K through each unit flow to all other active users.

See Definition 3.6 for definitions of links and outbranchings. □

[15] The additional factor u is because of taking the determinant.

Theorem 3.4 *Given a source Z_V with a dependency hypergraph $H = H_2$ in Definition 3.4 and $A = V$, the routing solution in Definition 3.9 achieves the secrecy capacity $C_{\mathrm{s}}^V(\log q)$ bits with*

$$C_{\mathrm{s}}^V = \max_{P_{\mathsf{H}^*}} \min_{B\subseteq V: s\in B\not\supseteq V} \mathrm{E}\left[|\delta^-_{\mathsf{H}^*}(B)|\right] \tag{3.18a}$$

$$= \min_{\mathcal{P}\in\Pi(A)} \frac{\sum_{C\in\mathcal{P}}|\delta^+_{H^*}(C)|}{|\mathcal{P}|-1} \tag{3.18b}$$

where H^* is a random star hypergraph of H with random root functions, while H^* is an arbitrary deterministic star hypergraph. This is independent of $s \in V$ and H^*, and is attained with delay (namely the extension n) at most $|V| - 1$. □

Proof (3.18) is the maximum number of unit flows by Edmond's branching theorem [3, Theorem 4.2]. The facts that (3.18) is the secrecy capacity and $n \leq |V| - 1$ follow from the same argument as in the proof of Theorem 3.3. ■

3.3 Potential of multisource network coding

In summary, we have shown that the secret key agreement problem has a practical solution under a linear source model. With the help of public discussion, the private source can be converted effectively into a private channel by selecting a base for the channel inputs. One active user can be designated as the source node to generate a random key, which can then be communicated perfectly secretly to other active users by a linear network multicast. Compared to the secret key agreement by communication for omniscience in [12], this is more practical since perfect secrecy and recoverability can be achieved without requiring the constraint length to go to infinity. In addition, all the processing involves only linear operations whose complexity is polynomial in the constraint length, for any network with constant size.

When some users need not share the secret key, we do not have a proof that the network coding approach is optimal, nor do we have a counter-example that suggests otherwise. Further investigation is also needed for a practical solution that scales with the size of the network. With a linear source model, however, it is quite reasonable to think that a linear coding scheme suffices. It may require a more complicated scheme such as the multisource network coding: have multiple sources generate and communicate independent secret keys to others through multiple multicast sessions. While further research is needed for a more concrete statement, we will consider Example 2.2 in Section 2.2 again to conveys the potential of this approach.

Recall that $V = [6]$, $A = [3]$, and Z_V is a finite linear source with Z_4, Z_5 and Z_6 independent and uniformly distributed over $\mathbb{F}_2$ while $\mathsf{Z}_1 = \mathsf{Z}_5 \oplus \mathsf{Z}_6$, $\mathsf{Z}_2 = \mathsf{Z}_4 \oplus \mathsf{Z}_6$ and $\mathsf{Z}_3 = \mathsf{Z}_4 \oplus \mathsf{Z}_5$. By Proposition 2.1, the secrecy capacity is $\frac{3}{4}$ bits. This can be attained by the following multisource network code with a delay of 4 time units, as illustrated in Figure 3-8.

Time 1: The source is first converted to a channel with inputs $(\mathsf{X}_1, \mathsf{X}_3, \mathsf{X}_6) := (\mathsf{Z}_1, \mathsf{Z}_3, \mathsf{Z}_6)$. Then, have user 1 acts as a source to generate a uniformly random key bit K_1. Assigning the inputs as $(\mathsf{X}_1, \mathsf{X}_3, \mathsf{X}_6) \leftarrow (\mathsf{K}_1, 0, 0)$, the key bit K_1 is then communicated to all users perfectly secretly except 3 and 6.

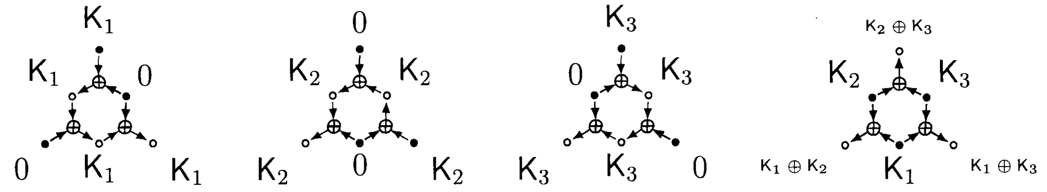
Time 2: The inputs are chosen as $(\mathsf{X}_1, \mathsf{X}_2, \mathsf{X}_4) := (\mathsf{Z}_1, \mathsf{Z}_2, \mathsf{Z}_4)$ and assigned the values $(0, \mathsf{K}_2, 0)$ after user 2 acts as a source to generate a uniformly random key bit K_2 independent of K_1. Everyone except 1 and 4 learns K_2.

Time 3: The inputs are chosen as $(\mathsf{X}_1, \mathsf{X}_2, \mathsf{X}_5) := (\mathsf{Z}_1, \mathsf{Z}_2, \mathsf{Z}_5)$ and assigned the values $(\mathsf{K}_3, 0, 0)$ after user 1 generates a uniformly random key bit K_3 independent of $(\mathsf{K}_1, \mathsf{K}_2)$. Everyone except 2 and 5 learns K_3.

Time 4: The inputs are chosen as $(\mathsf{X}_4, \mathsf{X}_5, \mathsf{X}_6) := (\mathsf{Z}_4, \mathsf{Z}_5, \mathsf{Z}_6)$ and assigned the values $(\mathsf{K}_1, \mathsf{K}_2, \mathsf{K}_3)$. User 1 observes $\mathsf{K}_2 \oplus \mathsf{K}_3$ from the channel and recover K_2 using his knowledge of K_3. Similarly, user 2 and 3 can recover K_3 and K_1 respectively.

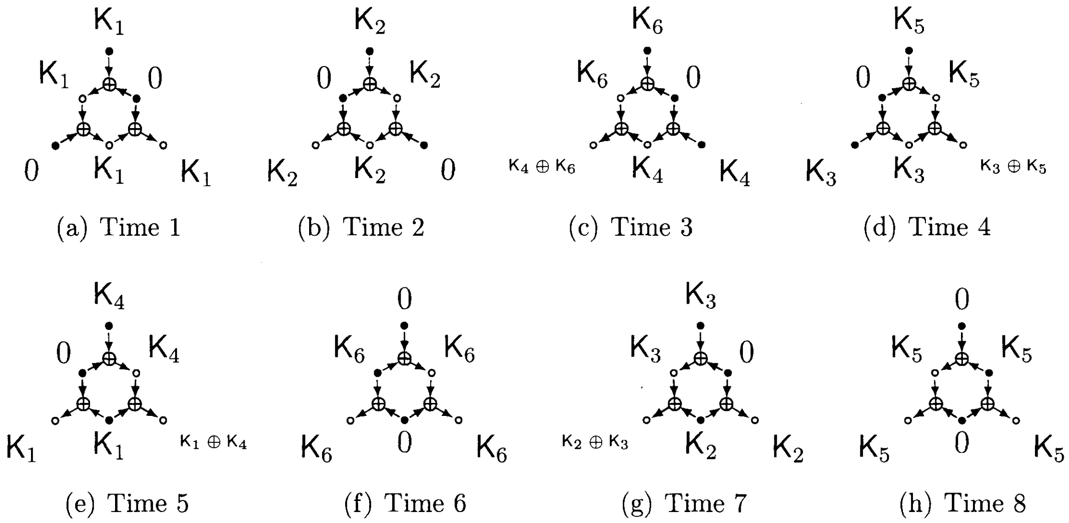
Since everyuser obtain 3 bits of secret key with 4 uses of the private source, the key rate is $\frac{3}{4}$ as desired.

Indeed, the secret key capacity can also be attained by the single-source network coding with a delay of 8 time units. User 1 is the source node and generates all the independent key bits $\mathsf{K}_1, \dots, \mathsf{K}_6$. The other active users (user 2 and 3) can recover the key bits after eight channel uses shown in Figure 3-9.



(a) Time 1 (b) Time 2 (c) Time 3 (d) Time 4

Figure 3-8: Multisource network code for Example 2.2



(a) Time 1 (b) Time 2 (c) Time 3 (d) Time 4

(e) Time 5 (f) Time 6 (g) Time 7 (h) Time 8

Figure 3-9: Single-source network code for Example 2.2

User 2: K_1, K_4, K_6, K_2 and K_5 are recovered at Time 1, 3, 6, 7 and 8 respectively. K_3 can be recovered from $K_3 \oplus K_5$ observed at Time 4.

User 3: K_2, K_3, K_1, K_6 and K_5 are recovered at Time 2, 4, 5, 6 and 8 respectively. K_4 can be recovered from $K_4 \oplus K_6$ observed at Time 4.

Multisource network coding gives more flexibility in optimizing the orientation of the channel inputs and the choice of the sources for the secret key bits. It is unclear whether the benefit can go beyond the delay performance to strictly improve the secret key rate, and whether the key rate reaches the secrecy capacity in the presence of helpers. Perhaps the secrecy capacity may also help characterize the solution of the multisource network coding problem in certain special case.

Part II

Multiterminal Secret Key Agreement

Chapter 4

General Multiterminal Network

Consider the secret agreement game described in Section 1.1. If the users have some time to whisper quietly to their neighboring users, how can they cooperate to enhance their correlation so that they can generate as much common secret as possible? The source model of [12] considered in Part I fails to address this because the correlation of the users is already given as it is by the private sources. The users do not take part in generating or enhancing their correlation; they simply turns it into a common secret.

A less passive model is given by Csiszár and Narayan [13]. They proposed a broadcast channel model where one user controls the input to a discrete memoryless private channel (DMC) while others observe from it. The optimal way to correlate the channel input and outputs turns out to be very simple: have the sender generate independent and identically distributed (iid) input symbols to turn the private channel into a DMMS for secret key agreement. The secrecy capacity for the channel model is simply the secrecy capacity of the source model maximized over the input distribution. There is essentially no cooperation needed in generating the correlation.

In contrast with the previous work [12, 13], we consider in Part II of this monograph a general private discrete memoryless multiterminal channel (DMMC) in which every terminal can both transmit and receive. It covers the source model in [12] as a special case when the terminals may receive but not transmit. It also covers the broadcast channel model in [13] as a special case when there is at most one transmitting terminal, which is not allowed to receive. The more general DMMC allows us to address the following new questions:

1. How should the transmitting terminals coordinate their channel inputs to enhance the correlation of their private observations?
2. How can the terminals which can both receive and transmit enhance the correlation by adapting their channel inputs to the channel outputs?

The proposed model will also cover continuous channels, which arise in practical scenarios such as wireless communication. More generally, we allow every channel input and output symbol to be a mixture of discrete and continuous random variables [45]. The channel input sequences may also be subject to additional constraints such as the average power constraint, which applies especially for mobile terminals with limited power in wireless communication.

4.1 Channel model

We first give an overview of the main ideas through a simplified model. As described in Chapter 2, there is a set V of users who want to generate and share a common secret key. They have access to a private channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ instead of a source P_{Z_V}. Each user $i \in V$ controls the input X_i and observes the output Y_i. They publicly discuss by broadcasting public messages at any time based on their accumulated knowledge, which includes their randomization, private observations from the channel, and the previous public messages. They can choose their channel inputs as functions of their accumulated knowledge as well. After n channel uses and some public discussion, each active users $i \in A \subseteq V$ generates a key K_i that equals almost surely to some common secret key K, which has to be independent of the public messages asymptotically as $n \to \infty$. The secrecy capacity is the maximum achievable key rate as usual.

4.2 Bounds on secrecy capacity

Recall from (2.1) that the secrecy capacity for the source model Z_V is

$$C_{\mathrm{s}} = \min_{\lambda} \overbrace{\left[H(\mathsf{Z}_V) - \sum_{B} \lambda_B H(\mathsf{Z}_B|\mathsf{Z}_{B^c}) \right]}^{\beta(\lambda, P_{\mathsf{Z}_V}) :=}$$

where λ is the fractional partition. If given a channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ instead, the users can emulate a source by sending iid inputs. The secret key rate achievable with the resulting DMMS Z_V with $\mathsf{Z}_i = (\mathsf{X}_i, \mathsf{Y}_i)$ for $i \in V$ is

$$\begin{aligned} C_{\mathrm{pse}} &= \max_{P_{\mathsf{X}_V} = \prod_{i \in V} P_{\mathsf{X}_i}} \min_{\lambda} \beta(\lambda, P_{\mathsf{Z}_V}) \\ &= \max_{P_{\mathsf{X}_1 \mathsf{X}_2} = P_{\mathsf{X}_1} P_{\mathsf{X}_2}} I(\mathsf{X}_1 \mathsf{Y}_1 \wedge \mathsf{X}_2 \mathsf{Y}_2) \qquad \text{if } V = A = [2] \end{aligned}$$

where the last expression is for the two-user case. This gives a lower bound on the secrecy capacity for the channel model. An upper bound can be obtained using some general properties of information as follows,

$$\begin{aligned} C_{\mathrm{su}} &= \min_{\lambda} \max_{P_{\mathsf{X}_V} \in \mathscr{P}(X_V)} [\beta(\lambda, P_{\mathsf{Z}_V}) - \beta(\lambda, P_{\mathsf{X}_V})] \\ &= \max_{P_{\mathsf{X}_1 \mathsf{X}_2}} [I(\mathsf{X}_1 \mathsf{Y}_1 \wedge \mathsf{X}_2 \mathsf{Y}_2) - I(\mathsf{X}_1 \wedge \mathsf{X}_2)] \qquad \text{if } V = A = [2] \end{aligned}$$

The maximization in the upper bound is over the simplex $\mathscr{P}(X_V)$ of all input distributions over the input alphabet set X_V. Compared to the lower bound, it has the additional term $-\beta(\lambda, P_{\mathsf{X}_V})$ instead of the independence condition $\prod_{i \in V} P_{\mathsf{X}_i}$ to discount the correlation among the inputs. Since the independence condition implies that the additional term is 0, the upper bound is clearly no smaller than the lower bound. Furthermore, the order of the maximization and minimization is different for the upper and lower bounds. This can make the upper bound even larger in the case with more than two users because the input distribution can be chosen as a function of λ.

The gap between the upper and lower bounds would be reduced if we can show that the lower bound remains achievable even with the order of the maximization and minimization reversed. This is indeed possible, borrowing the idea of mixed strategy from game theory. Consider a virtual zero-sum game where $\beta(\lambda, P_{\mathsf{Z}_V})$ is the payoff to the player who chooses the action $P_{\mathsf{X}_V} = \prod_{i \in V} P_{\mathsf{X}_i}$ and the cost to the opponent who chooses the action λ. With a pure strategy where the player chooses the best action with probability one, the guaranteed payoff is C_{pse}. However, just like the game of paper-scissor-stone, mixing between different actions can potentially increase the expected payoff. The player can pick a valid action from some finite set $\{P_{\mathsf{X}_V|\mathsf{Q}}(\cdot|q) : q \in Q\}$ randomly according to some distribution P_{Q}. The expected payoff is

$$\begin{aligned} C_{\text{mse}} &= \max_{P_{\mathsf{Q}}, P_{\mathsf{X}_V|\mathsf{Q}} = \prod_{i \in V} P_{\mathsf{X}_i|\mathsf{Q}}} \min_{\lambda} \mathrm{E}_{\mathsf{Q}} \left[\beta(\lambda, P_{\mathsf{Z}_V|Q}) \right] \\ &= \min_{\lambda} \max_{P_{\mathsf{X}_V} = \prod_{i \in V} P_{\mathsf{X}_i}} \beta(\lambda, P_{\mathsf{Z}_V}) \end{aligned}$$

where the last equality is by the minimax-type lemma in game theory. This is indeed an achievable key rate because we can have one user generate Q^n iid as P_{Q} in public and have each user $i \in V$ generate the channel input X_{it} iid over time $t \in [n]$ as $P_{\mathsf{X}_i}(\cdot|\mathsf{Q}_t)$ based on Q^n.

4.3 Tightness condition

With some simple algebra, the improved lower bound can be expressed in a form similar to the upper bound

$$\begin{aligned} C_{\text{mse}} &= \min_{\lambda} \max_{P_{\mathsf{X}_V} = \prod_{i \in V} P_{\mathsf{X}_i}} \alpha(\lambda, P_{\mathsf{Z}_V}) \\ C_{\text{su}} &= \min_{\lambda} \max_{P_{\mathsf{X}_V} \in \mathscr{P}(X_V)} \alpha(\lambda, P_{\mathsf{Z}_V}) \end{aligned}$$

where

$$\begin{aligned} \alpha(\lambda, P_{\mathsf{Z}_V}) &:= \beta(\lambda, P_{\mathsf{Z}_V}) - \beta(\lambda, P_{\mathsf{X}_V}) \\ &= \sum_B \lambda_B H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - \left(\sum_B \lambda_B - 1 \right) H(\mathsf{Y}_V|\mathsf{X}_V) \end{aligned}$$

This gives a simple condition for tightness: $C_{\text{mse}} = C_{\text{su}}$ if $\alpha(\lambda, P_{\mathsf{Z}_V})$ is maximized by some product distribution $\prod_{i\in V} P_{\mathsf{X}_i}$ (possibly as a function of λ). For example, the broadcast channel model $P_{\mathsf{Y}_{V\setminus\{1\}}|\mathsf{X}_1}$ in [13] satisfies this trivially. A wide range of channels with multiple inputs also satisfies the condition, including the interference free channel $\prod_{\ell\in V} P_{\mathsf{Y}_{\ell V}|\mathsf{X}_\ell}$ which consists of a set of independent broadcast channels with inputs from different users, and the finite linear network in which the channel outputs are linear combinations of the input in some finite field as described in Section 3.1.

In the subsequent chapters, we will describe the model, the bounds on secrecy capacity, and the tightness results in greater details. The general model is formulated in Chapter 5, the upper and lower bounds on the secrecy capacity under various model assumptions are derived in Chapter 6, and the tightness of the bounds is analyzed in Chapter 7.

Chapter 5

Secret Key Agreement Protocol

The model consists of a wiretapper and a set of users/terminals, which are categorized as active terminals, untrusted terminals and helpers.

Terminals: Let $V := [m]$ be the set of all m terminals, denoted by T_i for $i \in V$. The terminals are further partitioned into the following disjoint sets.

Active terminals: Let $A \subseteq V : |A| \geq 2$ denote the set of (at least two) active terminals who want to share a secret key.

Untrusted terminals: Let $D \subseteq A^c$ be a (possibly empty) set of untrusted terminals. They are untrusted in the sense that their knowledge can be revealed to the wiretapper. They, however, must follow the protocol to help generate secret.

Helpers: the remaining (possibly empty) subset $V \setminus (A \cup D)$ of terminals are called helpers. They are trusted but need not share a secret key.

Wiretapper: it is a malevolent entity who attempts to learn the key from the public information and any knowledge of the untrusted terminals. However, it is not allowed to inject any fraudulent messages nor intercept any private observations of the trusted terminals.

The terminals follow the protocol outlined in Figure 5-1 to generate a secret key for the active terminals. It is divided into three main phases: 1) randomization, 2) transmission, and 3) key generation. The terminals randomize in the randomiza-

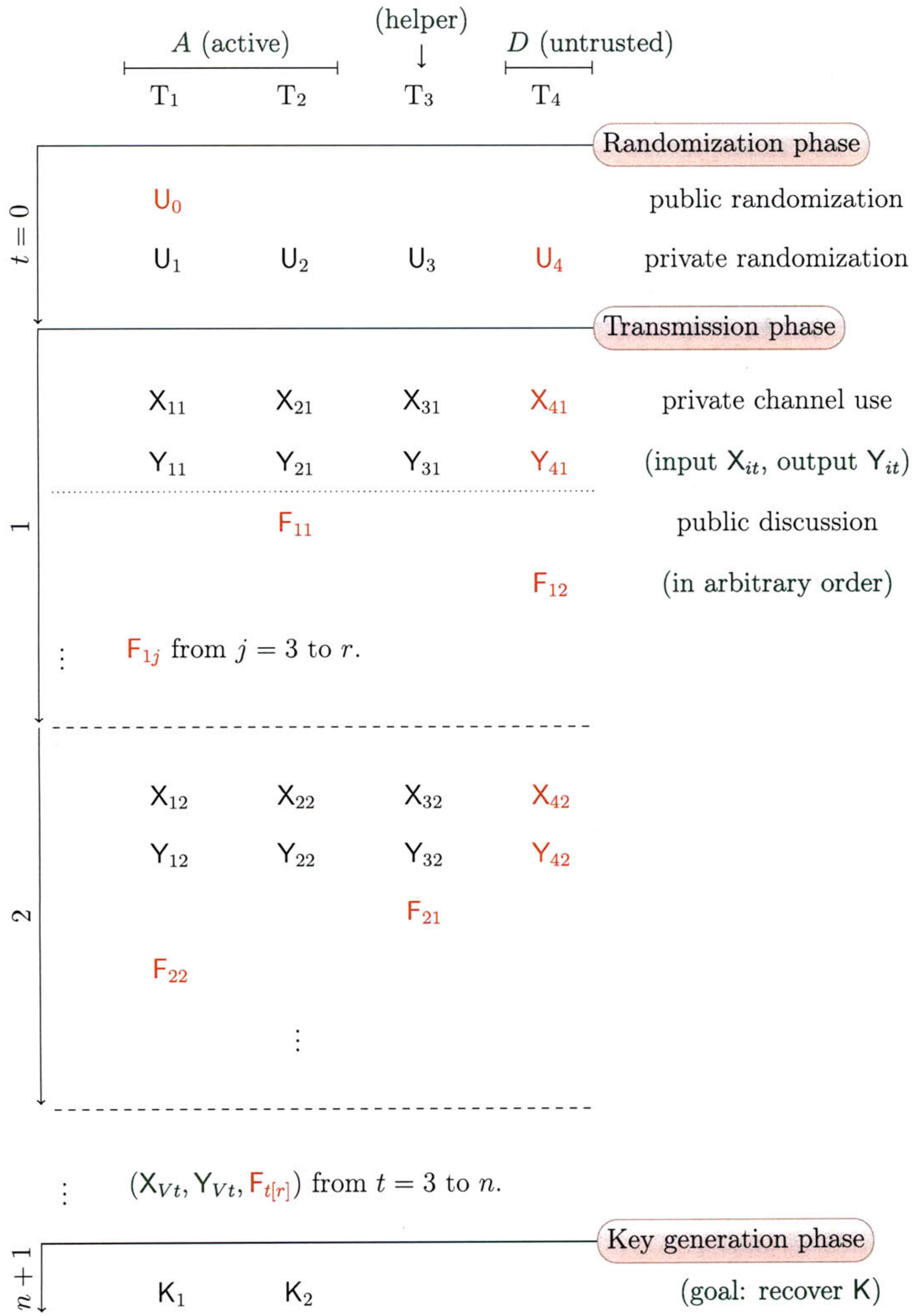


Figure 5-1: Timeline for the multiterminal secrecy protocol: $A = [2]$, $D = \{4\}$ and $V = [3]$. Entries in red are observed directly by the wiretapper.

tion phase for the purpose of generating random channel input sequences and public messages later in the transmission phase. In the transmission phase, the terminals use a private DMMC n times, and publicly discuss after every channel use for an arbitrary finite number r of rounds to decide on the next channel input and to generate secret keys in the final key generation phase. In the key generation phase, each active terminal generates an individual secret key based on the accumulated knowledge of its public and private observations. The goal is to maximize the key rate (bits per private channel use) under two constraints: 1) the recoverability condition that the individual keys are the same with high probability, and 2) the secrecy condition that the keys are close to uniformly random even given the accumulated knowledge of the wiretapper. In the following, we will describe the precise mathematical formulation of the protocol first, and then justify it in Section 5.2.

At time $t = 0$, the terminals randomize by generating continuous random variables publicly and privately as shown in Figure 5-2.

Randomization phase:

Public randomization: At time $t = 0$, the terminals publicly randomize by agreeing on a public continuous random variable U_0, known also to the wiretapper. For definiteness, we can have terminal 1 generate U_0 without loss of generality.

Private randomization: every terminal $i \in V$ then generates privately a continuous random variable U_i based on the public randomization U_0. Since U_i's are privately generated, they are unknown to the wiretapper. The catch is that they have to be conditionally independent, i.e.

$$P_{\mathsf{U}_0\mathsf{U}_V} = P_{\mathsf{U}_0} \prod_{i \in V} P_{\mathsf{U}_i|\mathsf{U}_0} \tag{5.1}$$

The transmission phase follows. Each terminal can transmit and receive over the private DMMC, and then engage in a public discussion. There are n public discussion sessions interleaving with n private channel uses.

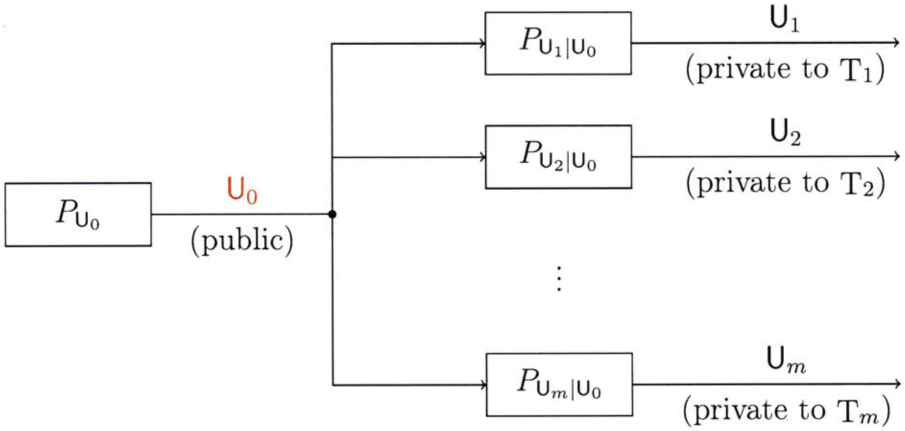

P_{U_0}
U_0
(public)
$P_{\mathsf{U}_1|\mathsf{U}_0}$
U_1
(private to T_1)
$P_{\mathsf{U}_2|\mathsf{U}_0}$
U_2
(private to T_2)
$\vdots$
$P_{\mathsf{U}_m|\mathsf{U}_0}$
U_m
(private to T_m)

Figure 5-2: Randomization

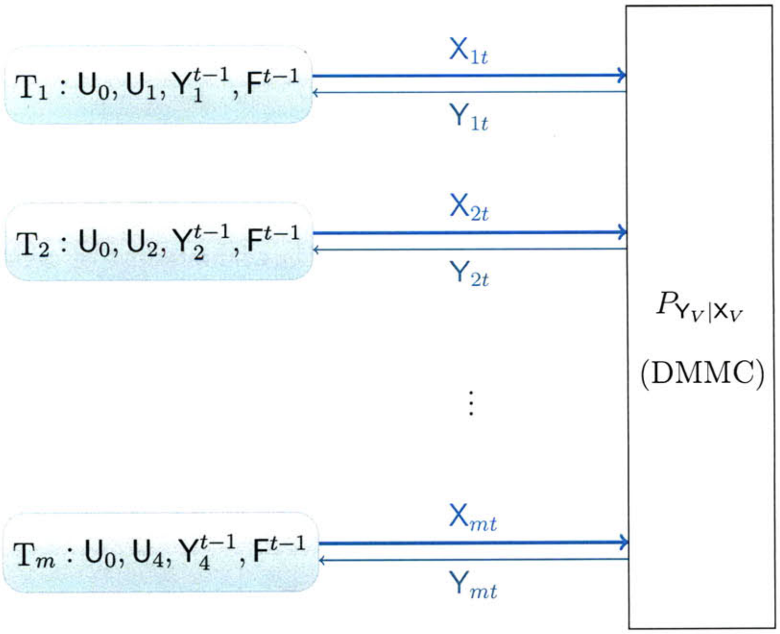

$\mathrm{T}_1 : \mathsf{U}_0, \mathsf{U}_1, \mathsf{Y}_1^{t-1}, \mathsf{F}^{t-1}$
X_{1t}
Y_{1t}
$\mathrm{T}_2 : \mathsf{U}_0, \mathsf{U}_2, \mathsf{Y}_2^{t-1}, \mathsf{F}^{t-1}$
X_{2t}
Y_{2t}
$\vdots$
$\mathrm{T}_m : \mathsf{U}_0, \mathsf{U}_4, \mathsf{Y}_4^{t-1}, \mathsf{F}^{t-1}$
X_{mt}
Y_{mt}
$P_{\mathsf{Y}_V|\mathsf{X}_V}$
(DMMC)

Figure 5-3: Private channel use: it is non-interactive in the sense the channel return the output symbols Y_{Vt} only after all the input symbols X_{Vt} are completely specified. The rounded boxes denote deterministic operation performed by T_i for $i \in V$ using the accumulated knowledge $(\mathsf{U}_0, \mathsf{U}_i, \mathsf{Y}_i^{t-1}, \mathsf{F}^{t-1})$.

Transmission phase – private channel use:

1. Terminal $i \in V$ chooses at every time $t \in [n]$ a *valid* input X_{it} as a function, denoted as $X_{it}(\cdots)$, of its accumulated knowledge. i.e.

$$\mathsf{X}_{it} = X_{it}(\mathsf{U}_0, \mathsf{U}_i, \mathsf{Y}_i^{t-1}, \mathsf{F}^{t-1}) \in X_i \tag{5.2}$$

   where $\mathsf{Y}_i^{t-1} := (\mathsf{Y}_{i\tau} : \tau \in [t-1])$ and $\mathsf{F}^{t-1} := (\mathsf{F}_\tau : \tau \in [t-1])$ are vectors of accumulated observations and public messages respectively.
2. After the channel input X_{Vt} is completely specified, the channel generates Y_{Vt} according to $P_{\mathsf{Y}_V|\mathsf{X}_V}$. It then returns Y_{it} privately to each terminal $i \in V$.

The channel input alphabet X_i needs not be finite. It may be infinitely-valued by having discrete components with unbounded support or continuous components with *absolutely continuous* probability measures. It may also be subject to the following type of constraint on the sample average.

Definition 5.1 (Inequality constraint on sample average) The *sample average constraint* is characterized by a finite set of functions $\phi_i : X_i \mapsto \mathbb{R}^\ell$ indexed by $i \in V$. It is satisfied by the channel input sequence X_V^n if for every n,

$$\Pr\left\{ \frac{1}{n} \sum_{t \in [n]} \phi_i(\mathsf{X}_{it}) \leq \delta_n \cdot \mathbf{1} \text{ for all } i \in V \right\} = 1 \tag{5.3}$$

for some $\delta_n \to 0$, where $\mathbf{1}$ is a vector of ℓ ones and $\leq$ denotes the elementwise inequality. This specializes to the usual *average power constraint* if ϕ_V gives certain powers of the real-valued components of the channel input. More precisely, let $\zeta_{ij}(\mathsf{X}_i)$ be the j-th real-valued component of the input symbol X_i. Then, (5.3) can be specialized to

$$\frac{1}{n} \sum_{t \in [n]} |\zeta_{ij}(\mathsf{X}_{it})|^{s_{ijk}} \leq c_{ijk} + \delta_n \qquad \text{(average power)} \tag{5.4}$$

with probability one for all (j, k) over some given finite set, $s_{ijk} \geq 1$ and $c_{ijk} \geq 0$. □

Note that the sample average constraint is particularly important for the infinitely-valued model, since the optimization over input distributions with unbounded support set may be ill-posed without such constraint.

We call X_{it} in (5.2) the *channel input function.*[1] It expresses the *causal relation* from the accumulated knowledge to the channel input. There is no benefit in further randomizing the input given the accumulated knowledge. i.e. we can have the channel input function deterministic rather than stochastic without loss of generality. This is because any additional randomization, say $\mathsf{U}_{\mathsf{Z}} \sim P_{\mathsf{U}|\mathsf{Z}}(\cdot|\mathsf{Z})$, generated based on the accumulated knowledge, say $\mathsf{Z} \in Z$, can be regarded as a function of a randomization, namely $(\mathsf{U}(z) : z \in Z)$, that can be generated without observing Z.[2]

Transmission phase – public discussion:

Right after every private channel use, the terminals engage in an *interactive authenticated public discussion.* Let r be the total number of rounds, and i_j be the terminal that speaks at the j-th round. At the j-th round of time t, terminal i_j generates and broadcasts the following public message noiselessly as a *finitely-valued* function of its accumulated knowledge.

$$\mathsf{F}_{tj} := F_{tj}(\mathsf{U}_0, \mathsf{U}_{i_j}, \mathsf{Y}^t_{i_j}, \mathsf{F}^{t-1}, \mathsf{F}_{t[j-1]}) \tag{5.5}$$

where $\mathsf{F}^{t-1} := (\mathsf{F}_\tau : \tau \in [t-1])$ and $\mathsf{F}_t := \mathsf{F}_{t[r]}$ is the vector of public messages at time t after the t-th private channel use.

We call F_{jt} in (5.5) the *public discussion function.*[3] Unlike the private channel described earlier, public discussion is interactive, noise-free and unlimited in rate. It is interactive in the sense that F_{tj} can be a function of the previous messages $\mathsf{F}_{t[j-1]}$.

[1] Even though X_i^{t-1} is part of the accumlated knowledge of terminal i, it needs not be included in the R.H.S. of (5.2) because it can be derived from the rest of the accumulated knowledge.

[2] One may think that randomizing at a latter time can allow one to adapt to more observations. This is unnecessary as argued because such randomization can be expressed as a deterministic function of the observations and a randomization before the observations. This deterministic function may not be a bijection however.

[3] It is unnecessary to include $\mathsf{X}^t_{i_j}$ since it is completely determined by the rest of the accumulated knowledge of terminal i_j.

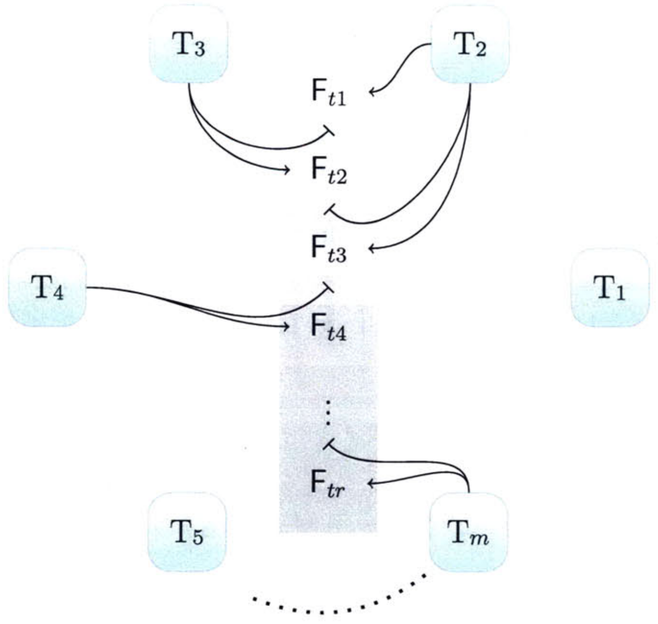


Figure 5-4: Public discussion: the accumulated knowledge of terminal i before the discussion session is $(\mathsf{U}_0, \mathsf{U}_1, \mathsf{X}_i^{t-1}, \mathsf{Y}_i^{t-1})$. In this specific case, 1) T_2 first broadcasts F_{t1}; 2) T_3 observes F_{t1} and broadcasts F_{t2}; 3) T_2 observes both F_{t1} and F_{t2} and then reply with F_{t3}; 4) T_4 transmits at the next time; 5) the final message $\mathsf{F}_{t[r]}$ is sent by T_m. T_1 and T_5 remain silent during the entire public discussion.

It is authenticated in the sense that every terminal knows the sender of every public message. The wiretapper cannot inject any fradulent messages. This is illustrated in Figure 5-4. The public messages are also finitely-valued for both practical and technical reasons.

Finally, the active terminals attempt to generate a common secret key as much as possible based on their accumulated knowledge. The key has be recoverable almost surely and remain nearly uniformly random to the wiretapper.

Key generation phase:

At time $n+1$, every active terminal $i \in A$ generates a finitely-valued *individual key* $\mathsf{K}_i \in K$ from their accumulated knowledge

$$\mathsf{K}_i := K_i(\mathsf{U}_0, \mathsf{U}_i, \mathsf{Y}_i^n, \mathsf{F}^n) \in K \tag{5.6}$$

such that K_i's are asymptotically the same and secure. More precisely, there exists a random variable $\mathsf{K} \in K$ satisfying the following conditions:

$$\Pr\{\exists i \in A, \mathsf{K}_i \neq \mathsf{K}\} \leq \epsilon_n \qquad \text{for some } \epsilon_n \to 0 \qquad \text{(recoverability)} \tag{5.7}$$

$$s_{\text{div}} \leq \delta_n \qquad \text{for some } \delta_n \to 0 \qquad \text{(secrecy)} \tag{5.8}$$

where s_{div} is the *secret leakage rate* defined as

$$s_{\text{div}} := \frac{1}{n} D(P_{\mathsf{K}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0} \| \mathbb{U}_K) \tag{5.9a}$$

$$= \frac{1}{n}\left[\log|K| - H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0)\right] \tag{5.9b}$$

$\mathbb{U}_K$ denotes the uniform distribution over K, $D(\cdot\|\cdot)$ is the information divergence and $H(\cdot|\cdot)$ is the conditional entropy. (See Section A.1 or [8].)

We can also have different levels of recoverability and secrecy by imposing additional constraints on the convergence rates of ϵ_n and δ_n. These conditions are the same as the ones given in [12] except for a weakening of the secrecy condition by an additional factor of $1/n$.[4] The motivation will be further elaborated in Section 5.2.

In summary, the complete knowledge of terminal $i \in V$ is

$$(\mathsf{U}_0, \mathsf{U}_i, \mathsf{X}_i^n, \mathsf{Y}_i^n, \mathsf{F}^n, \mathsf{K}_i) \qquad \text{(Knowledge of T}_i\text{)}$$

and the complete knowledge of the wiretapper is

$$(\mathsf{U}_0, \mathsf{U}_D, \mathsf{X}_D^n, \mathsf{Y}_D^n, \mathsf{F}^n) \qquad \text{(Knowledge of wiretapper)}$$

We can remove the channel input X_i^n and individual key K_i from the knowledge of T_i without loss of generality since they can be determined by the causal relations (5.2) and (5.6).

We can now define the performance metrics in terms of the exponential growth rate of the key cardinality.

[4] The secret leakage rate is equal to $1/n$ times the security index defined in [12].

Definition 5.2 (Multiterminal secrecy capacity) We use the term *secrecy scheme* to refer to the choice of the sequence (in n) of

1. distributions P_{U_0} and $(P_{\mathsf{U}_i|\mathsf{U}_0} : i \in D^c)$ for the randomizations,
2. private channel input functions $(X_{it} : i \in V, t \in [n])$,
3. public discussion functions $(F_{tj} : t \in [n], j \in [r])$, including the choice of r and the order $(i_j \in V : j \in [r])$ of discussion for each session, and
4. key functions $(K_i : i \in A)$, including the choice of the set K of possible keys.

For notational simplicity, we have made the dependence on the constraint length n implicit. K for instance is growing exponentially in n for the case of interest. The key rate is defined as the asymptotic growth rate

$$R := \liminf_{n\to\infty} \frac{1}{n} \log|K|$$

Any non-negative rate $R' \leq R$ is said to be *achievable* provided that there exists a secrecy scheme that achieves the key rate R and satisfies the recoverability (5.7) and secrecy constraints (5.8). It is said to be *strongly achievable* if $\epsilon_n, \delta_n \to 0$ exponentially in n. It is said to be *perfectly* achievable if $\epsilon_n, \delta_n = 0$ for sufficiently large n. The largest achievable key rate is called the *secrecy capacity*, denoted as C_{s} or more explicitly $C_{\mathrm{s}}^{A|D}|_{P_{\mathsf{Y}_V|\mathsf{X}_V}}$.[5] Upper and lower bounds on the secrecy capacity will be referred to as the *secrecy upper and lower bounds* respectively. □

One may question the need for continuous-valued randomization and public randomization. Intuitively, one thinks that continuous random variable can be well-approximated by finitely-valued random variables using fine enough quantization. Furthermore, public randomization is known to the wiretapper, and does not appear to improve any secrecy. This is indeed true for some special cases where we can give a capacity-achieving scheme that does not rely on such randomizations. Unfortunately, we are not able to make such conclusion concrete for the general case. It is unclear whether a model without continuous randomization suffices.

[5] In [12], this secrecy capacity is also called the private key capacity and secret key capacity in the respective cases with and without untrusted terminals.

5.1 Specialized models

We will specialize the model in various ways, by imposing additional constraints on the secrecy scheme or the private DMMC. The motivations are to

1) introduce the different proof techniques involved in a systematic way,
2) study the optimality of different secrecy schemes under different channels, and
3) give some important special cases covered by the current model.

A different proof technique is required for each of the following models.

Finitely-valued model: The channel input and output alphabets are all finitely-valued. This allows us to use the method of types in [11] directly.

Infinitely-valued source model: The terminals can receive infinitely-valued output but cannot send any channel input. This allows us to focus on the quantization trick that turns the model to the finitely-valued model.

Channel model with finite-input alphabet only: The channel input alphabets are all finite but the output alphabets are not necessarily finite. This is general enough to study the proposed cooperative schemes.

The solutions to these special cases and the finitely-valued source model in [12] will be used to compose the solution to the general infinitely-valued channel model with sample average constraint.

The different secrecy schemes we consider here can be categorized as follows.

Source emulation: The channel input sequences are iid generated over time. Since the channel is memoryless, the DMMC turns into a DMMS effectively.

Public input adaptation: The channel input sequences are generated independently over time given the public information, namely, $(\mathsf{U}_0, \mathsf{F}^{t-1})$.

Non-interactive public discussion: F_t is null for $t < n$ and the dependence on $\mathsf{F}_{t[j-1]}$ is removed from the R.H.S. of (5.5) for $t = n$. In other words, there is only one public discussion session, in which terminals do not reply to previous messages.

The model also covers the following cases of practical interests, such as fading in wireless communications.

Simultaneous independent channel: The channel consists of a finite set of independent channels, which are simultaneous in the sense that the output symbols are available only after the channel inputs to all component channels are specified.
Channel with publicly controllable channel state: At each time $t \in [n]$, the terminals can decide publicly one out of a given set of channels to use.
Channel with publicly observable channel state: At each time, the channel randomly realizes into one of a given set of channels. The state of the channel can be publicly observed before or after transmission. The sequence of channel states in time can be slowly varying as in slow fading, and ergodic as in fast fading.

If the terminals have access to a finite set $\{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}} : j \in L\}$ of simultaneous independent channels where T_i controls $\mathsf{X}_{Li} := (\mathsf{X}_{ji} : j \in L)$ and observes Y_{Li}, it is equivalent to having access to the composite channel

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = \prod_{j \in L} P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}$$

where $\mathsf{Y}_i := \mathsf{Y}_{Li}$ and $\mathsf{X}_i := \mathsf{X}_{Li}$ for all $i \in V$. For example, one component channel $P_{\mathsf{Y}_{1V}|\mathsf{X}_{1V}}$ can be a link $P_{\mathsf{Y}_{12}|\mathsf{X}_{11}}$ from T_1 to T_2, and another component channel $P_{\mathsf{Y}_{2V}|\mathsf{X}_{2V}}$ can be a link $P_{\mathsf{Y}_{23}|\mathsf{X}_{22}}$ from T_2 to T_3. The composite channel is $P_{\mathsf{Y}_{LV}|\mathsf{X}_{LV}} = P_{\mathsf{Y}_{12}|\mathsf{X}_{11}} P_{\mathsf{Y}_{23}|\mathsf{X}_{22}}$ consisting of two independent links.

To consider channel $P_{\mathsf{Y}_V|\mathsf{X}_V\mathsf{Q}}$ with publicly controllable state Q, we simply create a dummy untrusted terminal $m+1$ controlling Q. i.e. we consider the channel

$$P_{\mathsf{Y}_{\tilde{V}}|\mathsf{X}_{\tilde{V}}}(y_{\tilde{V}}|x_{\tilde{V}}) = P_{\mathsf{Y}_V|\mathsf{X}_V\mathsf{Q}}(y_V|x_V, x_{m+1})$$

where $\tilde{V} = [m+1]$, $\tilde{A} = A$, $\tilde{D} = D \cup \{m+1\}$ are the modified sets of terminals, active terminals and untrusted terminals respectively after adding the dummy terminal.[6]

[6] We can regard the entire knowledge of the dummy terminal as public. Indeed, it does not loose optimality to reveal in public the knowledge of any untrusted terminals since it is already known to

To consider channels with ergodic or iid channel states publicly observable *before* transmission, we can apply the law of large number to conclude that the resulting secrecy capacity is the expectation of the secrecy capacities over all possible realizations of the channel states. If the channel states are iid observable only after transmission, we can add a dummy untrusted terminal that observes the state as a channel output that is independent of the channel inputs. This is analogous to the usual idea of *ergodic capacity* in [55].

For slowly-varying channel states, we can define the usual outage event as the event that the instantaneous secrecy capacity given the channel state goes below the target key rate.

5.2 Asymptotic secrecy requirements

In this section, we will explain the use of the asymptotic secrecy condition (5.8) as a measure of security, which is introduced in [12] except for a factor of $1/n$.[7] Readers can refer to [9, 40, 51, 61] for more discussions on the asymptotic notion of secrecy. To motivate the definition, we first show that the key is provably secure in the *non-asymptotic* case when $\delta_n = 0$ in (5.8) for all n sufficiently large.

Perfect secrecy:

Suppose $\delta_n = 0$ in (5.8). It follows that

$$\log|K| = H(\mathsf{K}) = H(\mathsf{K}|\overbrace{\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0}^{\mathsf{W}:=})$$

by the additional fact that uniform distribution maximizes entropy under the finite-alphabet constraint. The first equality implies that the key K is uniformly distributed and the second equality implies that it is independent of the wiretapper's knowledge W. The key is therefore *perfectly secret* in the information-theoretic sense [51].

When δ_n is allowed to be positive while converging to zero, (5.8) becomes an

the wiretapper and there is no cost for public discussion.

[7] The additional factor of $1/n$ accommodates a weaker notion of secrecy and a stronger converse.

asymptotic notion of perfect secrecy, which has the merit of being more practical and allowing more interesting behavior. We will show that this convergence in divergence (5.9a) means that the key appears to be almost uniformly random to the wiretapper with probability converging to 1. To do so, we first relate the convergence in divergence to the convergence in variational distance.

Convergence in variational distance:

Define the variational distance as

$$s_{\text{var}} := \mathrm{E}_{\mathsf{W}} \left[\| P_{\mathsf{K}|\mathsf{W}}(\cdot|\mathsf{W}) - \mathbb{U}_{K} \|_1 \right] \tag{5.10}$$

where $\mathsf{W} := (\mathsf{U}_0, \mathsf{U}_D, \mathsf{Y}_D, \mathsf{F}^n)$ denotes the knowledge of the wiretapper. This is related to s_{div} by [12, Lemma 1] that

$$\frac{\log e}{2n} s_{\text{var}}^2 \leq s_{\text{div}} \leq \frac{s_{\text{var}}}{n} \log \frac{|K|}{s_{\text{var}}} \tag{5.11}$$

For the case of interest that K grows exponentially in n with strictly positive rate, it follows that

$$\begin{aligned} s_{\text{div}} = 0 &\iff s_{\text{var}} = 0 \\ s_{\text{div}} \xrightarrow{n\to\infty} 0 &\iff s_{\text{var}} \xrightarrow{n\to\infty} 0 \\ -\limsup_{n\to\infty} \frac{1}{n} \log s_{\text{div}} &= -\limsup_{n\to\infty} \frac{1}{n} \log s_{\text{var}} \end{aligned}$$

Thus, the secrecy condition (5.8) remains the same even if we replace s_{div} by s_{var} in each of the following cases:

$$\delta_n = 0 \qquad \text{(perfect secrecy)} \tag{5.12a}$$

$$\delta_n \to 0 \qquad \text{(weak secrecy)} \tag{5.12b}$$

$$\delta_n \leq 2^{-n\delta} \quad \text{for some } \delta > 0 \qquad \text{(strong secrecy)} \tag{5.12c}$$

which are the three different notions of secrecy of interest here. We can therefore consider the secrecy condition (5.8) with the secret leakage rate s_{div} replaced by the more explicit distance measure s_{var} in distributions.

Proposition 5.1 *Define the secrecy condition in variational distance as*

$$s_{\text{var}} \leq \delta_n \tag{5.13}$$

and the secrecy condition in probability as

$$P_{\mathsf{K}|\mathsf{W}}(k|w) = (1 \pm \epsilon_n)\frac{1}{|K|} \tag{5.14}$$

for all $w \in W_{\text{typ}}$ and $k \in K(w)$ where

- *w is a realization of the knowledge $\mathsf{W} := (\mathsf{U}_0, \mathsf{U}_D, \mathsf{Y}_D, \mathsf{F}^n)$ of the wiretapper in some typical set W_{typ} that has probability*

$$\Pr\{\mathsf{W} \in W_{\text{typ}}\} \geq 1 - \epsilon_n^2 \tag{5.15a}$$

- *k is a key in some subset $K(w) \subseteq K$ of size at least a factor $(1 - \epsilon_n)$ of the total number $|K|$ of keys. i.e.*

$$|K(w)| \geq (1 - \epsilon_n)|K| \tag{5.15b}$$

Then, we have for n sufficiently large that

(A) (5.13) *implies* (5.14) *with $\epsilon_n = \delta_n^{\frac{1}{4}}$, and*

(B) (5.14) *implies* (5.13) *with $\delta_n = \epsilon_n^2$.* □

Corollary 5.1 *For each type of secrecy in* (5.12), *the secrecy condition* (5.8) *is the same as* (5.14) *with $\epsilon_n = \delta_n$.* □

This is the desired conclusion that the key appears nearly uniformly distributed to the wiretapper over the set of possible keys except for a negligible amount and some atypical realizations of the knowledge of the wiretapper with negligible probability. (5.8) is therefore an intuitive notion of asymptotic secrecy.

Proof (Proof of Proposition 5.1 Part A) We first prove the implication from (5.13) to (5.14) using the Markov inequality for any non-negative random variable Z and constant $\alpha > 0$ that

$$\Pr(\mathsf{Z} > \alpha) < \frac{\mathrm{E}(\mathsf{Z})}{\alpha} \tag{5.16}$$

With Z set to $\|P_{\mathsf{K}|\mathsf{W}}(\cdot|\mathsf{W}) - \mathbb{U}_K\|_1$ and α set to $\delta_n^{\frac{1}{2}}$, we have from (5.13) that $\mathrm{E}(\mathsf{Z}) = s_{\text{var}} \leq \delta_n$. Thus,

$$\Pr\left\{\|P_{\mathsf{K}|\mathsf{W}}(\cdot|\mathsf{W}) - \mathbb{U}_K\|_1 > \delta_n^{\frac{1}{2}}\right\} < \delta_n^{\frac{1}{2}} \tag{5.17}$$

where the only randomness involved comes from W. We can define the typical set as follows to satisfy (5.15a) with $\epsilon_n = \delta_n^{\frac{1}{2}}$.

$$W_{\text{typ}} := \left\{w \in W : \|P_{\mathsf{K}|\mathsf{W}}(\cdot|w) - \mathbb{U}_K\|_1 \leq \delta_n^{\frac{1}{2}}\right\} \tag{5.18}$$

Let $\mathsf{K}_{\mathbb{U}}$ be a dummy random variable uniformly distributed over K. Then, for the typical case $w \in W_{\text{typ}}$, we have

$$\begin{aligned} \mathrm{E}_{\mathsf{W}}\left[\left|P_{\mathsf{K}|\mathsf{W}}(\cdot|\mathsf{W}) - \frac{1}{|K|}\right|\right] &= \frac{1}{|K|}\|P_{\mathsf{K}|\mathsf{W}}(\mathsf{K}_{\mathbb{U}}|\mathsf{W}) - \mathbb{U}_K\|_1 \\ &< \frac{\delta_n^{\frac{1}{2}}}{|K|} \end{aligned}$$

where the last inequality is by the definition of typicality in (5.18). Applying the Markov inequality (5.16) again with Z set to $|P_{\mathsf{K}|\mathsf{W}}(\cdot|\mathsf{W}) - |K|^{-1}|$ and α set to $\delta_n^{\frac{1}{4}}|K|^{-1}$, we have for all $w \in W_{\text{typ}}$ that

$$\Pr\left\{|P_{\mathsf{K}|\mathsf{W}}(\mathsf{K}_{\mathbb{U}}|\mathsf{W}) - |K|^{-1}| > \frac{\delta_n^{\frac{1}{4}}}{|K|}\right\} < \delta_n^{\frac{1}{4}}$$

where the only randomness involved comes from the dummy random variable $\mathsf{K}_{\mathbb{U}}$. Since $\mathsf{K}_{\mathbb{U}}$ is uniformly distributed, the last inequality says that at least a fraction $(1 - \delta_n^{\frac{1}{4}})$ of the keys satisfy (5.14) with $\epsilon_n = \delta_n^{\frac{1}{4}}$, which gives (5.15b) as desired. ■

PROOF (PROOF OF PROPOSITION 5.1 PART B) From (5.14), we have for all $w \in W_{\text{typ}}$ that

$$\left|P_{\mathsf{K}|\mathsf{W}}(k|w) - \frac{1}{|K|}\right| \leq \begin{cases} \frac{\epsilon_n}{|K|} & , k \in K(w) \\ 2 & \text{otherwise} \end{cases}$$

where the last case $k \notin K(w)$ is due to the fact $P_{\mathsf{K}|\mathsf{W}}(k|w), |K|^{-1} \in [0,1]$. Summing over $k \in K$, we have

$$\left\|P_{\mathsf{K}|\mathsf{W}}(\cdot|w) - \mathbb{U}_K\right\|_1 \leq \begin{cases} 3\epsilon_n & , w \in W_{\text{typ}} \\ 2 & \text{otherwise} \end{cases}$$

where the first case uses (5.15b), and the last case $w \notin W_{\text{typ}}$ is due to triangle inequality of 1-norm that

$$\left\|P_{\mathsf{K}|\mathsf{W}}(\cdot|w) - \mathbb{U}_K\right\|_1 \leq \left\|P_{\mathsf{K}|\mathsf{W}}(\cdot|w)\right\|_1 + \|\mathbb{U}_K\|_1 \leq 2$$

Finally, averaging over W gives

$$s_{\text{var}} = 3\epsilon_n(1-\epsilon_n^2) + 2\epsilon_n^2 = O(\epsilon_n^2)$$

This establishes (5.13) with $\delta_n = \epsilon_n^2$ for n large enough as desired. ■

PROOF (PROOF OF COROLLARY 5.1) (5.13) with $\delta_n = 0$ implies (5.14) with $\epsilon_n = 0$ by Proposition 5.1A. The converse is also true by Proposition 5.1B. This establishes the desired equivalence between (5.8) and (5.14) for perfect secrecy (5.12a) when $\delta_n = 0$, since (5.13) is equivalent to (5.8) as argued earlier. The equivalence for weak (5.12b) and strong secrecy (5.12c) can be proved similarly. ■

Chapter 6

Secrecy Capacity

Under the multiterminal network model formulated in Chapter 4, what is the maximum achievable secret key rate? We will derive upper and lower bounds on the secrecy capacity in Section 6.1 and Section 6.2 respectively using tools from information theory in Appendix A. The lower bound is achieved by a new cooperation strategy called the mixed source emulation, which is shown in Section 6.3 to be superior in to the conventional pure source emulation approach.

We strongly recommend skipping Section 6.1.2, 6.2.2 and 6.2.3 for the first reading. The main ideas can be understood by focusing only on the finitely-valued model in Section 6.1.1 and Section 6.2.1 without the sample average constraint.

6.1 Secrecy upper bound

We first derive single-letter upper bounds on the secrecy capacity C_{s} in Definition 5.2 using the Shearer-type lemma in Section A.2 and the expressions in Section A.4.

Theorem 6.1 (Finite-input-alphabet constraint) *If the channel inputs can be chosen arbitrarily from some given finite support sets, then*

$$C_{\mathrm{s}} \leq \min_{\lambda \in \Lambda_{A|D}} \max_{P_{\mathsf{X}_V} \in \mathscr{P}(X_V)} \alpha(\lambda, P_{\mathsf{X}_V}) \tag{6.1}$$

where α is defined in (A.17), *and $\lambda \in \Lambda_{A|D}$ is defined in* (A.9). □

Example 6.1 Consider $V = [3]$, $A = [2]$, $D = \{3\}$, and the private DMMC $P_{\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2}$. The active terminals T_1 and T_2 control the finitely-valued channel input X_1 and X_2 respectively, and the untrusted terminal 3 observes Y. By (6.1), the secrecy upper bound simplifies to

$$\max_{P_{\mathsf{X}_1\mathsf{X}_2} \in \mathscr{P}(X_1 \times X_2)} [I(\mathsf{X}_1 \wedge \mathsf{X}_2|\mathsf{Y}) - I(\mathsf{X}_1 \wedge \mathsf{X}_2)] \tag{6.2}$$

n.b. the minimization in (6.1) is trivial since $\Lambda_{A|D}$ is a singleton, containing only one fractional partition λ, namely the one with $\lambda_{\{1\}} = \lambda_{\{2\}} = 1$.

Consider, in particular, the binary multiple access channel

$$\mathsf{Y} = \mathsf{X}_1 \oplus \mathsf{X}_2$$

Since $I(\mathsf{X}_1 \wedge \mathsf{X}_2|\mathsf{Y}) \leq H(\mathsf{X}_1) \leq 1$ in (6.2), the secrecy upper bound is 1 bit with the optimal distribution $P_{\mathsf{X}_1,\mathsf{X}_2}(x_1, x_2) = \mathrm{Bern}_{\frac{1}{2}}(x_1)\, \mathrm{Bern}_{\frac{1}{2}}(x_2)$. □

For the more general case when the input can be a mixture of continuous and discrete random variables as described in Section A.1 subject to the sample average constraint in (5.3), we will derive a weaker[1] secrecy upper bound that uses the following single-letter *moment constraint.*

Definition 6.1 (Moment constraint) X_V satisfies the *moment constraint* $\phi_i : X_i \mapsto \mathbb{R}^\ell$ for $i \in V$ if

$$\mathrm{E}\left[\phi_i(\mathsf{X}_i)\right] \leq \mathbf{0} \qquad \text{for all } i \in V \tag{6.3}$$

With ζ_{ij}, s_{ijk} and c_{ijk} as defined in (5.4), it specializes to the *power constraint,*

$$\mathrm{E}\left[|\zeta_{ij}(\mathsf{X}_i)|^{s_{ijk}}\right] \leq c_{ijk} \qquad \text{(power)} \tag{6.4}$$

for all $i \in V$ and (j, k) in some given finite sets. □

[1] As described at the end of this section, there are examples for which (6.1) is a tighter bound.

Theorem 6.2 (Sample average constraint) *With the sample average constraint in* (5.3) *but not necessarily the finite-input-alphabet constraint, i.e. allowing the input to be continuous, we have the following secrecy upper bound,*

$$C_{\mathrm{s}} \leq \inf_{\lambda \in \Lambda_{A|D}} \sup_{P_{\mathsf{X}_V}} \min_{i \in D^c} \alpha_i(\lambda, P_{\mathsf{X}_V}) \tag{6.5}$$

where α_i is defined in (A.20) *and the input distribution is subject to the* moment constraint *in* (6.3). □

Example 6.2 Consider as in Example 6.1 the case $V = [3]$, $A = [2]$, $D = \{3\}$ and the DMMC $P_{\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2}$, where the active terminals T_1 and T_2 control X_1 and X_2 respectively under the sample average constraint $\phi_{[2]}$ in (5.3), and the untrusted terminal 3 observes Y. By (6.5), the secrecy upper bound simplifies to

$$\max_{P_{\mathsf{X}_1\mathsf{X}_2} \in \mathscr{P}(X_1 \times X_2)} \min\{I(\mathsf{X}_1 \wedge \mathsf{Y}|\mathsf{X}_2), I(\mathsf{X}_2 \wedge \mathsf{Y}|\mathsf{X}_1)\} \tag{6.6}$$

where $P_{\mathsf{X}_1\mathsf{X}_2}$ is subject to the moment constraint

$$\mathrm{E}(\phi_1(\mathsf{X}_1)) \leq \mathbf{0} \quad \text{and} \quad \mathrm{E}(\phi_2(\mathsf{X}_2)) \leq \mathbf{0}$$

Consider, in particular, the gaussian multiple access channel $\mathsf{Y} = \mathsf{X}_1 + \mathsf{X}_2 + \mathsf{N}$ where $\mathsf{N} \sim \mathscr{N}_{0,1}$ is a zero-mean unit-variance gaussian channel noise. The input X_i is subject to the average power constraint $\phi_i(x_i) = x_i^2 - P_i$ for some given $P_i > 0$. The secrecy upper bound becomes $\log(1+\min\{P_1, P_2\})$, with the optimal distribution $P_{\mathsf{X}_1,\mathsf{X}_2} = \mathscr{N}_{0,P_1}\mathscr{N}_{0,P_2}$. □

We will break down the proofs of the above theorems as follows. In Section 6.1.1, we consider the case with only the finite-input-alphabet constraint. The channel output, however, can be a mixture of continuous and discrete random variables. Then, we will consider the more general case with infinitely-valued input and the sample average constraint in Section 6.1.2.

6.1.1 Finite-input-alphabet constraint

In this section, we will prove the secrecy upper bound (6.1) in Theorem 6.1. From the secrecy condition (5.8) and expression (5.9b), we have

$$\frac{1}{n}\log|K| \leq \frac{1}{n}H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0) + \delta_n$$

Our goal is to derive the secrecy upper bound by turning the R.H.S. to an expression that is universal to any secrecy schemes. In particular, we will replace the dependence on K, F^n, U_0 and U_D by the dependence on the channel statistics $P_{\mathsf{Y}_V|\mathsf{X}_V}$. Let us first eliminate the dependence on K.

$$H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0) = H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_V^n\mathsf{U}_V\mathsf{U}_0) + I(\mathsf{K}\wedge\mathsf{Y}_{D^c}^n\mathsf{U}_{D^c}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0)$$

The first term on the right is negligible (sublinear in n) by the Fano's inequality [8].

Fano's inequality:

Since $(\mathsf{F}^n, \mathsf{Y}_i^n, \mathsf{U}_i, \mathsf{U}_0)$ determines the individual key K_i for every active terminal $i \in A$ by (5.6), and the individual keys equal K almost surely by (5.7), we have

$$H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_i^n\mathsf{U}_i\mathsf{U}_0) \leq h(\epsilon_n) + \epsilon_n\log|K| \qquad \forall i \in A \tag{6.7}$$

by the Fano's inequality, where

$$h(p) := -(1-p)\log(1-p) - p\log p \tag{6.8}$$

is the binary entropy function.

We now have

$$H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0) = I(\mathsf{K}\wedge\mathsf{Y}_{D^c}^n\mathsf{U}_{D^c}|\mathsf{F}^n\mathsf{Y}_D^n\mathsf{U}_D\mathsf{U}_0) + o(n)$$

To eliminate the dependence on K in the mutual information term, we apply the

Shearer-type lemma as follows.

$$\begin{aligned}
&I(\mathsf{K}\wedge \mathsf{Y}^n_{D^c}\mathsf{U}_{D^c}|\mathsf{F}^n\mathsf{Y}^n_D\mathsf{U}_D\mathsf{U}_0)\\
&\qquad = H(\mathsf{U}_{D^c}\mathsf{Y}^n_{D^c}|\mathsf{F}^n\mathsf{Y}^n_D\mathsf{U}_D\mathsf{U}_0) - H(\mathsf{U}_{D^c}\mathsf{Y}^n_{D^c}|\mathsf{K}\mathsf{F}^n\mathsf{Y}^n_D\mathsf{U}_D\mathsf{U}_0)\\
&\qquad \leq H(\mathsf{U}_{D^c}\mathsf{Y}^n_{D^c}|\mathsf{F}^n\mathsf{Y}^n_D\mathsf{U}_D\mathsf{U}_0) - \sum_{B\in\mathcal{H}_{A|D}} \lambda_B H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{K}\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)
\end{aligned}$$

The last inequality is by the weak form of the Shearer-type lower bound (A.10a). We restrict to $\mathcal{H}_{A|D}$ to ensure that B^c in the resulting expression intersects A, which in turn ensures that the conditions in the conditional entropy expressions determine at least one individual key by (5.6). As a result, the Fano's inequality (6.7) applies as follows.

$$\begin{aligned}
&\sum_{B\in\mathcal{H}_{A|D}} \lambda_B H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{K}\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)\\
&\qquad = \sum_B \lambda_B \Big[H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) - I(\mathsf{K}\wedge \mathsf{U}_B\mathsf{Y}^n_B|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)\Big]\\
&\qquad \geq \sum_B \lambda_B \Big[H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) - |D^c| H(\mathsf{K}|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)\Big]
\end{aligned}$$

To explain the last inequality, note that K is discrete-valued, which implies positivity of its conditional entropy (A.5). By the constraint (A.9) on fractional partitions that $\sum_{B\ni i}\lambda_B = 1$, we have

$$\sum_B \lambda_B \leq \sum_{i\in D^c}\left(\sum_{B\ni i}\lambda_B\right) = |D^c|$$

The last entropy term is negligible by the Fano's inequality (6.7). Thus, we have eliminated the dependence on K as desired. i.e. for some $\delta'_n \to 0$,

$$\frac{1}{n}\log|K| \leq \frac{1}{n}\Bigg[H(\mathsf{U}_{D^c}\mathsf{Y}^n_{D^c}|\mathsf{F}^n\mathsf{Y}^n_D\mathsf{U}_D\mathsf{U}_0) - \sum_{B\in\mathcal{H}_{A|D}} \lambda_B H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)\Bigg] + \delta'_n$$

It remains to eliminate the dependence on the randomizations and public messages. We will do so using the causal relations (5.2) and (5.5) in the model. First of all, the last inequality can be rewritten as

$$\frac{1}{n}\log|K| \leq \frac{1}{n}\left[H(\mathsf{U}_{D^c}\mathsf{Y}^n_{D^c}\mathsf{F}^n|\mathsf{U}_0) - H(\mathsf{U}_D\mathsf{Y}^n_D\mathsf{F}^n|\mathsf{U}_0) - \left(\sum_B \lambda_B - 1\right) H(\mathsf{U}_V\mathsf{Y}^n_V|\mathsf{U}_0)\right] + \delta'_n \tag{6.9}$$

using the expansion that

$$H(\mathsf{U}_B\mathsf{Y}^n_B|\mathsf{F}^n\mathsf{Y}^n_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) = H(\mathsf{U}_V\mathsf{Y}^n_V|\mathsf{U}_0) - H(\mathsf{U}_{B^c}\mathsf{Y}^n_{B^c}\mathsf{F}^n|\mathsf{U}_0)$$

and the same expression with B replaced by D^c and B^c replaced by D. This equation follows from the chain rule and the fact that

$$H(\mathsf{U}_V\mathsf{Y}^n_V\mathsf{F}^n|\mathsf{U}_0) = H(\mathsf{U}_V\mathsf{Y}^n_V|\mathsf{U}_0)$$

since $(\mathsf{U}_0, \mathsf{U}_V, \mathsf{Y}^n_V)$ completely determines F^n by an inductive argument on (5.5). Further simplification is possible with the following causal expansion.

Causal expansion:

$$H(\mathsf{U}_{B^c}\mathsf{Y}^n_{B^c}\mathsf{F}^n|\mathsf{U}_0) \overset{(a)}{=} H(\mathsf{U}_{B^c}|\mathsf{U}_0) + \sum_{t\in[n]}\left[H(\mathsf{Y}_{B^ct}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) + H(\mathsf{F}_t|\mathsf{F}^{t-1}\mathsf{Y}^t_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0)\right]$$

$$H(\mathsf{U}_D\mathsf{Y}^n_D\mathsf{F}^n|\mathsf{U}_0) \overset{(b)}{=} H(\mathsf{U}_D|\mathsf{U}_0) + \sum_{t\in[n]}\left[H(\mathsf{Y}_{Dt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_D\mathsf{U}_D\mathsf{U}_0) + H(\mathsf{F}_t|\mathsf{F}^{t-1}\mathsf{Y}^t_D\mathsf{U}_D\mathsf{U}_0)\right]$$

$$H(\mathsf{U}_V\mathsf{Y}^n_V\mathsf{F}^n|\mathsf{U}_0) \overset{(c)}{=} H(\mathsf{U}_V|\mathsf{U}_0) + \sum_{t\in[n]}\left[H(\mathsf{Y}_{Vt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_V\mathsf{U}_V\mathsf{U}_0)\right]$$

(a) by the chain rule expansion in the causal order illustrated in Figure 5-1.

(b) same as (a) with B^c replaced by D.

(c) same as (a) with B^c replaced by V. We have also used (5.5) that $(\mathsf{U}_0, \mathsf{U}_V, \mathsf{Y}^t_V)$ completely determines F^t, which implies that $H(\mathsf{F}_t|\mathsf{F}^{t-1}\mathsf{Y}^t_V\mathsf{U}_V\mathsf{U}_0) = 0$.

After applying these causal expansions to (6.9), we can regroup similar terms together and simplify them using the Shearer-type lemma as follows.

Applying Shearer-type lemma:

$$\sum_B \lambda_B H(\mathsf{U}_{B^c}|\mathsf{U}_0) - H(\mathsf{U}_D|\mathsf{U}_0) - \left(\textstyle\sum_B \lambda_B - 1\right) H(\mathsf{U}_V|\mathsf{U}_0)$$
$$\overset{(a)}{=} H(\mathsf{U}_{D^c}|\mathsf{U}_0) - \sum_B \lambda_B H(\mathsf{U}_B|\mathsf{U}_{B^c}\mathsf{U}_0) \overset{(b)}{=} 0$$

$$\sum_B \lambda_B H(\mathsf{F}_t|\mathsf{F}^{t-1}\mathsf{Y}^t_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) - H(\mathsf{F}_t|\mathsf{F}^{t-1}\mathsf{Y}^t_D\mathsf{U}_D\mathsf{U}_0) \overset{(c)}{\leq} 0$$

(a) by the conditional independence (5.1) of U_i's given U_D.

(b) by the equality case (A.10b) of the Shearer-type lemma.

(c) by the Shearer-type lemma (A.10c) for the causal relation (5.5).

Putting these together, (6.9) becomes

$$\frac{1}{n}\log|K| \leq \frac{1}{n}\sum_{t\in[n]} \Big[H(\mathsf{Y}_{B^c t}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) - H(\mathsf{Y}_{Dt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_D\mathsf{U}_D\mathsf{U}_0)$$
$$- \left(\textstyle\sum_B \lambda_B - 1\right) H(\mathsf{Y}_{Vt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_V\mathsf{U}_V\mathsf{U}_0) \Big] + \delta'_n \qquad \textbf{(6.10)}$$

To transform the dependence on the randomization and public discussion functions to the dependence on the channel statistics, we insert the channel input as an additional condition in the entropy terms.

Inserting channel input:

Define

$$\mathsf{Q}_t := (\mathsf{F}^{t-1}, \mathsf{Y}^{t-1}_D, \mathsf{U}_D, \mathsf{U}_0) \qquad \textbf{(6.11)}$$

Then, the entropy terms in (6.10) become

$$H(\mathsf{Y}_{B^c t}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) \overset{(a)}{=} H(\mathsf{Y}_{B^c t}|\mathsf{X}_{B^c t}\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_{B^c}\mathsf{U}_{B^c}\mathsf{U}_0) \overset{(b)}{\leq} H(\mathsf{Y}_{B^c t}|\mathsf{X}_{B^c t}\mathsf{Q}_t)$$
$$H(\mathsf{Y}_{Dt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_D\mathsf{U}_D\mathsf{U}_0) \overset{(c)}{=} H(\mathsf{Y}_{Dt}|\mathsf{X}_{Dt}\mathsf{Q}_t)$$
$$H(\mathsf{Y}_{Vt}|\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_V\mathsf{U}_V\mathsf{U}_0) \overset{(d)}{=} H(\mathsf{Y}_{Vt}|\mathsf{X}_{Vt}\mathsf{F}^{t-1}\mathsf{Y}^{t-1}_V\mathsf{U}_V\mathsf{U}_0) \overset{(e)}{=} H(\mathsf{Y}_{Vt}|\mathsf{X}_{Vt})$$

(a) by the causal relation (5.2).

(b) by the fact that conditioning reduces entropy (A.4).

(c) same as (a) with B^c replaced by D.

(d) same as (a) with B^c replaced by V.

(e) by the memorylessness assumption of the DMMC.

Substituting these into (6.10), we have

$$\frac{1}{n}\log|K| \leq \frac{1}{n}\sum_{t\in[n]} [H(\mathsf{Y}_{B^ct}|\mathsf{X}_{B^ct}\mathsf{Q}_t) - H(\mathsf{Y}_{Dt}|\mathsf{X}_{Dt}\mathsf{Q}_t) - (\textstyle\sum_B \lambda_B - 1)\, H(\mathsf{Y}_{Vt}|\mathsf{X}_{Vt})] + \delta'_n$$

Tightening the bound by minimizing over $\lambda \in \Lambda_{A|D}$ and then optimizing over all achievable schemes, we have

$$\begin{aligned}\sup\frac{1}{n}\log|K| \leq \sup_{P_{\mathsf{X}_V^n|\mathsf{Q}^n}, P_{Q^n}} \inf_{\lambda} \frac{1}{n}\sum_{t\in[n]} \Big[&H(\mathsf{Y}_{B^ct}|\mathsf{X}_{B^ct}\mathsf{Q}_t) - H(\mathsf{Y}_{Dt}|\mathsf{X}_{Dt}\mathsf{Q}_t) \\ &- (\textstyle\sum_B \lambda_B - 1)\, H(\mathsf{Y}_{Vt}|\mathsf{X}_{Vt})\Big] + \delta'_n\end{aligned}$$

where in the supremum on the R.H.S., we restrict $P_{\mathsf{X}_V^n|\mathsf{Q}^n}$ to a collection of valid input distributions, and relax Q^n from (6.11) to any mixture of discrete and continuous random variables that satisfy the Markov chains

$$\mathsf{Q}_t \leftrightarrow \mathsf{X}_{Vt} \leftrightarrow \mathsf{Y}_{Vt} \qquad \forall t \in [n] \tag{6.12}$$

These Markov chains come from the memorylessness property of the DMMC with the original definition (6.11) of Q_t. Exchanging the sup and inf can only increase the bound, and so

$$\begin{aligned}\sup\frac{1}{n}\log|K| &\leq \inf_{\lambda} \sup_{P_{Q^n}, P_{\mathsf{X}_V^n|\mathsf{Q}^n}} \frac{1}{n}\sum_{t\in[n]} \Big[H(\mathsf{Y}_{B^ct}|\mathsf{X}_{B^ct}\mathsf{Q}_t) - H(\mathsf{Y}_{Dt}|\mathsf{X}_{Dt}\mathsf{Q}_t) \\ &\qquad\qquad - (\textstyle\sum_B \lambda_B - 1)\, H(\mathsf{Y}_{Vt}|\mathsf{X}_{Vt})\Big] + \delta'_n \\ &= \inf_{\lambda} \sup_{P_{Q^n}, P_{\mathsf{X}_V^n|\mathsf{Q}^n}} \frac{1}{n}\sum_{t\in[n]} E\left[\alpha(\lambda, P_{\mathsf{X}_{Vt}|\mathsf{Q}^n}(\cdot|\mathsf{Q}_t))\right] + \delta'_n\end{aligned}$$

by the definition of α in (A.17a) and the Markov property (6.12). It is optimal to choose Q^n deterministic by the trivial fact that the supremum of α over Q^n is always no less than any averaging over Q^n. In summary, we have

$$\sup \frac{1}{n} \log|K| \leq \inf_{\lambda} \sup_{P_{\mathsf{X}_V^n}} \frac{1}{n} \sum_{t\in[n]} \alpha(\lambda, P_{\mathsf{X}_{Vt}}) + \delta'_n \tag{6.13}$$

We now specialize to the case where the channel input symbols are subject to the finite-alphabet constraint only.

Specializing to finite-input-alphabet constraint:

$$\begin{aligned}
\sup \frac{1}{n} \log|K| &\overset{\text{(a)}}{=} \inf_{\lambda} \frac{1}{n} \sum_{t\in[n]} \sup_{P_{\mathsf{X}_{Vt}}} \alpha(\lambda, P_{\mathsf{X}_{Vt}}) + \delta'_n \\
&\overset{\text{(b)}}{=} \inf_{\lambda} \sup_{P_{\mathsf{X}_V}} \alpha(\lambda, P_{\mathsf{X}_V}) + \delta'_n \\
&\overset{\text{(c)}}{=} \min_{\lambda} \max_{P_{\mathsf{X}_V}} \alpha(\lambda, P_{\mathsf{X}_V}) + \delta'_n
\end{aligned}$$

(a) We can push the supremum inside the summation in (6.13) because,

- the t-th summand $\alpha(\lambda, P_{\mathsf{X}_{Vt}})$ depends on $P_{\mathsf{X}_V^n}$ only through $P_{\mathsf{X}_{Vt}}$, and
- the finite-alphabet constraint on the support of $P_{\mathsf{X}_V^n}$ is separable into independent finite-alphabet constraints on the support of $P_{\mathsf{X}_{Vt}}$.

(b) By symmetry, $\alpha(\lambda, P_{\mathsf{X}_{Vt}})$ has the same supremum independent of t.

(c) The infimum and supremum can be replaced by the minimum and maximum since α is a continuous function over the compact set $\Lambda_{A|D} \times \mathscr{P}(X_V)$. See [11] for a detailed derivation of the continuity of information measures.

Finally, taking $\limsup_{n\to\infty}$ on both sides of (c) gives the desired bound (6.1).

6.1.2 Sample Average Constraint

We now prove the secrecy upper bound (6.5) in Theorem 6.2 for the more general case with possibly infinitely-valued input subject to the sample average constraint. We first weaken the bound (6.13) for quasi-concavity, replacing α by α_i as follows.

$$\alpha(\lambda, P_{\mathsf{X}_V}) \leq \min_{i\in D^c} \alpha_i(\lambda, P_{\mathsf{X}_V}) \qquad \text{by (A.24b)}$$

Applying this to (6.13) gives,

$$\sup \frac{1}{n}\log|K| \leq \inf_{\lambda} \sup_{P_{\mathsf{X}_V^n}} \overbrace{\min_{i\in D^c} \frac{1}{n} \sum_{t\in[n]} \alpha_i(\lambda, P_{\mathsf{X}_{Vt}})}^{f(\lambda,P_{\mathsf{X}_V^n}):=} + \delta_n' \tag{6.14}$$

f defined above is concave in the input distribution because α_i is concave according to Corollary A.1.[2] The weaker quasi-concavity will allow us to use a *mixing argument* to prove existence of an optimal (f-maximizing) input distribution (as a function λ) with identical marginal distributions $P_{\mathsf{X}_{Vt}}$ over time $t \in [n]$. In other words, we can make every t-th summand $\alpha_i(\lambda, P_{\mathsf{X}_{Vt}})$ in f independent of t, giving rise to the desired single-letter bound.

More precisely, starting with an arbitrary optimal solution $P^*_{\mathsf{X}_V^n}$,[3] define the following time-permuted input distribution $P^g_{\mathsf{X}_V^n}$ for every permutation $g : [n] \mapsto [n]$ in the *symmetric group*, Sym($[n]$), of all permutation functions on $[n]$:

$$P^g_{\mathsf{X}_V^n}(x_{V1}, x_{V2}, \dots, x_{Vn}) := P^*_{\mathsf{X}_V^n}(x_{Vg(1)}, x_{Vg(2)}, \dots, x_{Vg(n)}) \tag{6.15}$$

In other words, if we have X_V^n distributed as $P^*_{\mathsf{X}_V^n}$, then $P^g_{\mathsf{X}_V^n}$ is the statistics of a sequence obtained by moving X_{Vt} from time t to $g(t)$.[4] $P^g_{\mathsf{X}_V^n}$ is also a valid optimal input distribution in maximizing f by the following symmetry arguments:

- Since f (6.14) is symmetric over the marginal distributions for different t, we have

$$f(\lambda, P^g_{\mathsf{X}_V^n}) = f(\lambda, P^*_{\mathsf{X}_V^n}) \qquad \text{for all } g \in \text{Sym}([n])$$

- $P^g_{\mathsf{X}_V^n}$ also satisfies the sample average constraint (5.3) since the constraint is symmetric over any permutations in t.

[2] The minimum and average of concave functions are concave [5].

[3] If the optimal solution does not exist within the valid set of input distributions, we can instead consider a sequence of valid distributions that asymptotically achieve the supremum.

[4] There is a minor subtlety that $P^g_{\mathsf{X}_V^n}$ may not belong to a filtered probability space in t due to the fact that shuffling in time may disrupt causality. Nonetheless, we can relax the causality constraint for the purpose of obtaining an upper bound.

The desired distribution is the average over all time-permutated distribution

$$\bar{P}_{\mathsf{X}_V^n} := \mathrm{E}_{\mathsf{G}}\left(P_{\mathsf{X}_V^n}^{\mathsf{G}}\right) \tag{6.16}$$

where G is a random variable uniformly distributed over $\mathrm{Sym}([n])$. $\bar{P}_{\mathsf{X}_V^n}$ is optimal by the following quasi-concavity and linearity arguments:

- Applying Jensen's inequality [5] on the quasi-concave function $f(\lambda, \cdot)$, we have

  $$f\left(\lambda, \mathrm{E}\left(P_{\mathsf{X}_V^n}^{\mathsf{G}}\right)\right) \geq \min_g f(\lambda, P_{\mathsf{X}_V^n}^g) = f(\lambda, P_{\mathsf{X}_V^n}^*)$$

  and so $\bar{P}_{\mathsf{X}_V^n}$ is optimal.
- $\bar{P}_{\mathsf{X}_V^n}$ satisfies the sample average constraint because every $P_{\mathsf{X}_V^n}^g$ satisfies (5.3) with probability one by the earlier symmetry argument.

It remains to show that $\bar{P}_{\mathsf{X}_V^n}$ has identical marginal distributions $\bar{P}_{\mathsf{X}_{Vt}}$ that satisfies the moment constraint (6.3) asymptotically. By the definitions (6.16) and (6.15),

$$\bar{P}_{\mathsf{X}_{Vt}} := E\left(P_{\mathsf{X}_{Vt}}^{\mathsf{G}}\right) = \frac{1}{n}\sum_{\tau \in [n]} P_{\mathsf{X}_{V\tau}}^* =: \bar{P}_{\mathsf{X}_V} \tag{6.17}$$

which is independent of t as desired. To show that $\bar{P}_{\mathsf{X}_V}$ satisfies the moment constraint (6.3) asymptotically as $n \to \infty$, consider X_i distributed as $\bar{P}_{\mathsf{X}_i}$ and X_{it} distributed as $P_{\mathsf{X}_{it}}^*$ for $t \in [n]$. Then, by (6.17),

$$\mathrm{E}\left[\phi_i(\mathsf{X}_i)\right] = \mathrm{E}\left[\frac{1}{n}\sum_{t \in [n]} \phi(\mathsf{X}_{it})\right] \leq \delta_n \cdot \mathbf{1}$$

by the sample average constraint (5.3) as desired.[5] We can now complete the mixing argument by applying the optimal solution $\bar{P}_{\mathsf{X}_V^n}$ to (6.14) as follows.

[5] This should not be confused with the fact that convergence in probability is weaker than convergence in expectation. The sample average constraint requires the average to be upper bounded with probability one, instead of probability converging to one.

Mixing argument:

$$
\begin{aligned}
\sup \frac{1}{n} \log|K| &\leq \inf_{\lambda} \min_{i \in D^c} \frac{1}{n} \sum_{t \in [n]} \alpha_i(\lambda, \bar{P}_{\mathsf{X}_{Vt}}) + \delta'_n \\
&\overset{\text{(a)}}{=} \inf_{\lambda} \min_{i} \alpha_i(\lambda, \bar{P}_{\mathsf{X}_V}) + \delta'_n && \text{by (6.17)} \\
&\overset{\text{(b)}}{\leq} \inf_{\lambda} \sup_{P_{\mathsf{X}_V}} \min_{i} \alpha_i(\lambda, P_{\mathsf{X}_V}) + \delta'_n
\end{aligned}
$$

where P_{X_V} in the last supremum is subject to the moment constraint (6.3).

(a) by (6.17) that the marginal distributions of $\bar{P}_{\mathsf{X}_{Vt}}$ are identical over t.

(b) This is because the marginal distribution $\bar{P}_{\mathsf{X}_{Vt}}$ satisfies the moment contraint (6.3) as shown earlier. The supremum is placed between the two minimizations because $\bar{P}_{\mathsf{X}_V}$ is a function of λ but not i.

Finally, taking $\limsup_{n \to \infty}$ on both sides give the desired bound (6.5). We can also rewrite the infimum in λ as the minimum since $\sup_{P_{\mathsf{X}_V}} \min_i \alpha_i(\lambda, P_{\mathsf{X}_V})$ is continuous in λ over the compact set $\Lambda_{A|D}$. We conclude this section with the following problem concerning the tightness of this bound.

Problem:

Can the secrecy upper bound for the general case be improved to (6.1) with the input distribution subject to the moment constraint in (6.3)? n.b. this holds if one could prove quasi-concavity for α in the input distribution.

The improvement, if possible, is strict since there exists examples for which the weakening from α to α_i in the current proof is strict. e.g. consider the DMMC,

$$
\mathsf{Y}_3 = (\mathsf{X}_1, \mathsf{X}_2) \in \{0, 1\}^2
$$

with active terminals T_1 and T_2 being the only transmitters, and untrusted terminal 3 being the only receiver. The secrecy upper bound (6.1) gives 0 but (6.5) gives 1.

6.2 Secrecy lower bound

In this section, we will derive a single-letter lower bound on the secrecy capacity using a new type of secrecy schemes called the *mixed source emulation approach*, where

1. we effectively turn the DMMC into a DMMS by generating the channel input independently over time, and
2. we mix between different DMMS's by publicly randomizing the input distributions at each time t.

This is motivated partly by the optimality of the *pure source emulation approach* in [13], and partly by the idea of mixed strategy [56] in zero-sum games. The pure source emulation approach is a special case of the mixed source emulation with a fixed input distribution. It achieves the secrecy capacity for the broadcast-type DMMC in [13], where only one channel input is allowed. For the more general DMMC with multiple channel inputs from different terminals, however, we will prove that the mixed source emulation approach strictly improves over the pure source emulation approach. As will be illustrated in Section 6.3 with a concrete example, mixing over different input distributions allows the terminals to coordinate with each other by correlating their channel inputs through public discussion. This additional coordination gives rise to a larger secret key rate.

For simplicity, we first consider the finitely-valued model when all the input and output symbols of the private DMMC are subject to the finite-alphabet constraint only. We then extend the result to the more general infinitely-valued model with sample average constraint (5.3) by the usual quantization trick.

6.2.1 Finitely-valued model

Theorem 6.3 (Finitely-valued model) *For the finitely-valued case where all channel inputs and outputs are subject to the finite-alphabet constraint only, we have the* *secrecy lower bound expressed in terms of* $\tilde{\beta}$ *in* (A.18) *as*

$$C_{\mathrm{s}} \geq \max_{\substack{P_{\mathsf{Q}\mathsf{X}_V}=P_{\mathsf{Q}\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D\mathsf{Q}}}} \min_{\lambda\in\Lambda_{A|D}} \mathrm{E}_{\mathsf{Q}}\left[\tilde{\beta}(\lambda, P_{\mathsf{X}_V|\mathsf{Q}}(\cdot|\mathsf{Q}))\right] \tag{6.18a}$$

$$= \min_{\lambda\in\Lambda_{A|D}} \max_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \qquad \textit{(mse)} \tag{6.18b}$$

$$\geq \max_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \min_{\lambda\in\Lambda_{A|D}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \qquad \textit{(pse)} \tag{6.18c}$$

The bound (6.18b) *is the largest (strongly) achievable key rate for the mixed source emulation approach in Definition 6.2 below, and more generally, any public input adaptation scheme in Definition 6.3. Furthermore, it is admissible to have the alphabet set Q of the auxiliary source component* Q *satisfy the cardinality bound in* (A.12a). *The weakened bound* (6.18c) *is the largest (strongly) achievable key rate for the pure source emulation where* Q *is chosen to be deterministic.* □

Example 6.3 Consider the same model defined in Example 6.1 for two transmitting active users and one receiving untrusted terminal. Assume in addition that Y is finitely-valued. Then the secrecy lower bound (6.18b) simplifies to

$$\max_{P_{\mathsf{X}_1\mathsf{X}_2}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} I(\mathsf{X}_1 \wedge \mathsf{X}_2|\mathsf{Y})$$

n.b. this is similar to the secrecy upper bound in Example 6.1 except that we require the input to be independent instead of subtracting $I(\mathsf{X}_1 \wedge \mathsf{X}_2)$ from the conditional mutual information expression. For the binary multiple access channel $\mathsf{Y} = \mathsf{X}_1 \oplus \mathsf{X}_2$, the optimal input distribution is $P_{\mathsf{X}_1\mathsf{X}_2} = \mathrm{Bern}_{\frac{1}{2}}\,\mathrm{Bern}_{\frac{1}{2}}$. The secrecy lower bound is 1 bit which matches the upper bound. □

We now describe mixed source emulation approach in greater details.

Definition 6.2 (Source emulation for finitely-valued model) The mixed source emulation approach for the finitely-valued model is the following specialization of the secrecy scheme in Definition 5.2.

Mixed source emulation:

Public randomization: Terminal 1 publicly randomizes

$$\mathsf{U}_0 = (\mathsf{Q}^n, \mathsf{X}_D^n)$$

where each symbol $(\mathsf{Q}_t, \mathsf{X}_{Dt})$ is iid as $P_{\mathsf{Q},\mathsf{X}_D}$ over $t \in [n]$. Q^n is called an auxiliary iid source component, taking values from an arbitrary finite alphabet.

Private randomization: Every trusted terminal $i \in D^c$ privately generates X_i^n such that its symbol X_{it} is iid as $P_{\mathsf{X}_i|\mathsf{X}_D,\mathsf{Q}}(\cdot|\mathsf{X}_{Dt}, \mathsf{Q}_t)$ over $t \in [n]$.

Private channel use: X_{it} is sent as the input to the private DMMC at time $t \in [n]$ from terminal $i \in V$.

Since the channel input does not adapt to the accumulated knowledge by definition, it is unnecessary to perform any public discussion before the last private channel use at time n. The key generation phase proceeds as usual.

The pure source emulation approach is a special case of the mixed source emulation approach with Q being deterministic. □

This is called the source emulation approach since the channel input is chosen to be memoryless, which effectively turns the DMMC into a DMMS. The auxiliary random variable Q acts as a mixing random variable that mixes different conditional input distributions $P_{\mathsf{X}_V|\mathsf{Q}}$ in time. It can also be regarded as an auxiliary component source of a dummy untrusted terminal since Q^n is known in public. It gives an additional correlation among the input sequences privately generated by the trusted terminals.

$$P_{\mathsf{X}_V} = \mathrm{E}_{\mathsf{Q},\mathsf{X}_D}\left[P_{\mathsf{X}_D} \prod_{i \in V} P_{\mathsf{X}_i|\mathsf{X}_D,\mathsf{Q}}(\cdot|\cdot, \mathsf{Q})\right] \not\equiv P_{\mathsf{X}_D} \prod_{i \in V} P_{\mathsf{X}_i|\mathsf{X}_D}$$

Definition 6.3 (Public input adaptation) If we allow the channel input to adapt to any public information, we have the public input adaptation approach.

Public input adaptation:

Set the private randomization as

$$\mathsf{U}_i := (\mathsf{U}_{it} : t \in [n])$$

such that U_{it}'s are independent over i and t. Then, the input is chosen as

$$\mathsf{X}_{it} = X_{it}(\mathsf{U}_0, \mathsf{U}_{it}, \mathsf{F}^{t-1}) \tag{6.19}$$

Thus, X_{it} are conditionally independent over $i \in D^c$ and of the past given the accumulated public information $(\mathsf{U}_0, \mathsf{F}^{t-1})$. Mixed source emulation is a special case of public input adaption without the dependence on F^{t-1} in (6.19). □

Proof (Proof of Theorem 6.3) The mixed source emulation effectively turns the DMMC into a DMMS where terminal $i \in V$ observes the source $(\mathsf{X}_i, \mathsf{Y}_i)$ and a dummy untrusted terminal observes the auxiliary source Q. By [12], the secrecy capacity for this specialized model is given by (6.18a), which is also strongly achievable. This can be used as a lower bound for the secrecy capacity of the general model. Since $\tilde{\beta}(\lambda, P_{\mathsf{X}_V})$ is linear and continuous in λ over the compact set $\Lambda_{A|D}$, we can apply the minimax-type Lemma A.2 to obtain the secrecy lower bound

$$\sup \liminf_{n\to\infty} \frac{1}{n} \log|K| \geq \min_{\lambda \in \Lambda_{A|D}} \sup_{\substack{P_{\mathsf{X}_V} = P_{\mathsf{X}_D} \times \\ \times \prod_{i \in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})$$

Since $\tilde{\beta}(\lambda, P_{\mathsf{X}_V})$ is continuous in P_{X_V} over the compact set $\mathscr{P}(X_V)$ due to the finitely-valued model assumption, we can replace sup by max to obtain (6.18b) as desired. With the additional fact that P_{X_V} is connected, the range

$$\{\tilde{\beta}(\lambda, P_{\mathsf{X}_V}) : P_{\mathsf{X}_V} \in \mathscr{P}(X_V)\}$$

is also connected, and so by the Support Lemma A.3, it is admissible to bound the cardinality of Q as in (A.12).

It remains to show that the lower bound is the maximum key rate achievable by a public input adaptation scheme. We do so by showing that the secrecy upper bound (6.1) matches the lower bound under the additional constraint (6.19). First, note that it does not loose optimality to reveal $(\mathsf{X}_D^n, \mathsf{Y}_D^n)$ in public since they are known to the wiretapper. Thus, we can assume F^{t-1} determines $(\mathsf{U}_D, \mathsf{Y}_D^{t-1})$, and redefine Q_t in (6.11) as

$$\mathsf{Q}_t := (\mathsf{F}^{t-1}, \mathsf{U}_0)$$

By (6.19), X_{it}'s are conditionally independent given Q_t. This allows us to impose the additional (conditional) independence condition (A.26) in the secrecy upper bound (6.1), which then matches the lower bound as desired by equivalence relation (a) in Proposition A.4. ■

6.2.2 Infinitely-valued model

The proof for Theorem 6.3 does not immediately extend to the case when some of the channel inputs and outputs can be infinitely-valued. This is because the secrecy capacity for the source model is achieved in [12] by first attaining omniscience at the active terminals. i.e. the active terminals recover all the source components asymptotically losslessly. This cannot be done for the continuous-valued component unless one uses an appropriate fidelity criteria [52]. The method of types [10] arguments for the strong achievability result also rely on the finite-alphabet constraint. It does not apply directly to infinitely-valued random variables.

Fortunately, it is not essential to attain omniscience for the purpose of generating a secret key. We will simply convert the infinitely-valued model to a finitely-valued model by quantization. Given that the quantization is fine enough, and the entropy measure is well-defined for the source, we can asymptotically capture the correlation among the sources needed for generating the secret key. Indeed, we will show that uniform quantization suffices.

In the sequel, we first extend the secrecy capacity of the finitely-valued source model in [12] to the infinitely-valued source model. Then, in the next section, we derive the secrecy lower bound to the infinitely-valued channel model with sample average constraint (5.3).

Theorem 6.4 (Infinitely-valued source model) *For the source model where the terminals observe the private DMMS* Y_V*, the secrecy capacity is*

$$C_{\mathrm{s}} = \min_{\lambda \in \Lambda_{A|D}} \beta(\lambda, P_{\mathsf{Y}_V}) \tag{6.20}$$

and is strongly *achievable, where* β *is defined in* (A.19) *and* Y_V *can be a mixture of discrete and continuous random variables as described in Section A.1 such that the entropy measure is well-defined* (A.2). □

We convert the infinitely-valued model to a finitely-valued model using the following quantizer.

Quantizer:

For $b > \Delta > 0 : b \setminus \Delta \in \mathbb{P}$, define the quantization function $f_{\Delta,b} : \mathbb{R} \mapsto \{0, \dots, 2b/\Delta\}$ as follows,

$$f_{\Delta,b}(y) := \begin{cases} 0 & \text{if } y \notin [-b, b) \\ j & \text{if } y \in [-b + (j-1)\Delta, -b + j\Delta) \end{cases} \tag{6.21a}$$

The range of the quantizer is finite as desired,

$$\|f_{\Delta,b}\| = 1 + \frac{2b}{\Delta} < \infty \tag{6.21b}$$

Figure 6-1 illustrates how the quantization turns a continuous random variable into a finitely-valued random variable.

We can apply this quantizer to each output symbol of the infinitely-valued component of the source,[6] leading to a finitely-valued component. For notational simplicity,

[6] For discrete component with unbounded support, we can assume the support set is the set $\mathbb{Z}$ of integers without loss of generality.

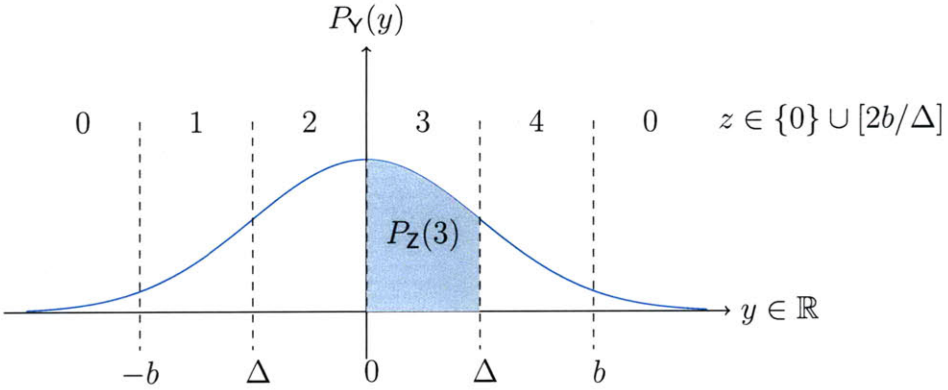


Figure 6-1: Quantization $Z := f_{\Delta,b}(\mathsf{Y})$ of a continuous random variable Y. The boundaries are at $\{j\Delta \in [-b,b] : j \in \mathbb{Z}\} \cup \{-b,b\}$

we use $f_{\Delta,b}(\mathsf{Y}_V)$ to denote the output with all infinitely-valued components in Y_V quantized by $f_{\Delta,b}$ but leaving the finitely-valued components intact.

Proof (Proof of Theorem 6.4) The converse follows immediately from Theorem 6.1 because having the private DMMS Y_V is equivalent to having the private DMMC with $P_{\mathsf{Y}_V|\mathsf{X}_V} = P_{\mathsf{Y}_V}$, which implies by definitions (A.17) and (A.19) that

$$\alpha(\lambda, P_{\mathsf{X}_V}) = \beta(\lambda, P_{\mathsf{Y}_V})$$

Substituting this into (6.1) gives (6.20) as desired.

To show that (6.20) is achievable, consider the specific scheme:

1. Each terminal $i \in V$ quantizes its private component source Y_i by $f_{\Delta,b}$ to

$$\mathsf{Z}_i := f_{\Delta,b}(\mathsf{Y}_i)$$

2. Terminal $i \in V$ broadcasts the indicator $\mathbb{1}\{\mathsf{Z}_i \neq 0\}$ in public. This is equivalent to having a dummy untrusted terminal 0 that observes the vector of indicators

$$\mathbf{Z}_0 := (\mathbb{1}\{\mathsf{Z}_i \neq 0\} : i \in V)$$

Let $b \to \infty$ slowly as $\Delta \to 0$ slowly as $n \to \infty$ such that the cardinality of the output of the quantizer, i.e. $2b/\Delta$ from (6.21b), grows sufficiently slowly for the achievability scheme in [12] to apply. The resulting strongly achievable key rate is[7]

$$\begin{aligned}\min_{\lambda}\lim_{b\to\infty}\lim_{\Delta\to 0}\mathrm{E}_{\mathbf{Z}_0}\left[\beta(\lambda, P_{\mathsf{Z}_V|\mathbf{Z}_0}(\cdot|\mathbf{Z}_0))\right]&\\ \geq \min_{\lambda}\lim_{b\to\infty}\lim_{\Delta\to 0} P_{\mathbf{Z}_0}(\mathbf{1})\,\mathrm{E}_{\mathbf{Z}_0}&\left[\beta(\lambda, P_{\mathsf{Z}_V|\mathbf{Z}_0}(\cdot|\mathbf{Z}_0=\mathbf{1}))\right]\end{aligned}$$

since β is non-negative by (A.10a) of the Shearer-type lemma. Note that $P_{\mathbf{Z}_0}(\mathbf{1})\to 1$ in the above limit by (A.2).[8] It suffices to show that the conditional expectation converges to $\beta(\lambda, P_{\mathsf{Y}_V})$.

Let $\ell(C)$ for $C\subseteq V$ be the number of continuous components in Y_C, which satisfies

$$\ell(B_1)+\ell(B_2)=\ell(B_1\cap B_2)+\ell(B_1\cup B_2)\qquad\text{for all } B_1,\, B_2\subseteq V.$$

The *modularity* implies the equality case (A.10b) of the Shearer-type lemma

$$0=\sum_B \lambda_B\ell(B^c)-\ell(D)-\left(\sum\nolimits_B\lambda_B-1\right)\ell(V)$$

Together with the definition of β in (A.19), we have

$$\begin{aligned}\mathrm{E}_{\mathbf{Z}_0}\left[\beta(\lambda, P_{\mathsf{Z}_V|\mathbf{Z}_0}(\cdot|\mathbf{Z}_0=\mathbf{1}))\right]=&\sum_B\lambda_B\left[H(\mathsf{Z}_{B^c}|\mathbf{Z}_0=\mathbf{1})+\ell(B^c)\log\Delta\right]\\ &-\left[H(\mathsf{Z}_D|\mathbf{Z}_0=\mathbf{1})+\ell(D)\log\Delta\right]\\ &-\left(\sum\nolimits_B\lambda_B-1\right)\left[H(\mathsf{Z}_V|\mathbf{Z}_0=\mathbf{1})+\ell(V)\log\Delta\right]\end{aligned}$$

By Corollary 6.1 stated below,

$$\lim_{b\to\infty}\lim_{\Delta\to 0}\left[H(\mathsf{Z}_C|\mathbf{Z}_0=\mathbf{1})+\ell(C)\log\Delta\right]=H(\mathsf{Y}_C)\tag{6.22}$$

Applying (6.22) to the previous expression gives the desired convergence. ■

The above proof uses the following technical Lemma, which is analogous to Theorem 9.3.1 in [8].

[7] By making $n\to\infty$ fast compared to $\Delta\to 0$ and $b\to\infty$, we can approximate the rate by taking the limit in n, followed by Δ and b.

[8] See the proof of Lemma 6.1 for a detailed derivation.

Lemma 6.1 (Quantization) *Let* Y *be a real-valued random variable with density function* P_{Y} *such that*

$$\int_{-\infty}^{\infty} |P_{\mathsf{Y}}(y)\log P_{\mathsf{Y}}(y)|\, \mathrm{d}y < \infty \tag{6.23}$$

then we have for $\mathsf{Z} := f_{\Delta,b}(\mathsf{Y})$ *and* $\mathsf{Z}_0 := \mathbb{1}\{\mathsf{Z} \neq 0\}$ *that*

$$\lim_{b\to\infty}\lim_{\Delta\to\infty} [H(\mathsf{Z}|\mathsf{Z}_0 = 1) + \log\Delta] = H(\mathsf{Y}) \tag{6.24}$$

where $f_{\Delta,b}$ *is a quantizer defined in* (6.21). □

Corollary 6.1 *Given* $\mathsf{Y} = (\mathsf{Y}_i : i \in [\ell+1])$ *is a mixture of a discrete random variable* $\mathsf{Y}_{\ell+1}$ *and a continuous random vector* $\mathsf{Y}_{[\ell]}$ *with* ℓ *continuous real-valued components, such that the joint densities* $P_{\mathsf{Y}_{[\ell+1]}}(\cdot, y_{\ell+1})$ *for* $y_{\ell+1} \in Y_{\ell+1}$ *are absolutely continuous and*

$$\sum_{y_{\ell+1}\in Y_{\ell+1}} \int_{\mathbb{R}^\ell} \left|P_{\mathsf{Y}_{[\ell]}}(y_{[\ell]})\log P_{\mathsf{Y}_{[\ell]}}(y_{[\ell]})\right| \mathrm{d}y_{[\ell]} < \infty \tag{6.25}$$

We have for $\mathsf{Z} := f_{\Delta,b}(\mathsf{Y})$ *and* $\mathbf{Z}_0 := (\mathbb{1}\{\mathsf{Z}_i \neq 0\} : i \in [\ell])$ *that*

$$\lim_{b\to\infty}\lim_{\Delta\to\infty} [H(\mathsf{Z}|\mathbf{Z}_0 = \mathbf{1}) + \ell\log\Delta] = H(\mathsf{Y}) \tag{6.26}$$

where $f_{\Delta,b}$ *is a quantizer defined in* (6.21). □

PROOF Consider proving Lemma 6.1 first. We first relate the conditional distributions of Z and Y given $\mathsf{Z}_0 = 1$ using the mean-value theorem, and then prove the desired convergence (6.24) under the condition (6.23) for the entropy measure to be well-defined. From the definition (6.21), we have for all $j \in [2b/\Delta]$ such that $P_{\mathsf{Y}|\mathsf{Z}_0}(\cdot|1)$ is continuous over the interval $[-b + (j-1)\Delta, -b + j\Delta]$,

$$P_{\mathsf{Z}}(j) = \int_{-b+(j-1)\Delta}^{-b+j\Delta} P_{\mathsf{Y}|\mathsf{Z}_0}(y|1)\,\mathrm{d}y = P_{\mathsf{Y}|\mathsf{Z}_0}(y_j|1)\Delta \tag{6.27}$$

for some tag $y_j \in [-b + (j-1)\Delta, -b + j\Delta]$ by the mean-value theorem. The desired convergence can be proved in two stages as follows.

$$\begin{aligned}
H(\mathsf{Z}|\mathsf{Z}_0=1)+\log\Delta &= \sum_{j\in[2b/\Delta]} P_{\mathsf{Z}|\mathsf{Z}_0}(j|1)\log\frac{\Delta}{P_{\mathsf{Z}|\mathsf{Z}_0}(j|1)}\\
&\xrightarrow{(a)\ \Delta\to 0} \int_{-b}^{b} P_{\mathsf{Y}|\mathsf{Z}_0}(y|1)\log\frac{1}{P_{\mathsf{Y}|\mathsf{Z}_0}(y|1)}\,\mathrm{d}y\\
&\xrightarrow{(b)\ b\to\infty} H(\mathsf{Y})
\end{aligned}$$

(a) This is by (6.23) that $P_{\mathsf{Y}|\mathsf{Z}_0}(\cdot|1)\log\frac{1}{P_{\mathsf{Y}|\mathsf{Z}_0}(\cdot|1)}$ is Riemann-integrable over $[-b,b]$. More explicitly, we have the convergence that

$$\begin{aligned}
\int_{-b}^{b} P_{\mathsf{Y}|\mathsf{Z}_0}(y|1)\log\frac{1}{P_{\mathsf{Y}|\mathsf{Z}_0}(y|1)}\,\mathrm{d}y &= \lim_{\Delta\to 0}\sum_{j\in[2b/\Delta]} \Delta P_{\mathsf{Y}|\mathsf{Z}_0}(y_j|1)\log\frac{1}{P_{\mathsf{Y}|\mathsf{Z}_0}(y_j|1)}\\
&= \lim_{\Delta\to 0}\sum_{j\in[2b/\Delta]} P_{\mathsf{Z}|\mathsf{Z}_0}(j|0)\log\frac{\Delta}{P_{\mathsf{Z}|\mathsf{Z}_0}(j|1)}
\end{aligned}$$

where the last equality is by (6.27).[9]

(b) Since (6.23) implies

$$\lim_{b\to\infty}\int_{b}^{\infty}\left[P_{\mathsf{Y}}(y)+P_{\mathsf{Y}}(-y)\right]\,\mathrm{d}y = 0$$

we have $\lim_{b\to\infty}\Pr(\mathsf{Z}_0)=0$. Since $\lim_{x\downarrow 0}x\ln x=0$, we also have $P_{\mathsf{Z}_0}(0)\log P_{\mathsf{Z}_0}(0)\to 0$ and $H(\mathsf{Z}_0)\to 0$ as $b\to\infty$. By the chain rule,

$$\begin{aligned}
H(\mathsf{Y})-H(\mathsf{Y}|\mathsf{Z}_0) &= H(\mathsf{Z}_0)-H(\mathsf{Z}_0|\mathsf{Y})\\
&\leq H(\mathsf{Z}_0)\xrightarrow{b\to\infty} 0
\end{aligned}$$

Thus, R.H.S. of (a) can be expressed as

$$H(\mathsf{Y}|\mathsf{Z}_0=1)\approx\frac{1}{P_{\mathsf{Z}_0}(1)}\left[H(\mathsf{Y})-P_{\mathsf{Z}_0}(0)H(\mathsf{Y}|\mathsf{Z}_0=0)\right]$$

with equality in the limit as $b\to\infty$. To show the desired convergence to $H(\mathsf{Y})$,

[9] More precisely, the mean value theorem applies almost everywhere as $\Delta\to 0$ due to the fact that a function is Riemann integrable iff it is continuous almost everywhere [32]. Absolute continuity of the probability measure should not be confused with continuity of the density function.

it suffices to show that the following term converges to 0.

$$\begin{aligned}
P_{\mathsf{Z}_0}(0)H(\mathsf{Y}|\mathsf{Z}_0=0) &= P_{\mathsf{Z}_0}(0)\int_{(-\infty,b]\cup[b,\infty)} \frac{P_{\mathsf{Y}}(y)}{P_{\mathsf{Z}_0}(0)}\log\frac{P_{\mathsf{Z}_0}(0)}{P_{\mathsf{Y}}(y)}\,\mathrm{d}y \\
&= \int_{(-\infty,b]\cup[b,\infty)} P_{\mathsf{Y}}(y)\log\frac{P_{\mathsf{Z}_0}(0)}{P_{\mathsf{Y}}(y)}\,\mathrm{d}y \\
&= P_{\mathsf{Z}_0}(0)\log P_{\mathsf{Z}_0}(0) + \int_{(-\infty,b]\cup[b,\infty)} P_{\mathsf{Y}}(y)\log\frac{1}{P_{\mathsf{Y}}(y)}\,\mathrm{d}y
\end{aligned}$$

which goes to zero as $b \to \infty$ by (6.23)

This complete the proof of Lemma 6.1. Corollary 6.1 is a straight-forward extension to the vector case. We again relate the distributions of Z and Y through the mean-value theorem, with an ℓ-fold integral and a factor of Δ^ℓ instead of Δ in (6.27). This gives the $\ell \log \Delta$ terms in (6.26). ■

6.2.3 Sample Average Constraint

We now incorporate the sample average constraint (5.3) into the input distribution. To do so, we consider a modified mixed source emulation approach.

Definition 6.4 The modified mixed source emulation approach for the sample average constraint is the same as the mixed source emulation approach in Definition 6.2 but with the following modifications.

Modifications to mixed source emulation:

i) $P_{\mathsf{X}_V}(\cdot|q)$ is chosen to satisfy the moment constraint (6.3) for all $q \in Q$.

ii) Before the transmission phase, if terminal $i \in V$ finds that the sample average constraint is violated for its input sequence X_i^n, it declares an *outage*.

iii) If any terminal declares an outage, they skip the transmission phase entirely. In this case, the active terminals simply generate the individual keys K_i's for $i \in A$ independently and uniformly randomly.

□

Theorem 6.5 (Sample average constraint) *With the sample average constraint* (5.3) *and the infinitely-valued model, we have the following secrecy lower bound*

$$
\begin{aligned}
C_{\mathrm{s}} &\geq \sup_{\substack{P_{\mathsf{QX}_V}=P_{\mathsf{QX}_D}\times \\ \times\prod_{i\in D^c}P_{\mathsf{X}_i|\mathsf{X}_D\mathsf{Q}}}} \min_{\lambda\in\Lambda_{A|D}} \mathrm{E}_{\mathsf{Q}}\Big[\tilde{\beta}(\lambda, P_{\mathsf{X}_V|\mathsf{Q}}(\cdot|\mathsf{Q}))\Big] && \textbf{(6.28a)}\\
&= \min_{\lambda\in\Lambda_{A|D}} \sup_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c}P_{\mathsf{X}_i|\mathsf{X}_D}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \qquad \textit{(mse)} && \textbf{(6.28b)}\\
&\geq \sup_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c}P_{\mathsf{X}_i|\mathsf{X}_D}}} \min_{\lambda\in\Lambda_{A|D}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \qquad \textit{(pse)} && \textbf{(6.28c)}
\end{aligned}
$$

where the input distribution is also subject to the moment constraint (6.3). (6.28b) *is the largest achievable key rate for the modified mixed source emulation approach in Definition 6.4. It is admissible to have the alphabet set Q of the auxillary source component satisfy the cardinality bound in* (A.12b). (6.28c) *is the largest achievable key rate for the corresponding pure source emulation where* Q *is chosen to be deterministic.* □

Example 6.4 Consider the same model defined in Example 6.2 for two transmitting active users and one receiving untrusted terminal. Then, by (6.28b), the secrecy lower bound simplifies to

$$\sup_{P_{\mathsf{X}_1\mathsf{X}_2}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} I(\mathsf{X}_1\wedge\mathsf{X}_2|\mathsf{Y})$$

where the input may be subject to certain moment constraint that corresponds to the sample average constraint.

Consider, in particular, the gaussian multiple access channel $\mathsf{Y}=\mathsf{X}_1+\mathsf{X}_2+\mathsf{N}$, with channel noise $\mathsf{N}\sim\mathscr{N}_{0,1}$ and average power constraint $f_i(x_i)=x_i^2-P_i$ for $i=1,2$. We can set the channel input distribution to be gaussian, $P_{\mathsf{X}_1\mathsf{X}_2}=\mathscr{N}_{0,P_1}\mathscr{N}_{0,P_2}$, which satisfies the power constraints. This gives the following secrecy lower bound

$$
\begin{aligned}
I(\mathsf{X}_1\wedge\mathsf{X}_2|\mathsf{Y}) &= H(\mathsf{Y}|\mathsf{X}_1)+H(\mathsf{Y}|\mathsf{X}_2)-H(\mathsf{Y})-H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2)\\
&= H(\mathsf{X}_2+\mathsf{N})+H(\mathsf{X}_1+\mathsf{N})-H(\mathsf{Y})-H(\mathsf{N})\\
&= \log\left(1+\frac{P_1P_2}{P_1+P_2+1}\right)
\end{aligned}
$$

As P_2/P_1 or $P_1/P_2 \to \infty$, this approaches the secrecy upper bound $\log(1+\min\{P_1, P_2\})$ derived in Example 6.2 □

PROOF (THEOREM 6.5) Consider the modifications i, ii and iii in Definition 6.4. The idea is that imposing the moment constraint in i ensures the sample average constraint is satisfied with high probability such that the outage event in ii almost surely does not occur. For the purpose of the proof, the terminals do not take further action in case of an outage, and simply generate individual random keys in iii that are independent of everything else, including the wiretapper's knowledge.[10] In other words, the modification does not affect the secrecy condition (5.8) in the sense that if any scheme satisfies the secrecy condition without modification ii and iii, it must also satisfy the condition with the modifications.

Next, we will show that the recoverability condition is unaffected and so we can ignore modifications ii and iii for the purpose of computing the largest achievable key rate. Let $\mathcal{E}_{\mathrm{r}}$ and $\tilde{\mathcal{E}}_{\mathrm{r}}$ be the events that the active terminals fail to agree on the secret key with and respectively without the modifications. Let $\mathcal{E}_{\mathrm{p}}$ and $\tilde{\mathcal{E}}_{\mathrm{p}}$ be the outage event in ii for the case with and respectively without the modifications. Then,

$$\Pr(\tilde{\mathcal{E}}_{\mathrm{r}}) \geq \Pr(\tilde{\mathcal{E}}_{\mathrm{r}}|\tilde{\mathcal{E}}_{\mathrm{p}}^{c})\Pr(\tilde{\mathcal{E}}_{\mathrm{p}}^{c}) \geq \Pr(\mathcal{E}_{\mathrm{r}}|\mathcal{E}_{\mathrm{p}}^{c})\Pr(\mathcal{E}_{\mathrm{p}}^{c}) \tag{6.29}$$

because $\Pr(\tilde{\mathcal{E}}_{\mathrm{r}}|\tilde{\mathcal{E}}_{\mathrm{p}}^{c}) = \Pr(\mathcal{E}_{\mathrm{r}}|\mathcal{E}_{\mathrm{p}}^{c})$ due to the fact that the modifications are ineffective if there is no outage, and $\Pr(\tilde{\mathcal{E}}_{\mathrm{p}}) = \Pr(\mathcal{E}_{\mathrm{p}})$ by definition.[11]

$$\Pr(\mathcal{E}_{\mathrm{r}}) \leq \Pr(\mathcal{E}_{\mathrm{p}}) + \Pr(\mathcal{E}_{\mathrm{r}} \cap \mathcal{E}_{\mathrm{p}}^{c}) \leq \Pr(\mathcal{E}_{\mathrm{p}}) + \Pr(\tilde{\mathcal{E}}_{\mathrm{r}}) \qquad \text{by (6.29)}$$

If $\Pr(\tilde{\mathcal{E}}_{\mathrm{r}}) \to 0$ exponentially for any scheme without the modifications, then $\Pr(\mathcal{E}_{\mathrm{r}}) \to 0$ exponentially since $\Pr(\mathcal{E}_{\mathrm{p}})$ decays to zero exponentially by the Chernoff bound [8]. Thus, the recoverability condition is unaffected as desired. We can therefore ignore modifications ii and iii for the purpose of computing the largest achievable key rate.

[10] In practice, the terminals can regenerate the input repeatedly until the sample average constraint is satisfied.

[11] The probability that the outage event occurs is not affected by any action performed after it occurs.

Applying Theorem 6.4, the maximum key rate strongly achievable is (6.28b) as desired by maximizing over the input distribution satisfying the corresponding moment constraint. The admissibility condition on $|Q|$ follows from Lemma A.3. ■

Problem:

For the gaussian multiple access channel considered in Example 6.4, is the gaussian input distribution globally optimal for the maximization in (6.28b)? It can be shown that gaussian is a local maximum. To see this, note that X_1 independent of X_2 implies the following after some algebra.

$$\begin{aligned} I(\mathsf{X}_1 \wedge \mathsf{X}_2|\mathsf{Y}) &= H(\mathsf{X}_1|\mathsf{X}_1 + \mathsf{X}_2 + \mathsf{N}) - H(\mathsf{N}|\mathsf{X}_1 + \mathsf{N}) \\ &= H(\mathsf{X}_2|\mathsf{X}_2 + \mathsf{X}_1 + \mathsf{N}) - H(\mathsf{N}|\mathsf{X}_2 + \mathsf{N}) \end{aligned}$$

From the first equality, it is optimal to have $P_{\mathsf{X}_2} = \mathcal{N}_{0,P_2}$ given $P_{\mathsf{X}_1} = \mathcal{N}_{0,P_1}$. A similar statement follows from the second equality, which gives the desired local maximality.

6.3 Mixed vs pure source emulations

We will give an example for which the secrecy lower bound obtained from mixed source emulation is 1) strictly larger than the secrecy lower bound from pure source emulation, and 2) strictly smaller than the secrecy upper bound.

To make the first point concrete, we will also describe a variant of the source emulation scheme that illustrates why the mixed source emulation approach outperforms the pure source emulation approach. The public discussion functions and key functions will be completely specified for this variant scheme, unlike the optimal pure or mixed source emulation scheme.

The second point implies two *possibilities*:

1. the secrecy lower bound may be loose, in which case one can somehow improve the key rate by

   Private input adaptation: adapting the channel input to the accumulated observations over time rather than generating them all at the beginning before any private channel use, or

Interactive public discussion: adapting the public message to the previous public messages in multiple rounds of interactions instead of generating the public message completely from the private observations.

2. the secrecy upper bound can perhaps be strictly improved by new techniques.

We will follow up on this later with a simpler example in Section 7.3. For now, consider three terminals consisting of two active terminals and one trusted helper, i.e. $A = [2] \subseteq D^c = V = [3]$, and the DMMC defined below.

Coupling channel:

terminal	1	2	3
input	$\mathsf{X}_1 \in \{0,1\}$	$\mathsf{X}_2 \in \{0,1\}$	
output	$\mathsf{Y}_1 \in \{0,1\}$	$\mathsf{Y}_2 \in \{0,1\}$	$\mathsf{Y}_3 \in \{0,1\}$

The output bits are defined as follows,

$$\mathsf{Y}_3 := \mathsf{N}_3 \tag{6.30a}$$

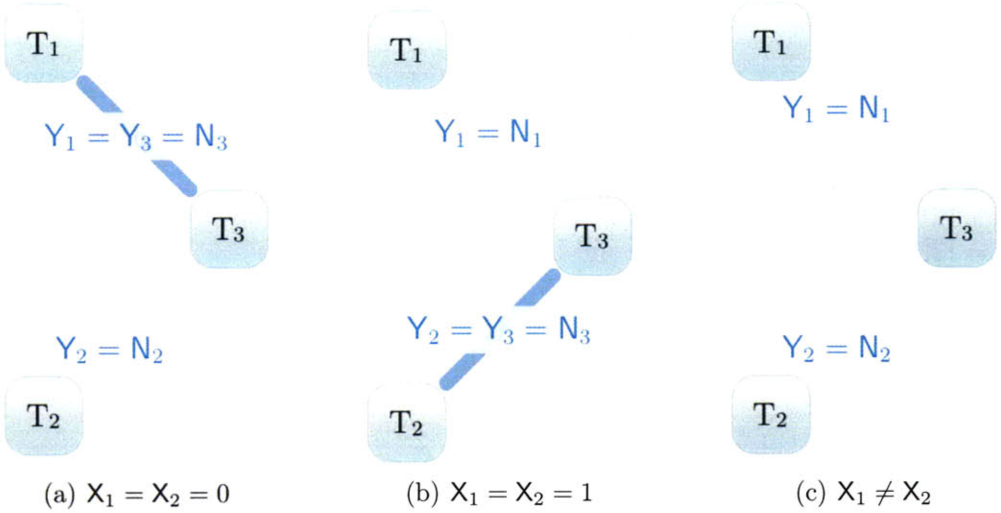
$$\mathsf{Y}_1 := \begin{cases} \mathsf{N}_3 & \text{if } \mathsf{X}_1 = \mathsf{X}_2 = 0 \\ \mathsf{N}_1 & \text{otherwise} \end{cases} \tag{6.30b}$$

$$\mathsf{Y}_2 := \begin{cases} \mathsf{N}_3 & \text{if } \mathsf{X}_1 = \mathsf{X}_2 = 1 \\ \mathsf{N}_2 & \text{otherwise} \end{cases} \tag{6.30c}$$

where $\mathsf{N}_1, \mathsf{N}_2, \mathsf{N}_3$ are uniformly independent random bits mutually independent of the channel input bits $\mathsf{X}_1, \mathsf{X}_2$. As illustrated in Figure 6-2, the active terminals control jointly the coupling of the observations:

1. Y_1 couples with Y_3 if the input bits are 0;
2. Y_2 couples with Y_3 if the input bits are 1;
3. the output bits are all independent if the input bits disagree.

It is beneficial for the active terminals to coordinate their inputs to enhance their correlation.



(a) $X_1 = X_2 = 0$ (b) $X_1 = X_2 = 1$ (c) $X_1 \neq X_2$

Figure 6-2: Coupling channel: each terminal, denoted as T_1, T_2, T_3, observes one of the independent random bits N_1, N_2, N_3 depending on the channel input.

Table 6.1: Secret key rates for the coupling channel

	key rate (bits)
optimal pure source emulation	$C_{\text{pse}} \approx 0.41$
variant of source emulation	$R_{\text{mse}} = 0.5$
optimal mixed source emulation	$C_{\text{mse}} \approx 0.54$
secrecy upperbound	$C_{\text{su}} \approx 0.60$

Table 6.1 summarizes the achievable key rates for the following schemes. The detailed computation can be found in Section C.1.

Optimal pure source emulation:

The active terminals generate the channel input sequence iid according to

$$P_{\mathsf{X}_1\mathsf{X}_2}(x_1, x_2) = \text{Bern}_p(x_1)\,\text{Bern}_{1-p}(x_2), \quad \text{with } p \approx 0.44$$

i.e. X_1 and X_2 are independent Bernoulli random variables.

Optimal mixed source emulation:

The auxillary source and channel input sequence are generated iid according to

$$P_{\mathsf{Q}} = \text{Bern}_{\frac{1}{2}}$$

$$P_{\mathsf{X}_1,\mathsf{X}_2|\mathsf{Q}}(x_1, x_2|q) = \begin{cases} \text{Bern}_0(x_1)\,\text{Bern}_{\frac{2}{17}}(x_2) & \text{if } q = 0, \\ \text{Bern}_{\frac{15}{17}}(x_1)\,\text{Bern}_1(x_2) & \text{if } q = 1 \end{cases}$$

i.e. X_1 and X_2 are conditionally independent Bernoulli random variables given the uniformly random bit Q.

The specific choices of the public discussion and key functions are not known even for this particular example. To help understand more concretely why the optimal mixed source emulation approach outperforms the pure source emulation approach, we consider the following variant scheme for which the public discussion and key functions are completely specified.

Variant scheme:

The active terminals set their channel inputs equal to the parity of the time t, i.e.

$$\mathsf{X}_{1t} = \mathsf{X}_{2t} = \begin{cases} 0 & \text{if } t \text{ is odd} \\ 1 & \text{if } t \text{ is even} \end{cases}$$

This can be considered as a trivial public input adaptation approach defined in Definition 6.3 since the terminals adapt the input only to trivial public information, namely t. During the public discussion, the helper T_3 broadcasts the XOR bits

$$\mathsf{F} = (\mathsf{Y}_{3t} \oplus \mathsf{Y}_{3(t+1)} : t \text{ is odd})$$

The key is set to be

$$\mathsf{K} = (\mathsf{Y}_{3t} : t \text{ is odd})$$

which is uniformly distributed and independent of F as desired. Furthermore, K is observed by T_1 through Y_{1t} for odd t, and perfectly recoverable by T_2 using the bitwise XOR operation,

$$\mathsf{F} \oplus (\mathsf{Y}_{2(t+1)} : t \text{ is odd}) = \mathsf{K}$$

because $\mathsf{Y}_{2(t+1)} = \mathsf{Y}_{3(t+1)}$ for odd t. The key rate achieved is therefore 0.5 bits as shown in Table 6.1.

We now relate this variant scheme to each source emulation approach to show that *mixed source emulation outperforms pure source emulation by the additional coordination through the auxillary component source.* Consider the 2-extended coupling channel defined as

$$P_{\bar{\mathsf{Y}}_V|\bar{\mathsf{X}}_V}((y_V^{\text{odd}}, y_V^{\text{even}})|(x_V^{\text{odd}}, x_V^{\text{even}})) = P_{\mathsf{Y}_V|\mathsf{X}_V}(y_V^{\text{odd}}|x_V^{\text{odd}})P_{\mathsf{Y}_V|\mathsf{X}_V}(y_V^{\text{even}}|x_V^{\text{even}})$$

i.e. each channel use corresponds to two simultaneous uses of the original coupling channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$. The variant scheme can then be considered as a pure source emulation scheme with $n/2$ uses of the 2-extended coupling channel, and the trivial iid input $\bar{\mathsf{X}} = (0, 1)$ with probability 1. The improvement on the original pure source emulation approach comes from the additional coordination through this 2-block memory.

The same coordination can come from the auxillary component source instead

$$\mathsf{X}_1 = \mathsf{X}_2 = \mathsf{Q} \sim \mathrm{Bern}_{\frac{1}{2}}$$

By the large deviation theory, the fraction of time where the input bits are 0 is arbitrarily close to 1/2 with probability exponentially converging to 1. The same holds for the condition where the input bits are 1. Let $\mathsf{Y}^{(q)}$ for $q \in \{0, 1\}$ be vectors of $1/2 - \delta_n$ output bits Y_{3t} at disjoint time t, with as many of them satisfying $\mathsf{X}_{1t} = \mathsf{X}_{2t} = q$ as possible, and $\delta_n \to 0$ at sufficiently slow rate, say $\omega(1/n)$.[12] T$_3$ reveals the following elementwise XOR bits in public,

$$\mathsf{F} = \mathsf{Y}^{(0)} \oplus \mathsf{Y}^{(1)}$$

The key is set to be $\mathsf{K} = \mathsf{Y}^{(0)}$, which is independent of F and Q^n. By the large deviation theory, $\mathsf{Y}^{(q)}$ almost surely consists of bits at time t where $\mathsf{X}_{1t} = \mathsf{X}_{2t} = q$. Thus, it is almost surely observed by terminal 1 through Y_{1t} at time t where $\mathsf{X}_{1t} = \mathsf{X}_{2t} = 0$ and recoverable by terminal 2 from the public message and its private observation Y_{2t} at time t where $\mathsf{X}_{1t} = \mathsf{X}_{2t} = 0$. This mixed source emulation is therefore almost surely the same as the variant scheme under a reordering of the time index. It achieves the same coordination that improves the pure source emulation approach, but with the auxillary component source instead of the 2-block memory in the channel input.

Problem:

Is the maximum key rate achievable by pure source emulation with block memory the same as that achievable by mixed source emulation?

[12] $\omega(1/n)$ refers to the set of function that dominates $1/n$ asymptotically. See Bachmann-Lantau notation for details.

Chapter 7

Tightness of Secrecy Bounds

If a secrecy upper bound in section 6.1 matches a secrecy lower bound in section 6.2, we have a single-letter characterization of the secrecy capacity. We will give a necessary condition for tightness in Section 7.1, some general sufficient conditions for tightness in Section 7.2, and illustrate how the bounds can be loose in Section 7.3.

7.1 Optimality of single-letter form

Before studying the tightness of the secrecy bounds, we will introduce a weaker notion of optimality, called the optimality of single-letter form, without which the secrecy bounds cannot be tight. Roughly speaking, we say that a single-letter bound is single-letter optimal if multiletterizing it does not improve the bound.

Multiletterization by channel extension:

Given a function f on the DMMC $P_{\mathsf{Y}_V|\mathsf{X}_V}$ and a positive integer $k \in \mathbb{P}$, the *k-letter form* of f is defined as

$$f^{(k)}(P_{\mathsf{Y}_V|\mathsf{X}_V}) := \frac{1}{k} f(P^k_{\mathsf{Y}_V|\mathsf{X}_V}) \tag{7.1}$$

where $P^k_{\mathsf{Y}_V|\mathsf{X}_V}$ is the *k-extension* of $P_{\mathsf{Y}_V|\mathsf{X}_V}$ defined as

$$P^k_{\mathsf{Y}_V|\mathsf{X}_V}(y^k_V|x^k_V) := \prod_{\tau\in[k]} P_{\mathsf{Y}_V|\mathsf{X}_V}(y_{V\tau}|x_{V_\tau}) \tag{7.2}$$

with any additional constraints on the input X_V such as the moment constraint (6.3) translated directly to the constraints on $\mathsf{X}_{V[k]}$.

Definition 7.1 (Single-letter optimality) A function f of DMMC $P_{\mathsf{Y}_V|\mathsf{X}_V}$ is called *single-letter maximal* or a *single-letter optimal lower bound* (*single-letter minimal* or a *single-letter optimal upper bound*) if f is no smaller (no larger) than its k-letter form $f^{(k)}$ defined in (7.1) for all $k \in \mathbb{P}$. □

Any of the secrecy expressions α, $\tilde{\beta}$, α_i and γ defined in Section A.4 maximized over any set of input distributions is single-letter minimal because the k-letter form is equal to the single-letter form when we impose an additional memorylessness constraint on the k-letter input distribution $P_{\mathsf{X}_{V[k]}}$ that

$$P_{\mathsf{X}_{V[k]}} = \prod_{\tau \in [k]} P_{\mathsf{X}_{V\tau}} \tag{7.3}$$

Thus, the single-letter secrecy upper bounds characterized in α and α_i are no larger than their multiletter form. i.e. the secrecy upper bounds are single-letter optimal, and are therefore potentially tight.

Theorem 7.1 (Single-letter minimality) *The secrecy upper bounds in Theorem 6.1 and 6.2 are single-letter optimal.* □

PROOF With the additional memorylessness constraint (7.3) on the k-letter input distribution, we have for all $\lambda \in \Lambda_{A|D}$ and $i \in D^c$ that

$$\begin{aligned}
\sup_{P_{\mathsf{X}_{V[k]}}} \frac{1}{k}\alpha_i(\lambda, P_{\mathsf{X}_{V[k]}})\Big|_{P^k_{\mathsf{Y}_V|\mathsf{X}_V}} &\overset{\text{(a)}}{=} \max_{P_{\mathsf{X}_{V[k]}}} \sum_{\tau\in[k]} \frac{1}{k}\alpha_i(\lambda, P_{\mathsf{X}_{V\tau}}) \\
&\overset{\text{(b)}}{=} \frac{1}{k}\sum_{\tau\in[k]} \max_{P_{\mathsf{X}_{V\tau}}} \alpha_i(\lambda, P_{\mathsf{X}_{V\tau}}) \\
&= \max_{P_{\mathsf{X}_V}} \frac{1}{k}\alpha_i(\lambda, P_{\mathsf{X}_V})
\end{aligned}$$

where (a) follows from the definitions (A.20) and (7.2), and (b) follows from the fact that the moment constraint (6.3) is imposed on each marginal input distribution

$P_{\mathsf{X}_{V_\tau}}$. Taking the infimum on both sides over λ and i gives the desired equivalence of the k-letter form on the left and the single-letter form on the right for the secrecy upper bound in Theorem 6.2. Applying similar arguments for α instead of α_i, the secrecy upper bound in Theorem 6.1 is also single-letter optimal. ■

Theorem 7.2 (Single-letter maximality) *The secrecy lower bounds in Theorem 6.3 and 6.5 are single-letter optimal if the DMMC $P_{\mathsf{Y}_V|\mathsf{X}_V}$ satisfies the single-leakage condition* (A.27) *that $P_{\mathsf{Y}_D|\mathsf{X}_V} = P_{\mathsf{Y}_D|\mathsf{X}_{D\cup\{s\}}}$ for some $s \in D^c$.* □

Proof The input distribution $P_{\mathsf{X}_{V[k]}}$ for the k-letter form (7.1) of the secrecy lower bounds (6.18b) and (6.28b) satisfies the conditional independence condition (A.26)

$$P_{\mathsf{X}_{V[k]}} = P_{\mathsf{X}_{D[k]}} \prod_{i\in D^c} P_{\mathsf{X}_{i[k]}}$$

Since the k-extension of the DMMC also satisfies the single-leakage condition, we have, by Proposition A.4, that

$$\begin{aligned}\frac{1}{k}\tilde{\beta}(\lambda, P_{\mathsf{X}_{V[k]}})\Big|_{P^k_{\mathsf{Y}_V|\mathsf{X}_V}} &= \frac{1}{k}\gamma(\lambda, P_{\mathsf{X}_{V[k]}})\Big|_{P^k_{\mathsf{Y}_V|\mathsf{X}_V}}\\ &\overset{\text{(a)}}{=} \frac{1}{k}\sum_{\tau\in[k]}\gamma(\lambda, P_{\mathsf{X}_{V_\tau}})\\ &\overset{\text{(b)}}{=} \frac{1}{k}\sum_{\tau\in[k]}\tilde{\beta}(\lambda, P_{\mathsf{X}_{V_\tau}})\end{aligned}$$

where (a) is by Proposition A.5, (b) is again by Proposition A.4. Maximizing over $P_{\mathsf{X}_{V[k]}}$ as a function of λ and minimizing over λ on both sides give the desired equivalence of the k-letter form on the left and single-letter form on the right, as in the proof of Theorem 7.1. ■

Problem:

Construct an example, if any, for which the secrecy lower bounds are not single-letter optimal.

7.2 Optimality conditions for source emulation

In this section we will show that the secrecy lower bound by source emulation in Section 6.2 matches the upper bound in Section 6.1 for a wide range of channels. In other words, source emulation is optimal and gives the desired single-letter characterization of the secrecy capacity. In the following, we will first derive some general sufficient conditions for tightness of the secrecy bounds. Then, we will study specific classes of channels that satisfy the conditions.

Theorem 7.3 (Finitely-valued model) *For finitely-valued private channel, the secrecy lower bound* (6.18b) *in Theorem 6.3 matches the secrecy upper bound* (6.1) *in Theorem 6.1 if the channel* $P_{\mathsf{Y}_V|\mathsf{X}_V}$ *satisfies*

$$\max_{P_{\mathsf{X}_V}} \alpha(\lambda, P_{\mathsf{X}_V}) = \max_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \alpha(\lambda, P_{\mathsf{X}_V}) \qquad \text{for all } \lambda \in \Lambda_{A|D} \tag{7.4}$$

i.e. α *is maximized by conditionally independent channel inputs given the inputs of the untrusted terminals.* □

Proof With conditionally independent channel input, $\alpha = \tilde{\beta}$ by Proposition A.4 and so

$$\max_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \alpha(\lambda, P_{\mathsf{X}_V}) = \max_{\substack{P_{\mathsf{X}_V}=P_{\mathsf{X}_D}\times \\ \times\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})$$

After minimizing over $\lambda \in \Lambda_{A|D}$, the R.H.S. gives the secrecy lower bound (6.18b), which equals the secrecy upper bound (6.1) given by the L.H.S. of (7.4). ■

The following example illustrates the use of the tightness condition.

Example 7.1 (Binary MAC) Consider two active transmitting terminals and one untrusted receiving terminal who observes $\mathsf{Y} = \mathsf{X}_1 \oplus \mathsf{X}_2$. i.e. $A = [2] = D^c \subsetneq V = [3]$.

We will show that pure source emulation achieves the secrecy capacity of 1 bit. First, $\boldsymbol{\lambda} = \mathbf{1}$ (i.e. $\lambda_{\{1\}} = \lambda_{\{2\}} = 1$) is the only fractional partition in $\Lambda_{A|D}$, and so we have equality for (6.18c). That means pure source emulation is optimal if mixed source emulation is. Next, we prove the sufficient condition (7.4) as follows. By definition (A.17b),

$$\begin{aligned}\alpha(\mathbf{1}, P_{\mathsf{X}_1\mathsf{X}_2}) &= H(\mathsf{Y}|\mathsf{X}_1) + H(\mathsf{Y}|\mathsf{X}_2) - H(\mathsf{Y}) - \overbrace{H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2)}^{=0} \\ &= H(\mathsf{Y}|\mathsf{X}_1) - I(\mathsf{X}_2 \wedge \mathsf{Y}) \leq H(\mathsf{Y}) \leq 1\end{aligned}$$

The inequalities are achievable with equalities by independently and uniformly distributed inputs

$$P_{\mathsf{X}_1\mathsf{X}_2} = \mathrm{Bern}_{\frac{1}{2}}(x_1)\,\mathrm{Bern}_{\frac{1}{2}}(x_2)$$

By Theorem 7.3, we have the desired optimality of source emulation. Furthermore, there is a practical way to attain the secrecy capacity non-asymptotically with $n = 1$: have terminal 3 reveal Y in pubic and choose X_1 as the key. Terminal 2 can perfectly recover $\mathsf{X}_1 = \mathsf{Y} \oplus \mathsf{X}_2$, which is perfectly secret since it is independent of Y. □

For the infinitely-valued model with sample average constraints, we have the following sufficient condition instead.

Theorem 7.4 (sample average constraint) *The secrecy lower bound* (6.28b) *matches the upper bound* (6.5) *if the channel* $P_{\mathsf{Y}_V|\mathsf{X}_V}$ *satisfies*

$$P_{\mathsf{Y}_D|\mathsf{X}_V} = P_{\mathsf{Y}_D|\mathsf{X}_{D\cup\{s\}}} \qquad \textit{for some } s \in D^c \tag{7.5a}$$

$$\sup_{P_{\mathsf{X}_V}} \gamma(\lambda, P_{\mathsf{X}_V}) = \sup_{P_{\mathsf{X}_V} = P_{\mathsf{X}_D}\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}} \gamma(\lambda, P_{\mathsf{X}_V}) \qquad \textit{for all } \lambda \in \Lambda_{A|D} \tag{7.5b}$$

where the input distribution can be subject to the sample average constraint (5.3). □

Proof Consider bounding the supremum in the secrecy upper bound (6.5) as follows,

$$\begin{aligned}\sup_{P_{\mathsf{X}_V}} \min_{i\in D} \alpha_i(\lambda, P_{\mathsf{X}_V}) \overset{\text{(a)}}{=} \sup_{P_{\mathsf{X}_V}} \gamma(\lambda, P_{\mathsf{X}_V}) &\overset{\text{(b)}}{=} \sup_{P_{\mathsf{X}_V} = P_{\mathsf{X}_D}\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}} \gamma(\lambda, P_{\mathsf{X}_V}) \\ &\overset{\text{(c)}}{=} \sup_{P_{\mathsf{X}_V} = P_{\mathsf{X}_D}\prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})\end{aligned}$$

(a) Under the single-leakage condition (7.5a), $\alpha_i = \gamma$ by Proposition A.4.

(b) This is by condition (7.5b).

(c) With conditionally independent input, $\gamma = \tilde{\beta}$ again by Proposition A.4.

Minimizing over $\lambda \in \Lambda_{A|D}$ gives the secrecy upper bound on the L.H.S. and the lower bound on the R.H.S. as desired. ■

Example 7.2 (Gaussian channel) Consider the following two-user gaussian channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ with $A = [2] = D^c = V$,

$$\mathsf{Y}_1 = h_{11}\mathsf{X}_1 + h_{12}\mathsf{X}_2 + \mathsf{N}_1 \tag{7.6a}$$

$$\mathsf{Y}_2 = h_{21}\mathsf{X}_1 + h_{22}\mathsf{X}_2 + \mathsf{N}_2 \tag{7.6b}$$

where all variables are real-valued, and N_1 and N_2 are arbitrary zero-mean jointly gaussian noises normalized to have unit variance. i.e.

$$P_{\mathsf{N}_1\mathsf{N}_2}(n_1, n_2) = \mathscr{N}_{\mathbf{0}, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]}(\boldsymbol{n}) \qquad \text{for all } \boldsymbol{n} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \in \mathbb{R}^2 \tag{7.7}$$

where $\mathscr{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ denotes the jointly gaussian distribution [8] with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$,

$$\mathscr{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) := \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{7.8}$$

In addition, the channel input sequences X_1^n and X_2^n for T_1 and T_2 are subject to the average power constraints,

$$\frac{1}{n} \sum_{t \in [n]} \mathsf{X}_{1t}^2 \leq P_1 \quad \text{and} \quad \frac{1}{n} \sum_{t \in [n]} \mathsf{X}_{2t}^2 \leq P_2$$

which translate to the following power constraints on the secrecy bounds in Theorem 6.2 and Theorem 6.5.

$$E(\mathsf{X}_1^2) \leq P_1 \quad \text{and} \quad E(\mathsf{X}_2^2) \leq P_2 \tag{7.9}$$

We will show that pure source emulation is optimal and compute the secrecy capacity.

Since $D = \emptyset$, the condition (7.5a) is satisfied trivially. Evaluating (A.21b) with the only fractional partition $\boldsymbol{\lambda} = \mathbf{1}$, we have

$$\gamma(\mathbf{1}, P_{\mathsf{X}_V}) = H(\mathsf{Y}_1|\mathsf{X}_1) + H(\mathsf{Y}_2|\mathsf{X}_2) - H(\mathsf{Y}_1\mathsf{Y}_2|\mathsf{X}_1\mathsf{X}_2)$$

The last entropy term is

$$H(\mathsf{Y}_1\mathsf{Y}_2|\mathsf{X}_1\mathsf{X}_2) = H(\mathsf{N}_1\mathsf{N}_2) = \log 2\pi e(1-\rho^2)$$

The remaining terms can be bounded as follows,

$$\begin{aligned} H(\mathsf{Y}_1|\mathsf{X}_1) &= H(h_{12}\mathsf{X}_2 + \mathsf{N}_1|\mathsf{X}_1) \\ &\leq H(h_{12}\mathsf{X}_2 + \mathsf{N}_1|\mathsf{X}_1) \\ &\leq \frac{1}{2}\log 2\pi e(h_{12}^2 P_2 + 1) \end{aligned}$$

by the fact that gaussian distribution maximizes entropy for a given variance [8]. Similarly,

$$H(\mathsf{Y}_2|\mathsf{X}_2) \leq \frac{1}{2}\log 2\pi e(h_{21}^2 P_1 + 1)$$

All the inequalities are satisfied with equality by the gaussian input distribution

$$P_{\mathsf{X}_1\mathsf{X}_2}(x_1, x_2) = \mathcal{N}_{0,P_1}(x_1)\mathcal{N}_{0,P_2}(x_2) \qquad \forall\, x_1, x_2 \in \mathbb{R} \tag{7.10}$$

which therefore maximizes γ. Hence, Theorem 7.4 implies that the secrecy capacity is given by (6.28b),

$$C_{\mathrm{GC}} := \gamma(\mathbf{1}, \mathcal{N}_{0,P_1}\mathcal{N}_{0,P_2}) = \frac{1}{2}\ln\frac{(h_{12}^2 P_2 + 1)(h_{21}^2 P_1 + 1)}{(1-\rho^2)^2}$$

This is also equal to (6.28c) since the optimal input distribution is independent of λ. Pure source emulation turns out to be optimal.

It is possible to attain this secrecy capacity without public discussion if $\rho = 0$. To argue this, note that the secrecy capacity can be rewritten as a sum of two channel capacities as follows,

$$C_{\mathrm{GC}} = \overbrace{\frac{1}{2}\ln(h_{12}^2 P_2 + 1)}^{C_2} + \overbrace{\frac{1}{2}\ln(h_{21}^2 P_1 + 1)}^{C_1}$$

where C_1 and C_2 are the capacities of the component channels from T_1 to T_2 and T_2 to T_1 respectively after removing the interference. Hence, the terminals can directly transmit independent secret key bits at the capacities of the respective channels.[1] □

Note that Theorem 7.4 also applies to the finitely-valued model as a special case with or without sample average constraints. The sufficient condition is not as general as that in Theorem 7.3 because (7.5) implies (7.4) but the converse is not true.[2] The additional single-leakage condition (7.5a) essentially turns condition (7.4) to (7.5b) by the equivalence relation (b) of (A.28) in Proposition A.4. (7.5b) is easier to work with, however, because of the concavity of γ in the input distribution by Proposition A.2. For instance, we can use this to derive the following tightness condition for simultaneous independent channels.

Theorem 7.5 (Simultaneous independent channels) *Suppose the channel consists of a finite set $L := [\ell]$ of simultaneous independent channels in* (A.29), *i.e.*

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = \prod_{j \in L} P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}$$

Then, the secrecy lower bound (6.28b) *matches the upper bound* (6.5) *if the channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ satisfies*

$$P_{\mathsf{Y}_D|\mathsf{X}_V} = P_{\mathsf{Y}_D|\mathsf{X}_{D\cup\{s\}}} \qquad \text{for some } s \in D^c \tag{7.11a}$$

$$\sup_{P_{\mathsf{X}_{jV}}} \gamma(\lambda, P_{\mathsf{X}_{jV}}) = \sup_{\substack{P_{\mathsf{X}_{jV}} = P_{\mathsf{X}_{jD}} \times \\ \times \prod_{i \in D^c} P_{\mathsf{X}_{ji}|\mathsf{X}_{jD}}}} \gamma(\lambda, P_{\mathsf{X}_{jV}}) \qquad \text{for all } \lambda \in \Lambda_{A|D},\ j \in L \tag{7.11b}$$

where the input distribution is subject to the sample average constraint (5.3). *Fur-*

[1] This may not belong to the source emulation approach since the channel inputs may not be iid over time.

[2] For instance, Example 7.1 does not satisfy the single-leakage condition (7.5a).

thermore, the secrecy capacity can be achieved by the modified mixed source emulation in Section 6.2.3 with conditionally independent inputs for different channels given an auxiliary source. i.e.

$$P_{\mathsf{X}_V|\mathsf{Q}} = \prod_{j\in L} P_{\mathsf{X}_{jV}|\mathsf{Q}} \tag{7.12}$$

where Q *is the auxiliary source.* □

PROOF We first show that it is optimal to have the following independence constraint

$$P_{\mathsf{X}_V} = \prod_{j\in L} P_{\mathsf{X}_{jV}} \tag{7.13}$$

for the maximization in the secrecy upper bound (6.5).

$$\begin{aligned}
\sup_{P_{\mathsf{X}_V}} \min_{i\in D} \alpha_i(\lambda, P_{\mathsf{X}_V}) &\overset{\text{(a)}}{=} \sup_{P_{\mathsf{X}_V}} \gamma(\lambda, P_{\mathsf{X}_V}) \\
&\overset{\text{(b)}}{\leq} \sum_{j\in L} \sup_{P_{\mathsf{X}_{jV}}} \gamma(\lambda, P_{\mathsf{X}_j V})\big|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}} \\
&\overset{\text{(c)}}{=} \sum_{j\in L} \sup_{\substack{P_{\mathsf{X}_{jV}} = P_{\mathsf{X}_{jD}} \times \\ \times \prod_{i\in D^c} P_{\mathsf{X}_{ji}|\mathsf{X}_{jD}}}} \underbrace{\gamma(\lambda, P_{\mathsf{X}_j V})\big|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}}}_{\overset{\text{(d)}}{=} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}}} \\
&\overset{\text{(e)}}{\leq} \sup_{P_{\mathsf{X}_V} = P_{\mathsf{X}_D} \prod_{i\in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})
\end{aligned}$$

(a) by (b) of (A.28) in Proposition A.4 due to the sufficient condition (7.11a).

(b) by (A.30b) in Proposition A.5. Equality is achievable by independent inputs (7.13).

(c) by the sufficient condition (7.11b).

(d) by (c) of (A.28).

(e) because independent inputs (7.13) achieves

$$\tilde{\beta}(\lambda, P_{\mathsf{X}_V}) = \sum_{j\in L} \tilde{\beta}(\lambda, P_{\mathsf{X}_V})\Big|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}}$$

by the definition (A.18) of $\tilde{\beta}$. Relaxing this independence gives the upper bound. Finally, minimizing over $\lambda \in \Lambda_{A|D}$ gives the desired tightness because the L.H.S. of

(a) and the R.H.S. of (e) become the secrecy upper and lower bounds respectively. Furthermore, the equality for (e) implies the optimality of (7.13) in maximizing $\tilde{\beta}$. ■

Example 7.3 (Noise-free network) Consider the following finitely-valued noise-free network for three terminals that are all active. i.e. $A = V = [3]$.

terminal	1	2	3
input	$\mathsf{X}_1 \in X_1$	$\mathsf{X}_2 \in X_2$	
output		$\mathsf{Y}_2 = \mathsf{X}_1$	$\mathsf{Y}_3 = \mathsf{X}_2$

There is a noiseless channel from T_1 to T_2, and an independent one from T_2 to T_3. $D = \emptyset$ implies (7.11a). Since each component channel has only one sender, conditional independent input trivially maximizes γ for each channel, i.e. (7.11b). Hence, source emulation is optimal by Theorem 7.5.

By (6.18b), the secrecy capacity is

$$\begin{aligned} C_{\mathrm{N}} &:= \min_{\lambda \in \Lambda_{A|D}} \max_{P_{\mathsf{X}_V} = P_{\mathsf{X}_1} P_{\mathsf{X}_2}} \sum_{B \in \mathcal{H}_{A|D}} \lambda_B H(\mathsf{Y}_{B^c} | \mathsf{X}_{B^c}) \\ &= \min_{\lambda \in \Lambda_{A|D}} \sum_{B \in \mathcal{H}_{A|D}} \lambda_B \log |X_{B^c}| \\ &= \min\{\log |X_1|, \log |X_2|\} \end{aligned}$$

Since the optimal input distribution, namely the uniform distribution, is independent of λ, pure source emulation is optimal. There is also a practical scheme to achieve the capacity by directly transmitting the secret key from terminal 1 to 2 and relaying it from 2 to 3.[3] □

In general, there is a superadditive gain in secrecy capacity for simultaneous independent channels, i.e. the secrecy capacity of the composite channel is no less than that of each component channel. Example 7.3 for instance has 0 secrecy capacity for each component channel (since at least one terminal is isolated from the others in each case), but the secrecy capacity for the composite channel is positive. Note that we can

[3] This does not belong to the source emulation approach since it involves memory in the input for relaying. However, it is also optimal to first convert it to a source by the result in Section 7.2.2, and then attain the secrecy capacity by the network coding approach in Chapter 3.

also combine very different channels together, such as adding the continuous-valued channel in Example 7.2 to the finitely-valued channel in Example 7.3. We cannot add the channel from Example 7.1 however since it does not satisfy the single-leakage condition.

A trivial condition for tightness is when the secrecy upper bound is 0.

Theorem 7.6 (Zero secrecy capacity) *The secrecy capacity is zero if the channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ is such that there exists a bipartition $\{C_1, C_2\}$ of D^c through A, i.e.*

$$C_1, C_2 \not\supseteq A, \quad C_1 \cap C_2 = \emptyset, \quad C_1 \cup C_2 = D^c$$

such that

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = P_{\mathsf{Y}_{C_1\cup D}|\mathsf{X}_{C_1\cup D}} P_{\mathsf{Y}_{C_2}|\mathsf{X}_{C_2\cup D}} \tag{7.14}$$

In other words, we have the Markov chain, $\mathsf{Y}_{C_1\cup D} \leftrightarrow \mathsf{X}_{C_1\cup D} \leftrightarrow \mathsf{X}_{C_2\cup D} \leftrightarrow \mathsf{Y}_{C_2}$, regardless of the input distribution P_{X_V}. □

PROOF Given (7.14) is satisfied, consider some $i' \in C_1$ and $\lambda' \in \Lambda_{A|D}$ with

$$\lambda'_B = \begin{cases} 1 & \text{if } B = C_1 \text{ or } C_2 \\ 0 & \text{otherwise} \end{cases}$$

By (6.5), the secrecy capacity is upper bounded by

$$\begin{aligned} C_{\text{su}} &= \min_{\lambda \in \Lambda_{A|D}} \sup_{P_{\mathsf{X}_V}} \min_{i \in D^c} \alpha_i(\lambda, P_{\mathsf{X}_V}) \\ &\leq \sup_{P_{\mathsf{X}_V}} \alpha_{i'}(\lambda', P_{\mathsf{X}_V}) \\ &= \sup_{P_{\mathsf{X}_V}} \left[H(\mathsf{Y}_{C_2}|\mathsf{X}_{C_2\cup D}\mathsf{Y}_D) + H(\mathsf{Y}_{C_1\cup D}|\mathsf{X}_{C_1\cup D}) - H(\mathsf{Y}_V|\mathsf{X}_V) \right] \end{aligned}$$

where the last equality is by (A.20c). Applying (7.14) to the first two entropy terms,

$$\begin{aligned} H(\mathsf{Y}_{C_2}|\mathsf{X}_{C_2\cup D}\mathsf{Y}_D) &= H(\mathsf{Y}_{C_2}|\mathsf{X}_V\mathsf{Y}_{C_1\cup D}) \\ H(\mathsf{Y}_{C_1\cup D}|\mathsf{X}_{C_1\cup D}) &= H(\mathsf{Y}_{C_1\cup D}|\mathsf{X}_V) \end{aligned}$$

The sum of the entropies above is $H(\mathsf{Y}_V|\mathsf{X}_V)$, and so the secrecy upper bound is at most zero as desired. ■

In the special case when we have a source model instead, i.e. no channel input, the condition in Theorem 7.6 is also necessary.

Theorem 7.7 (ZSC for source model) *For the (possibly infinitely-valued) source model, secrecy capacity is zero iff the source P_{X_V} is such that there exists a bipartition $\{C_1, C_2\}$ of D^c through A with*

$$P_{\mathsf{X}_{D^c}|\mathsf{X}_D} = P_{\mathsf{X}_{C_1}|\mathsf{X}_D} P_{\mathsf{X}_{C_2}|\mathsf{X}_D} \tag{7.15}$$

or equivalently the Markov chain $\mathsf{X}_{C_1} \leftrightarrow \mathsf{X}_D \leftrightarrow \mathsf{X}_{C_2}$. □

PROOF Sufficiency follows from Theorem 7.6 by treating X_V as the output of a channel that does not admit any input. Consider proving the converse. Let λ^* be the optimal solution to the minimization in the secrecy capacity (6.20)

$$C_{\mathrm{s}} = \min_{\lambda \in \Lambda_{A|D}} \beta(\lambda, P_{\mathsf{X}_V})$$

We may assume that there exists a bipartition $\{C_1, C_2\}$ of D^c through A such that $\lambda^*_{C_1} > 0$ because $\sum_{B \in \mathcal{H}_{A|D}} \lambda^*_B \geq 1$ by (A.9).

If the secrecy capacity is zero, i.e. $\beta(\lambda, P_{\mathsf{X}_V}) = 0$, then we have by (A.19) that

$$H(\mathsf{X}_{D^c}|\mathsf{X}_D) = \sum_B \lambda^*_B H(\mathsf{X}_B|\mathsf{X}_{B^c})$$

Note that the L.H.S. is no smaller than the R.H.S. in general by the weak form (A.10a) of the Shearer-type lemma. From the proof of the lemma, the above equality implies that, for all $B \in \mathcal{H}_{A|D}$ with $\lambda^*_B > 0$,

$$\sum_{i \in B} H(\mathsf{X}_i|\mathsf{X}_{[i-1]}\mathsf{X}_D) = \sum_{i \in B} H(\mathsf{X}_i|\mathsf{X}_{[i-1] \cap B}\mathsf{X}_{B^c})$$

This holds in particular for C_1 that

$$\sum_{i \in C_1} H(\mathsf{X}_i | \mathsf{X}_{[i-1]} \mathsf{X}_D) = \sum_{i \in C_1} H(\mathsf{X}_i | \mathsf{X}_{[i-1] \cap C_1} \mathsf{X}_{C_2 \cup D})$$

Without loss of generality, we can re-index the terminals such that $C_1 = [c] \subseteq D^c$ for some positive integer c. Then, we have, $[i-1] \subseteq C_1$ for all $i \in C_1$, and so the above equality simplifies to

$$H(\mathsf{X}_{C_1} | \mathsf{X}_D) = H(\mathsf{X}_{C_1} | \mathsf{X}_{C_2 \cup D})$$

which gives (7.15) as desired. ■

Problem:

Give an example, if any, for which pure source emulation is strictly suboptimal even though one of the tightness conditions described above is satisfied. In particular, consider the channel given in Proposition 7.3. Is it possible to correlate the noise such that pure source emulation is suboptimal?

7.2.1 Interference-free Channels

In the following, we show that the secrecy bounds are tight for channels that are interference-free. We first consider DMMC's that has any number of untrusted senders but at most one trusted sender. The trusted sender can also receive a channel output as an immediate feedback from the channel. Furthermore, the channel can be infinitely-valued with sample average constraints on the input. We will show that pure source emulation achieves the secrecy capacity in this case, which extends the result of [13].

Proposition 7.1 (Single trusted sender) *If the channel has at most one trusted sender, i.e. for some $s \in D^c$,*

$$P_{\mathsf{Y}_V | \mathsf{X}_V} = P_{\mathsf{Y}_V | \mathsf{X}_{D \cup \{s\}}} \qquad \textit{(Single trusted sender)} \qquad \mathbf{(7.16)}$$

then pure source emulation is optimal and the secrecy capacity is given by (6.28c) *in Theorem 6.5.* □

Proof This follows immediately from Theorem 7.4 that (6.28b) is the secrecy capacity since the single-trusted-sender condition (7.16) trivially implies both the single-leakage condition (7.5a) and the optimality of conditionally independent input distribution (7.5b). To prove the stronger result that pure source emulation is optimal, i.e. the secrecy capacity is (6.28c), it suffices to consider the finitely-valued model because the achieving scheme for Theorem 6.5 first convert the infinitely-valued model to finitely-valued model by quantization (6.21). By the minimax theorem [53],

$$\min_{\lambda \in \Lambda_{A|D}} \max_{P_{\mathsf{X}_{D\cup\{s\}}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_{D\cup\{s\}}}) = \max_{P_{\mathsf{X}_{D\cup\{s\}}}} \min_{\lambda \in \Lambda_{A|D}} \tilde{\beta}(\lambda, P_{\mathsf{X}_{D\cup\{s\}}})$$

since $\tilde{\beta}$ (or equivalently γ by Proposition A.4) is concave in the input distribution $P_{\mathsf{X}_{D\cup\{s\}}}$ by Proposition A.2 and linear in λ over convex compact sets.[4] The L.H.S. is the secrecy capacity by Theorem 7.4 while the R.H.S. is the secrecy lower bound (6.18c) achievable by pure source emulation. ■

Roughly speaking, we can think of the channel equivalently as a broadcast channel with immediate feedback to the sender, and with a channel state, namely X_D, publicly controllable by the untrusted terminals. Since coordination is trivial with just one sender, even pure source emulation achieves the secrecy capacity. By Theorem 7.5, we can extend the result further to channels that can be decomposed into simultaneous independent broadcast-type channels as follows.

Proposition 7.2 (Single trusted sender per channel) *If every simultaneous independent channel has one trusted sender (not necessarily the same one), and at most one channel has output observable by the untrusted terminals, i.e.*

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = P_{\mathsf{Y}_{1V}|\mathsf{X}_{D\cup\{s_1\}}} \prod_{j \in L\setminus\{1\}} P_{\mathsf{Y}_{jD^c}|\mathsf{X}_{j\,D\cup\{s_j\}}} \qquad \textbf{(7.17)}$$

[4] The set of valid input distributions remains convex under the sample average constraints. If there were more than one trusted sender, however, the set of conditionally independent input would not be convex.

with $s_j \in D^c$ for all $j \in L$, then pure source emulation is optimal and achieves the secrecy capacity (6.28c). *n.b. Proposition 7.1 is a special case when $|L| = 1$.* □

PROOF This follows immediately from Theorem 7.5 with the same argument as in the proof of Proposition 7.1. Since each simultaneous independent channel has at most one trusted sender, conditionally independent input is trivially optimal for each channel, giving (7.11b). (7.11a) follows from the definition (7.17) that $\mathsf{Y}_D = \mathsf{Y}_{1D}$ which depends on the input P_{X_V} only through X_{s_1}. To prove that pure source emulation is optimal, it suffices to consider the finitely-valued model as argued in the proof of Proposition 7.1. Denote the vector $(P_{\mathsf{X}_{j\,D\cup\{s_j\}}} : j \in L)$ by $(P_{\mathsf{X}_{j\,D\cup\{s_j\}}})_{j\in L}$, and define

$$
\begin{aligned}
f(\lambda, (P_{\mathsf{X}_{j\,D\cup\{s_j\}}})_{j\in L}) :=\ & \tilde{\beta}(\lambda, P_{\mathsf{X}_{1\,D\cup\{s_1\}}})\Big|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{1\,D\cup\{s_1\}}}} \\
& + \sum_{j\in L\setminus\{1\}} \tilde{\beta}(\lambda, P_{\mathsf{X}_{j\,D\cup\{s_j\}}})\Big|_{P_{\mathsf{Y}_{jD^c}|\mathsf{X}_{j\,D\cup\{s_j\}}}}
\end{aligned}
$$

which is linear in λ over the convex compact set $\Lambda_{A|D}$ and concave in $(P_{\mathsf{X}_{j\,D\cup\{s_j\}}} : j \in L)$ over the convex compact set of vectors of valid input distributions for the independent channels. By the minimax theorem [53],

$$
\min_{\lambda\in\Lambda_{A|D}} \max_{P_{\mathsf{X}_{j\,D\cup\{s_j\}}}:j\in L} f(\lambda, (P_{\mathsf{X}_{j\,D\cup\{s_j\}}})_{j\in L}) = \max_{P_{\mathsf{X}_{j\,D\cup\{s_j\}}}:j\in L} \min_{\lambda\in\Lambda_{A|D}} f(\lambda, (P_{\mathsf{X}_{j\,D\cup\{s_j\}}})_{j\in L})
$$

The L.H.S. is the secrecy capacity by Theorem 7.5, while the R.H.S. is the secrecy lower bound (6.18c). ■

Proposition 7.2 can be extended to the following three-user case with correlated noises for the component channels.

Proposition 7.3 (Three correlated channels) *Consider the three-user case where $A \subseteq D^c \subseteq V = [3]$. One of the terminals can be a helper or an untrusted terminal. i.e. we may have $A = [2]$ and $D = \emptyset$ or $\{3\}$. Suppose the channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ satisfies*

$$
\mathsf{Y}_i = f_i(\mathsf{X}_{s_i}, \mathsf{N}_i) \qquad \forall i \in V \tag{7.18}
$$

where, for all $i \in V$, $s_i \in V$, the channel noise N_i's are independent of X_i's and recoverable from the corresponding channel input and output in the sense that there exists functions g_i's with

$$\mathsf{N}_i = g_i(\mathsf{X}_{s_i}, \mathsf{Y}_i) \qquad \forall i \in V \tag{7.19}$$

Then, mixed source emulation is optimal and the secrecy capacity is given by (6.28b). *If* $|Y_D| > 1$, *then pure source emulation is optimal. If* $|Y_D| \leq 1$ *and the channel noises are independent of each other, then pure source emulation is optimal with input distribution* $P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_2} = P^*_{\mathsf{X}_1} P^*_{\mathsf{X}_2} P^*_{\mathsf{X}_3}$ *where* $P^*_{\mathsf{X}_{s_i}}$ *maximizes* $H(\mathsf{Y}_i)$ *under the corresponding moment constraints.* □

PROOF The single-leakage condition (7.5a) is satisfied trivially if $D = \emptyset$ or $|Y_D| \leq 1$. If not, consider for definiteness that $D = \{3\}$. By (7.18), $\mathsf{Y}_3 = f_3(\mathsf{X}_{s_3}, \mathsf{N}_3)$ which is independent of X_V given X_{s_3} since N_3 is independent of X_V. We have the desired Markov chain for (7.5a) that $\mathsf{Y}_3 \leftrightarrow \mathsf{X}_{s_3} \leftrightarrow \mathsf{X}_V$.

To show (7.5b), consider the case $D = \emptyset$ first. Then, by (A.21b),

$$\gamma(\lambda, P_{\mathsf{X}_V}) = \sum_B \lambda_B H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_V|\mathsf{X}_V)$$

By (7.18) and (7.19), $H(\mathsf{Y}_V|\mathsf{X}_V) = H(\mathsf{N}_V|\mathsf{X}_V) = H(\mathsf{N}_V)$ independent of P_{X_V}. Thus, it suffices to show as follows that $H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c})$ is maximized by independent inputs for every choice of $B \in \mathcal{H}_{A|D}$.

I) Consider the case when $|B| = 2$. For definiteness, suppose $B = \{1, 2\}$. If $s_3 = 3$, we have

$$\begin{aligned} H(\mathsf{Y}_3|\mathsf{X}_3) &= H(\mathsf{Y}_3, \mathsf{N}_3|\mathsf{X}_3) && \text{by (7.19)} \\ &= H(\mathsf{N}_3|\mathsf{X}_3) && \text{by (7.18)} \\ &= H(\mathsf{N}_3) && \text{by independence} \end{aligned}$$

which is trivially maximized by independent inputs. If $s_3 \in \{1, 2\}$ instead, then

$$H(\mathsf{Y}_3|\mathsf{X}_3) \leq H(\mathsf{Y}_3)$$

with equality if X_{s_3} is independent of X_3 by the data processing theorem and the Markov chain $\mathsf{Y}_3 \leftrightarrow \mathsf{X}_{s_3} \leftrightarrow \mathsf{X}_3$ from (7.18) that

$$I(\mathsf{Y}_3 \wedge \mathsf{X}_3) \leq I(\mathsf{X}_{s_3} \wedge \mathsf{X}_3) = 0$$

Thus, we have

$$\sup_{P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3}} H(\mathsf{Y}_3|\mathsf{X}_3) \leq \sup_{P_{\mathsf{X}_{s_3}}} H(\mathsf{Y}_3)$$

since P_{Y_3} depends on P_{X_V} only through $P_{\mathsf{X}_{s_3}}$. Thus, it is optimal to choose $P^*_{\mathsf{X}_{s_3}}$ that maximizes $H(\mathsf{Y}_3)$.

II) Consider the case $|B| = 1$. For definiteness, suppose $B = \{1\}$.

(a) Suppose $s_2 = s_3 = 1$, which gives $\mathsf{Y}_2\mathsf{Y}_3 \leftrightarrow \mathsf{X}_1 \leftrightarrow \mathsf{X}_2\mathsf{X}_3$. Then, by the same argument as before,

$$\sup_{P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3}} H(\mathsf{Y}_2\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3) = \sup_{P_{\mathsf{X}_1}} H(\mathsf{Y}_2\mathsf{Y}_3)$$

and so it is optimal to have independent inputs.

(b) Suppose $s_2, s_3 \in \{2, 3\}$ instead. Again by (7.18) and (7.19),

$$\begin{aligned} H(\mathsf{Y}_2\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3) &= H(\mathsf{Y}_2\mathsf{Y}_3\mathsf{N}_2\mathsf{N}_3|\mathsf{X}_2\mathsf{X}_3) \\ &= H(\mathsf{N}_2\mathsf{N}_3|\mathsf{X}_2\mathsf{X}_3) \\ &= H(\mathsf{N}_2\mathsf{N}_3) \end{aligned}$$

which is trivially maximized by independent inputs.

(c) The remaining case has exactly one of s_2 and s_3 equal to 1. For definiteness, suppose $s_2 \in \{2, 3\}$ and $s_3 = 1$.

$$\begin{aligned} H(\mathsf{Y}_2\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3) &= H(\mathsf{Y}_2|\mathsf{X}_2\mathsf{X}_3) + H(\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3\mathsf{Y}_2) \\ &= H(\mathsf{N}_2) + H(\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3\mathsf{N}_2) \end{aligned} \tag{7.20}$$

again by (7.18) and (7.19). The second entropy can be bounded as follows,

$$H(\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3\mathsf{N}_2) \leq H(\mathsf{Y}_3|\mathsf{N}_2) \tag{7.21}$$

with equality if X_1 is independent of $(\mathsf{X}_2, \mathsf{X}_3)$. To argue this, note that we have the Markov chain $\mathsf{Y}_3\mathsf{N}_2 \leftrightarrow \mathsf{X}_1 \leftrightarrow \mathsf{X}_2\mathsf{X}_3$ since $(\mathsf{Y}_3, \mathsf{N}_2)$ is a function of $(\mathsf{X}_1, \mathsf{N}_2, \mathsf{N}_3)$ by (7.18). By the data processing theorem,

$$\begin{aligned} I(\mathsf{Y}_3 \wedge \mathsf{X}_2\mathsf{X}_3|\mathsf{N}_2) &\leq I(\mathsf{Y}_3\mathsf{N}_2 \wedge \mathsf{X}_2\mathsf{X}_3) \\ &\leq I(\mathsf{X}_1 \wedge \mathsf{X}_2\mathsf{X}_3) = 0 \end{aligned}$$

Hence, applying (7.21) to (7.20) and maximizing over the input distribution, we have

$$\sup_{P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3}} H(\mathsf{Y}_2\mathsf{Y}_3|\mathsf{X}_2\mathsf{X}_3) = H(\mathsf{N}_2) + \sup_{P_{\mathsf{X}_1}} H(\mathsf{Y}_3|\mathsf{N}_2)$$

If N_3 is independent of N_2 in addition, then $H(\mathsf{Y}_3|\mathsf{N}_2) = H(\mathsf{Y}_3)$ and so it is optimal to choose $P^*_{\mathsf{X}_1}$ that maximizes $H(\mathsf{Y}_3)$.

Thus, by Theorem 7.4, the secrecy capacity is given by (6.28b) and so mixed source emulation is optimal. If the channel noises are independent, it is optimal to choose $P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3} = P^*_{\mathsf{X}_1}P^*_{\mathsf{X}_2}P^*_{\mathsf{X}_3}$, which is independent of the choice of λ. (6.28c) is therefore satisfied with equality and so pure source emulation is optimal.

Consider proving (7.5b) for the remaining case when $|Y_D| > 1$. Assume for definiteness that $A = [2]$ and $D = \{3\}$. Since the only fractional partition in $\Lambda_{A|D}$ is $\boldsymbol{\lambda} = \mathbf{1}$, we need only consider the following by (A.21b).

$$\gamma(\mathbf{1}, P_{\mathsf{X}_V}) = H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) + H(\mathsf{Y}_2|\mathsf{X}_2\mathsf{X}_3\mathsf{Y}_3) - H(\mathsf{Y}_1\mathsf{Y}_2|\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3\mathsf{Y}_3)$$

By (7.18) and (7.19),

$$H(\mathsf{Y}_1\mathsf{Y}_2|\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3\mathsf{Y}_3) = H(\mathsf{N}_1\mathsf{N}_2|\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3\mathsf{N}_3) = H(\mathsf{N}_1\mathsf{N}_2|\mathsf{N}_3)$$

which is trivially maximized by independent inputs. It suffices to show that $H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3)$ is maximized by independent inputs as follows. The same conclusion will apply to $H(\mathsf{Y}_2|\mathsf{X}_2\mathsf{X}_3\mathsf{Y}_3)$ by symmetry.

1. Consider the case $s_1 = s_3 = 2$. This gives the Markov chain $\mathsf{Y}_1\mathsf{Y}_3 \leftrightarrow \mathsf{X}_2 \leftrightarrow \mathsf{X}_1\mathsf{X}_3$. Thus,
$$H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) \leq H(\mathsf{Y}_1|\mathsf{Y}_3)$$
with equality if X_2 is independent of $(\mathsf{X}_1, \mathsf{X}_3)$.
2. If $s_1, s_3 \in \{1, 3\}$ instead, then
$$H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) = H(\mathsf{N}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{N}_3) = H(\mathsf{N}_1|\mathsf{N}_3)$$
which is trivially maximized by independent inputs.
3. If $s_1 = 2$ and $s_3 \in \{1, 3\}$ instead, then
$$H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) = H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{N}_3) \leq H(\mathsf{Y}_1|\mathsf{N}_3)$$
We have equality by choosing X_2 independent of $(\mathsf{X}_1, \mathsf{X}_3)$ because of the Markov chain $\mathsf{Y}_1\mathsf{N}_3 \leftrightarrow \mathsf{X}_2 \leftrightarrow \mathsf{X}_1\mathsf{X}_3$.
4. If $s_3 = 2$ and $s_1 \in \{1, 3\}$ instead, then
$$H(\mathsf{Y}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) = H(\mathsf{N}_1|\mathsf{X}_1\mathsf{X}_3\mathsf{Y}_3) \leq H(\mathsf{N}_1|\mathsf{Y}_3)$$
We have equality by choosing X_2 independent of $(\mathsf{X}_1, \mathsf{X}_3)$ because of the Markov chain $\mathsf{Y}_3\mathsf{N}_1 \leftrightarrow \mathsf{X}_2 \leftrightarrow \mathsf{X}_1\mathsf{X}_3$.

Hence, mixed source emulation is optimal by Theorem 7.4. Indeed, pure source emulation is optimal since $\Lambda_{A|D}$ is a singleton and so (6.28b) equals (6.28c). ■

For example, consider the channel defined as follows.

$$\mathsf{Y}_2 = \mathsf{X}_1 + \mathsf{N}_2$$
$$\mathsf{Y}_3 = \mathsf{X}_2 + \mathsf{N}_3$$
$$\mathsf{Y}_1 = \mathsf{X}_3 + \mathsf{N}_1$$

where N_1, N_2 and N_3 are arbitrarily correlated noise independent of the channel inputs. The channel satisfies (7.18) with $(s_1, s_2, s_3) = (3, 1, 2)$, and (7.19) with $g_i(x, y) = y - x$ for all $i \in [3]$. Thus, by Theorem 7.3, mixed source emulation is optimal. If the channel noises are independent, pure source emulation is optimal. In the special case when the channel noises N_i's are correlated and jointly gaussian (7.8), and the channel inputs X_i's are subject to the power constraints P_i's, it can be shown that it is optimal to have independent gaussian input, i.e. $P_{\mathsf{X}_1\mathsf{X}_2\mathsf{X}_3} = \mathscr{N}_{\mathbf{0},\operatorname{diag}(P_1,P_2,P_3)}$. Thus, pure source emulation is optimal in this case even though the noises are not necessarily independent.

7.2.2 Finite homomorphic network

In this section, we consider the following type of finitely-valued channels where the channel outputs are *group homomorphisms* of the channel inputs and noises. It can be considered as a generalization of finite linear channels where the outputs are linear combinations of the inputs. We will show that pure source emulation is optimal with uniform input distribution, which gives an explicit expression for the secrecy capacity. For finite linear channels, in particular, Chapter 3 gives a practical way to attain secrecy capacity using linear network coding.

Definition 7.2 (Finite homomorphic network) The channel output Y_V depends on the input X_V as follows,

$$\mathsf{Y}_i = M_i(\mathsf{X}_V, \mathsf{N}_0) + \mathsf{N}_i \qquad \text{for all } i \in V \tag{7.22}$$

with the following assumptions. n.b. in the special case when $D = \emptyset$ and N_0 deterministic, Assumption 2, 4 and 5 below are automatically satisfied. Those assumptions are technicalities for the case when $D \neq \emptyset$.

1. The channel input X_i and output Y_i for terminal $i \in V$ take values from the finite abelian group $(X_i, +)$ and $(Y_i, +)$ respectively. The individual channel noise N_i for terminal $i \in V$ takes values from Y_i.

2. The common channel noise N_0 is uniformly distributed over the finite abelian group $(N_0, +)$. i.e.

$$P_{N_0}(n_0) = \frac{1}{|N_0|} \qquad \text{for all } n_0 \in N_0 \tag{7.23}$$

3. M_i is a homomorphism for all $i \in V$. i.e. for all $x_V, x'_V \in X_V$ and $n_0, n'_0 \in N_0$,

$$M_i(x_V + x'_V, n_0 + n'_0) = M_i(x_V, n_0) + M_i(x'_V, n'_0) \tag{7.24}$$

4. N_D is determined by Y_D. i.e. for all $n_D, n'_D \in Y_D$ such that $n_D \neq n'_D$ and $P_{\mathsf{N}}(n_D), P_{\mathsf{N}}(n'_D) > 0$,

$$n'_D - n_D \notin M_i(X_V, N_0) := \{M_i(x_V, n_0) : x_V \in X_V, n_0 \in N_0\} \tag{7.25}$$

   In other words, the support of N_D has at most one element from each coset of the subgroup $M_D(X_V, N_0)$ of Y_D.

5. There exists a special terminal $s \in D^c$ such that

$$P_{\mathsf{N}_0, \mathsf{N}_V | \mathsf{X}_V} = P_{\mathsf{N}_0} \cdot P_{\mathsf{N}_{D^c}} \cdot P_{\mathsf{N}_D | \mathsf{N}_s} \tag{7.26}$$

   Furthermore, uniform P_{X_V} implies

$$I(\mathsf{Y}_D \wedge \mathsf{X}_{D^c \setminus \{s\}} | \mathsf{X}_D) = 0 \tag{7.27}$$

   Roughly speaking, uniformly distributed input for terminal s completely jams the channel from the trusted terminals to the untrusted terminals. □

Theorem 7.8 *For the finite homomorphic network defined above, pure source emu-*

lation with uniform P_{X_V} attains the secrecy capacity

$$C_{\mathrm{FHN}} := \min_{\lambda \in \Lambda_{A|D}} \sum_{B \in \mathcal{H}_{A|D}} \lambda_B \left[\log \left|M^B_{B^c}(X_B, \mathbf{0})\right| + H(\mathsf{S}^B)\right]$$
$$- \left[\log \left|M^{D^c}_D(X_D, \mathbf{0})\right| + H(\mathsf{S}^{D^c})\right] - \left(\textstyle\sum_B \lambda - 1\right) H(M_V(\mathbf{0}, \mathsf{N}_0) + \mathsf{N}_V)$$

where $M^B_i(x_B, x_{B^c}) := M_i(x_V, 0)$ and S^B is a random variable which denotes the coset of $M^B_{B^c}(X_B, \mathbf{0})$ that contains $M_{B^c}(\mathbf{0}, \mathsf{N}_0) + \mathsf{N}_{B^c}$. □

The proof relies on the group structure that can be captured by the following simple finite homomorphic channel.

Simple finite homomorphic channel:

We say $P_{\mathsf{Y}|\mathsf{X}}$ is a (single-input single-output) finite homomorphic channel if

$$\mathsf{Y} = M(\mathsf{X}) + \mathsf{N} \tag{7.28}$$

for some M and N such that

- X and Y take values from the finite abelian group $(X, +)$ and $(Y, +)$ respectively,
- $\mathsf{N} \in Y$ is independent of X, and
- M is a homomorphism. i.e.

$$M(x + x') = M(x) + M(x') \qquad \forall x, x' \in X$$

We write

- $M(X) := \{M(x) : x \in X\}$ as a *subgroup* of Y,
- $M(X) + n := \{M(x) + n : x \in X\}$ for $n \in Y$ as a *coset* of $M(X)$ in Y,
- $\mathcal{P}(M(X)) := \{M(X) + n : n \in Y\}$ as the partition of Y into cosets of $M(X)$, and
- $\mathrm{Null}(M) := \{x \in X : M(x) = 0\}$ as the *kernelcoset* of M.

Lemma 7.1 *For the simple finite homomorphic channel defined above,*

$$H(\mathsf{Y}) \leq \log|M(X)| + H(\mathsf{S}) \tag{7.29}$$

where S *is the unique coset from* $\mathcal{P}(M(X))$ *that contains* N. *i.e.*

$$\mathsf{N} \in \mathsf{S} \in \mathcal{P}(M(X)) \tag{7.30}$$

Equality holds for (7.29) *if* P_{X} *is uniform.* □

Proof $\mathsf{Y} \in \mathsf{S}$ iff $\mathsf{N} \in \mathsf{S}$ since $\mathsf{Y}-\mathsf{N} \in M(X)$. In other words, S is not only determined by N, but also by S.[5] Thus,

$$\begin{aligned} H(\mathsf{Y}) = H(\mathsf{Y},\mathsf{S}) &= H(\mathsf{Y}|\mathsf{S}) + H(\mathsf{S}) \\ &\leq \log|\mathsf{S}| + H(\mathsf{S}) \end{aligned}$$

where (a) is because Y determines S, and (b) is because $\mathsf{Y} \in \mathsf{S}$. This gives (7.29) because $|\mathsf{S}| = |M(X)|$.

To show the equality case, suppose P_{X} is uniform. For all $n \in Y$ and $y \in M(X)+n$,

$$P_{\mathsf{Y}|\mathsf{N}}(y|n) \overset{\text{(a)}}{=} \frac{|\text{Null}(M)|}{|X|} \overset{\text{(b)}}{=} \frac{1}{|M(X)|} \tag{7.31}$$

where the R.H.S. of (a) is the probability that $\mathsf{X} - x \in \text{Null}(M)$ for some particular solution x to $y = M(x)+n$, and (b) is a well-known identity in linear algebra. Hence, for all $y \in S$,

$$\begin{aligned} P_{\mathsf{Y}|\mathsf{S}}(y|S) &\overset{\text{(a)}}{=} \sum_{n \in S} P_{\mathsf{YN}|\mathsf{S}}(y,n|S) \\ &\overset{\text{(b)}}{=} \sum_{n \in S} P_{\mathsf{N}|\mathsf{S}}(n|S) P_{\mathsf{Y}|\mathsf{N}}(y|n) = \frac{1}{|M(X)|} \end{aligned}$$

where the summation in (a) is over S since $\mathsf{N} \in \mathsf{S}$ by (7.30), (b) is because N determines S, and (c) follows from (7.31). Thus, $H(\mathsf{Y}|\mathsf{S}) = \log|M(X)|$ as desired. ■

Finite homomorphic channels need not be symmetric in general. Nonetheless, the lemma says that the output entropy is maximized by uniform input distribution. The essence of the proof lies in the property that the effective channel $P_{\mathsf{Y}|\mathsf{XS}}(\cdot|\cdot,S)$ conditioned on the maximum common function S of Y and N is strongly symmetric [8].

[5] Indeed, S is the maximum common function defined in [22]. It is the common function of Y and N with maximum entropy.

In particular, for all $x \in X, y \in S$,

$$\begin{aligned} P_{\mathsf{Y}|\mathsf{XS}}(y|x,S) &= P_{\mathsf{Y}|\mathsf{XS}}(y + M(x'-x)|x',S) \qquad &\forall x' \in X \\ &= P_{\mathsf{Y}|\mathsf{XS}}(y'|x+\bar{x}(y'-y),S) \qquad &\forall y' \in S \end{aligned}$$

where $\bar{x}(y'-y)$ denotes a particular solution $\Delta \in X$ to $y'-y = M(\Delta)$. Thus, uniform input leads to uniform output, which maximizes $H(\mathsf{Y}|\mathsf{S}=S)$, the only component of $H(\mathsf{Y})$ that depends on P_{X}. We can now prove Theorem 7.8 using this special structure of finite homomorphic channels.

Proof (Theorem 7.8) $\alpha(\lambda, P_{\mathsf{X}_V})$ can be expressed as follows by (A.17c).

$$\alpha(\lambda, P_{\mathsf{X}_V}) = \sum_{B \ni s} \lambda_B \overbrace{[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_D|\mathsf{X}_D)]}^{\textcircled{1}} + \sum_{B \not\ni s} \lambda_B \left[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_V|\mathsf{X}_V)\right]$$

Using the result from Lemma 7.1, it is straightforward to show that the secrecy capacity claimed in the theorem is obtained from $\min_{\lambda \in \Lambda_{A|D}} \alpha(\lambda, P_{\mathsf{X}_V})$ with uniform P_{X_V}. By Theorem 7.3, it suffices to show that uniform P_{X_V} maximizes $\alpha(\lambda, P_{\mathsf{X}_V})$.

We can ignore the last entropy term $H(\mathsf{Y}_V|\mathsf{X}_V)$ since it is independent of P_{X_V}. More precisely,

$$H(\mathsf{Y}_V|\mathsf{X}_V) = H(\mathsf{Y}_V - M_V(\mathsf{X}_V, 0)|\mathsf{X}_V) = H(M_V(\mathbf{0}, \mathsf{N}_0) + \mathsf{N}_V)$$

by (7.22), (7.24) and (7.26).

Given $\mathsf{X}_{B^c} = x_{B^c} \in X_{B^c}$, we have from (7.22) that

$$\begin{aligned} \mathsf{Y}_{B^c} &= M^B_{B^c}(\mathsf{X}_B, x_{B^c}) + M_{B^c}(\mathbf{0}, \mathsf{N}_0) + \mathsf{N}_{B^c} \\ &= M^B_{B^c}(\mathsf{X}_B, \mathbf{0}) + \mathsf{N} \end{aligned}$$

with $\mathsf{N} := M^B_{B^c}(\mathbf{0}, x_{B^c}) + M_{B^c}(\mathbf{0}, \mathsf{N}_0) + \mathsf{N}_{B^c}$. This is a simple homomorphic finite channel, and so by Lemma 7.1,

$$H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c} = x_{B^c}) \leq \log|M^B_{B^c}(X_B, \mathbf{0})| + H(\mathsf{S}^B)$$

with equality if P_{X_B} is uniform. Thus, uniform P_{X_V} achieves the maximum

$$H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) = \log|M_{B^c}^B(X_B, \mathbf{0})| + H(\mathsf{S}^B)$$

The remaining term ① in $\alpha(\lambda, P_{\mathsf{X}_V})$ can be bounded as follows.

$$①= H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) - I(\mathsf{Y}_D \wedge \mathsf{X}_{B^c\setminus D}|\mathsf{X}_D) \leq H(\mathsf{Y}_C|\mathsf{X}_{B^c}\mathsf{Y}_D) \text{ with } C := B^c \setminus D$$

with equality if P_{X_V} is uniform by (7.27). It remains to show that uniform P_{X_V} maximizes $H(\mathsf{Y}_C|\mathsf{X}_{B^c}\mathsf{Y}_D)$ where $B \in \mathcal{H}_{A|D} : s \in B$ and $C := B^c \setminus D$.

Indeed, we need only consider the special case without N_0. More precisely, let

$$\tilde{\mathsf{X}}_i = \begin{cases} (\mathsf{X}_i, \mathsf{N}_0) & \text{if } i = s \\ \mathsf{X}_i & \text{otherwise} \end{cases}$$

Then, we have $\mathsf{Y}_i = \tilde{M}_i(\tilde{\mathsf{X}}_i) + \mathsf{N}_i$ for some homomorphism $\tilde{M}_i$ by (7.22) and (7.24). Furthermore,

$$I(\mathsf{Y}_D \wedge \tilde{\mathsf{X}}_{D^c\setminus\{s\}}|\tilde{\mathsf{X}}_D) = I(\mathsf{Y}_D \wedge \mathsf{X}_{D^c\setminus\{s\}}|\mathsf{X}_D)$$

which equals 0 by (7.27) if $P_{\tilde{\mathsf{X}}_V}$ is uniform (which happens iff P_{X_V} is uniform by (7.23)). Since $s \notin B^c$, we have

$$H(\mathsf{Y}_C|\tilde{\mathsf{X}}_{B^c}, \mathsf{Y}_D) = H(\mathsf{Y}_C|\mathsf{X}_{B^c}, \mathsf{Y}_D)$$

If uniform $P_{\tilde{\mathsf{X}}_V}$ maximizes $H(\mathsf{Y}_C|\tilde{\mathsf{X}}_{B^c}\mathsf{Y}_D)$, uniform P_{X_V} also maximizes $H(\mathsf{Y}_C|\mathsf{X}_{B^c}, \mathsf{Y}_D)$. We can therefore focus on the case without N_0, i.e.

$$\mathsf{Y}_i = M_i(\mathsf{X}_V) + \mathsf{N}_i \qquad \forall i \in V \tag{7.32}$$

We now simplify the condition on Y_D as a condition on X_B. Let

- $\bar{n}_D(y_D)$ be the value of N_D given $\mathsf{Y}_D = y_D \in Y_D$ by (7.25),
- $S_{\mathsf{X}_B}(x_{B^c}, y_D) := \{x_B \in X_B : P_{\mathsf{Y}_D|\mathsf{X}_V}(y_D|x_V) > 0\}$,

- $S_{\mathsf{X}_{B^c}\mathsf{Y}_D} := \{(x_{B^c}, y_D) \in X_{B^c} \times Y_D : S_{\mathsf{X}_B}(x_{B^c}, y_D) \neq \emptyset\}$, and
- $\bar{x}_B(x_{B^c}, y_D)$ be a particular solution $x_B \in X_B$ to $y_D = M_D(x_V) + \bar{n}_D(y_D)$ for $(x_{B^c}, y_D) \in S_{\mathsf{X}_{B^c}\mathsf{Y}_D}$.

It follows from (7.32) that

$$S_{\mathsf{X}_B}(x_{B^c}, y_D) = \bar{x}_B(x_{B^c}, y_D) + \mathrm{Null}(M_D^B(\cdot, \mathbf{0}))$$

Furthermore, for $(x_{B^c}, y_D) \in S_{\mathsf{X}_{B^c}\mathsf{Y}_D}$,

$$\begin{aligned} H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{Y}_D = y_D) &\overset{\text{(a)}}{=} H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{Y} = y_D, \mathsf{N}_D = \bar{n}_D(y_D)) \\ &\overset{\text{(b)}}{=} H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D), \mathsf{N}_D = \bar{n}_D(y_D)) \\ &\overset{\text{(c)}}{=} H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D)) \end{aligned}$$

(a) $\mathsf{Y}_D = y_D$ implies $\mathsf{N}_D = \bar{n}_D(y_D)$.

(b) Conditioned on $(\mathsf{X}_{B^c}, \mathsf{N}_D) = (x_{B^c}, \bar{n}_D(y_D))$, we have $\mathsf{Y}_D = y_D$ iff $\mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D)$.

(c) $(\mathsf{Y}_C, \mathsf{X}_V, \mathsf{N}_C)$ is independent of N_D by (7.26).[6]

It suffices now to show that uniform P_{X_V} maximizes $H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D))$ to a constant independent of $(x_{B^c}, y_D) \in S_{\mathsf{X}_{B^c}\mathsf{Y}_D}$.

Conditioned on $\mathsf{X}_{B^c} = x_{B^c}$ and $\mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D)$,

$$\begin{aligned} \mathsf{Y}_C = M_C^B(\mathsf{X}_B, x_{B^c}) + \mathsf{N}_C = M_C^B(\mathsf{X}, \mathbf{0}) + \mathsf{N} \qquad &\text{where} \\ \mathsf{X} := \mathsf{X}_B - \bar{x}_B(x_{B^c}, y_D)& \\ \mathsf{N} := M_C^B(-\bar{x}_B(x_{B^c}, y_D), x_{B^c}) + \mathsf{N}_C& \end{aligned}$$

The condition $\mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D)$ implies $\mathsf{X} \in \mathrm{Null}(M_D^B(\cdot, \mathbf{0}))$. Viewing X and Y_C as an input and output to a simple finite homomorphic channel, we have by Lemma 7.1 that uniform P_{X_V} attains the maximum

$$H(\mathsf{Y}_C|\mathsf{X}_{B^c} = x_{B^c}, \mathsf{X}_B \in S_{\mathsf{X}_B}(x_{B^c}, y_D)) = \log\left|M_C^B(\mathrm{Null}(M_D^B(\cdot, \mathbf{0})), \mathbf{0})\right| + H(\mathsf{S})$$

for some S where $H(\mathsf{S})$ depends only on P_{N_C} but not (x_{B^c}, y_D) as desired. ■

[6] More precisely, $s \notin C$ and so $(\mathsf{X}_V, \mathsf{N}_C)$ is independent of N_D. Y_C is a function of X_V and N_C.

7.3 Suboptimality of source emulation

In this section, we study how source emulation approach can be strictly suboptimal. In particular, we will construct as follows an example where one can achieve a key rate strictly larger than the secrecy lower bound in Theorem 6.3.

1. First, we construct a simple DMMC for two active terminals, called the *consensus channel*, such that the secrecy upper bound is strictly larger than the secrecy lower bound by source emulation.
2. Then, we construct a DMMS from the optimal input distribution that gives the secrecy upper bound.
3. Finally, we combine the DMMC and the DMMS as simultaneous independent components of a composite DMMC, called the *augmented consensus channel.*

We will show that the secrecy lower bound of the composite channel is strictly smaller than the upper bound, but one can achieve the secrecy upper bound by having the active terminal directly feed its last observation from the component DMMS to the input of the component DMMC. The source emulation approach is strictly suboptimal because it cannot provide the optimal correlation between the channel input symbols that are readily available from the component DMMC.

<u>Consensus channel:</u>

Consider two active terminals T_1 and T_2. i.e. $A = [2] = D^c = V$. The DMMC is a *binary consensus channel* defined as follows:

terminal	1	2
input	$\mathsf{X}_1 \in \{0,1\}$	$\mathsf{X}_2 \in \{0,1\}$
output	$\mathsf{Y} \in \{0,1\}$	Y

The output bit Y is defined as

$$\mathsf{Y} := \begin{cases} \mathsf{X}_1 & \text{if } \mathsf{X}_1 = \mathsf{X}_2 \\ \mathsf{N} & \text{otherwise} \end{cases} \tag{7.33}$$

where N is a uniformly random bit independent of the channel input bits X_1 and X_2. We can think of the channel as a way to reach consensus Y by a random coin flip if the terminals do not provide identical input bits. More practically, we can think of Y as a noisy observation that depends on the inputs iff they adds coherently.

By definition (A.8), we have $\mathcal{H}_{A|D} = \{\{1\},\{2\}\}$. There is only one possible fractional partition λ in $\Lambda_{A|D}$, namely the one with $\lambda_{\{1\}} = \lambda_{\{2\}} = 1$. All the minimization over fractional partitions in the secrecy bounds become trivial. Furthermore, the secrecy lower bounds by pure and mixed source emulations are the same.

From the computations in Section C.2, the secrecy lower bound C_{se} by source emulation is strictly smaller than the secrecy upper bound C_{su}.

$$C_{\text{se}} \approx 1.12 < C_{\text{su}} \approx 1.17$$

Using this, we will prove that private input adaptation strictly outperforms the mixed source emulation (or any public input adaptation) for the DMMC below.

<u>Augmented consensus channel:</u>

Let $P_{\mathsf{Y}_{1V}|\mathsf{X}_V}$ be the consensus channel defined in (7.33) and $P_{\mathsf{Y}_{2V}}$ be a DMMS with

$$P_{\mathsf{Y}_{2V}} := \begin{bmatrix} P_{\mathsf{Y}_{21}\mathsf{Y}_{22}}(0,0) & P_{\mathsf{Y}_{21}\mathsf{Y}_{22}}(0,1) \\ P_{\mathsf{Y}_{21}\mathsf{Y}_{22}}(1,0) & P_{\mathsf{Y}_{21}\mathsf{Y}_{22}}(1,1) \end{bmatrix} := \begin{bmatrix} \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} \end{bmatrix}$$

By the computation in Section C.2.2, $P_{\mathsf{Y}_{2V}}$ is the optimal input distribution that gives the secrecy upper bound for the consensus channel. The augmented consensus channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ is defined as

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = P_{\mathsf{Y}_{1V}|\mathsf{X}_V} P_{\mathsf{Y}_{2V}}$$

which corresponds to the simultaneous and independent use of the consensus channel $P_{\mathsf{Y}_{1V}|\mathsf{X}_V}$ and the DMMS $P_{\mathsf{Y}_{2V}}$.

By Theorem 6.3, the secrecy lower bound by source emulation is

$$\begin{aligned}
\tilde{C}_{\text{se}} &:= \min_{\lambda\in\Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_V}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \\
&\overset{\text{(a)}}{=} \max_{P_{\mathsf{X}_V}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \gamma(\mathbf{1}, P_{\mathsf{X}_V}) \\
&\overset{\text{(b)}}{=} \max_{P_{\mathsf{X}_V}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \gamma(\mathbf{1}, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_{1V}|\mathsf{X}_V}} + \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}}) \\
&\overset{\text{(c)}}{=} C_{\text{se}} + \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}}) \approx 2.04
\end{aligned}$$

(a) The minimization is trivial because there is only one possible fractional partition, namely $\mathbf{1}$. Furthermore, $\tilde{\beta} = \gamma$ by Proposition A.4 since $D = \emptyset$ satisfies the single-leakage condition (A.27).

(b) This is by the equality case of (A.30b) and the equalities that

$$\gamma(\mathbf{1}, 1)|_{P_{\mathsf{Y}_{2V}}} = \tilde{\beta}(\mathbf{1}, 1)\Big|_{P_{\mathsf{Y}_{2V}}} = \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}})$$

which follows from Proposition A.4 and the definition (A.19) of β.

(c) This is by definition of C_{se} in Section C.2.2.

By Theorem 6.1, the secrecy upper bound is

$$\begin{aligned}
\tilde{C}_{\text{su}} &:= \min_{\lambda\in\Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_V}} \alpha(\mathbf{1}, P_{\mathsf{X}_V}) \\
&\overset{\text{(a)}}{=} \max_{P_{\mathsf{X}_V}} \alpha_1(\mathbf{1}, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_{1V}|\mathsf{X}_V}} + \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}}) \\
&\overset{\text{(b)}}{=} C_{\text{su}} + \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}}) \approx 2.09
\end{aligned}$$

(a) The minimization is trivial because there is only one possible fractional partition. (a) follows from the equivalence $\alpha = \alpha_1$ by Proposition A.4 due to the single-leakage condition ($\because D = \emptyset$), the equality case of (A.30a), and the equalities

$$\alpha_1(\mathbf{1}, 1)|_{P_{\mathsf{Y}_{2V}}} = \tilde{\beta}(\mathbf{1}, 1)\Big|_{P_{\mathsf{Y}_{2V}}} = \beta(\mathbf{1}, P_{\mathsf{Y}_{2V}})$$

which follows from Proposition A.4 and the definition (A.19) of β.

(b) This is by the definition of C_{su} in Section C.2.2 and $\alpha = \alpha_1$ by Proposition A.4.

Since $C_{\mathrm{se}} < C_{\mathrm{su}}$, we have $\tilde{C}_{\mathrm{se}} < \tilde{C}_{\mathrm{su}}$ as desired. It remains to show that $\tilde{C}_{\mathrm{su}}$ is achievable by some other scheme. Consider setting the input X_{it} from terminal $i \in V$ at time $t \in [n]$ to the previous observation $\mathsf{Y}_{2i(t-1)}$ from the DMMS at time $t-1$. i.e.

$$\mathsf{X}_{it} \leftarrow \mathsf{Y}_{2i(t-1)} \qquad \text{for all } i = 1, 2$$

Asymptotically, it is equivalent to having a DMMS $\mathsf{Y}_V = \mathsf{Y}_{[2]V}$ with $P_{\mathsf{Y}_{2V}|\mathsf{Y}_{1V}} = P_{\mathsf{Y}_{2V}|\mathsf{X}}$. It is straightforward to show that $\tilde{C}_{\mathrm{su}}$ is the key rate achievable by Theorem 6.4.Hence, the private input adaptation scheme is strictly better than the source emulation approach. The secrecy lower bound in Theorem 6.3 is loose.

Conjecture:

We conjecture that the secrecy upper bound in Theorem 6.1 is loose for a variant of the consensus channel, called the *public consensus channel*, where the channel output is publicly observable. More precisely, we have

$$A = [2] = D^c \subseteq V = [3]$$

terminal	1	2	3
input	$\mathsf{X}_1 \in \{0,1\}$	$\mathsf{X}_2 \in \{0,1\}$	
output			$\mathsf{Y} \in \{0,1\}$

The dummy untrusted terminal 3 observes Y defined in (7.33). This is equivalent to revealing Y in public, since doing so does not lose optimality. It can be shown that the secrecy lower bound by source emulation and the upper bound are $C_{\mathrm{se}} - 1$ and $C_{\mathrm{su}} - 1$ respectively, with C_{se} and C_{su} defined in Section C.2 for the original consensus channel. It follows that the secrecy bounds do not match. However, with Y already revealed in public, it does not seem plausible to increase the key rate by any private input adaptation approach. Thus, we conjecture that the secrecy upper bound is loose, and the stronger statement that the secrecy lower bound is the secrecy capacity.

Chapter 8

Conclusion

Consider the secret agreement game in Section 1.1. The maximum amount of secret the users can generate corresponds to the secrecy capacity for the secret key agreement problem. We can now think in mathematical terms that the secrecy capacity is the mutual dependence of the users, and intuitively the maximum amount of consensus that needs not be publicly discussed.

Under a special source model when the dependence structure is linear, or can be modeled by a dependency hypergraph, the secrecy capacity can be interpreted as network information flow or partition connectivity. There is also a systematic and practical way to generate the secret by network coding.

When the users are given time to discuss privately, they are indeed given a private channel to generate secret. We now know that a mixed strategy in choosing the channel inputs can strictly outperform a pure strategy in some scenario. While such cooperation scheme is unnecessary for many classes of channels, it may not be enough for others. In some cases, the users should consider adapting their channel input to previous private observations.

Through this work, we are beginning to capture the fundamentals of information in the multiuser setting. Its intimate connection with combinatorics such as graph theory and matroid theory allows us to discover new identities and generalizations in those areas. Although the focus here is in the secrecy framework, we attempt to develop general techniques or insights that may potentially drive new applications.

Appendices

Appendix A

Information Theory

In this section, we introduce the mathematical tools for deriving the upper and lower bounds on the secrecy capacities in Chapter 6. Section A.1 defines the basic measures of randomness when dealing with a mixture of continuous and discrete random variables. Section A.2 is a brief summary of the Shearer-type lemma useful for deriving the secrecy upper bounds. Section A.3 applies the von Neumann's minimax theorem [35, 53] and Eggleston-Carathéodory theorem [19] for the construction of the achieving scheme that gives the secrecy lower bounds. Finally, in Section A.4, we define the expressions that characterize the secrecy upper and lower bounds, and derive some useful properties for studying the tightness of the bounds in Chapter 7.

A.1 Information measures

In addition to the basic definitions for entropy, mutual information and divergence, we will point out certain caveats and technicalities needed when handling a mixture of discrete and continuous random variables, since the model in Chapter 5 consists of both discrete and continuous random variables. For example, the randomizations are continuous-valued while the public messages and secret keys are finitely-valued.

To measure the randomness of purely discrete random variables, we have the classical definition of *entropy* by Shannon [52],

$$H(\mathsf{X}) := \mathop{\mathrm{E}}_{\mathsf{X}\sim P_{\mathsf{X}}}\left[\log\frac{1}{P_{\mathsf{X}}(\mathsf{X})}\right] = \sum_{x\in X} P_{\mathsf{X}}(x)\log\frac{1}{P_{\mathsf{X}}(x)}$$

where P_{X} is the probability mass function of X. For continuous random variable $Y \in \mathbb{R}$ with probability measure absolutely continuous with respect to the Lebesgue measure[1], *differential entropy* is used, i.e.

$$H(\mathsf{Y}) := \mathrm{E}\left[\log\frac{1}{P_{\mathsf{Y}}(\mathsf{Y})}\right] = \int_{\mathbb{R}} P_{\mathsf{Y}}(y)\log\frac{1}{P_{\mathsf{Y}}(y)}\,\mathrm{d}y$$

For a mixed-pair $\mathsf{Z} := (\mathsf{X},\mathsf{Y})$ of discrete and continuous random variables X and respectively Y, a natural "combination" of the above entropy measures is

$$H(\mathsf{Z}) := \mathrm{E}\left[\log\frac{1}{P_{\mathsf{X},\mathsf{Y}}(\mathsf{X},\mathsf{Y})}\right] = \sum_{x\in X}\int_{\mathbb{R}} P_{\mathsf{X},\mathsf{Y}}(x,y)\log\frac{1}{P_{\mathsf{X},\mathsf{Y}}(x,y)}\,\mathrm{d}y$$

where $P_{\mathsf{X},\mathsf{Y}}(x,y)$ is the probability density function of $\Pr(\mathsf{X}=x,\mathsf{Y}\leq y)$ in y. The classical entropy measures can be considered as special cases.

This is considered formally by Nair et al. [45], where the above definition is also extended in the same way to the multivariate case as follows.[2]

Entropy:

Consider a random vector $\mathbf{Z} = (\mathsf{Z}_i : i \in [\ell+1])$ where $\ell \in \mathbb{P}$ is a positive integer, $\mathsf{Z}_{\ell+1}$ is a discrete random variable with possibly unbounded support $Z_{\ell+1}$ and Z_i's for $i \in [\ell]$ are real-valued continuous random variables with absolutely continuous probability measure. Let

$$P_{\mathbf{Z}}(z_{[\ell+1]}) = P_{\mathsf{Z}_{\ell+1}}(z_{\ell+1})P_{\mathsf{Z}_{[\ell]}|\mathsf{Z}_{\ell+1}}(z_{[\ell]}|z_{\ell+1})$$

for all $z_{[\ell+1]} \in \mathbb{R}^{\ell}\times Z_{\ell+1}$, where $P_{\mathsf{Z}_{\ell+1}}$ is the probability mass function for $\mathsf{Z}_{\ell+1}$ and

[1] Absolute continuity is the technical condition needed in the fundamental theorem of calculus for the Lebesgue integral [32].

[2] For generality, [45] shows that the definition also applies to mixed random variables [47], which is discrete-valued with a probability strictly within $(0,1)$.

$P_{\mathsf{Z}_{[\ell]}|\mathsf{Z}_{\ell+1}}(\cdot|z_{\ell+1})$ is the conditional probability density function for $\mathsf{Z}_{[\ell]}$ given $\mathsf{Z}_{\ell+1} = z_{\ell+1}$. The entropy is defined as

$$\begin{aligned} H(\mathbf{Z}) &:= \mathrm{E}\left[\log \frac{1}{P_{\mathbf{Z}}(\mathbf{Z})}\right] \\ &= \sum_{z_{\ell+1} \in Z_{\ell+1}} \int_{\mathbb{R}^\ell} P_{\mathbf{Z}}(z_{[\ell+1]}) \log \tfrac{1}{P_{\mathbf{Z}}(z_{[\ell+1]})} \, \mathrm{d}z_{[\ell]} \end{aligned} \qquad \text{(Entropy)} \qquad \textbf{(A.1)}$$

The conditional entropy is defined in the usual way,

$$H(\mathbf{Z}|\mathsf{U}) = \mathrm{E}\left[-\log P_{\mathbf{Z}|\mathsf{U}}(\mathbf{Z}|\mathsf{U})\right] \qquad \text{(Conditional entropy)}$$

For the entropy to be well-defined, the following constraint is imposed [45],

$$\sum_{z_{\ell+1} \in Z_{\ell+1}} \int_{\mathbb{R}^\ell} \left|P_{\mathbf{Z}}(z_{[\ell+1]}) \log P_{\mathbf{Z}}(z_{[\ell+1]})\right| \, \mathrm{d}z_{[\ell]} < \infty \qquad \textbf{(A.2)}$$

We require further that the above holds also for *Riemann integral* [32], which is a technicality needed for Lemma 6.1.

The usual properties of the classical entropy measures follow immediately from these definitions. For X, Y and Z that are mixtures of discrete and continuous random variables, we have the chain rule expansion

$$H(\mathsf{XY}) = H(\mathsf{X}) + H(\mathsf{Y}|\mathsf{X}) = H(\mathsf{Y}) + H(\mathsf{X}|\mathsf{Y}) \qquad \textbf{(A.3)}$$

and the fact that successive conditioning reduces entropy (or equivalently positivity of mutual information),

$$\begin{aligned} I(\mathsf{X} \wedge \mathsf{Y}|\mathsf{Z}) &:= H(\mathsf{X}|\mathsf{Z}) - H(\mathsf{X}|\mathsf{YZ}) \\ &= H(\mathsf{Y}|\mathsf{Z}) - H(\mathsf{Y}|\mathsf{XZ}) \geq 0 \end{aligned} \qquad \text{(Mutual information)} \qquad \textbf{(A.4)}$$

with equality iff X and Y are conditionally independent given Z. For discrete random variable X, we have the additional positivity property that

$$H(\mathsf{X}|\mathsf{Y}) \geq 0 \tag{A.5}$$

with equality iff Y, which may be continuous-valued, completely determines X. This entropy is also preserved under bijection. For general mixed-pair that has a continuous component, the entropy need not be positive, just like the differential entropy, nor does it have to be preserved under bijection. In particular, it is not invariant under scaling of the random variable.[3]

The information divergence between two distributions P_{Z} and P'_{Z} is defined as

$$D(P_{\mathsf{Z}} \| P'_{\mathsf{Z}}) := \underset{\mathsf{Z} \sim P_{\mathsf{Z}}}{\mathrm{E}} \left[\log \frac{P_{\mathsf{Z}}(\mathsf{Z})}{P'_{\mathsf{Z}}(\mathsf{Z})} \right] \qquad \text{(Divergence)} \tag{A.6}$$

where the expectation is taken over Z distributed as P_{Z} (not P'_{Z}). $D(p\|q)$ satisfies the usiual positivity and convexity in (p,q) by Jensen's inequality [8].

We also consider the following generalization $\tilde{D}(f\|g)$ of the divergence operation to non-negative but not necessarily stochastic functions $f, g : \mathbb{R}^{\ell} \times Z_{\ell+1} \mapsto \mathbb{R}_+$ on the mixture of discrete set $Z_{\ell+1}$ and finite-dimensional Euclidean space $\mathbb{R}^{\ell}$.

$$\tilde{D}(f\|g) := \sum_{Z_{\ell+1}} \int_{\mathbb{R}^{\ell}} f(z_{[\ell+1]}) \log \frac{f(z_{[\ell+1]})}{g(z_{[\ell+1]})} \, \mathrm{d}z_{[\ell]} \tag{A.7}$$

$\tilde{D}(f\|g)$ is also convex in (f,g) by the log-sum inequality [8] but may not be positive.

A.2 Shearer-type lemma

In this section, we introduce a set of inequalities, collectively referred to as Shearer-type lemma, from [38] and [13] that is useful in deriving the secrecy upper bounds in Section 6.1. They require the following notion of fractional partitions from fractional graph theory [49].

[3] It is possible, however, to add a factor involving an invariant measure as in [31] to ensure scale invariance for the case of continuous random variable.

Fractional partition:

For finite sets A, D and V satisfying

$$A \subseteq D^c \subseteq V : |A| \geq 2$$

where $D^c := V \setminus D$ denotes the *complement* w.r.t. V, we define the following *set system/hypergraph* with no multiple edges,

$$\mathcal{H}_{A|D} := \{B \subsetneq D^c : B \neq \emptyset \text{ and } B \not\supseteq A\} \tag{\textbf{A.8}}$$

Each element in $\mathcal{H}_{A|D}$ is called a *hyperedge*, which is just a subset of the vertices in D^c. The corresponding set of all *fractional (edge) partitions* is defined as

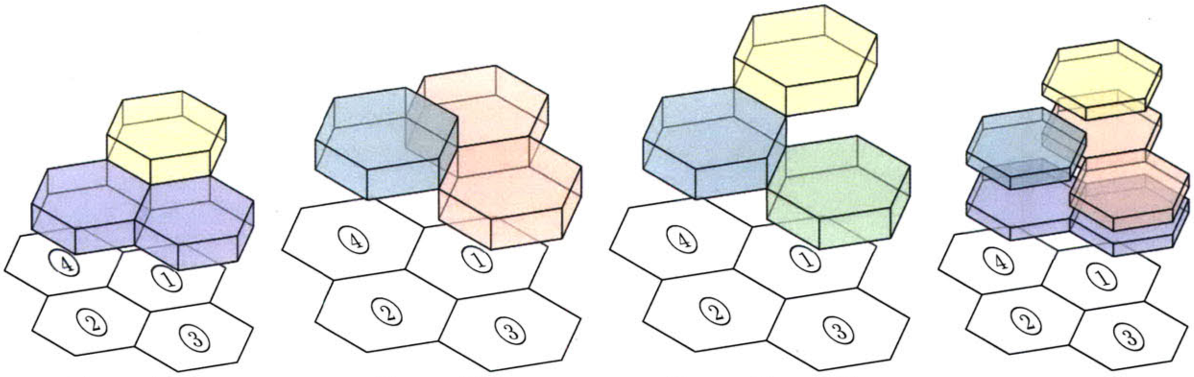
$$\Lambda_{A|D} := \left\{ \lambda = (\lambda_B : B \in \mathcal{H}_{A|D}) : \lambda_B \geq 0 \text{ and } \sum_{B \ni i} \lambda_B = 1 \qquad , \forall i \in D^c \right\} \tag{\textbf{A.9}}$$

where $\sum_{B \ni i}$ is a shorthand notation for $\sum_{B \in \mathcal{H}_{A|D} : i \in B}$. This is illustrated in Figure A-1. The name *fractional partition* comes from the constraint that every vertex in D^c is covered by its incident edges to a total weight of one. Any vertices in D, such as vertex 4 in Figure A-1, are not covered at all.

We say that a fractional partition is *basic* if it is not a convex combination of other fractional partitions. For instance, the fractional partitions in Figure A-1(a), A-1(b) and A-1(c) are basic but the one in A-1(d) is not. It is straightforward to show that $\Lambda_{A|D}$ is a convex set, and in particular, the convex hull of the basic fractional partitions.

In the derivation of the secrecy upper bound, we use the following specialized versions of the Shearer-type lemma.

Lemma A.1 (Shearer-type lemma for entropy functions) *Consider any fractional partition $\lambda \in \Lambda_{A|D}$ defined in* (A.9) *over the hypergraph $\mathcal{H}_{A|D}$. For any random vector Z_V that is a mixture of discrete and continuous random variables, we have the*



(a) $\lambda^{(1)}_{\{1\}} = \lambda^{(1)}_{\{2,3\}} = 1$ (b) $\lambda^{(2)}_{\{2\}} = \lambda^{(2)}_{\{1,3\}} = 1$ (c) $\lambda^{(3)}_{\{1\}} = \lambda^{(3)}_{\{2\}} = \lambda^{(3)}_{\{3\}} = 1$ (d) $\frac{1}{2}(\lambda^{(1)} + \lambda^{(2)})$

Figure A-1: Fractional partitions of $\mathcal{H}_{A|D}$ where $A = [2]$, $D = \{4\}$ and $V = [4]$, i.e. $\mathcal{H}_{A|D} = \{\{1\},\{2\},\{3\},\{1,3\},\{2,3\}\}$. $\lambda^{(k)}$ for $k \in [3]$ defined in (a), (b) and (c) respectively are all the basic fractional partitions of $\mathcal{H}_{A|D}$. (d) is a non-basic fractional partition with weight $\frac{1}{2}$ over the hyperedges in $\mathcal{H}_{A|D} \setminus \{3\}$.

weak form of the Shearer-type lower bound *that*

$$H(\mathsf{X}_{D^c}|\mathsf{X}_D) \geq \sum_{B \in \mathcal{H}_{A|D}} \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}) \qquad \textit{(weak)} \tag{A.10a}$$

With the conditional independence constraint that $P_{\mathsf{X}_{D^c}|\mathsf{X}_D} = \prod_{i \in D^c} P_{\mathsf{X}_i|\mathsf{X}_D}$, *we have the* equality case of the Shearer-type lemma *that*

$$H(\mathsf{X}_{D^c}|\mathsf{X}_D) = \sum_{B} \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}) \qquad \textit{(equality)} \tag{A.10b}$$

For discrete random variable $\mathsf{F}_{[r]}$ *that satisfies the casual relation that*

$$\mathsf{F}_j := F_j(\mathsf{X}_D, \mathsf{X}_{i_j}, \mathsf{F}_{[j-1]})$$

for some $i_j \in V$ *and all* $j \in [r]$, *we have the* Shearer-type lower bound

$$H(\mathsf{F}_{[r]}|\mathsf{X}_D) \geq \sum_{B} \lambda_B H(\mathsf{F}_{[r]}|\mathsf{X}_{B^c}) \qquad \textit{(causal)} \tag{A.10c}$$

□

PROOF By the chain rule (A.3) and the constraint (A.9) on fractional partitions that $\sum_{B\ni i}\lambda_B = 1$ for all $i \in D^c$, we have

$$\begin{aligned}
H(\mathsf{X}_{D^c}|\mathsf{X}_D) &= \sum_{i\in D^c}\left(\sum_{B\ni i}\lambda_B\right) H(\mathsf{X}_i|\mathsf{X}_{[i-1]}\mathsf{X}_D)\\
&= \sum_{B\in\mathcal{H}_{A|D}} \lambda_B \sum_{i\in B} H(\mathsf{X}_i|\mathsf{X}_{[i-1]}\mathsf{X}_D)\\
&\overset{\text{(a)}}{\geq} \sum_{B\in\mathcal{H}_{A|D}} \lambda_B \sum_{i\in B} H(\mathsf{X}_i|\mathsf{X}_{[i-1]\cap B}\mathsf{X}_{B^c})\\
&= \sum_{B} \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c})
\end{aligned}$$

where (a) is due to the fact that conditioning reduces entropy (A.4) since $B^c \supsetneq D$. This proves the weak form (A.10a). The equality case (A.10b) follows from the fact that (a) is satisfied with equality if X_i's are conditionally independent given X_D.

To prove the case with the causal relation, we again apply the expansion by chain rule and fractional partitioning as follows.

$$\begin{aligned}
H(\mathsf{F}_{[r]}|\mathsf{X}_D) &= \sum_{j\in[r]}\left(\sum_{B\ni i_j}\lambda_B\right) H(\mathsf{F}_j|\mathsf{F}_{[j-1]}\mathsf{X}_D)\\
&= \sum_{B\in\mathcal{H}_{A|D}} \lambda_B \sum_{j\in[r]:i_j\in B} H(\mathsf{F}_j|\mathsf{F}_{[j-1]}\mathsf{X}_D)\\
&\overset{\text{(a)}}{\geq} \sum_{B} \lambda_B \sum_{j\in[r]:i_j\in B} H(\mathsf{F}_j|\mathsf{F}_{[j-1]}\mathsf{X}_{B^c})\\
&\overset{\text{(b)}}{=} \sum_{B} \lambda_B \sum_{j\in[r]} H(\mathsf{F}_j|\mathsf{F}_{[j-1]}\mathsf{X}_{B^c})\\
&= \sum_{B} \lambda_B H(\mathsf{F}_{[r]}|\mathsf{X}_{B^c})
\end{aligned}$$

where (a) again follows from the fact that conditional reduces entropy (A.4), and (b) follows from consequence (A.5) of the given causal relation that $i_j \in B^c$ implies $H(\mathsf{F}_j|\mathsf{F}_{[j-1]}\mathsf{X}_{B^c}) = 0$. This proves the causal case (A.10c). ■

A.3 Minimax-type lemma and support lemma

The secrecy upper and lower bounds are both expressed in terms of some minimax optimization problems. This type of problems naturally arises in game theory in the study of two-person zero-sum games [56], where one player maximizes and the other minimizes some real-valued function of their actions. The important notion of mixed strategy in such problem indeed apply to the secrecy problem here. In particular, we describe a mixed strategy for the secrecy scheme in Section 6.2, the secret key rate of which can be characterized with the help of the following minimax-type lemma derived from von Neumann's minimax theorem [35, 53]. The proofs will be given at the end of this section.

Lemma A.2 (Minimax-type lemma) *For any function $f : \Lambda \times S \mapsto \mathbb{R}$ such that $f(\lambda, s)$ is quasi-convex [53] and continuous in λ over a compact convex set Λ, we have*

$$\lim_{|Q|\to\infty} \sup_{\substack{g:Q\to S,\\ P_{\mathsf{Q}}\in\mathscr{P}(Q)}} \min_{\lambda\in\Lambda} \mathrm{E}_{\mathsf{Q}}\left[f(\lambda, g(\mathsf{Q}))\right] = \min_{\lambda\in\Lambda} \sup_{s\in S} f(\lambda, s) \tag{A.11}$$

as long as the R.H.S. is finite. □

In the language of a two-person zero-sum game, $f(\lambda, s)$ is the payoff for Player 1 and the cost for Player 2 when they play s and λ respectively. The expectation on the L.H.S. of (A.11) is the guaranteed payoff when Player 1 plays a mixed strategy of choosing $g(q) \in S$ with probability $P_{\mathsf{Q}}(q)$, followed by Player 2 playing the best response $\lambda \in \Lambda$ to Player 1's strategy. We consider the specific case of interest where f is equal to α (A.17) or $\tilde{\beta}$ (A.18) to be defined in the next section, $\Lambda = \Lambda_{A|D}$ (A.9) and S is the set of valid input distributions P_{X_V}.

The R.H.S. of (A.11) gives us a simpler expression to work with than the L.H.S.. It is still important, however, to solve the L.H.S. of (A.11) directly for the optimal mixed strategy, i.e. the optimal choice of P_{Q} and g. This can be greatly facilitated by the following cardinality bound on Q derived from the Eggleston-Carathéodory theorem [19], as it avoids the complexity of taking the limit as $|Q| \to \infty$.

Lemma A.3 (Support lemma) *Consider the function $f : \Lambda_{A|D} \times S \mapsto \mathbb{R}$ such that $f(\lambda, s)$ is linear and continuous in λ. Let N_λ be the number of connected components of $\{f(\lambda, s) : s \in S\}$. Then, it is admissible to impose the following constraint on the L.H.S. of* (A.11) *without diminishing the optimal value.*

$$\begin{aligned} |Q| &\leq l && \textit{if } N_\lambda \leq l \textit{ for all } \lambda, \textit{ and} && \textbf{(A.12a)} \\ |Q| &\leq l+1 && \textit{otherwise.} && \textbf{(A.12b)} \end{aligned}$$

where $l := 2^{|V|-|D|} - 2^{|V|-|D|-|A|} - |V| + |D|$ is the dimension of $\Lambda_{A|D}$ plus one. □

For example, if $f = \tilde{\beta}$ and the channel input has the finite-alphabet constraint, then

$$\left\{\tilde{\beta}(\lambda, P_0 \textstyle\prod_{i \in D^c} P_i) : P_0 \in \mathscr{P}(X_D), P_i \in \mathscr{P}(X_i), i \in D^c\right\}$$

has only one connected component, i.e. $N_\lambda = 1$, because the map $\tilde{\beta}(\lambda, \cdot)$ is continuous [11, Lemma 2.7] on a connected compact set for every $\lambda \in \Lambda_{A|D}$. In this case, (A.12a) is admissible. If the channel input is infinitely-valued, we may use the slightly larger bound (A.12b) instead.

Proof (Proof of Lemma A.2) Consider proving $\leq$ for (A.11). Indeed, this holds more generally for any $f(\lambda, s)$ that needs not be continuous nor convex in λ as follows.

$$\sup_{g, P_{\mathsf{Q}}} \min_\lambda \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right] \overset{(a)}{\leq} \min_\lambda \sup_{g, P_{\mathsf{Q}}} \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right] \overset{(b)}{=} \min_\lambda \sup_{s \in S} f(\lambda, s)$$

where

(a) because (g, P_{Q}) can be chosen as a function of λ on the right but not left.

(b) It is optimal to have $g(\mathsf{Q})$ deterministic since $\max_{q \in Q} f(\lambda, g(q)) \geq \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right]$.

We now use the convexity and continuity assumption to prove the reverse inequality. First, we can assume without loss of generality that $f(\lambda, s)$ is finite for all λ and s. If not, we can simply clip the value of f to the finite value on the R.H.S. of (A.11). Doing so neither increases the value on the L.H.S. nor decreases the value on the R.H.S.. Fix $\delta > 0$ arbitrarily small. For $\lambda \in \Lambda$, let

$s_\lambda \in S$ be a response to λ that satisfies

$$\sup_{s \in S} f(\lambda, s) - f(\lambda, s_\lambda) < \delta \tag{\textbf{A.13a}}$$

$f_\lambda : \Lambda \mapsto \mathbb{R}$ be the continuous function

$$f_\lambda(\lambda') = f(\lambda', s_\lambda) \qquad \text{for all } \lambda' \in \Lambda \tag{\textbf{A.13b}}$$

$I_\lambda \subseteq \mathbb{R}$ be the open interval

$$\left(\sup_{s \in S} f(\lambda, s) - \delta \ , \ \sup_{s \in S} f(\lambda, s) + \delta \right) \tag{\textbf{A.13c}}$$

Since f_λ is continuous and I_λ is open, $f_\lambda^{-1}(I_\lambda)$ is open. By (A.13a) and (A.13b),

$$\sup_s f(\lambda, s) - \delta < f_\lambda(\lambda) \leq \sup_s f(\lambda, s)$$

which implies that

$$\lambda \in f_\lambda^{-1}(I_\lambda) \tag{\textbf{A.14}}$$

Thus, $\{f_\lambda^{-1}(I_\lambda) : \lambda \in \Lambda\}$ is an open cover of Λ. Since Λ is compact, there exists a finite subcover

$$\{f_q^{-1}(I_q) : q \in Q\} \qquad \text{with} \qquad \bigcup_{q \in Q} f_q^{-1}(I_q) = \Lambda$$

for some finite set $Q \subseteq \Lambda$. Let

$q_\lambda \in Q$ be a quantization of λ such that

$$\lambda \in f_{q_\lambda}^{-1}(I_{q_\lambda}) \tag{\textbf{A.15a}}$$

$g^* : Q \mapsto S$ be the function

$$g^*(q) = s_q \qquad \text{for all } q \in Q \tag{\textbf{A.15b}}$$

By (A.14) and (A.15a), q_λ and λ are both in $f_{q_\lambda}^{-1}(I_{q_\lambda})$, which implies under (A.13a) and (A.15b) that

$$|f(q_\lambda, g^*(q_\lambda)) - f(\lambda, g^*(q_\lambda))| < 2\delta \qquad \text{for all } \lambda \in \Lambda \text{, and so} \tag{A.16}$$

$$\begin{aligned}
\sup_{g,P_{\mathsf{Q}}} \min_{\lambda} \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right] &\overset{\text{(a)}}{=} \sup_{g} \min_{\lambda} \max_{P_{\mathsf{Q}}} \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right] \\
&\overset{\text{(b)}}{\geq} \min_{\lambda} f(\lambda, g^*(q_\lambda)) \\
&\overset{\text{(c)}}{\geq} \min_{\lambda} f(q_\lambda, g^*(q_\lambda)) - 2\delta \\
&= \min_{q\in Q} f(q, g^*(q)) - 2\delta \\
&\overset{\text{(d)}}{\geq} \min_{\lambda\in\Lambda} f(\lambda, s_\lambda) - 2\delta \\
&\overset{\text{(e)}}{\geq} \min_{\lambda} \sup_{s\in S} f(\lambda, s) - 3\delta
\end{aligned}$$

(a) by the Minimax Theorem [53, Theorem 3.4] because the expectation is linear in P_{Q} and quasi-convex in λ over compact convex sets $\mathscr{P}(Q)$ and Λ respectively.

(b) by the specific choice of $g = g^*$ and $\mathsf{Q} = q_\lambda$ with probability 1.

(c) by (A.16).

(d) because $Q \subseteq \Lambda$.

(e) by (A.13a).

Taking the limit as $|Q| \to\infty$ gives the result since δ can be made arbitrarily small.■

PROOF (PROOF OF LEMMA A.3) Let be. Since $f(\lambda, s)$ is linear in λ, it can be written in the matrix form

$$f(\lambda, s) = \boldsymbol{b}_s^T \boldsymbol{\lambda} \qquad \text{with } \boldsymbol{\lambda} \geq \boldsymbol{0} \text{ and } \boldsymbol{M}\boldsymbol{\lambda} = \boldsymbol{1}$$

where $\boldsymbol{b}_s$ is some column vector independent of λ, $\boldsymbol{\lambda}$ is the column-vector form of λ, and $\boldsymbol{M}$ is the incidence matrix of the hypergraph $\mathcal{H}_{A|D}$. The entry of $\boldsymbol{M}$ at row $i \in D^c$ and column $B \in \mathcal{H}_{A|D}$ is $M_{iB} := \mathbb{1}_B(i)$. Note that $\boldsymbol{M}$ is an $(|V| -| D|)$-by-$|\mathcal{H}_{A|D}|$ matrix with full row rank since the columns corresponding to the singleton edges $\{i\}$ for $i \in D^c$ are linearly independent. Thus, solving the above linear equation gives the following solution space in $\Lambda_{A|D}$,

$$\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}} + \boldsymbol{N}\boldsymbol{c} = \begin{bmatrix} \boldsymbol{N} & \hat{\boldsymbol{\lambda}} \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{1} \end{bmatrix}$$

$\hat{\boldsymbol{\lambda}}$ is a particular solution defined as $\hat{\lambda}_B := \mathbb{1}\{|B| = 1\}$ for $B \in \mathcal{H}_{A|D}$.

$\boldsymbol{N}$ is a $|\mathcal{H}_{A|D}|$-by-(l-1) matrix, whose columns form the kernel of $\boldsymbol{M}$, where

$$l := \underbrace{2^{|V|-|D|} - 2^{|V|-|D|-|A|} - 1}_{|\mathcal{H}_{A|D}|} - |V| + |D| + 1$$

$\boldsymbol{c}$ is restricted to the following set to ensure positivity of $\boldsymbol{\lambda}$

$$C := \{\boldsymbol{c} \in \mathbb{R}^{l-1} : \hat{\boldsymbol{\lambda}} + \boldsymbol{N}\boldsymbol{c} \geq 0\}$$

Thus, for every (λ, s), we can write

$$f(\lambda, s) = \underbrace{\boldsymbol{b}_s^T \begin{bmatrix} \boldsymbol{N} & \hat{\boldsymbol{\lambda}} \end{bmatrix}}_{\tilde{\boldsymbol{b}}_s^T :=} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{1} \end{bmatrix} \qquad \text{for some } \boldsymbol{c} \in C$$

Then, by the linearity of expectation,

$$\sup_{g, P_{\mathsf{Q}}} \min_{\lambda} \mathrm{E}\left[f(\lambda, g(\mathsf{Q}))\right] \overset{\text{(b)}}{=} \sup_{g, P_{\mathsf{Q}}} \min_{\boldsymbol{c} \in C} \underbrace{\mathrm{E}\left[\tilde{\boldsymbol{b}}_{g(\mathsf{Q})}\right]^T}_{\circledast} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{1} \end{bmatrix}$$

Note that $\circledast$ is a convex cover of the set $\mathscr{X} := \{\tilde{\boldsymbol{b}}_s : s \in S\} \subseteq \mathbb{R}^l$. By the Eggleston-Carathéodory theorem [19, p.35], every point in the convex cover is a convex combination of at most $l + 1$ points in $\mathscr{X}$, and at most l points if in addition that $\mathscr{X}$ has at most l components. It follows that restricting $|Q|$ to $l + 1$ and l points in the respective cases does not change the space of possible values for $\circledast$, and so the overall optimization gives the same value as desired.[4] ■

[4] This also relies on the observation that $\boldsymbol{c}$ can be chosen as a function of $\circledast$ instead of (g, P_{Q})

A.4 Secrecy expressions

In this section, we derive some important properties of the expressions that characterize the secrecy upper and lower bounds in Chapter 6.

For $\lambda \in \Lambda_{A|D}$ (A.9), distribution P_{X_V}, and DMMC $P_{\mathsf{Y}_V|\mathsf{X}_V}$, we define the following secrecy expression that characterizes the secrecy upper bound.

$$\begin{aligned}
\alpha(\lambda, P_{\mathsf{X}_V}) &= \alpha(\lambda, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \\
&:= H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - \sum_{B\in\mathcal{H}_{A|D}} \lambda_B H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) \\
&\qquad - H(\mathsf{X}_{D^c}|\mathsf{X}_D) + \sum_{B\in\mathcal{H}_{A|D}} \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}) \qquad \textbf{(A.17a)} \\
&= \sum_{B} \lambda_B H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_D|\mathsf{X}_D) - \left(\sum\nolimits_B \lambda_B - 1\right) H(\mathsf{Y}_V|\mathsf{X}_V) \qquad \textbf{(A.17b)} \\
&= \sum_{B\ni i} \lambda_B \left[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_D|\mathsf{X}_D)\right] + \sum_{B\not\ni i} \lambda_B \left[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_V|\mathsf{X}_V)\right] \qquad \textbf{(A.17c)}
\end{aligned}$$

where $i \in D^c$ for (A.17c). We provide different forms (A.17a), (A.17b) and (A.17c) for the same expression because some properties are easier to see in one form than the other. The equivalence will be proved in a bit.

For the secrecy lower bound, we define

$$\begin{aligned}
\tilde{\beta}(\lambda, P_{\mathsf{X}_V}) &= \tilde{\beta}(\lambda, P_{\mathsf{X}_V})\Big|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \\
&:= H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - \sum_{B} \lambda_B H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) \qquad \textbf{(A.18)}
\end{aligned}$$

When given a DMMS P_{X_V} instead of the DMMC $P_{\mathsf{Y}_V|\mathsf{X}_V}$, the secrecy lower bound can be characterized using the following expression,

$$\beta(\lambda, P_{\mathsf{X}_V}) := H(\mathsf{X}_{D^c}|\mathsf{X}_D) - \sum_{B} \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}) \qquad \textbf{(A.19)}$$

To account for the sample average constraint (5.3) in the input, we define for any

without loss of optimality.

trusted terminal $i \in D^c$ the expression

$$\begin{aligned}
\alpha_i(\lambda, P_{\mathsf{X}_V}) &= \alpha_i(\lambda, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \\
&:= H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - \sum_B \lambda_B H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) \\
&\qquad - H(\mathsf{X}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) + \sum_B \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}[\mathsf{Y}_D \mathbb{1}_B(i)]) && \textbf{(A.20a)} \\
&= \sum_B \lambda_B H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - \sum_{B\ni i} \lambda_B H(\mathsf{Y}_D|\mathsf{X}_{B^c}) - \left(\sum\nolimits_B \lambda_B - 1\right) H(\mathsf{Y}_V|\mathsf{X}_V) && \textbf{(A.20b)} \\
&= \sum_{B\ni i} \lambda_B H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) + \sum_{B\not\ni i} \lambda_B \left[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_V|\mathsf{X}_V)\right] && \textbf{(A.20c)}
\end{aligned}$$

where $[\mathsf{Y}_D \mathbb{1}_B(i)]$ equals Y_D if $i \in B$ and 0 otherwise.

To derive the tightness conditions under which the secrecy lower bound meets the upper bound, we define the expression

$$\begin{aligned}
\gamma(\lambda, P_{\mathsf{X}_V}) &= \gamma(\lambda, P_{\mathsf{X}_V})|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \\
&:= H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - \sum_B \lambda_B H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) \\
&\qquad - H(\mathsf{X}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) + \sum_B \lambda_B H(\mathsf{X}_B|\mathsf{X}_{B^c}\mathsf{Y}_D) && \textbf{(A.21a)} \\
&= \sum_B \lambda_B H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) - \left(\sum\nolimits_B \lambda_B - 1\right) H(\mathsf{Y}_{D^c}|\mathsf{X}_V\mathsf{Y}_D) && \textbf{(A.21b)} \\
&= \sum_{B\ni i} \lambda_B H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) + \sum_{B\not\ni i} \lambda_B \Big[H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) - H(\mathsf{Y}_V|\mathsf{X}_V) - I(\mathsf{Y}_D \wedge \mathsf{X}_B|\mathsf{X}_{B^c})\Big] \\
& && \textbf{(A.21c)}
\end{aligned}$$

where $i \in D^c$ for (A.21c).

Proposition A.1 (Equivalence) *The different forms* (A.17a), (A.17b) *and* (A.17c) *for α are equal, and similarly for α_i and γ in* (A.20) *and* (A.21) *respectively.* □

Proof (A.17b) can be obtained from (A.17a) using the identity

$$\begin{aligned} &H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) - H(\mathsf{X}_B|\mathsf{X}_{B^c}) \\ &\qquad = H(\mathsf{X}_V\mathsf{Y}_V) - H(\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) - H(\mathsf{X}_V) + H(\mathsf{X}_{B^c}) \\ &\qquad = H(\mathsf{Y}_V|\mathsf{X}_V) - H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) \end{aligned} \tag{A.22}$$

and the same identity with B replaced by D^c

$$H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - H(\mathsf{X}_{D^c}|\mathsf{X}_D) = H(\mathsf{Y}_V|\mathsf{X}_V) - H(\mathsf{Y}_D|\mathsf{X}_D)$$

Similarly, (A.20b) and (A.21b) can be obtained from (A.20a) and (A.21a) respectively with the additional identity that

$$\begin{aligned} &H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) - H(\mathsf{X}_B|\mathsf{X}_{B^c}\mathsf{Y}_D) \\ &\qquad = H(\mathsf{X}_V\mathsf{Y}_V) - H(\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) - H(\mathsf{X}_V\mathsf{Y}_D) + H(\mathsf{X}_{B^c}\mathsf{Y}_D) \\ &\qquad = H(\mathsf{Y}_{D^c}|\mathsf{X}_V\mathsf{Y}_D) - H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) \end{aligned} \tag{A.23}$$

and the same identity with B replaced by D^c

$$H(\mathsf{X}_{D^c}\mathsf{Y}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) - H(\mathsf{X}_{D^c}|\mathsf{X}_D\mathsf{Y}_D) = H(\mathsf{Y}_{D^c}|\mathsf{X}_V\mathsf{Y}_D)$$

Indeed, (A.23) is the same as (A.22) with the additional conditioning on Y_D.

(A.17b), (A.20b) and (A.21b) can be obtained from (A.17c), (A.20c) and (A.21c) respectively using the constraint (A.9) for fractional partitions that $\sum_{B\ni i}\lambda_B = 1$. Alternatively, using

$$\sum_B \lambda_B - 1 = \sum_{B\not\ni i} \lambda_B$$

we can derive the equivalence in the other direction. ■

All the secrecy expressions α, β, $\tilde{\beta}$, α_i and γ are linear in their first argument $\lambda \in \Lambda_{A|D}$,[5] and the marginal P_{X_D} of their second argument P_{X_V}. In addition, α_i

[5] Functions that are affine in $\lambda \in \Lambda_{A|D}$ are also linear in λ because any constant term can be written as a linear function of λ by the constraint (A.9) that $\sum_{B\ni i}\lambda_B = 1$.

and γ are both concave in the input distribution, while β and $\tilde{\beta}$ are non-negative by (A.10a) of the Shearer-type lemma.

Proposition A.2 (Concavity) *$\alpha_i(\lambda, P_{\mathsf{X}_V})$ and $\gamma(\lambda, P_{\mathsf{X}_V})$ are concave in P_{X_V} for all $\lambda \in \Lambda_{A|D}$.* □

This implies the following (quasi)-concavity that will be useful in deriving the secrecy upper bound under the sample average constraint (5.3).

Corollary A.1 *$\min_{i \in D^c} \alpha_i(\lambda, P_{\mathsf{X}_V})$ is concave (and therefore quasi-concave) in P_{X_V} for all $\lambda \in \Lambda_{A|D}$.* □

PROOF The corollary follows from the concavity of α_i and the fact that the minimum of concave functions is concave.

To prove Proposition A.2, consider each entropy term in (A.20c) and (A.21b). $H(\mathsf{Y}_V|\mathsf{X}_V)$ and $H(\mathsf{Y}_{D^c}|\mathsf{X}_V\mathsf{Y}_D)$ are both linear in P_{X_V}. The other terms can be expressed in terms of the generalized divergence in (A.7) as follows.[6]

$$\begin{aligned} H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) &= -\tilde{D}(P_{\mathsf{X}_{B^c}\mathsf{Y}_{B^c}} \| P_{\mathsf{X}_{B^c}}) \\ H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) &= -\tilde{D}(P_{\mathsf{X}_{B^c}\mathsf{Y}_{B^c}} \| P_{\mathsf{X}_{B^c}\mathsf{Y}_D}) \end{aligned}$$

The entropy terms are concave in P_{X_V} because

- $\tilde{D}(f\|g)$ is convex in the pair (f, g) by the log-sum inequality [8], and
- the arguments $P_{\mathsf{X}_{B^c}\mathsf{Y}_{B^c}}$, $P_{\mathsf{X}_{B^c}}$ and $P_{\mathsf{X}_{B^c}\mathsf{Y}_D}$ are all linear in P_{X_V} with $P_{\mathsf{Y}_V|\mathsf{X}_V}$ fixed.

Thus, α_i and γ have the desired concavity since they are positively weighted sums of concave functions. ■

Problem:

Is $\alpha(\lambda, P_{\mathsf{X}_V})$ quasi-concave in the input distribution P_{X_V}? An affirmative answer would strictly improve the secrecy upper bound with sample average constraint in Theorem 6.2.

[6] The definition of $\tilde{D}(f\|g)$ in (A.7) requires f and g to share the same domain. To do so, we have implicitly used the trivial extension $P_{\mathsf{Z}_1}(z_1, z_2) = P_{\mathsf{Z}_1}(z_1)$ for all z_2. Since this extension is not stochastic, we use the generalized divergence instead of (A.6).

The following identities relate the different secrecy expressions.

Proposition A.3 *For any* $\lambda \in \Lambda_{A|D}$ *and* $P_{\mathsf{X}_V} \in \mathscr{P}(X_V)$,

$$\alpha(\lambda, P_{\mathsf{X}_V}) = \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) - \beta(\lambda, P_{\mathsf{X}_V}) \tag{A.24a}$$

$$= \alpha_i(\lambda, P_{\mathsf{X}_V}) - \sum_{B \ni i} \lambda_B I(\mathsf{Y}_D \wedge \mathsf{X}_{B^c \setminus D} | \mathsf{X}_D) \tag{A.24b}$$

$$\gamma(\lambda, P_{\mathsf{X}_V}) = \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) - \mathrm{E}_{\mathsf{Y}_D}\left[\beta(\lambda, P_{\mathsf{X}_V|\mathsf{Y}_D}(\cdot|\mathsf{Y}_D))\right] \tag{A.25a}$$

$$= \alpha_i(\lambda, P_{\mathsf{X}_V}) - \sum_{B \not\ni i} \lambda_B I(\mathsf{Y}_D \wedge \mathsf{X}_B | \mathsf{X}_{B^c}) \tag{A.25b}$$

□

PROOF (A.24a) follows immediately from (A.17b), (A.18) and (A.19). Similarly, (A.25a) follows from (A.21b), (A.18) and (A.19). (A.24b) follows from (A.17c) and (A.20c) since

$$H(\mathsf{Y}_{B^c \setminus D} | \mathsf{X}_{B^c} \mathsf{Y}_D) - [H(\mathsf{Y}_{B^c} | \mathsf{X}_{B^c}) - H(\mathsf{Y}_D | \mathsf{X}_D)] = I(\mathsf{Y}_D \wedge \mathsf{X}_{B^c \setminus D} | \mathsf{X}_D)$$

Similarly, (A.25b) follows from (A.21c) and (A.20c). ■

From these identities, we can derive sufficient conditions under which the secrecy expressions are equivalent, implying that the secrecy bounds are tight. In particular, we consider the following conditions on the channel input distribution P_{X_V} and the channel statistics $P_{\mathsf{Y}_V|\mathsf{X}_V}$.

Conditional independence condition:

The input distribution P_{X_V} satisfies

$$P_{\mathsf{X}_V} = P_{\mathsf{X}_D} \prod_{i \in D^c} P_{\mathsf{X}_i|\mathsf{X}_D} \tag{A.26}$$

i.e. X_i's are independent over $i \in D^c$ given X_D.

Single-leakage condition:

The channel $P_{\mathsf{Y}_V|\mathsf{X}_V}$ satisfies the single-leakage condition that

$$\exists s \in D^c,\ P_{\mathsf{Y}_D|\mathsf{X}_V} = P_{\mathsf{Y}_D|\mathsf{X}_{D\cup\{s\}}} \tag{A.27}$$

which is independent of the input distribution P_{X_V}. Roughly speaking, the channel output symbols Y_D of the untrusted terminals are affected by the input symbol X_s of at most one trusted terminal (hence the name *single-leakage*) and the input symbols X_D of any untrusted terminals. In particular, this is satisfied if $D = \emptyset$ or $|Y_D| \leq 1$.

These conditions give the following equivalence relations.

Proposition A.4 (Conditions for equivalence) *Consider the following equalities for all $\lambda \in \Lambda_{A|D}$ and $s \in D^c$ that*

$$\tilde{\beta}(\lambda, P_{\mathsf{X}_V}) \overset{(a)}{=} \alpha(\lambda, P_{\mathsf{X}_V}) \overset{(c)}{=} \alpha_s(\lambda, P_{\mathsf{X}_V}) \overset{(b)}{=} \gamma(\lambda, P_{\mathsf{X}_V}) \tag{A.28}$$

The conditional independence condition (A.26) *on the input distribution implies (a), while the single-leakage condition* (A.27) *on the channel statistics implies (b). (c) holds if $|Y_D| \leq 1$ or if both conditions* (A.26) *and* (A.27) *hold.* □

PROOF To show (a), consider the identity (A.24a). By the conditional independence condition (A.26) on the channel input X_V, we have $\beta(\lambda, P_{\mathsf{X}_V}) = 0$ by the equality case (A.10b) of the Shearer-type lemma.

To show (b), consider the identity (A.25b). By the single-leakage condition (A.27) on the channel, we have

$$0 = I(\mathsf{Y}_D \wedge \mathsf{X}_{D^c\setminus\{s\}}|\mathsf{X}_{D\cup\{s\}}) \geq I(\mathsf{Y}_D \wedge \mathsf{X}_B|\mathsf{X}_{B^c}) \qquad \text{for all } B \in \mathcal{H}_{A|D} : B \not\ni s$$

where the last inequality is because $B^c \supseteq D \cup \{s\}$. This implies (b).

To show (c), consider the identity (A.24b). The case when $|Y_D| \leq 1$ is trivial. Consider the other case where both (A.26) and (A.27) hold. We have for all $B \in \mathcal{H}_{A|D}$

such that $B \ni s$ that

$$\begin{aligned} I(\mathsf{Y}_D \wedge \mathsf{X}_{B^c \setminus D}|\mathsf{X}_D) &\leq I(\mathsf{X}_s\mathsf{Y}_D \wedge \mathsf{X}_{D^c\setminus\{s\}}|\mathsf{X}_D) \\ &= \underbrace{I(\mathsf{X}_s \wedge \mathsf{X}_{D^c\setminus\{s\}}|\mathsf{X}_D)}_{\overset{\text{(i)}}{=}0} + \underbrace{I(\mathsf{Y}_D \wedge \mathsf{X}_{D^c\setminus\{s\}}|\mathsf{X}_{D\cup\{s\}})}_{\overset{\text{(ii)}}{=}0} \end{aligned}$$

where (i) and (ii) follow directly from (A.26) and (A.27) respectively.[7] This gives the desired equality (c). ■

Suppose the DMMC consists of a set of ℓ simultaneous[8] independent channels defined below.

Simultaneous independent channels:

$$P_{\mathsf{Y}_V|\mathsf{X}_V} = \prod_{j \in L} P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}} \tag{A.29}$$

where $L := [\ell]$ for some positive integer ℓ.

Then, α_i and γ satisfy the following maximality of independent input distribution, which is useful in studying the optimality of the secrecy bounds in Chapter 7.

Proposition A.5 (Maximality of independent input) *Given $P_{\mathsf{Y}_V|\mathsf{X}_V}$ consists of a set $\{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}} : j \in L\}$ of simultaneous independent channels* (A.29)*, we have*

$$\alpha_i(\lambda, P_{\mathsf{X}_{LV}})|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \leq \sum_{j \in L} \alpha_i(\lambda, P_{\mathsf{X}_{jV}})|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}} \tag{A.30a}$$

$$\gamma(\lambda, P_{\mathsf{X}_{LV}})|_{P_{\mathsf{Y}_V|\mathsf{X}_V}} \leq \sum_{j \in L} \gamma(\lambda, P_{\mathsf{X}_{jV}})|_{P_{\mathsf{Y}_{jV}|\mathsf{X}_{jV}}} \tag{A.30b}$$

with equality if X_{jV}'s are independent over $j \in L$. □

[7] Alternatively, one can show (c) using the identity (A.25a) and the fact that X_i's for $i \in D^c$ are conditionally independent given $(\mathsf{X}_D, \mathsf{Y}_D)$, which follows from both (A.26) and (A.27).

[8] Simultaneity means no one can observe any channel output symbol until all input symbols are specified.

PROOF We will bound each entropy term in (A.20c) and (A.21b) as follows.

$$\begin{aligned} H(\mathsf{Y}_V|\mathsf{X}_V) = H(\mathsf{Y}_{LV}|\mathsf{X}_{LV}) &\overset{\text{(a)}}{=} \sum_{j\in L} H(\mathsf{Y}_{jV}|\mathsf{X}_{LV}\mathsf{Y}_{[j-1]V}) \\ &\overset{\text{(b)}}{=} \sum_{j\in L} H(\mathsf{Y}_{jV}|\mathsf{X}_{jV}) \end{aligned}$$

where (a) is by the chain rule (A.3) and (b) is by the assumption (A.29) of simultaneity and independence of the component channels. Similarly,

$$\begin{aligned} H(\mathsf{Y}_{D^c}|\mathsf{X}_V\mathsf{Y}_D) = H(\mathsf{Y}_{LD^c}|\mathsf{X}_{LV}\mathsf{Y}_{LD}) &\overset{\text{(a)}}{=} \sum_{j\in L} H(\mathsf{Y}_{jD^c}|\mathsf{X}_{LV}\mathsf{Y}_{LD}\mathsf{Y}_{[j-1]V}) \\ &\overset{\text{(b)}}{=} \sum_{j\in L} H(\mathsf{Y}_{jD^c}|\mathsf{X}_{jV}\mathsf{Y}_{jD}) \end{aligned}$$

where (a) and (b) follow from the same reasoning as before. The remaining entropy terms can be bounded using (A.4) as follows.

$$\begin{aligned} H(\mathsf{Y}_{B^c\setminus D}|\mathsf{X}_{B^c}\mathsf{Y}_D) = H(\mathsf{Y}_{LB^c\setminus D}|\mathsf{X}_{LB^c}\mathsf{Y}_{LD}) &= \sum_{j\in L} H(\mathsf{Y}_{jB^c\setminus D}|\mathsf{X}_{LB^c}\mathsf{Y}_{LD}\mathsf{Y}_{[j-1]B^c\setminus D}) \\ &\leq \sum_{j\in L} H(\mathsf{Y}_{jB^c\setminus D}|\mathsf{X}_{jB^c}\mathsf{Y}_{jD}) \end{aligned}$$

$$\begin{aligned} H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) = H(\mathsf{Y}_{LB^c}|\mathsf{X}_{LB^c}) &= \sum_{j\in L} H(\mathsf{Y}_{jB^c}|\mathsf{X}_{LB^c}\mathsf{Y}_{[j-1]B^c}) \\ &\leq \sum_{j\in L} H(\mathsf{Y}_{jB^c}|\mathsf{X}_{jB^c}) \end{aligned}$$

Substituting these inequalities into (A.20c) and (A.21b) gives (A.30a) and (A.30b).∎

A.5 Proof of duality with source coding

In this section, we detail the proof of the duality in Theorem 2.3 between secret key agreement and maximum common randomness under public discussion rate constraint. The proof relies on the following error exponent from [11] for the source coding problem and a straight-forward extension of Lemma B.2 in [12].

Proposition A.6 (Source coding exponent) *Given that terminal $i \in A$ observes n samples of the finitely-valued DMS Z_i, there exists a decoder $\phi_i : F \times Z_i^n \mapsto$ for every $i \in A \setminus \{1\}$ satisfying the following error bound in expectation when the code $\theta : Z_1^n \mapsto F$ is uniformly randomly chosen.*

$$\begin{aligned}\Pr\{\exists i \in A \setminus \{1\}, \mathsf{Z}_1^n \neq \phi_i(\theta(\mathsf{Z}_1^n), \mathsf{Z}_i^n)\}&\\ &\leq |A|(n+1)^{3|Z_A|} 2^{-n\zeta\left(\frac{1}{n}\log|F| - \max_{i\in A\setminus\{1\}} H(\mathsf{Z}_1|\mathsf{Z}_i)\right)} \qquad \textbf{(A.31)}\end{aligned}$$

where $\zeta : \mathbb{R} \mapsto \mathbb{R}_+$ is some non-negative function that depends only on P_{Z_A} but not n, and satisfies $\zeta(\Delta) > 0$ for all $\Delta > 0$. The randomness comes from both the source Z_V and the random code θ. □

This guarantees a deterministic choice of the source code $(\theta, \phi_{A\setminus\{1\}})$ that attains an exponentially decaying error probability

$$\Pr\{\exists i \in A \setminus \{1\}, \mathsf{Z}_1^n \neq \phi_i(\theta(\mathsf{Z}_1^n), \mathsf{Z}_i^n)\} \leq 2^{-n\zeta(\Delta)(1-\Delta)}$$

with some rate $\frac{1}{n}\log|F| \leq \max_{i\in A\setminus\{1\}} H(\mathsf{Z}_1 \mid \mathsf{Z}_i) + \Delta$, provided that $|Z_A|$ grows slowly enough in n.

Proof Consider the *minimum entropy decoder* [11],

$$\phi_i(f, \boldsymbol{z}_i) = \arg \min_{\boldsymbol{z}_1 \in \phi^{-1}(f)} H(\boldsymbol{z}_1, \boldsymbol{z}_i)$$

where $H(\boldsymbol{z}_1, \boldsymbol{z}_i)$ is the joint *empirical entropy* of the sequences $(\boldsymbol{z}_1, \boldsymbol{z}_i)$. The error probability is at most

$$\sum_{i\in A\setminus\{1\}} \sum_{\boldsymbol{z}_1, \boldsymbol{z}_i} P^n_{\mathsf{Z}_1\mathsf{Z}_i}(\boldsymbol{z}_1, \boldsymbol{z}_i) \sum_{\tilde{\boldsymbol{z}}_1 \in \mathcal{E}(\boldsymbol{z}_1, \boldsymbol{z}_i)} \Pr\{\theta(\boldsymbol{z}_1) = \theta(\tilde{\boldsymbol{z}}_1)\} \qquad \textbf{(A.32)}$$

where $\mathcal{E}(\boldsymbol{z}_1, \boldsymbol{z}_i)$ is the set of erroneous decodings defined as

$$\mathcal{E}(\boldsymbol{z}_1, \boldsymbol{z}_i) := \{\tilde{\boldsymbol{z}}_1 \in Z_1^n \setminus \{\boldsymbol{z}_1\} : H(\tilde{\boldsymbol{z}}_1, \boldsymbol{z}_i) \leq H(\boldsymbol{z}_1, \boldsymbol{z}_i)\}$$

The decoder fails to recover the source $\boldsymbol{z}_1$ only if the encoder fails to distinguish between $\boldsymbol{z}_1$ and any sequence $\tilde{\boldsymbol{z}}_1 \in \mathcal{E}(\boldsymbol{z}_1, \boldsymbol{z}_i)$. This happens with probability $\Pr\{\theta(\boldsymbol{z}_1) = \theta(\tilde{\boldsymbol{z}}_1)\} = |F|^{-1}$ since θ is uniformly random. In the language of the method of types [11], The set $Z_1^n \times Z_i^n$ of sequences $(\boldsymbol{z}_1, \boldsymbol{z}_i)$ are partitioned by *type classes* $T_{\mathsf{Z}_1',\mathsf{Z}_i'}$ for different $P_{\mathsf{Z}_1',\mathsf{Z}_i'}$ in the set $\mathscr{P}_n(Z_1, Z_i)$ of possible *joint types.* From [11], we have

$$\begin{aligned}
P^n_{\mathsf{Z}_1\mathsf{Z}_i}(\boldsymbol{z}_1, \boldsymbol{z}_i) &= 2^{-n\left[D(P_{\mathsf{Z}_1'\mathsf{Z}_i'}\|P_{\mathsf{Z}_1\mathsf{Z}_i})+H(\mathsf{Z}_1'\mathsf{Z}_i')\right]} \\
|\mathcal{E}(\boldsymbol{z}_1, \boldsymbol{z}_i)| &\leq \frac{|\mathscr{P}_n(Z_1, Z_i)||T_{\mathsf{Z}_1'\mathsf{Z}_i'}|}{|T_{\mathsf{Z}_i'}|} \\
|\mathscr{P}_n(Z_1, Z_i)| &\leq (n+1)^{|Z_1||Z_i|} \\
|T_{\mathsf{Z}_1'\mathsf{Z}_i'}| &\leq 2^{nH(\mathsf{Z}_1'\mathsf{Z}_i')} \\
|T_{\mathsf{Z}_i'}| &\geq 2^{nH(\mathsf{Z}_i')}|\mathscr{P}_n(Z_1Z_i)|^{-1}
\end{aligned}$$

Applying these inequalities to (A.32) gives the desired error upper bound (A.31) with

$$\zeta(\Delta) := \min_{P_{\mathsf{Z}_1'\mathsf{Z}_i'}\in\mathscr{P}(Z,Z_i),i\in A\setminus\{1\}} \max\{D(P_{\mathsf{Z}_1'\mathsf{Z}_i'}\|P_{\mathsf{Z}_1\mathsf{Z}_i}) + \Delta, 0\}$$

where $\mathscr{P}(Z, Z_i)$ is the set of probability distributions over $Z \times Z_i$. ■

Lemma A.4 (Extension of [12, Lemma B.2]) *Given finitely-valued random variables L', Z' and F', and $d > 0$ such that*

$$\Pr\left\{P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > \frac{1}{d}\right\} \leq \epsilon^2 \qquad \textit{for some } 0 < \epsilon < \frac{1}{9} \tag{A.33}$$

the probability that a uniformly random mapping $\theta : L' \mapsto K'$ fails to satisfy

$$E\left[\sum_{k\in K'}\left|\sum_{l:\theta(l)=k} P_{\mathsf{L}'|\mathsf{F}'\mathsf{Z}'}(l|\mathsf{F}'\mathsf{Z}') - \frac{1}{|K'|}\right|\right] < 7\epsilon \tag{A.34}$$

is at most

$$|K'||F'||L'|2^{1-\frac{\epsilon^3 d}{3|K'||F'|}} \tag{A.35}$$

n.b. [12, Lemma B.2] is a special case when F *is a function of* L. □

Proof Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be the events defined as

$$\mathcal{E}_1 := \left\{ P_{\mathsf{F}'\mathsf{Z}'}(\mathsf{F}',\mathsf{Z}') \geq \frac{\epsilon}{|F'|} P_{\mathsf{Z}'}(\mathsf{Z}') \right\} \tag{A.36}$$

$$\mathcal{E}_2 := \left\{ \Pr\left\{ P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > \frac{1}{d} \middle| \mathsf{Z}' \right\} \leq \epsilon \right\} \tag{A.37}$$

$\mathcal{E}_1$ corresponds to the subset of realizations $(f,z) \in F' \times Z'$ such that $P_{\mathsf{F}'\mathsf{Z}'}(f,z) \geq \frac{\epsilon}{|F'|} P_{\mathsf{Z}'}(z)$. $\mathcal{E}_2$ corresponds to the subset of realizations $z \in Z'$ such that

$$\sum_{l \in L' : P_{\mathsf{L}'|\mathsf{Z}'}(l|z) > \frac{1}{d}} P_{\mathsf{Z}'}(z) \leq \epsilon$$

It follows that the two events are typical as $\epsilon \to 0$ in the sense that

$$\begin{aligned}
\Pr(\mathcal{E}_1^c) &\overset{(a)}{<} \sum_{f \in F', z \in Z'} \frac{\epsilon}{|F'|} P_{\mathsf{Z}'}(z) = \epsilon \\
\Pr(\mathcal{E}_2^c) &\overset{(b)}{<} \frac{\Pr\left\{ P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > \frac{1}{d} \right\}}{\Pr\left\{ P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > \frac{1}{d} \middle| \mathcal{E}_2^c \right\}} \overset{(c)}{\leq} \epsilon
\end{aligned}$$

where (a) is by (A.36) that every possible realization $(\mathsf{F}',\mathsf{Z}') = (f,z)$ for $\mathcal{E}_1^c$ must satisfy $P_{\mathsf{F}'\mathsf{Z}'}(f,z) < \frac{\epsilon}{|F'|} P_{\mathsf{Z}'}(z)$; (b) is by the Bayes' rule; and (c) is by (A.33) and (A.37). Furthermore, given $\mathcal{E}_1 \cap \mathcal{E}_2$, we have

$$\begin{aligned}
P_{\mathsf{L}'|\mathsf{F}'\mathsf{Z}'}(\mathsf{L}'|\mathsf{F}',\mathsf{Z}') &:= \frac{P_{\mathsf{L}'\mathsf{Z}'\mathsf{F}'}(\mathsf{L}',\mathsf{Z}',\mathsf{F}')}{P_{\mathsf{Z}'\mathsf{F}'}(\mathsf{Z}',\mathsf{F}')} \\
&\leq \frac{|F'|}{\epsilon} \frac{P_{\mathsf{L}'\mathsf{Z}'}(\mathsf{L}',\mathsf{Z}')}{P_{\mathsf{Z}'}(\mathsf{Z}')} && \text{by (A.36) and } P_{\mathsf{F}'|\mathsf{L}'\mathsf{Z}'} \leq 1 \\
&= \frac{|F'|}{\epsilon} P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}')
\end{aligned}$$

The probability on the left is larger than $\frac{|F'|}{\epsilon d}$ only if the probability on the right is larger than $\frac{1}{d}$. i.e.

$$\Pr\left\{ P_{\mathsf{L}'|\mathsf{F}'\mathsf{Z}'}(\mathsf{L}'|\mathsf{F}',\mathsf{Z}') > \frac{|F'|}{\epsilon d} \middle| \mathcal{E}_1 \cap \mathcal{E}_2 \right\} \leq \Pr\left\{ P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > \frac{1}{d} \middle| \mathcal{E}_1 \cap \mathcal{E}_2 \right\} \leq \epsilon$$

where the last inequality is by (A.37). (A.34) can then be upper bounded as follows since the expression inside the expectation is at most 2.

$$2\Pr(\mathcal{E}_1^c \cup \mathcal{E}_2^c) + E\left[\sum_{k\in K'}\left|\sum_{l:\theta(l)=k} P_{\mathsf{L}'|\mathsf{F}'\mathsf{Z}'}(l|\mathsf{F},'\mathsf{Z}') - \frac{1}{|K'|}\right| \middle| \mathcal{E}_1 \cap \mathcal{E}_2\right]$$

This gives the desired result since $\Pr(\mathcal{E}_1^c \cup \mathcal{E}_2^c) < 2\epsilon$ as argued previously and, by Lemma B.1 of [12], the conditional expectation above is larger than 3ϵ with probability upper bounded by (A.35). ■

We now break down the proof of Theorem 2.3 into two parts, proving $\leq$ and $\geq$ separately for (2.21).

PROOF (THEOREM 2.3, PART 1) We will show that

$$C(R) \leq C_{\mathrm{s}}(R) + R$$

by showing that a secret key rate of $C(R) - R$ is strongly achievable given a solution that achieves the common randomness capacity $C(R)$ (but not necessarily strongly).

Let $(\dot{\mathsf{U}}_{D^c}, \dot{\mathsf{F}}, \dot{\mathsf{L}}_A, \dot{\mathsf{L}})$ in $\dot{n}$ be an optimal solution to the rate-constrained MCR problem in Definition 2.3 such that

$$C(R) - \Delta_{\dot{n}} \leq \frac{H(\dot{\mathsf{L}}|\mathsf{Z}_D^{\dot{n}})}{\dot{n}} \tag{A.38a}$$

$$\leq \frac{\log|\dot{L}|}{\dot{n}} \leq C(R) + \Delta_{\dot{n}} \tag{A.38b}$$

$$\frac{1}{\dot{n}}\log|\dot{F}| \leq R + \Delta_{\dot{n}} \tag{A.38c}$$

$$\Pr\left\{\exists i \in A, \dot{\mathsf{L}} \neq \dot{\mathsf{L}}_i\right\} \leq \dot{\epsilon}_{\dot{n}} \to 0 \tag{A.38d}$$

where $\Delta_{\dot{n}} \to 0$ arbitrarily slowly. (A.38a) and (A.38b) follow from the definition of the capacity in (2.19) and the uniformity constraint in (2.20b). (A.38c) and (A.38d) follow from the discussion rate constraint in (2.20c) and recoverability constraint in (2.20a) respectively. Based on this optimal solution to the MCR problem, we want to construct a solution to the SKA problem that achieves the rate $C(R) - R$ strongly.

We do so in two steps: first, we use an additional source code in Proposition A.6 to guarantee that the common randomness is strongly recoverable; we then extract a secret key from the common randomness using Lemma A.4.

Assume $1 \in A$ without loss of generality, and consider $((\dot{\mathsf{L}}_i, \mathsf{Z}_D^{\dot{n}}) : i \in A)$ as a DMMS by counting the time in $\dot{n}$-blocks. i.e. we group the first $\dot{n}$ time units into a frame, and the next $\dot{n}$ time units into another frame, etc. The coding is then performed by regarding the samples in each frame as a symbol. By Proposition A.6, there exists a source code $(\ddot{F}, \ddot{\phi}_{A\setminus\{1\}})$ in $\ddot{n}$ for the component source $\dot{\mathsf{L}}_1$ with $\dot{n} = \lceil 1/\Delta_{\ddot{n}} \rceil$[9] such that the error probability decays exponentially as follows,[10]

$$\begin{aligned} &\Pr\left\{\exists i \in A \setminus \{1\}, \dot{L}_1^{\ddot{n}} \neq \ddot{\phi}_i(\ddot{F}(\dot{L}_1^{\ddot{n}}, \mathsf{Z}_D^{n}), \dot{\mathsf{L}}_i^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}})\right\} \\ &\qquad \leq |A|(\ddot{n}+1)^{3(2)^{\dot{n}[|A|(C(R)+\Delta_{\dot{n}})+\log|Z_D|]}} 2^{-\ddot{n}\zeta(\Delta_{\ddot{n}})} \leq 2^{-\dot{n}\ddot{n}\Delta_{\ddot{n}}\zeta(\Delta_{\ddot{n}})(1-\Delta_{\ddot{n}})} \end{aligned} \tag{A.39}$$

while the discussion rate can be made arbitrarily small as follows,

$$\begin{aligned} \frac{1}{\dot{n}\ddot{n}} \log|\ddot{F}| &\leq \frac{1}{\dot{n}}\left[\Delta_{\ddot{n}} + \max_{i\in A\setminus\{1\}} H(\dot{\mathsf{L}}_1|\dot{\mathsf{L}}_i\mathsf{Z}_D^{\dot{n}})\right] \\ &\overset{\text{(a)}}{\leq} \frac{\Delta_{\ddot{n}}}{\dot{n}} + |A|\left[\frac{h(\dot{\epsilon}_{\dot{n}})}{\dot{n}} + \dot{\epsilon}_{\dot{n}}(C(R)+\Delta_{\dot{n}})\right] \\ &\overset{\text{(b)}}{\leq} \Delta_{\dot{n}} \xrightarrow{\ddot{n}\to 0} 0 \end{aligned} \tag{A.40}$$

(a) is by the Fano's inequality [8], (A.38b) and (A.38d). (b) is because we can make $\Delta_{\ddot{n}} \to 0$ arbitrarily slowly. Let

$$\acute{\mathsf{L}} := \begin{cases} \dot{\mathsf{L}}_1 & \text{if } P_{\dot{\mathsf{L}}_1|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}}(\dot{\mathsf{L}}_1|\dot{\mathsf{F}}, \mathsf{Z}_D^{\dot{n}}) \geq |\dot{L}_1|^{-\frac{1}{\Delta_{\dot{n}}}} \geq 2^{-\frac{\dot{n}(C(R)+\Delta_{\dot{n}})}{\Delta_{\dot{n}}}} \\ 0 & \text{otherwise (for some symbol } 0 \in \dot{L}_1) \end{cases} \tag{A.41}$$

[9] As $\ddot{n}$ increases, $\dot{n}$ also increases but arbitrarily slowly. Thus, $\dot{n}$ can be viewed as a constant with respect to $\ddot{n}$.

[10] The first inequality uses the cardinality bound in (A.38b). For the second inequality, we have added the constant factor $\dot{n}\Delta_{\ddot{n}}$ in the exponent, and a factor $(1-\Delta_{\ddot{n}})$ to absorb the remaining terms that are subexponential in $\ddot{n}$ since $\dot{n}$ grows arbitrarily slowly in $\ddot{n}$.

Then, $\acute{\mathsf{L}}$ is determined by $(\dot{\mathsf{L}}_1, \dot{\mathsf{F}}, \mathsf{Z}_D^{\dot{n}})$ and

$$\begin{aligned}
\Pr\{\acute{\mathsf{L}} \neq \dot{\mathsf{L}}_1\} \left\lceil \frac{\log|\dot{L}_1|}{\Delta_{\dot{n}}} \right\rceil &\leq H(\dot{\mathsf{L}}_1|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}) \leq \log|\dot{L}_1| \\
\text{implying } \Pr\{\acute{\mathsf{L}} \neq \dot{\mathsf{L}}_1\} &\leq \Delta_{\dot{n}} \xrightarrow{\ddot{n}\to 0} 0 \qquad \textbf{(A.42)}
\end{aligned}$$

To apply Lemma A.4, let

$$\mathsf{L}' = \acute{\mathsf{L}}^{\ddot{n}}, \qquad \mathsf{Z}' = (\dot{F}^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}}) \quad \text{and} \quad \mathsf{F}' = \ddot{\mathsf{F}}$$

$\log P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}')$ can be expanded into a sum of $\ddot{n}$ independent random variables identically distributed as the random variable $\log P_{\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}}(\acute{\mathsf{L}}|\dot{\mathsf{F}}, \mathsf{Z}_D^{\dot{n}})$, which takes values from the finite interval $[\min \log P_{\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}}, 0]$.[11] By Hoeffding's inequality [29],

$$\Pr\{\log P_{\mathsf{L}'|\mathsf{Z}'}(\mathsf{L}'|\mathsf{Z}') > -\log d\} \leq \epsilon^2 \qquad \text{with} \qquad \textbf{(A.43a)}$$

$$\log d := \ddot{n} \left\lfloor H(\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}) - \Delta_{\ddot{n}} \right\rceil \overset{\text{(a)}}{\geq} \dot{n}\ddot{n}(C(R) - R - \sqrt{\Delta_{\dot{n}}}) \qquad \textbf{(A.43b)}$$

$$-\log \epsilon := \ddot{n} \left[\frac{\Delta_{\ddot{n}}^2}{-\min \log P_{\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}}} \right] \overset{\text{(b)}}{\geq} \dot{n}\ddot{n}\Delta_{\ddot{n}}^5 \qquad \textbf{(A.43c)}$$

where (a) follows from the Fano's inequality, (A.38b), (A.43b) and (A.42) that

$$\begin{aligned}
H(\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}) &\geq H(\dot{\mathsf{L}}_1|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}) - \underbrace{H(\dot{\mathsf{L}}_1|\acute{\mathsf{L}}\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}})}_{\leq h(\Delta_{\dot{n}})+\Delta_{\dot{n}}(C(R)+\Delta_{\dot{n}})\dot{n}} \\
H(\dot{\mathsf{L}}_1|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}) &\geq H(\dot{\mathsf{L}}_1|\mathsf{Z}_D^{\dot{n}}) - \underbrace{H(\dot{\mathsf{F}}|\mathsf{Z}_D^{\dot{n}})}_{\leq \dot{n}(R+\Delta_{\dot{n}})} \\
H(\dot{\mathsf{L}}_1|\mathsf{Z}_D^{\dot{n}}) &\geq H(\dot{\mathsf{L}}|\mathsf{Z}_D^{\dot{n}}) - \underbrace{H(\dot{\mathsf{L}}|\dot{\mathsf{L}}_1\mathsf{Z}_D^{\dot{n}})}_{\leq h(\dot{\epsilon}_{\dot{n}})+\dot{\epsilon}_{\dot{n}}(C(R)+\Delta_{\dot{n}})\dot{n}}
\end{aligned}$$

and (b) follows from the definition of $\acute{\mathsf{L}}$ in (A.41) that

$$-\min \log P_{\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}} \leq -\min \log P_{\dot{\mathsf{L}}_1|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}} \leq \frac{\dot{n}(C(R) + \Delta_{\dot{n}})}{\Delta_{\dot{n}}}$$

[11] $P_{\acute{\mathsf{L}}|\dot{\mathsf{F}}\mathsf{Z}_D^{\dot{n}}}$ is non-zero for every possible realization of $(\acute{\mathsf{L}}, \dot{\mathsf{F}}, \mathsf{Z}_D^{\dot{n}})$.

and the definition $\dot{n} = \lceil 1/\Delta_{\ddot{n}} \rceil$. (A.43a) trivially implies (A.33) and so, by Lemma A.4, there exists $\mathsf{K} = \mathsf{K}'$ as a function of L' such that (A.34) is satisfied with (A.35) strictly smaller than 1, say

$$\frac{1}{\dot{n}\ddot{n}} \log|K| = \frac{1}{\dot{n}\ddot{n}} \log \frac{\epsilon^3 d}{|F'|} - \Delta_{\ddot{n}} \geq C(R) - R - 2\sqrt{\Delta_{\dot{n}}} \tag{A.44}$$

by (A.40), (A.43b) and (A.43c). By (5.11) from [12, Lemma 1] and (A.34),

$$\frac{1}{\dot{n}\ddot{n}} \left[\log|K| - H(\mathsf{K}|\mathsf{Z}_D^{\dot{n}\ddot{n}}\dot{\mathsf{F}}^{\ddot{n}}\ddot{\mathsf{F}})\right] \leq \frac{7\epsilon}{\dot{n}\ddot{n}} \log \frac{|K|}{7\epsilon} \leq 7C(R)2^{-\ddot{n}\dot{n}\Delta_{\ddot{n}}^5} \tag{A.45}$$

With $n := \dot{n}\ddot{n}$, K is a valid secret key that attains the desired rate $C(R) - R$ by (A.44) strongly with exponentially decaying ϵ_n and δ_n for the recoverability and secrecy constraints in (2.18) respectively given by (A.39) and (A.45). ■

Proof (Theorem 2.3, Part 2) We will show that

$$C(R) \geq C_{\mathrm{s}}(R) + R$$

by showing that a common randomness of rate $C_{\mathrm{s}}(R) + R$ is strongly achievable given a solution that achieves the secrecy capacity $C_{\mathrm{s}}(R)$ (but not necessarily strongly). We will outline the main idea of the proof. Some details are similar to Part 1 of the proof above, and are therefore omitted.

Let $(\dot{\mathsf{U}}_{D^c}, \dot{\mathsf{F}}, \dot{\mathsf{K}}_A, \dot{\mathsf{K}})$ in $\dot{n}$ be an optimal solution to the rate-constrained SKA problem in Definition 2.2 such that, for some $\Delta_{\dot{n}} \to 0$ arbitrarily slowly,

$$\begin{gathered} C_{\mathrm{s}}(R) - \Delta_{\dot{n}} \leq \frac{H(\dot{\mathsf{K}}|\mathsf{Z}_D^{\dot{n}}\dot{\mathsf{F}})}{n} \leq \frac{\log|\dot{K}|}{\dot{n}} \leq C_{\mathrm{s}}(R) + \Delta_{\dot{n}} \\ \frac{1}{\dot{n}} \log|\dot{F}| \leq R + \Delta_{\dot{n}} \\ \Pr\left\{\exists i \in A, \dot{\mathsf{K}} \neq \dot{\mathsf{K}}_i\right\} \leq \dot{\epsilon}_{\dot{n}} \to 0 \end{gathered}$$

Assume $1 \in A$ without loss of generality, and consider $\dot{\mathsf{K}}_1$ as a DMMS by counting the time in $\dot{n}$-blocks. By Proposition A.6, there exists a source code $(\bar{F}, \bar{\phi}_{A\setminus\{1\}})$ in $\ddot{n}$

for the sequence $\dot{\mathsf{K}}_1^{\ddot{n}}$ with $\dot{n} = \lceil 1/\Delta_{\ddot{n}} \rceil$ such that

$$\Pr\{\exists i \in A \setminus \{1\}, \dot{K}_1^{\ddot{n}} \neq \bar{\phi}_i(\bar{F}(\dot{K}_1^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}}), \dot{\mathsf{K}}_i^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}})\} \leq 2^{-\ddot{n}\zeta(\Delta_{\ddot{n}})(1-\Delta_{\ddot{n}})}$$

while the discussion rate can be made arbitrarily small,

$$\begin{aligned}
\frac{1}{\dot{n}\ddot{n}} \log|\bar{F}| &\leq \frac{1}{\dot{n}} \left[\Delta_{\ddot{n}} + \max_{i \in A\setminus\{1\}} H(\dot{\mathsf{K}}_1 | \dot{\mathsf{K}}_i \mathsf{Z}_D^{\dot{n}}) \right] \\
&\leq \frac{\Delta_{\ddot{n}}}{\dot{n}} + |A| \left[\frac{h(\dot{\epsilon}_{\dot{n}})}{\dot{n}} + \dot{\epsilon}_{\dot{n}}(C_{\mathrm{s}}(R) + \Delta_{\dot{n}}) \right] \\
&\leq \Delta_{\dot{n}} \xrightarrow{\bar{n}\to\infty} 0
\end{aligned}$$

The reasoning is analogous to that of (A.40) in the previous proof using Fano's inequality and the assumption that the individual keys K_i's agree with small error probability $\dot{\epsilon}_{\dot{n}}$.

To efficiently use the public discussion channel, we further compress the public messages as follows. Consider $\dot{\mathsf{F}}_{[\dot{r}]}$ as a DMMS by counting the time in $\dot{n}$-blocks. By Proposition A.6, there exists a source code $(\ddot{F}_j, \ddot{\phi}_j)$ in $\ddot{n}$ for every sequence $\dot{\mathsf{F}}_j^{\ddot{n}}$ for $j \in [\dot{r}]$ such that the error probability decays exponentially,

$$\Pr\{\dot{\mathsf{F}}_j^{\ddot{n}} \neq \ddot{\phi}_j(\ddot{F}_j(\dot{\mathsf{F}}_{[j]}^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}}), \dot{\mathsf{F}}_{[j-1]}^{\ddot{n}}, \mathsf{Z}_D^{\dot{n}\ddot{n}})\} \leq 2^{-\ddot{n}\zeta(\Delta_{\ddot{n}})(1-\Delta_{\ddot{n}})}$$

and, with $\ddot{F} := \ddot{F}_{[\dot{r}]}$, the total rate is

$$\begin{aligned}
\frac{1}{\dot{n}\ddot{n}} \log|\ddot{F}| &= \frac{1}{\dot{n}} \sum_{j\in[\dot{r}]} H(\dot{\mathsf{F}}_j | \dot{\mathsf{F}}_{[j-1]} \mathsf{Z}_D^{\dot{n}}) + \Delta_{\ddot{n}} \\
&= \frac{1}{\dot{n}} H(\dot{F} | \mathsf{Z}_D^{\dot{n}}) + \Delta_{\ddot{n}}
\end{aligned}$$

n.b. the above source code is successive in the sense that the encoding and decoding of $\dot{\mathsf{F}}_j^{\ddot{n}}$ require the decodings of $\dot{\mathsf{F}}_{[j-1]}^{\ddot{n}}$.

We now modify the public discussion as follows to replace $\dot{\mathsf{F}}^{\ddot{n}}$ by its compressed version $\ddot{\mathsf{F}}$.

1. Given $\dot{\mathsf{F}}_{[j-1]}^{\ddot{n}}$ and $\mathsf{Z}_D^{\dot{n}\ddot{n}}$, the terminal $i_j \in V$ compresses $\dot{\mathsf{F}}_j^{\ddot{n}}$ maximally to a nearly

(conditionally) uniformly distributed message $\ddot{\mathsf{F}}_j$ by the successive source code described earlier with exponentially decaying error in $\ddot{n}$.

2. Terminal i_j then publicly reveal $\ddot{\mathsf{F}}_j$ instead of $\dot{\mathsf{F}}_j^{\ddot{n}}$ for a more efficient use of the public discussion channel. Since every terminal can recover $\dot{\mathsf{F}}_j^{\ddot{n}}$ from $\ddot{\mathsf{F}}_j$, the procedure can continue until $j = \dot{r}$ with exponentially decaying error probability in $\ddot{n}$.[12]
3. Terminal 1 then reveals publicly a private randomization U_1 that is uniformly distributed at rate $R - \frac{1}{\dot{n}}H(\dot{\mathsf{F}}|\mathsf{Z}_D^{\dot{n}})$ and independent of everything else. Thus, the entire public message is uniformly distributed at rate R.

Since $\ddot{\mathsf{F}}^{\ddot{n}}$ can be recovered, the terminals can apply the source code $\bar{F}$ described earlier to encode $\dot{\mathsf{K}}_1^{\ddot{n}}$. As a result, the active terminals can recover $\dot{\mathsf{K}}_1^{\ddot{n}}$, which is nearly uniformly distributed and independent of $\dot{\mathsf{F}}^{\ddot{n}}$, and therefore $\ddot{\mathsf{F}}$, which is also nearly uniformly distributed. Thus, the active terminal attains a common randomness of rate approaching $C_{\mathrm{s}}(R) + R$ as desired. At this point, however, the uniformity condition may not be satisfied strongly. To strengthen it, we can again count the time in $\ddot{n}$-blocks and apply the above scheme independently to each of the $\bar{n}$ blocks with $\ddot{n} = \lceil 1/\Delta_{\bar{n}} \rceil$. Similar to the Proof of Part 1, we can apply another source coding step to have every active terminal agree on a sequence of independent and nearly uniformly distributed random variables at rate $C_{\mathrm{s}}(R) + R$ with error probability decaying exponentially in $\bar{n}$. Then, Lemma A.4 can be applied to obtain the desired common randomness that satisfies the uniformity condition strongly. ■

It is clear from the proof that the duality relation holds also for the capacities defined without private randomizations. It is unknown, however, whether randomization strictly improves the common randomness capacity or secrecy capacity. Indeed, as shown in [12], randomization is unnecessary when the public discussion rate can be larger than the smallest rate of communication for omniscience. Furthermore, as shown in [13], the secret key can be purely a function of Z_i^n for any active terminal $i \in A$. This leads to the following conjecture.

[12] This is by the union bound and the fact that $\dot{r}$ is linear in $\dot{n}$ since $2^{\dot{n}(R+\Delta_{\dot{n}})} \geq \prod_{j \in [\dot{r}]} |\dot{F}_j| \geq 2^{\dot{r}}$.

Conjecture:

The secrecy and common randomness capacities defined in Definition 2.2 and 2.3 can be attained without private randomization, i.e. with U_{D^c} deterministic. Furthermore, the key can be a function of the component source of an active terminal.

Appendix B

Combinatorics of Information Flow

This section contains some identities that are essential in the development of the mutual dependence expression in Chapter 2 and the proof of optimality of the network coding approach in Chapter 3. We first derive in Section B.1 a general identity in submodular function optimization using some generalized notions of partitions. We then introduce a general framework for matroids in Section B.2 that captures the notion of information flow in the secret key agreement problem. In Section B.3, we strengthen the results by assuming a specific dependence structure that can be captured by a dependency hypergraph. In Section B.4, we briefly summarize some related work in understanding the fundamentals of information.

B.1 Submodular function

In the following, we will introduce an identity in submodular function optimization that establishes the mutual dependence expression in Section 2.1. To do so, we need the following generalized notion of partitions.

Definition B.1 Given a finite set V, consider subsets $A \subseteq V : |A| \geq 2$. Define

$$\mathscr{F}(A) := \{B \subseteq V : B \not\supseteq A\} \tag{B.1a}$$

$$\mathscr{F} := \mathscr{F}(V) = 2^V \setminus \{V\} \tag{B.1b}$$

Define $\Phi(A)$ as the collection of families $\mathcal{F} \subseteq 2^V \setminus \{V\}$ that satisfy

$$\forall B \in \mathcal{F},\ B \not\supseteq A \qquad \text{and} \tag{B.2a}$$

$$\forall B' \in \mathcal{F}, B \cup B' \not\supseteq A \implies B \cap B', B \cup B' \in \mathcal{F} \tag{B.2b}$$

We say that $\mathcal{F}$ is an *A-co-intersecting family.*[1] It follows that $\Phi(A) \subsetneq \Phi(A')$ for all $A \subsetneq A'$. In particular, $\mathscr{F}(A') \in \Phi(A') \setminus \Phi(A)$ and $\mathscr{F} \in \Phi(V)$.

Denote the complement of a family $\mathcal{F}$ as $\bar{\mathcal{F}} := \{B^c : B \in \mathcal{F}\}$. Define $\Pi(\mathcal{F}, U)$ for $\mathcal{F} \in \Phi(V)$ and $U \subseteq V$ as the collection of all families $\mathcal{P}$ such that $\{C \cap U : C \in \mathcal{P}\}$ is a set-partition of U into at least 2 non-empty disjoint sets in $\bar{\mathcal{F}}$, i.e. $\mathcal{P} \subseteq \bar{\mathcal{F}} : |\mathcal{P}| \geq 2$ such that

$$\forall C \in \mathcal{P}, C \cap U \neq \emptyset \qquad \text{and} \qquad \forall i \in U, \exists! C \in \mathcal{P} : i \in C \tag{B.3}$$

We say that $\mathcal{P}$ is a partition of U with respect to $\bar{\mathcal{F}}$. It follows that $\Pi(\mathcal{F}, U) \supseteq \Pi(\mathcal{F}, U')$ for all $U \subseteq U'$. For convenience, we write

$$\Pi(A) := \Pi(\mathscr{F}(A), A) \tag{B.4a}$$

$$\Pi := \Pi(\mathscr{F}, V) \tag{B.4b}$$

Π is the set of all set-partitions of V into at least 2 non-empty disjoint subsets of V.

Define $\Lambda(\mathcal{F}, U)$ as the set of $\lambda := (\lambda_B : B \in \mathcal{F})$ satisfying

$$\forall B \in \mathcal{F}, \lambda_B \geq 0 \qquad \text{and} \qquad \forall i \in U, \sum_{B \in \mathcal{F} : i \in B} \lambda_B = 1 \tag{B.5}$$

We say that λ is a fractional partition of U w.r.t. $\mathcal{F}$.[2] It follows that $\Lambda(\mathcal{F}, U) \supseteq \Lambda(\mathcal{F}, U')$ for all $U \subseteq U'$. □

[1] See also the related definitions of intersecting family in [3] and crossing family in [50, p.838].

[2] See the related definition of fractional edge partition in (A.9) and [49].

We say that a function $h : \mathcal{F} \mapsto \mathbb{R}$ is *supermodular* if

$$h(B_1) + h(B_2) \geq h(B_1 \cup B_2) + h(B_1 \cap B_2) \tag{B.6}$$

for all $B_1, B_2, B_1 \cup B_2, B_1 \cap B_2 \in \mathcal{F}$. $-h$ is called a *submodular* function.

Theorem B.1 *Given a finite set $V : |V| \geq 2$, we have for all $A \subseteq V : |A| \geq 2$, $\mathcal{F} \in \Phi(A)$, and supermodular function $h : \mathcal{F} \mapsto \mathbb{R}$ that*

$$\max_{\lambda \in \Lambda(\mathcal{F},A)} \sum_{B \in \mathcal{F}} \lambda_B h(B) = \max_{\mathcal{P} \in \Pi(\mathcal{F},A)} \frac{1}{|\mathcal{P}| - 1} \sum_{C \in \mathcal{P}} h(C^c) \tag{B.7}$$

with the convention that max. *over an empty set is* $-\infty$.[3] □

Proof By the strong duality theorem [15, Table 5.2], the maximization in (B.7) is equal to its linear programming dual

$$\begin{aligned}
&\text{minimize} && \sum_{i \in A} r_i && && \text{(B.8a)}\\
&\text{subject to} && \sum_{i \in B} r_i \geq h(B) && \text{for all } B \in \mathcal{F} && \text{(B.8b)}\\
& && r_i \leq 0 && \text{for all } i \in A^c && \text{(B.8c)}
\end{aligned}$$

The supermodularity property of h translates to the following property on the tight relations of the dual problem.

Subclaim B.1A *For any feasible solution r to the dual linear program* (B.8), *and $B_1, B_2 \in \mathcal{F} : B_1 \cap B_2, B_1 \cup B_2 \in \mathcal{F}$, if B_1 and B_2 are tight constraints, i.e.*

$$\sum_{i \in B_j} r = h(B_j) \qquad \textit{for } j = 1, 2 \tag{B.9a}$$

[3] This gives as a corollary that $\Lambda(\mathcal{F}, A) = \emptyset$ iff $\Pi(\mathcal{F}, A) = \emptyset$.

then $B_1 \cup B_2$ *and* $B_1 \cap B_2$ *are also a tight constraint,*

$$\sum_{i \in B_1 \cup B_2} r_i = h(B_1 \cup B_2) \tag{B.9b}$$

n.b. we need only the tightness of the union $B_1 \cup B_2$ *for the proof of Theorem B.1.* $\triangleleft$

PROOF Since $B_1 \cup B_2 \in \mathcal{F}$, we immediately have $\sum_{i \in B_1 \cup B_2} r_i \geq h(B_1 \cup B_2)$ by (B.8b). The reverse inequality can be proved as follows.

$$\begin{aligned}\sum_{i \in B_1 \cup B_2} r_i &= \sum_{i \in B_1} r_i + \sum_{i \in B_2} r_i - \sum_{i \in B_1 \cap B_2} r_i \\ &\overset{\text{(a)}}{\leq} h(B_1) + h(B_2) - h(B_1 \cap B_2) \\ &\overset{\text{(b)}}{\leq} h(B_1 \cup B_2)\end{aligned}$$

where (a) is by (B.9a) and (B.8b) on $B_1 \cap B_2 \in \mathcal{F}$, and (b) is by the supermodularity of h. With a similar argument, we also have $\sum_{i \in B_1 \cap B_2} r_i = h(B_1 \cap B_2)$.[4] ◀

For any $\mathcal{P} \in \Pi(\mathcal{F}, A)$, we can construct $\lambda \in \Lambda(\mathcal{F}, A)$ with $\lambda_B = \frac{\mathbb{1}\{B^c \in \mathcal{P}\}}{|\mathcal{P}|-1}$. Thus, $\Lambda(\mathcal{F}, A) = \emptyset$ implies $\Pi(\mathcal{F}, A) = \emptyset$, in which case both sides of (B.7) are $-\infty$ by convention. Consider the non-trivial case when $\Lambda(\mathcal{F}, A)$ is non-empty. Let λ^* be an optimal solution to the maximization in (B.7). Define its support set as

$$\mathcal{B} := \{B \in \mathcal{F} : \lambda^*_B > 0\} \tag{B.10}$$

and the corresponding partition of A as

$$\mathcal{P}^* := \left\{ \left(\bigcup \{B \in \mathcal{B} : B \not\ni i\} \right)^c : i \in A \right\} \tag{B.11}$$

Subclaim B.1B $\mathcal{P}^*$ *in* (B.11) *belongs to* $\Pi(\mathcal{F}, A)$. $\triangleleft$

[4] It is valid to have $h(\emptyset) < 0$. In that case, by Subclaim B.1A, we have $B_1, B_2 \in \mathcal{F}$ being tight constraints implies either $B_1 \cup B_2 = V$ or $B_1 \cap B_2 \neq \emptyset$. Otherwise, it would lead to the contradiction that $0 = h(\emptyset)$ by the tightness of $B_1 \cap B_2$.

PROOF Define the relation R on A as

$$i \sim_R j \iff i \in C_j \qquad \text{for } i,j \in A$$

where $C_i := (\bigcup\{B \in \mathcal{B} : B \not\ni i\})^c$. By definition (B.11), $\mathcal{P}^* = \{C_i : i \in A\}$. To show that $\mathcal{P}^*$ partitions A, it suffices to show that $\sim_R$ is an equivalence relation on A as follows.

$$\begin{aligned} i \sim_R j &\iff \{B \in \mathcal{B} : B \not\ni i\} \supseteq \{ B \in \mathcal{B} : B \not\ni j\} \\ &\iff \{B \in \mathcal{B} : i \in B\} \subseteq \{ B \in \mathcal{B} : j \in B\} \end{aligned}$$

i.e. any set in $\mathcal{B}$ that contains i also contains j. Using this simplification, it is easy to see that $\sim_R$ satisfies the defining properties of an equivalence relation:

- *Reflexivity:* R is reflexive since $i \in C_i$ trivially for $i \in A$.
- *Transitivity:* Suppose $i \sim_R j$ and $j \sim_R k$ for some $i,j,k \in A$. Then,

  $$\{B \in \mathcal{B} : i \in B\} \subseteq \{ B \in \mathcal{B} : j \in B\} \subseteq \{ B \in \mathcal{B} : k \in B\}$$

  which implies $i \sim_R k$ as desired.
- *Symmetry:* suppose to the contrary that $i \sim_R j$ but $j \not\sim_R i$ for some $i,j \in A$. Then,

  $$\{B \in \mathcal{B} : i \in B\} \subsetneq \{B \in \mathcal{B} : j \in B\}$$

  This implies, by definition (B.10) of $\mathcal{B}$ that

  $$\sum_{B \ni i} \lambda^*_B < \sum_{B \ni j} \lambda^*_B$$

  which is the desired contradiction since both sides equal 1 by the definition of $\Lambda(\mathcal{F}, A)$.

Finally, to argue that $|\mathcal{P}^*| \geq 2$, note that $\mathcal{B} \neq \emptyset$ as $\sum_{B \in \mathcal{F}} \lambda^*_B > 0$. Since any $B \in \mathcal{F} \in \Phi(A)$ satisfies $B \not\supseteq A$, we have $C_i \not\supseteq A$ for all $i \in A$ as desired. ◀

The supermodularity of h implies the following property on every part of $\mathcal{P}^*$.

Subclaim B.1C *For any optimal r^* to the dual problem* (B.8),

$$\sum_{i\in C^c} r_i^* = h(C^c) \qquad \textit{for all } C \in \mathcal{P}^* \tag{B.12}$$

$\lhd$

Proof By the complementary slackness theorem [15, Theorem 5.4], $\sum_{i\in B} r_i^* = h(B)$ for all $B \in \mathcal{B}$. By Subclaim B.1A, for all $i \in A$, we have

$$\sum_{i\in\bigcup\{B\in\mathcal{B}:B\not\ni i\}} r_i^* = h\left(\bigcup\{B\in\mathcal{B}:B\not\ni i\}\right)$$

which gives the desired equality (B.12) under (B.11). ◀

It follows that

$$\sum_{i\in A} r_i^* = \frac{1}{|\mathcal{P}^*|-1}\sum_{C\in\mathcal{P}^*} h(C^c)$$

which completes the proof since the primal/dual optimality criteria [15, Theorem 5.5] implies that $(\mathbb{1}\{B^c \in \mathcal{P}^*\}/(|\mathcal{P}^*|-1) : B \in \mathscr{F})$ is an optimal solution in $\Lambda(\mathcal{F}, A)$. ■

B.2 Matroid partitioned by vertices

In this section, we will give a general identity for matroids that prove the optimality of the single-source network coding approach to secret key agreement in Section 3.1. We first give some preliminaries on matroid theory [50].

Definition B.2 (Matroid) A matroid is characterized by a finite *ground set* S and a *rank function* $r : 2^S \mapsto \mathbb{N}$ with

$$T \subseteq U \implies r(T) \leq r(U) \leq |U| \tag{B.13a}$$

$$r(T) + r(U) \geq r(T\cup U) + r(T\cap U) \tag{B.13b}$$

for all $T, U \subseteq S$. The *conditional rank* of $T \subseteq S$ given $U \subseteq S$ is defined as

$$r(T|U) := r(T\cup U) - r(U)$$

and the *span* of $T \subseteq S$ with respect to r is defined as

$$\langle T\rangle := \{e \in S : r(\{e\}|T) = 0\}$$

$I \subseteq S$ is called *independent* if $|I| = r(I)$. It is called a *base* if in addition that $|I| = r(I) = r(S)$. We will use $\mathcal{I}$ and $\mathcal{X}$ to denote the sets of independent sets and bases respectively. □

It can be shown that every independent set is a subset of a base. i.e. $\mathcal{I} = \{I \subseteq X : X \in \mathcal{X}\}$. Indeed, there are various alternative ways of describing and understanding what a matroid is, using properties of its span function, independent sets or bases. The following is an equivalent characterization of matroids in terms of their span functions. Other characterizations can be found in [50].

Proposition B.1 *$\langle\cdot\rangle : 2^S \mapsto 2^S$ is the* span function *of a matroid iff for all $T, U \subseteq S$ and $e, f \in S \setminus T$, we have*

$$T \subseteq U \implies T \subseteq \langle T\rangle \subseteq \langle U\rangle \tag{B.14a}$$

$$e \in \langle T \cup \{f\}\rangle \setminus \langle T\rangle \implies f \in \langle T \cup \{e\}\rangle \setminus \langle T\rangle \tag{B.14b}$$

A convenient necessary condition is

$$e \in \langle T\rangle \implies \langle T\rangle = \langle T \cup \{e\}\rangle \tag{B.15}$$

which is a simple consequence of (B.14). □

PROOF While we need only the necessity part for the subsequent results, we also include the proof of the sufficiency part for completeness.

Necessity: $r(T|T) = 0$ implies $T \subseteq \langle T\rangle$. Suppose $f \in \langle T\rangle$ and $e \in S$. By (B.13b),

$$\begin{aligned} r(T \cup \{e, f\}) &\leq r(T \cup \{e\}) + r(T \cup \{f\}) - r(T) \\ &= r(T \cup \{e\}) \end{aligned}$$

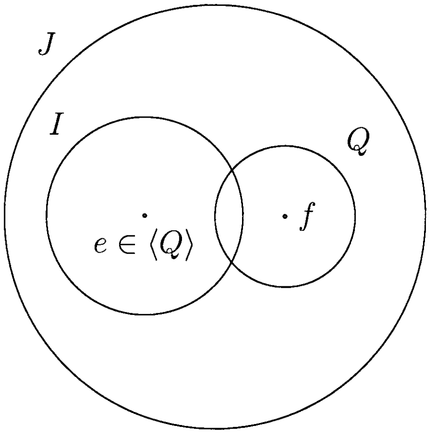


Figure B-1: Venn diagram for the proof Proposition B.1

The reverse inequality holds by (B.13a) and so $f \in \langle T \cup \{e\} \rangle$. It follows that $\langle T \rangle \subseteq \langle T \cup \{e\} \rangle$, which establishes (B.14a).

Assume the premise of (B.14b), i.e.

$$r(T) \overset{\text{(a)}}{<} r(T \cup \{e\}) \overset{\text{(b)}}{\leq} r(T \cup \{e, f\}) \overset{\text{(c)}}{=} r(T \cup \{f\})$$

where (a) is by the premise that $e \notin \langle T \rangle$; (b) is by (B.13a); and (c) is by the premise that $e \in \langle T \cup \{f\} \rangle$. The inequality $r(T) < r(T \cup \{f\})$ implies $f \notin \langle T \rangle$. By (B.13), $r(T \cup \{e\}) \leq r(T) + r(\{e\}) \leq r(T) + 1$, which are satisfied with equalities under (a). (b) is also satisfied with equality due to (c), implying $f \in \langle T \cup \{e\} \rangle$ as desired.

To prove (B.15), suppose to the contrary that there exists $e \in \langle T \rangle$ and $f \in \langle T \cup \{e\} \rangle \setminus \langle T \rangle$. This contradicts (B.14b).[5]

Sufficiency: Given the span function, define the rank as

$$r(T) := \min\{|I| : I \subseteq T, \langle I \rangle = \langle T \rangle\} \qquad \text{with } T \subseteq S$$

It follows immediately that $r(U) \leq |U|$. For $T \subseteq U \subseteq S$, we have $\langle I \rangle = \langle U \rangle$ imply $\langle I \rangle \supseteq \langle T \rangle$ by (B.14a). Thus $r(T) \leq r(U)$, which gives (B.13a).

To prove (B.13b), consider for $T, U \subseteq S$, a minimal subset I of $T \cap U$ with $\langle I \rangle = \langle T \cap U \rangle$, a minimal set J with $I \subseteq J \subseteq T \cup U$ and $\langle J \rangle = \langle T \cup U \rangle$. It follows

[5] This contradicts the reverse implication of (B.14b) by symmetry.

that $r(I) = |I|$ because otherwise $e \in \langle I \setminus \{e\} \rangle \cap I$ implies that $\langle I \setminus \{e\} \rangle = \langle I \rangle$ under (B.15), contradicting the minimality of I.

We now argue that $r(J) = |J|$ as well. Suppose to the contrary that there exists $e \in \langle J \setminus \{e\} \rangle \cap J$. Then, $e \in I$ by the minimality of J. Let Q be a minimal subset of $J \setminus \{e\}$ with $e \in \langle Q \rangle$. Then, there exists $f \in Q \setminus I$ by the minimality of I because otherwise $Q \subseteq I \setminus \{e\}$ implies the contradiction $e \in \langle I \setminus \{e\} \rangle \cap I$ by (B.14a) and the fact that $e \in \langle Q \rangle \cap I$. This is illustrated in Figure B-1. By (B.14b) and the minimality of Q that $e \in \langle Q \rangle \setminus \langle Q \setminus \{f\} \rangle$, we have $f \in \langle Q \cup \{e\} \setminus \{f\} \rangle$. This implies $f \in \langle J \setminus \{f\} \rangle \cap J$ by (B.14a) and the fact $Q \cup \{e\} \subseteq J$. This contradicts the minimality of J as desired.

Indeed, given $r(J) = |J|$, any subset J' of J also satisfies $r(J') = |J'|$ because otherwise $e \in \langle J' \setminus \{e\} \rangle \cap J'$ implies the contradiction $e \in \langle J \setminus \{e\} \rangle \cap J$ as argued before. We can now obtain (B.13b) as follows,

$$\begin{aligned} r(T) + r(U) &\overset{\text{(a)}}{\geq} r(J \cap T) + r(J \cap U) \\ &\overset{\text{(b)}}{=} |J \cap T| + |J \cap U| \\ &\overset{\text{(c)}}{=} |J \cap (T \cup U)| + |J \cap (T \cap U)| \\ &\overset{\text{(d)}}{\geq} |J| + |I| \\ &\overset{\text{(e)}}{\geq} r(T \cup U) + r(T \cap U) \end{aligned}$$

where (a) is by (B.13a); (b) is because $r(J') = |J'|$ for all $J' \subseteq J$; (c) is by modularity of the cardinality function; (d) is because $J \subseteq T \cup U$ and $I \subseteq J \cap (T \cap U)$; and (e) is by the definitions of I and J that $\langle I \rangle$ and $\langle J \rangle$ equals $\langle T \cap U \rangle$ and $\langle T \cup U \rangle$ respectively.

n.b. the fact that independent sets are subsets of bases follows easily from the above arguments. Let I be an independent set, which satisfies $|I| = r(I)$. With $T = I$ and $U = S$, J defined above satisfies $J \supseteq I$ and $|J| = r(J) = r(S)$. Thus, I is a subset of a base, namely J. Subsets of bases are independent since $J' \subseteq J$ satisfies $|J'| = r(J')$. ■

In the following, we will introduce a framework to view matroids as a generalization of edges in graphs.

Definition B.3 (Vertex-partitioned matroid) A *matroid* partitioned by a *vertex set* V is denoted by the pair (Z_V, r) of finite set $Z_V := \bigcup\{Z_i : i \in V\}$ and rank $r : Z_V \mapsto \mathbb{N}$ satisfying (B.13). $\mathcal{X}$ is the set of *bases* $X_V := \bigcup\{X_i \subseteq Z_i : i \in V\}$ with

$$|X_V| = r(X_V) = r(Z_V) \qquad \text{and} \qquad X_i \cap X_j = \emptyset \qquad \text{for all } i \neq j \in V$$

i.e. X_V has *disjoint* X_i's and maximum rank $r(Z_V)$. □

The dependence structure of the matroid relates the nodes in V like edges in a graph, but in a more general way. We think of a base of the matroid as an orientation of a graph in a way that we can define graph-theoretic notion such as directed paths or flows. To do so, we first consider the following simple property of the span function.

Proposition B.2 *For any matroid (S, r), $T \cup U$ being independent implies that*

$$\langle T \rangle \cap \langle U \rangle \subseteq \langle T \cap U \rangle$$

The reverse inclusion holds more generally for arbitrary subsets T, U of S. □

PROOF Suppose to the contrary that there exists $e \in \langle T \rangle \cap \langle U \rangle \setminus \langle T \cap U \rangle$ with $r(T \cup U) = |T \cup U|$. Then,

$$r(T \cup \{e\}) \overset{(a)}{=} r(T) \overset{(b)}{\geq} r(T \cup U) - r(U \setminus T) \overset{(c)}{\geq} |T|$$

where (a) is by $e \in \langle T \rangle$; (b) is by (B.13b) and $r(\emptyset) = 0$; and (c) is by $r(U \setminus T) \leq |U \setminus T|$ in (B.13a). It follows that $r(T \cup \{e\}) = |T|$. Similarly, we have $r(U \cup \{e\}) = |U|$ and $r(T \cup U \cup \{e\}) = |T \cup U|$. Now, $e \notin \langle T \cap U \rangle$ implies that

$$\begin{aligned} r((T \cap U) \cup \{e\}) > |T \cap U| &= |U| + |T| - |T \cup U| \\ &= r(U \cup \{e\}) + r(T \cup \{e\}) - r(T \cup U \cup \{e\}) \end{aligned}$$

which contradicts (B.13b). ■

Definition B.4 (Directed vertex-partitioned matroid) Given a matroid $M = (Z_V, r)$ partitioned by V and a base $X_V \in \mathcal{X}$ in Definition B.3, we say (M, X_V) is a *directed (vertex-partitioned) matroid*.[6] Define the support function

$$\operatorname{supp}(e, X_V) := \min\{I \subseteq X_V : e \in \langle I \rangle\} \qquad \text{with } e \in Z_V, X_V \in \mathcal{X}$$

as the inclusionwise minimum subset of X_V that spans e. It is well-defined by Proposition B.2 that each element is in the span of a unique minimal subset of every base.[7]

In a directed matroid (M, X_V), incut $\delta^- : 2^S \mapsto \mathbb{N}$ and outcut δ^+ are defined as

$$\delta^-(C) = \delta^+(C^c) := \bigcup_{f \in Z_C} \operatorname{supp}(f, X_V) \cap X_{C^c} \tag{B.16}$$

for $C \subseteq V$. The value of a cut is simply its cardinality

$$\begin{aligned} |\delta^-(C)| = |\delta^+(C^c)| = r(Z_C|X_C) &= r(Z_C \cup X_C) - r(X_C) \\ &= r(Z_C) - |X_C| \end{aligned}$$

which is a submodular function in $C \subseteq V$.

A (unit) *antiflow* is a sequence

$$u_1, (e_1, f_2), u_2, (e_2, f_3), \dots, u_{l+1} \tag{B.17}$$

where $l \in \mathbb{N}$ is the length, u_i's are distinct nodes from V, and (e_i, f_{i+1})'s are directed edges in $X_{u_i} \times Z_{u_{i+1}}$ that satisfy

$$e_i \in \operatorname{supp}(f_{i+1}, X_V) \qquad \text{for all } i \in [l] \tag{B.18a}$$

$$e_j \notin \operatorname{supp}(f_{i+1}, X_V) \qquad \text{for all } j \in [i-1] \tag{B.18b}$$

or equivalently $f_{i+1} \in \langle X_V \setminus \{e_j : j \in [i-1]\} \rangle \setminus \langle X_V \setminus \{e_i\} \rangle$ for all $i \in [l]$.

[6] This is not the same as the oriented matroid described in [50].

[7] Suppose to the contrary that there exists two distinct minimal sets $T, U \subseteq X_V$ such that $e \in \langle T \rangle \cap \langle U \rangle$. Then, $e \in \langle T \cap U \rangle$ by Proposition B.2 since $T \cup U$ is independent, contradicting the minimality of T and U.

A (unit) *flow* is a sequence in (B.17) defined like an antiflow except that (B.18) is replaced by

$$e_i \in \operatorname{supp}(f_{i+1}, X_V) \qquad \text{for all } i \in [l] \tag{B.19a}$$

$$e_j \notin \operatorname{supp}(f_{i+1}, X_V) \qquad \text{for all } j > i \tag{B.19b}$$

or equivalently $f_{i+1} \in \langle X_V \setminus \{e_j : j > i\} \rangle \setminus \langle X_V \setminus \{e_i\} \rangle$ for all $i \in [l]$. Antiflows and flows are collectively called (directed) paths. □

Given a graph $G = (V, E, \theta)$ with vertex set V, edge set E and edge function $\theta : E \mapsto \binom{V}{2}$, we can define the corresponding vertex-partitioned matroid (Z_V, r) by setting $e \in Z_i$ for $e \in E$ and $i \in \theta(e)$, and setting $r(T) = |T|$ for all $T \subseteq E$. A base X_V corresponds to the choice of a root node $\rho(e) \in V$ for every edge $e \in E$ under the mapping $e \in X_i \iff \rho(e) = i$. A directed matroid (Z_V, r, X_V) therefore corresponds to a digraph (V, E, ϕ, ρ). Flows or antiflows in the directed matroid correspond to directed paths in the digraph.

Directed matroid captures more general notion of digraphs such as the star hypergraphs in [3]. In its full generality, a flow can be different from an antiflow. For example, consider the linear matroid (Z_V, r) partitioned by three nodes as follows,

$Z_1 := \{e_1\}$	$Z_2 := \{e_1, f_2\}$	$Z_3 := \{f_3\}$
$e_1 := (0,1)$	$e_2 := (1,0)$	
	$f_2 := (0,1)$	$f_3 := (1,1)$

where e_1 and e_2 are binary vectors, and r is the corresponding rank function in $\mathbb{F}_2^2$. Choosing $X_1 := \{e_1\}, X_2 := \{e_2\}, X_3 := \emptyset$ as the base X_V, the sequence $1, (e_1, f_2), 2, (e_2, f_3), 3$ is a flow since $f_2 = e_1$ and $f_3 = e_1 + e_2$. i.e. $e_1 \in \operatorname{supp}(f_2, X_V)$ and $e_2 \in \operatorname{supp}(f_3, X_V) \setminus \operatorname{supp}(f_2, X_V)$, satisfying (B.19). However, this is not an antiflow since $e_1 \in \operatorname{supp}(f_3, X_V)$ violates (B.18b).

Despite their difference, antiflows and flows are closely related. Indeed, it is easy to see that the reverse sequence $3, (f_3, e_2), 2, (f_2, e_1), 1$ is an antiflow in the directed matroid with $X_1 := \emptyset, X_2 := \{f_2\}, X_3 := \{f_3\}$ chosen as the base. More generally, antiflows and flows are related by the following reversal operation.

Definition B.5 Given a directed path (flow or antiflow) $u_1, (e_1, f_2), \dots, u_{l+1}$ in a directed matroid (M, X_V), we say that $\dot{X}_V$ is obtained by reversing the path if

$$\dot{X}_{u_i} = \begin{cases} X_{u_i} \setminus \{e_i\} & i = 1 \\ X_{u_i} \cup \{f_i\} \setminus \{e_i\} & i = 2, \dots, l \\ X_{u_i} \cup \{f_i\} & i = l+1 \end{cases} \tag{B.20}$$

$u_{l+1}, (f_{l+1}, e_l), \dots, u_1$ is the reverse of the path. □

Proposition B.3 *The reverse of an antiflow in a directed matroid* (M, X_V) *is a flow in* $(M, \dot{X}_V)$ *with* $\dot{X}_V$ *obtained by reversing the antiflow according to* (B.20)*. Thus,*

$$r(Z_C|\dot{X}_C) - r(Z_C|X_C) = \mathbb{1}_C(u_1) - \mathbb{1}_C(u_{l+1}) \tag{B.21}$$

for all $C \subseteq V$*. Similarly, the reverse of a flow gives an antiflow.* □

Proof Consider an antiflow denoted as (B.17). For $k \in [l]$, let X_V^k be obtained by reversing the antiflow from u_k to u_{l+1} as in (B.20). X_V^{k+1} and X_V^k differ by the elements e_k and f_{k+1} as shown below.

	$\dots$	$Z_{u_{k-1}}$	Z_{u_k}	$Z_{u_{k+1}}$	$Z_{u_{k+2}}$	$\dots$
$X_V = X_V^{l+1}$	$\dots$	e_{k-1}	e_k	e_{k+1}	e_{k+2}	$\dots$
$\vdots$					$\iddots$	
X_V^{k+1}	$\dots$	e_{k-1}	e_k	□	f_{k+2}	$\dots$
X_V^k	$\dots$	e_{k-1}	□	f_{k+1}	f_{k+2}	$\dots$
$\vdots$		$\iddots$				
$\dot{X}_V = X_V^1$	$\dots$	f_{k-1}	f_k	f_{k+1}	f_{k+2}	$\dots$

It will be helpful to refer to this for the subsequent arguments.

Let $S(k)$ be the statement that

$$X_V^k \in \mathcal{X} \tag{B.22a}$$

$$e_k \in \langle X_V^k \setminus \{e_j : j \in [k-1]\}\rangle \tag{B.22b}$$

$$e_k \notin \langle X_V^1 \setminus \{f_{k+1}\}\rangle \tag{B.22c}$$

It suffices to prove that $S(k)$ is true for $k \in [l]$ by induction because (B.22c) with $k = 1$ implies $\dot{X}_V \in \mathcal{X}$, while (B.22b) and (B.22c) imply (B.18b) and (B.18a) respectively for the reverse path.

Assume as an inductive hypothesis that $S(k+1), \dots, S(l)$ are true. By (B.18b), $f_j \in \langle X_V \setminus \{e_k\}\rangle$ for all $j \geq k+2$. By (B.15),

$$\begin{aligned}\langle X_V \setminus \{e_k\}\rangle &= \langle X_V \cup \{f_j : j \geq k+2\} \setminus \{e_k\}\rangle \\ &\overset{\text{(a)}}{\supseteq} \langle X_V^{k+1} \setminus \{e_k\}\rangle \\ &\overset{\text{(b)}}{=} \langle X_V^k \setminus \{f_{k+1}\}\rangle\end{aligned}$$

where (a) is by (B.14a) and (b) is by the fact that

$$X_V^{k+1} \setminus \{e_k\} = X_V^k \setminus \{f_{k+1}\}$$

By (B.18a), $f_{k+1} \notin \langle X_V \setminus \{e_k\}\rangle$. With (a) and (b) above,

$$\begin{aligned}f_{k+1} \notin \langle X_V \setminus \{e_k\}\rangle &\implies f_{k+1} \notin \langle X_V^k \setminus \{f_{k+1}\}\rangle \\ &\implies X_V^k \in \mathcal{X}\end{aligned}$$

by the hypothesis that $X_V^{k+1} \in \mathcal{X}$, which gives (B.22a).

Consider proving (B.22b). By the inductive hypothesis,

$$e_{k'} \in \langle \underbrace{X_V^{k'} \setminus \{e_j : j \in [k'-1]\}}_{\overset{\text{(c)}}{\subseteq} X_V^{k+1} \setminus \{e_j : j \in [k]\}} \rangle \qquad \text{for all } k' > k \tag{B.23}$$

where (c) follows directly from the definition of X_V^{k+1}. Thus, by (B.15),[8]

$$\langle X_V \setminus \{e_j : j \in [k-1]\}\rangle \subseteq \langle X_V^{k+1} \setminus \{e_j : j \in [k-1]\}\rangle$$

Since f_{k+1} is contained by the L.H.S., it is also contained by the R.H.S.. i.e.

$$f_{k+1} \in \langle X_V^{k+1} \setminus \{e_j : j \in [k-1]\}\rangle \tag{B.24}$$

We also have

$$f_{k+1} \notin \langle X_V^{k+1} \setminus \{e_j : j \in [k]\}\rangle \tag{B.25}$$

because otherwise $f_{k+1} \in \langle X_V^k \setminus \{f_{k+1}\}\rangle$, contradicting (B.22a) that $X_V^k \in \mathcal{X}$ argued previously. By (B.14b), we have (B.24) and (B.25) imply (B.22b).

Finally, suppose to the contrary that (B.22c) does not hold. i.e.

$$e_k \in \langle X_V^1 \setminus \{f_{k+1}\}\rangle \tag{B.26}$$

By Proposition B.2,

$$e_k \in \langle X_V^{k+1} \setminus \{e_j : j \in [k]\}\rangle$$

because the argument of the span function above is the intersection of those in (B.26) and (B.22b) proved earlier. This contradicts the hypothesis that $X_V^{k+1} \in \mathcal{X}$ as desired.

To complete the induction, the base case with $k = l$ can be proved by repeating the above arguments with the hypothesis $X_V^{k+1} \in \mathcal{X}$ replaced by $X_V^{l+1} := X_V \in \mathcal{X}$.

To prove (B.21), note that $X_V, \dot{X}_V \in \mathcal{X}$ implies

$$\begin{aligned} r(Z_C|\dot{X}_C) - r(Z_C|X_C) &= |X_C| - |\dot{X}_C| \\ &= \sum_{i \in C} |X_i| - |\dot{X}_i| \end{aligned}$$

[8] $e_{k'}$ for $k' > k$ present in X_V but not X_V^{k+1} is indeed in the span of the R.H.S. by (B.23).

The desired result follows then from the fact that

$$|X_i| - |\dot{X}_i| = \begin{cases} 1 & i = u_1 \\ -1 & i = u_{l+1} \\ 0 & \text{otherwise} \end{cases}$$

which is an immediate consequence of (B.20). ■

Antiflows originating from a node can be constructed as follows.[9]

Proposition B.4 *Given a matroid $M := (Z_V, r)$, a base $X_V \in \mathcal{X}$ and $t \in V$, construct $T \subseteq V$ by adding a sequence $v_1, v_2, \dots$ of distinct nodes with $v_1 = t$, and v_i for $i > 1$ chosen as any node in $V \setminus \{v_j : j < i\}$ that satisfies*

$$r(Z_{v_i} | X_{V \setminus \{v_j : j < i\}}) > 0 \tag{B.27}$$

Then, for all $v \in T \setminus \{t\}$, there is an antiflow from t to v in the directed matroid (M, X_V). We say that T is the set of nodes reachable from t by antiflows. □

PROOF Let $\rho(v_i) = i$ be the order that node v_i is added to T. We first construct a sequence $u_1, \dots, u_{l+1}$ starting from $u_{l+1} = v \in T \setminus \{t\}$ backwards until $u_1 = t$ as follows: for $i \geq 1$, define u_i as the node v_k with the smallest $k < \rho(u_{i+1})$ such that

$$r(Z_{u_{i+1}} | X_{V \setminus \{v_j : j \leq k\}}) > 0 \tag{B.28a}$$

Such k exists by (B.27). The minimality of k implies that

$$r(Z_{u_{i+1}} | X_{V \setminus \{v_j : j < k\}}) = 0 \tag{B.28b}$$

It follows that $\rho(u_i)$ decreases strictly as i decreases, and so we must eventually have $u_1 = t$ for an appropriate choice of $l \in \mathbb{P}$. Furthermore, it follows from (B.28b) that

[9] By Proposition B.3, flows ending in a node can also be constructed similarly.

any choice of $e_i \in X_{u_i}, f_{i+1} \in Z_{u_{i+1}}$ for $i \in [l-1]$ must satisfy

$$r(f_{i+1}|X_V \setminus \{e_j : j < i\}) = 0 \tag{B.29}$$

because $\{e_j : j < i\} \subseteq X_{\{v_j : j < \rho(u_i)\}}$. We now argue that we can choose e_i's and f_{i+1}'s in such a way that

$$r(f_{i+1}|X_V \setminus \{e_i\}) > 0 \qquad \text{for all } i \in [l-1] \tag{B.30}$$

This will complete the proof since $u_1, (e_1, f_2), \ldots, u_{l+1}$ is the desired antiflow from t to v as (B.29) trivially implies (B.18b) while (B.30) implies (B.18a). Suppose to the contrary that there exists $i \in [l-1]$ such that $f \in \langle X_v \setminus \{e\}\rangle$ for all $f \in Z_{u_{i+1}}$ and $e \in X_{u_i}$. Then,

$$\begin{aligned} Z_{u_{i+1}} &\subseteq \bigcap_{e \in X_{u_i}} \langle X_v \setminus \{e\}\rangle \\ &\overset{\text{(a)}}{\subseteq} \left\langle \bigcap \{X_V \setminus \{e\} : e \in X_{u_i}\} \right\rangle = \langle X_{V \setminus \{u_i\}}\rangle \end{aligned}$$

where (a) is by Proposition B.2. This contradicts (B.28a) as $u_i \in \{v_j : j \leq \rho(u_i)\}$. ■

We can now prove the following identity for matroids that establishes the optimality of the single-source network coding approach in Theorem 3.1.

Theorem B.2 *For $s \in A \subseteq V : |A| \geq 2$, matroid $M = (Z_V, r)$ partitioned by V in Definition B.3, and base $X_V \in \mathcal{X}$ of M, define*

$$d_M(A, s, X_V) := \min_{B \subseteq V : s \in B \not\supseteq A} r(Z_{B^c}|X_{B^c}) \tag{B.31}$$

$$p_M(A, X_V) := \min_{\mathcal{P} \in \Pi(A)} \frac{1}{|\mathcal{P}| - 1} \sum_{C \in \mathcal{P}} r(Z_C|X_C) \tag{B.32}$$

where $\Pi(A)$ is defined in (B.4a). *Then, we have*

$$\max_{X_V \in \mathcal{X}} d_M(A, s, X_V) = \left\lfloor \max_{X_V \in \mathcal{X}} p_M(A, X_V) \right\rfloor \tag{B.33}$$

independent of $s \in A$. When $A = V$, $p_M(V, X_V)$ becomes

$$\min_{\mathcal{P} \in \Pi(V)} \frac{1}{|\mathcal{P}| - 1} \left[\left(\sum_{C \in \mathcal{P}} r(Z_C) \right) - r(Z_V) \right] \tag{B.34}$$

independent of $X_V \in \mathcal{X}$. i.e. the maximization on the R.H.S. of (B.33) *is trivial when $A = V$.* □

PROOF (B.34) can be obtained by rewriting the summation

$$\begin{aligned} \sum_{C \in \mathcal{P}} r(Z_C | X_C) &= \sum_{C \in \mathcal{P}} \left(r(Z_C) - \sum_{i \in C} |X_i| \right) \\ &= \sum_{C \in \mathcal{P}} r(Z_C) - \sum_{i \in V} \sum_{C \ni i} |X_i| \end{aligned}$$

The last term equals $|X_V| = r(Z_V)$ when $A = V$ because there is a unique part $C \in \mathcal{P}$ that contains each element $i \in V$.

To prove (B.33), let $\mathcal{R}(A, X_V)$ be the set of bases obtained by reversing one or more antiflows in the directed matroid (M, X_V) in Definition B.4 between distinct nodes in A.[10] We will argue that

$$d_M(A, s, X_V) \leq p_M(A, X_V) \tag{B.35a}$$

$$\forall \dot{X}_V \in \mathcal{R}(A, X_V),\ p_M(A, \dot{X}_V) = p_M(A, X_V) \tag{B.35b}$$

$$\exists \dot{X}_V \in \mathcal{R}(A, X_V),\ d_M(A, s, \dot{X}_V) = \lfloor p_M(A, X_V) \rfloor \tag{B.35c}$$

(B.33) follows immediately from the last equality.

Let $d := d_M(A, s, X_V)$ for convenience. Then, (B.35a) follows from the fact that $r(Z_C | X_C) \geq d$ for every $C \in \mathcal{P} \in \Pi(A)$ such that $s \notin C$, and there are $|\mathcal{P}| - 1$ such distinct C for every $\mathcal{P}$.

To prove (B.35b), it suffices to show it for an arbitrary $\dot{X}_V$ obtained by reversing an antiflow from t to v for arbitrary nodes $t, v \in A$. Consider summing both sides of (B.21) over $C \in \mathcal{P}$ with $u_1 = t$ and $u_{l+1} = v$. The sum on the R.H.S. equals 0

[10] i.e. every antiflow being reversed must begin and end at distinct nodes in A, while the intermediate nodes can be outside A.

regardless of whether t and v are contained in the same part or not.[11] Thus, the sum on the L.H.S. is also 0, which gives (B.35b) as desired.

We now prove (B.35c) by generalizing the proof in [3, Theorem 5.1]. By (B.35a) and (B.35b), it suffices to show that

$$d_M(A, s, \dot{X}_V) \geq \lfloor p_M(A, X_V) \rfloor \tag{B.36}$$

for some $\dot{X}_V \in \mathcal{R}(A, X_V)$ with the additional constraint that there is no antiflow from any $t \in A\setminus\{s\}$ to s in $(M, \dot{X}_V)$. The additional constraint is admissible because if the optimal $\dot{X}_V$ has an antiflow from t to s, we can reverse the antiflow without diminishing $d_M(A, s, \dot{X}_V)$. Reversing such antiflow strictly increases $r(\dot{X}_s) \leq r(Z_s)$ and so doing so repeatedly eventually gives the desired $\dot{X}_V$ without any such antiflow.

Define the operation of adding a new element e to M over $\{s,t\}$ for some $t \in A\setminus\{s\}$ as follows

$$\begin{aligned} Z_i &\leftarrow Z_i \cup \{e\} && \text{for } i \in \{s,t\} \\ r(Z \cup \{e\}) &\leftarrow r(Z) + 1 && \text{for } Z \subseteq Z_V \end{aligned}$$

It is easy to see that by adding new elements repeatedly this way for any choices of t, the L.H.S. of (B.36) is bound to increase. Thus, to prove (B.36) by contradiction, suppose that at least one edge $\dot{e}$ needs to be added for some $t \in A \setminus \{s\}$ to obtain a matroid $\dot{M} := (\dot{Z}_V, r)$ with minimum $|\dot{Z}_V|$ such that

$$d_{\dot{M}}(A, s, \dot{X}_V) \geq \lfloor p_M(A, X_V) \rfloor \tag{B.37}$$

for a base $\dot{\mathsf{X}}_V \in \dot{\mathcal{X}}$ of $\dot{M}$ such that $\dot{\mathsf{X}}_V \cap Z_V \in \mathcal{R}(A, X_V)$ and there is no antiflow from t to s in $(\dot{M}, \dot{\mathsf{X}}_V)$. To come up with the desired contradiction, we will construct $\mathcal{P} \in \Pi(A)$ such that $\exists C \in \mathcal{P}, t \in C \not\ni s$ and

$$\frac{1}{|\mathcal{P}| - 1} \sum_{C \in \mathcal{P}} r(\dot{Z}_C | \dot{X}_C) = \lfloor p_M(A, X_V) \rfloor \tag{B.38}$$

[11] By the definition of $\mathcal{P} \in \Pi(A)$ in (B.4a), t and v are each contained in exactly one part of $\mathcal{P}$. Thus, they contribute to a $+1$ and -1 to the sum, which cancels out.

Since $\dot{e}$ contributes to the L.H.S., we have

$$\frac{1}{|\mathcal{P}|-1}\sum_{C\in\mathcal{P}} r(Z_C|\dot{X}_C \cap Z_V) < \lfloor p_M(A, X_V)\rfloor \tag{B.39}$$

which contradicts the equality (B.35b) by the assumption that $\dot{\mathsf{X}}_V \cap Z_V \in \mathcal{R}(A, X_V)$.

We now construct the desired $\mathcal{P}$. For convenience, define

$$p := \lfloor p_M(A, X_V)\rfloor \qquad \text{and} \qquad \mathscr{C}(A) := \{W \subseteq V : s \in W^c \not\supseteq A\}$$

Call any set $C \in \mathscr{C}(A)$ tight if $r(\dot{Z}_C|\dot{X}_C) = p$. Let $T \subseteq V$ be the set of nodes reachable from t in $(\dot{M}, \dot{X}_V)$ as defined in Proposition B.4. Define $\mathcal{P} := (C_1, \dots, C_k)$ where $(C_1, \dots, C_{k-1})$ is the collection of distinct maximal tight sets that overlap both T and A, and $C_k := T^c$. Thus,

$$r(\dot{Z}_{C_i}|\dot{X}_{C_i}) = \begin{cases} p & ,i \in [k-1] \\ 0 & ,i = k \end{cases} \tag{B.40}$$

This gives (B.38) as desired. It remains to show that $\mathcal{P} \in \Pi(A)$.

First, we argue that $C_i^c \in \mathscr{F}(A)$ for all $i \in [k]$. This is true for $i \in [k-1]$ because $C_i \cap T \cap A \neq \emptyset$ by definition. $C_k^c = T \in \mathscr{F}(A)$ follows from the fact that $s \not\in T$ because there is no antiflow from t to s by assumption. i.e. $C_k \cap A \supseteq \{s\}$.

Next, we argue that every node in A is contained by at most one part in $\mathcal{P}$. By the submodularity of r, we have for all $i, j \in [k]$ that

$$r(\dot{Z}_{C_i}|\dot{X}_{C_i}) + r(\dot{Z}_{C_j}|\dot{X}_{C_j}) \geq r(\dot{Z}_{C_i\cap C_j}|\dot{X}_{C_i\cap C_j}) + r(\dot{Z}_{C_i\cup C_j}|\dot{X}_{C_i\cup C_j}) \tag{B.41}$$

Suppose $A \cap C_i \cap C_j \neq \emptyset$ for some $i \neq j \in [k-1]$. Then, $C_i \cap C_j, C_i \cup C_j \in \mathscr{C}(A)$ and so the R.H.S. of (B.41) is at least $2p$ by (B.37). Since the L.H.S. equals $2p$ by (B.40), $C_i \cup C_j$ is a tight set that overlaps with T and A, which contradicts the maximality of C_i and C_j. Suppose $A \cap C_i \cap C_j \neq \emptyset$ for some $i \in [k-1]$ and $j = k$. Then, $C_i \cap C_j \in \mathscr{C}(A)$ and so the first term on the R.H.S. of (B.41) is at least p by (B.37).

Since the L.H.S. of (B.41) is p by (B.40), equality must hold for (B.41), and the first and second terms on the R.H.S. must equal p and 0 respectively. In particular, the second term equal to 0 implies that $t \in C_i$. Otherwise, it contradicts the fact that any $v \in A \cap C_i \setminus C_j$ is reachable by an antiflow from t.[12] With $t \in C_i \setminus C_j$ and $s \in C_j \setminus C_i$, the inequality (B.41) must be strict because removing $\dot{e}$ reduces the L.H.S. but not the R.H.S., while the inequality must remain to hold after this removal by the submodularity of r. This contradicts the earlier conclusion that (B.41) is satisfied with equality.

It remains to show that every node in A is contained by at least one part in $\mathcal{P}$. In particular, since $C_k = T^c$, we need only prove that every node in $T \cap A$ is contained in a tight set. Suppose to the contrary that there exists $v \in T \cap A$ with

$$r(\dot{Z}_C|\dot{X}_C) \geq p+1 \qquad \text{for all } C \subseteq V : v \in C, s \notin C \tag{B.42}$$

Let $\ddot{X}_V$ be a base obtained by reversing an antiflow from t to v. By (B.21) in Proposition B.3,

$$r(\dot{Z}_C|\ddot{X}_C) = \begin{cases} r(\dot{Z}_C|\dot{X}_C)+1 & ,t \in C \not\ni v \\ r(\dot{Z}_C|\dot{X}_C)-1 & ,v \in C \not\ni t \\ r(\dot{Z}_C|\dot{X}_C) & \text{otherwise} \end{cases}$$

It follows from (B.42) and (B.37) that

$$r(\dot{Z}_{B^c}|\ddot{X}_{B^c}) \geq \begin{cases} p+1 & ,t \notin B \\ p & \text{otherwise} \end{cases}$$

for all $B \subseteq V : s \in B \not\supseteq A$. However, this contradicts the minimality of $\dot{Z}_V$ since $\ddot{X}_{B^c} \cap Z_V \in \mathcal{R}(A, X_V)$ and $\dot{e}$ can be removed without violating (B.37) with $\ddot{X}_V$ as the base. This completes the proof. ■

[12] More precisely, suppose $t \notin C_i$, and that there is an antiflow from $u_1 := t$ to some $u_{l+1} := v \in A \cap C_i \setminus C_j = A \cap C_i \cap T$, which is non-empty by the definition of C_i for $i \in [k-1]$. By (B.21), the second term on the R.H.S. of (B.41) is positive by (B.27).

Corollary B.1 *Given $s \in A \subseteq V : |A| \geq 2$ and a matroid $M = (Z_V, r)$, let $\hat{\mathsf{X}}_V$ be a base randomly chosen from $\mathcal{X}$ according to the distribution $P_{\hat{\mathsf{X}}_V} \in \mathscr{P}(\mathcal{X})$. As a corollary to Theorem B.2, we have*

$$\max_{P_{\hat{\mathsf{X}}_V} \in \mathscr{P}(\mathcal{X})} \min_{B \subseteq V : s \in B \not\supseteq A} \mathrm{E}\left[r(Z_{B^c} | \hat{\mathsf{X}}_{B^c}) \right] \tag{B.43a}$$

$$= \max_{P_{\hat{\mathsf{X}}_V}} \min_{\mathcal{P} \in \Pi(A)} \frac{1}{|\mathcal{P}| - 1} \sum_{C \in \mathcal{P}} \mathrm{E}[r(Z_C | \hat{\mathsf{X}}_C)] \tag{B.43b}$$

As a corollary to Theorem B.1, the above can be expressed as

$$= \max_{P_{\hat{\mathsf{X}}_V}} \min_{\lambda \in \Lambda(\mathscr{F}(A), A)} \sum_{B \in \mathscr{F}(A)} \lambda_B \, \mathrm{E}[r(Z_{B^c} | \hat{\mathsf{X}}_{B^c})] \tag{B.43c}$$

$$= \min_{\lambda \in \Lambda(\mathscr{F}(A), A)} \max_{X_V \in \mathcal{X}} \sum_{B \in \mathscr{F}(A)} \lambda_B r(Z_{B^c} | X_{B^c}) \tag{B.43d}$$

$$\leq \min_{\lambda \in \Lambda(\mathscr{F}(A), V)} \sum_{B \in \mathscr{F}(A)} \lambda_B r(Z_{B^c} | X_{B^c}) \tag{B.43e}$$

where the last expression is independent of $X_V \in \mathcal{X}$.[13] □

Proof Consider the n-extension $M^n := (Z_V^n, r)$ defined as the union of n replica $(Z_V^{(i)}, r)$ for $i \in [n]$ of the matroid (Z_V, r). It follows from Theorem B.2 that

$$\frac{1}{n} \max_{X_V^n \in \mathcal{X}^n} d_{M^n}(A, s, X_V^n) = \frac{1}{n} \left\lfloor \max_{X_V^n \in \mathcal{X}^n} p_{M^n}(A, X_V^n) \right\rfloor$$

It is straightforward to show that the L.H.S. and R.H.S. equal (B.43a) and (B.43b) respectively in the limit as $n \to \infty$. (B.43c) follows from Theorem B.1 with $\mathcal{F} = \mathscr{F}(A) \in \Phi(A)$ because $\mathrm{E}[r(Z_C | \hat{\mathsf{X}}_C)]$ is submodular in $C \subseteq V$. (B.43d) follows from the minimax theorem [56]. (B.43e) follows from the fact that $\Lambda(\mathscr{F}(A), A) \supseteq \Lambda(\mathscr{F}(A), V)$

[13] It can be shown easily that (B.43e) is the secrecy capacity in [12] for the finite linear source in Definition 3.1. It is unclear if the bound is tight except for the case $A = V$ by Theorem B.1.

and that (B.43e) is independent of X_V shown below.

$$\begin{aligned}\sum_{B\in\mathscr{F}(A)} \lambda_B r(Z_{B^c}|X_{B^c}) = &\sum_{B\in\mathscr{F}(A)} \lambda_B \left[r(Z_{B^c}) - \sum_{i\in B^c} |X_i| \right] \\ =&\sum_{B\in\mathscr{F}(A)} \lambda_B r(Z_{B^c}) - \sum_{i\in V}\sum_{B\not\ni i} \lambda_B |X_i|\end{aligned}$$

For $\lambda \in \Lambda(\mathscr{F}(A), V)$, $\sum_{B\ni i} \lambda_B = 1$ for all $i \in B$. Thus, the last term is

$$\sum_{i\in V}\sum_{B\not\ni i} \lambda_B |X_i| = \left(\sum_B \lambda_B - 1\right) \sum_{i\in V} |X_i|$$

which equals $(\sum_B \lambda_B - 1)r(Z_V)$, independent of X_V. ■

B.3 Dependency hypergraph

While the notions of flows and cuts for directed matroids generalizes the corresponding notions for star hypergraphs in Definition 3.6, hypergraphs have more structure that leads to stronger results. For example, the following theorem extends Bang-Jensen and Thomassé's generalization of Menger's theorem from star hypergraphs to hypergraphs.

Theorem B.3 *Any hypergraph $H = (V, E, \phi)$ can be shrunk to a graph $G = (V, E, \theta)$such that $\theta(e) \subseteq \phi(e) : |\theta(e)| = 2$ for all $e \in E$ and*

$$\min_{B\subseteq V: s\in B\not\ni t} |\delta_H(B)| = \min_{B\subseteq V: s\in B\not\ni t} |\delta_G(B)|$$

Applying Menger's theorem to G, the above min-cut value is the maximum number of edge-disjoint unit flows in H as defined in Definition 3.6. □

Proof Consider the greedy approach of constructing G by sequentially removing nodes $v \in \phi(e)$ from edges $e \in E$ with $|\phi(e)| \geq 3$. Suppose to the contrary that this cannot be done without diminishing the min-cut value. Then, there exists an edge e

containing distinct vertices $v_1, v_2, v_3 \in V$ such that removing any v_i from e reduces the min-cut value.

Consider the first case that there exists $B_1, B_2 \subseteq V : s \in B_i \not\ni t$ that attains the min-cut value, say k, with

$$\phi(e) \setminus \{v_1\} \subseteq B_1 \not\ni v_1 \qquad \text{and} \qquad \phi(e) \setminus \{v_2\} \subseteq B_2 \not\ni v_2 \tag{B.44}$$

The assumption that v_i's are distinct implies that $v_1 \in B_2 \setminus B_1$, $v_2 \in B_1 \setminus B_2$ and $v_3 \in B_1 \cap B_2$. By submodularity of δ_H, we have

$$|\delta_H(B_1)| + |\delta_H(B_1)| \geq| \delta_H(B_1 \cap B_2)| + |\delta_H(B_1 \cup B_2)| \tag{B.45}$$

The L.H.S. is $2k$ by the optimality of B_i's, and the R.H.S. is at least $2k$ since $B_1 \cap B_2$, $B_1 \cup B_2$ both contain s but not t. Thus, (B.45) should be tight, i.e. satisfied with equality. However, by (B.44),

$$e \in \delta_H(B_1) \cap \delta_H(B_2) \cap \delta_H(B_1 \cap B_2) \setminus \delta_H(B_1 \cup B_2) \tag{B.46}$$

and so the (B.45) should be strict since the inequality must hold even with e removed from H but doing so reduces the L.H.S. of (B.45) more than the R.H.S. by (B.46). This contradicts the earlier conclusion that (B.45) should be tight. The same argument applies to the cases with v_i's permuted in (B.44) by symmetry.

Consider the other case with

$$\phi(e) \setminus \{v_1\} \subseteq B_1^c \not\ni v_1 \qquad \text{and} \qquad \phi(e) \setminus \{v_2\} \subseteq B_2^c \not\ni v_2 \tag{B.47}$$

instead of (B.44). The assumption that v_i's are distinct implies that $v_1 \in B_1 \setminus B_2$, $v_2 \in B_2 \setminus B_1$ and $v_3 \in (B_1 \cup B_2)^c$. Similar to the previous argument, (B.45) should be tight by the optimality of B_i's, while it should be strict because (B.47) implies that

$$e \in \delta_H(B_1) \cap \delta_H(B_2) \cap \delta_H(B_1 \cup B_2) \setminus \delta_H(B_1 \cap B_2)$$

This is the desired contradiction. The same argument applies to the remaining cases with v_i's permuted in (B.47), which completes the proof by contradiction. ■

Since the above proof mainly uses the submodularity of cut values, the result can be generalized further as follow.

Proposition B.5 *For any star hypergraph $H^* = (V, E, \phi, \rho)$, co-intersecting family $\mathcal{F} \in \Phi(V)$ (see Definition B.1) and submodular function $f : \mathcal{F} \mapsto \mathbb{R}$, we can shrink H^* greedily while preserving $\min_{B \in \mathcal{F}} \left[|\delta^-_{H^*}(B)| + f(B) \right]$ in the sense that for any $\dot{e} \in E : |\phi(\dot{e})| \geq 3$, there exists $\dot{v} \in \phi(\dot{e}) \setminus \{\rho(\dot{e})\}$ such that*

$$\min_{B \in \mathcal{F}} \left[|\delta^-_{\dot{H}^*}(B)| + f(B) \right] = \min_{B \in \mathcal{F}} \left[|\delta^-_{H^*}(B)| + f(B) \right] \qquad \textbf{(B.48)}$$

where the shrunk star hypergraph $\dot{H}^* = (V, E, \dot{\phi}, \rho)$ has $\dot{\phi} = \phi$ except for $\dot{e}$ where $\dot{\phi}(\dot{e}) = \phi(\dot{e}) \setminus \{\dot{v}\}$. □

Corollary B.2 *Given star hypergraph $H^* = H_1^* \sqcup H_2^*$ as defined in Definition 3.4, we can shrink H_2^* to $G^* = (V_2, E_2, \theta, \rho_2)$ such that*

$$\min_{B \in \mathscr{F}(A)} \left[|\delta^+_{H_1^*}(B)| + |\delta^-_{H_2^*}(B)| \right] = \min_{B \in \mathscr{F}(A)} \left[|\delta^+_{H_1^*}(B)| + |\delta^-_{G^*}(B)| \right] \qquad \textbf{(B.49)}$$

and θ satisfies $\rho_2(e) \in \theta(e) \subseteq \phi_2(e)$ and $|\theta(e)| = 2$. $\mathscr{F}(A)$ is defined in (B.1a). □

PROOF The corollary follows from an inductive argument using the proposition with $\mathcal{F} = \mathscr{F}(A)$ and $f(B) = |\delta^+_{H_1^*}(B)|$.

To prove the proposition, suppose to the contrary that H^* cannot be shrunk as stated. This means that there exists $\dot{e} \in E : |\dot{e}| \geq 3$ that cannot be shrunk. i.e. for any $v \in \phi(\dot{e}) \setminus \{\rho(\dot{e})\}$, there is $B_v \in \mathcal{F}$ with $\rho(\dot{e}) \in B_v^c, B_v \cap \phi(\dot{e}) = \{v\}$ and

$$|\delta^-_{H^*}(B_v)| + f(B_v) = \min_{B \in \mathcal{F}} \left[|\delta^-_{H^*}(B)| + f(B) \right] \qquad \textbf{(B.50)}$$

Consider distinct $v_1, v_2 \in \phi(\dot{e}) \setminus \{\rho(\dot{e})\}$. We have

$$\sum_{i \in \{1,2\}} |\delta^-_{H^*}(B_{v_i})| > |\delta^-_{H^*}(B_{v_1} \cap B_{v_2})| + |\delta^-_{H^*}(B_{v_1} \cup B_{v_2})|$$

by the submodularity of $\delta^-_{H^*}$. The strict inequality follows from the additional fact $\dot{e}$ contributes 1 to each term on the L.H.S. but only 1 to the last term on the R.H.S.. This is because $B_{v_i} \cap \phi(\dot{e}) = \{v_i\}$ for $i = 1, 2$ implies that $v_1, v_2 \notin B_{v_1} \cap B_{v_2}$. By the submodularity of f, we have

$$\sum_{i \in \{1,2\}} \left[|\delta^-_{H^*}(B_{v_i})| + f(B_{v_i})\right] > |\delta^-_{H^*}(B_{v_1} \cap B_{v_2})| + f(B_{v_1} \cap B_{v_2})$$
$$+ |\delta^-_{H^*}(B_{v_1} \cup B_{v_2})| + f(B_{v_1} \cup B_{v_2})$$

The fact that $\rho(\dot{e}) \notin B_{v_1} \cup B_{v_2}$ implies $B_{v_1} \cap B_{v_2}, B_{v_1} \cup B_{v_2} \in \mathcal{F}$ by the definition in (B.2). Thus, the R.H.S. is at least $2 \min_{B \in \mathcal{F}} \left[|\delta^-_{H^*}(B)| + f(B)\right]$, which equals the L.H.S.. This contradicts (B.50) as desired. ■

The secrecy capacity under the source model in Definition 3.4 with dependence structure captured by a dependency hypergraph gives a concrete operational meaning to the following notion of partition connectivity for hypergraphs from [3].

Definition B.6 (Partition connectivity) Given a dependency hypergraph $H = H_1 \sqcup H_2$, define

$$p_H := \min_{\mathcal{P} \in \Pi} \left[p^-_{H_1}(\mathcal{P}) + p^+_{H_2}(\mathcal{P})\right] \tag{B.51}$$

where

$$p^-_{H_1}(\mathcal{P}) := \frac{|\{C \cap \phi(e) : C \in \mathcal{P}, e \in E_1\} \setminus \{\emptyset\}| - |E_1|}{|\mathcal{P}| - 1} \tag{B.52a}$$

$$= \frac{\sum_{C \in \mathcal{P}} |\delta^-_{H_1^*}(C)|}{|\mathcal{P}| - 1} \tag{B.52b}$$

$$p^+_{H_2}(\mathcal{P}) := \frac{|\{e \in E_2 : \forall C \in \mathcal{P}, C \not\supseteq \phi(e)\}|}{|\mathcal{P}| - 1} \tag{B.52c}$$

$$= \frac{\sum_{C \in \mathcal{P}} |\delta^+_{H_2^*}(C)|}{|\mathcal{P}| - 1} \tag{B.52d}$$

H_1^* and H_2^* are arbitrary star hypergraphs of H_1 and H_2 respectively. □

Equalities (B.52b) and (B.52d) follows easily from the double counting principle,

$$\begin{aligned}\sum_{C\in\mathcal{P}}|\delta^-_{H_1^*}| &= \sum_{e\in E_1}\sum_{C\not\ni\rho_1(e)} \mathbb{1}\{C^c \not\supseteq \phi_1(e)\} \\ &= \sum_{e\in E_1}[|\{C\cap\phi(e) : C\in\mathcal{P}\}\setminus\{\emptyset\}|-1] \\ \sum_{C\in\mathcal{P}}|\delta^+_{H_2^*}| &= \sum_{e\in E_2}\sum_{C\ni\rho_2(e)} \mathbb{1}\{C \not\supseteq \phi_2(e)\}\end{aligned}$$

The minimizing partition $\mathcal{P}$ has the intuitive meaning of groups of highly connected nodes by the following Proposition.

Proposition B.6 *If there exists $e \in E$ such that removing e does not change p_H in* (B.51)*, then*

$$e \notin \delta_H(\mathcal{P}) := \{e \in H : \forall C \in \mathcal{P}, C \not\supseteq \phi(e)\}$$

for all minimizing $\mathcal{P} \in \Pi$ that attains p_H in (B.51). □

Proof Suppose to the contrary that there exists a minimizing $\mathcal{P}$ such that $e \in \delta_H(\mathcal{P})$. Then, there exists $\dot{C} \in \mathcal{P}$ such that $\dot{C} \not\supseteq \phi(e) \not\subseteq \dot{C}^c$. Let $H^* = (V, E, \phi, \rho)$ be a star hypergraph of H with $\rho(e) \in \dot{C}$ if e is an edge in H_2 while $\rho(e) \in \dot{C}^c$ otherwise. This implies that

$$e \in \delta^-_{H_1^*}(\dot{C}) \cup \delta^+_{H_2^*}(\dot{C}) \tag{B.53}$$

Let $\dot{H} = (V, E\setminus\{e\}, \phi)$ be the hypergraph H with e removed. Then,

$$\begin{aligned} p_{\dot{H}} \overset{\text{(a)}}{=} p_H &\overset{\text{(b)}}{=} \frac{1}{|\mathcal{P}|-1}\sum_{C\in\mathcal{P}}\left[|\delta^-_{H_1^*}(C)| + |\delta^+_{H_2^*}(C)|\right] \\ &\overset{\text{(c)}}{>} \frac{1}{|\mathcal{P}|-1}\sum_{C\in\mathcal{P}}\left[|\delta^-_{\dot{H}_1^*}(C)| + |\delta^+_{\dot{H}_2^*}(C)|\right] \overset{\text{(d)}}{\geq} p_{\dot{H}}\end{aligned}$$

where (a) is by assumption, (b) is by (B.51) and the optimality of $\mathcal{P}$, (c) is by (B.53) and the definition that e is absent in $\dot{H}$, and (d) is by (B.51). This completes the proof by contradiction. ■

B.4 Fundamentals of information

The framework described above captures the behavior of information through the submodularity of entropy. It was first pointed out by Fujishige in [21], and can be regarded as one of the many structural properties of joint distributions summarized in [38]. Yeung [64] developed a software that can verify any information inequalities derived from the submodularity of entropy, and collectively refer to them as Shannon-type inequalities. Zhang and Yeung [65] later discovered a non-Shannon-type inequality, which proves [66] that there are additional structure to the entropy function for a set of four or more random variables. Using this, Dougherty, Freiling and Zeger [18] showed that Shannon-type inequality is insufficient to compute network coding capacities. Although more and more non-Shannon-type inequalities have been discovered in [33, 39], a complete characterization of the structure of the entropy remains open.

Appendix C

Computations

C.1 Coupling channel

In this section, we will give the detailed computation for the key rates in Table 6.1 for the coupling channel defined in Section 6.3.

C.1.1 Preliminaries

We first carry out some preliminary calculations. With $V = [3]$, $A = [2]$ and $D = \emptyset$, the hypergraph $\mathcal{H}_{A|D}$ in (A.1) is

$$\mathcal{H}_{[2]|\emptyset} = \{\{1\}, \{2\}, \{3\}, \{1,3\}, \{2,3\}\}$$

$\Lambda_{[2]|\emptyset}$ in (A.9) is the convex hull of the following basic fractional partitions.[1]

$$\begin{array}{r|ccccc}
 & \lambda^{(k)}_{\{1\}} & \lambda^{(k)}_{\{2\}} & \lambda^{(k)}_{\{3\}} & \lambda^{(k)}_{\{2,3\}} & \lambda^{(k)}_{\{1,3\}} \\
\hline
\lambda^{(1)} := & (1, & 0, & 0, & 1, & 0) \\
\lambda^{(2)} := & (0, & 1, & 0, & 0, & 1) \\
\lambda^{(3)} := & (1, & 1, & 1, & 0, & 0)
\end{array} \tag{C.1}$$

[1] This is the same as the fractional partitions illustrated in Figure A-1 since the two hypergraphs have the same set of edges.

Consider an arbitrary joint distribution for the channel input

$$P_{\mathsf{X}_1\mathsf{X}_2} = \begin{bmatrix} P_{\mathsf{X}_1\mathsf{X}_2}(0,0) & P_{\mathsf{X}_1\mathsf{X}_2}(0,1) \\ P_{\mathsf{X}_1\mathsf{X}_2}(1,0) & P_{\mathsf{X}_1\mathsf{X}_2}(1,1) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{C.2}$$

where (a, b, c, d) is stochastic, i.e.

$$a, b, c, d \in [0, 1] : a + b + c + d = 1$$

From the definition (6.30) of the coupling channel, we can compute the entropies $H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c})$ for $B \in \mathcal{H}_{A|D} \cup \{\emptyset\}$. The results are summarized in Table C.1. We will show the computation for the case $B = \{1\}$ as an example. With h defined as the binary entropy function in (6.8),

$$\begin{aligned} H(\mathsf{Y}_2\mathsf{Y}_3|\mathsf{X}_2) &= H(\mathsf{N}_3|\mathsf{X}_2) + H(\mathsf{Y}_2|\mathsf{X}_2, \mathsf{N}_3) \\ &= \overbrace{H(\mathsf{N}_3)}^{=1} + \overbrace{H(\mathsf{Y}_2|\mathsf{X}_2 = 0, \mathsf{N}_3)}^{=H(\mathsf{N}_2)=1}(a + c) + [H(\mathsf{Y}_2|\mathsf{X}_2 = 1, \mathsf{N}_3) + H(\mathsf{Y}_2|\mathsf{X}_2 = 1, \mathsf{N}_3 = 1)]\, \tfrac{b+d}{2} \\ &= 1 + a + c + \big[h(\overbrace{P_{\mathsf{X}_1\mathsf{N}_2|\mathsf{X}_2\mathsf{N}_3}(0, 1|1, 0)}^{=\frac{b}{2(b+d)}}) + h(\underbrace{P_{\mathsf{X}_1\mathsf{N}_2|\mathsf{X}_2\mathsf{N}_3}(0, 0|1, 1)}_{=\frac{b}{2(b+d)}})\big]\, \tfrac{b+d}{2} \end{aligned}$$

which is the desired expression for $B = \{1\}$ in Table C.1.

Using the result in Table C.1, we evaluate α for each of the basic fractional partitions in (C.1) as follows.

$$\begin{aligned} \alpha(\lambda^{(1)}, P_{\mathsf{X}_V}) &= H(\mathsf{Y}_{23}|\mathsf{X}_2) + H(\mathsf{Y}_1|\mathsf{X}_1) - H(\mathsf{Y}_{123}|\mathsf{X}_{12}) \\ &= a - b + (b + d)h\left(\frac{b}{2(b + d)}\right) \end{aligned} \tag{C.3a}$$

$$\begin{aligned} \alpha(\lambda^{(2)}, P_{\mathsf{X}_V}) &= H(\mathsf{Y}_{13}|\mathsf{X}_1) + H(\mathsf{Y}_2|\mathsf{X}_2) - H(\mathsf{Y}_{123}|\mathsf{X}_{12}) \\ &= d - b + (a + b)h\left(\frac{b}{2(a + b)}\right) \end{aligned} \tag{C.3b}$$

Table C.1: Entropy terms for the coupling channel with correlated inputs (C.2)

B	$H(\mathsf{Y}_{B^c}\vert\mathsf{X}_{B^c})$
$\emptyset$	$H(\mathsf{Y}_{123}\vert\mathsf{X}_{12})^{\dagger} = 2+b+c$
$\{1\}$	$H(\mathsf{Y}_{23}\vert\mathsf{X}_{2}) = 1+a+c+(b+d)h\left(\frac{b}{2(b+d)}\right)$
$\{2\}$	$H(\mathsf{Y}_{13}\vert\mathsf{X}_{1}) = 1+c+d+(a+b)h\left(\frac{b}{2(a+b)}\right)$
$\{3\}$	$H(\mathsf{Y}_{12}\vert\mathsf{X}_{12}) = 2$
$\{1,3\}$	$H(\mathsf{Y}_{2}\vert\mathsf{X}_{2}) = 1$
$\{2,3\}$	$H(\mathsf{Y}_{1}\vert\mathsf{X}_{1}) = 1$

† Y_{123} is short for $(\mathsf{Y}_1,\mathsf{Y}_2,\mathsf{Y}_3)$ and similarly for others.

$$\begin{aligned}\alpha(\lambda^{(3)},P_{\mathsf{X}_V}) &= H(\mathsf{Y}_{23}|\mathsf{X}_2)+H(\mathsf{Y}_{13}|\mathsf{X}_1)+H(\mathsf{Y}_{12}|\mathsf{X}_{12})-2H(\mathsf{Y}_{123}|\mathsf{X}_{12})\\ &= a+d-2b+(b+d)h\left(\frac{b}{2(b+d)}\right)+(a+b)h\left(\frac{b}{2(a+b)}\right) \qquad \textbf{(C.3c)}\end{aligned}$$

From these, we can observe the equality that

$$\alpha(\lambda^{(3)},P_{\mathsf{X}_V}) = \alpha(\lambda^{(1)},P_{\mathsf{X}_V})+\alpha(\lambda^{(2)},P_{\mathsf{X}_V}) \qquad \textbf{(C.4)}$$

Consider the independence constraint on the input distribution that for some $p_1,p_2\in[0,1]$,

$$P_{\mathsf{X}_1} = \mathrm{Bern}_{p_1} \qquad \textbf{(C.5a)}$$

$$P_{\mathsf{X}_2} = \mathrm{Bern}_{p_2} \qquad \textbf{(C.5b)}$$

$$P_{\mathsf{X}_1\mathsf{X}_2} = P_{\mathsf{X}_1}P_{\mathsf{X}_2} \qquad \textbf{(C.5c)}$$

With the independence constraint, the entropy terms in Table C.1 become those in Table C.2, obtained by the following substitution

$$\begin{bmatrix} a & b\\ c & d\end{bmatrix} = \begin{bmatrix}(1-p_1)(1-p_2) & (1-p_1)p_2\\ p_1(1-p_2) & p_1p_2\end{bmatrix} \qquad \textbf{(C.6)}$$

Table C.2: Entropy terms for the coupling channel with independent input (C.5)

B	$H(\mathsf{Y}_{B^c}\vert\mathsf{X}_{B^c})$	
$\emptyset$	$H(\mathsf{Y}_{123}\vert\mathsf{X}_{12})$	$= 2 + p_1(1-p_2) + p_2(1-p_1)$
$\{1\}$	$H(\mathsf{Y}_{23}\vert\mathsf{X}_{2})$	$= 2 - p_2 + p_2 h\left(\frac{1-p_1}{2}\right)$
$\{2\}$	$H(\mathsf{Y}_{13}\vert\mathsf{X}_{1})$	$= 1 + p_1 + (1-p_1)h\left(\frac{p_2}{2}\right)$
$\{3\}$	$H(\mathsf{Y}_{12}\vert\mathsf{X}_{12})$	$= 2$
$\{1,3\}$	$H(\mathsf{Y}_{2}\vert\mathsf{X}_{2})$	$= 1$
$\{2,3\}$	$H(\mathsf{Y}_{1}\vert\mathsf{X}_{1})$	$= 1$

With independent input, α is equal to $\tilde{\beta}$ by the equivalence relation (a) of Proposition A.4. Thus, we have from (C.3) and (C.4) that

$$\tilde{\beta}(\lambda^{(1)}, P_{\mathsf{X}_V}) = (1-p_1)(1-2p_2) + p_2 h\left(\frac{1-p_1}{2}\right) \tag{C.7a}$$

$$\tilde{\beta}(\lambda^{(2)}, P_{\mathsf{X}_V}) = p_2(2p_1-1) + (1-p_1)h\left(\frac{p_2}{2}\right) \tag{C.7b}$$

$$\tilde{\beta}(\lambda^{(3)}, P_{\mathsf{X}_V}) = \tilde{\beta}(\lambda^{(1)}, P_{\mathsf{X}_V}) + \tilde{\beta}(\lambda^{(2)}, P_{\mathsf{X}_V}) \tag{C.7c}$$

C.1.2 Optimal pure source emulation

In the pure source emulation approach, T_1 and T_2 transmit independent inputs iid over time, i.e.

$$P_{\mathsf{X}_1^n\mathsf{X}_2^n}(x_1^n, x_2^n) = \prod_{t\in[n]} P_{\mathsf{X}_1}(x_{1t})P_{\mathsf{X}_2}(x_{2t})$$

Then, each terminal $i \in [3]$ broadcasts a public message at rate r_i such that T_1 and T_2 can recover the entire channel input and output sequences $(\mathsf{X}_1^n, \mathsf{X}_2^n, \mathsf{Y}_1^n, \mathsf{Y}_2^n, \mathsf{Y}_3^n)$. By minimizing the sum rate $\sum_{i\in[3]} r_i$, we maximizes the asymptotic rate of the extractible

Table C.3: Optimal public message rates for the coupling channel

terminal (i)	optimal public message rate† (r_i)
$1,2$	$p^2+h(p)+(1-p)\left[2-p-h\left(\frac{1-p}{2}\right)\right]$ ≈ 1.578
3	$p+(1-p)h\left(\frac{1-p}{2}\right)\approx 0.919$

† $p \approx 0.44$ is the optimal solution to (e) in (C.8).

key independent of the public messages [12]. The maximum key rate is

$$\begin{aligned}
C_{\text{pse}} &\overset{(a)}{=} \max_{P_{\mathsf{X}_1\mathsf{X}_2}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \min_{\lambda\in\Lambda_{[2]|\emptyset}} \tilde{\beta}(\lambda, P_{\mathsf{X}_1\mathsf{X}_2}) \\
&\overset{(b)}{=} \max_{P_{\mathsf{X}_1\mathsf{X}_2}=P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \min_{k\in[2]} \tilde{\beta}(\lambda^{(l)}, P_{\mathsf{X}_1\mathsf{X}_2}) \\
&\overset{(c)}{=} \max_{p_1,p_2\in[0,1]} \min\{g(p_1,p_2), g(1-p_2,1-p_1)\} \\
&\overset{(d)}{=} \max_{p\in[0,1]} \overbrace{(1-p)(p-1)+(1-p)h\left(\frac{1-p}{2}\right)}^{g(p,1-p)} \\
&\overset{(e)}{\approx} 0.41
\end{aligned} \tag{C.8}$$

where

$$g(p_1,p_2) := (1-p_1)(1-2p_2)+p_2 h\left(\frac{1-p_1}{2}\right)$$

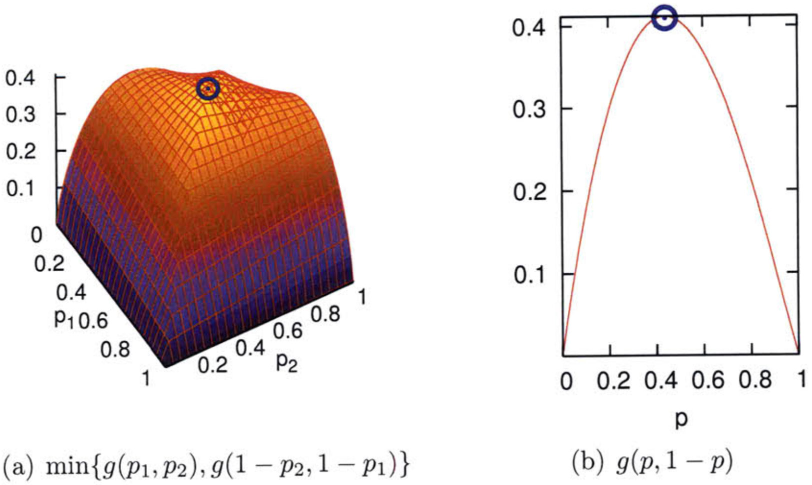
(a) The equality follows from (6.18c) with Q deterministic.

(b) Since $\min_\lambda \tilde{\beta}(\lambda, P_{\mathsf{X}_1\mathsf{X}_2})$ is a linear program, the optimal value is achieved at some basic fractional partition in (C.1). We can exclude the basic fractional partition $\lambda^{(3)}$ since it cannot achieve a smaller value than $\lambda^{(1)}$ (or $\lambda^{(2)}$) by (C.7c) and the positivity of $\tilde{\beta}$ by (A.10a) of the Shearer-type lemma.

(c) This is by (C.7a) and (C.7b) under (C.5).

(d) The maximum is achieved at $p_1 = 1-p_2$ as shown in Figure C-1(a).

(e) The maximum is achieved at $p\approx 0.44$ as shown in Figure C-1(b).



(a) $\min\{g(p_1,p_2), g(1-p_2,1-p_1)\}$

(b) $g(p,1-p)$

Figure C-1: Optimal input distribution for the pure source emulation scheme: $P_{\mathsf{X}_1\mathsf{X}_2}(x_1,x_2) = \operatorname{Bern}_{p_1}(x_1)\operatorname{Bern}_{p_2}(x_2)$ where $p_1 = 1-p_2 = p \approx 0.44$.

Although we do not know the optimal choice of the key and public discussion functions, we can compute the optimal choice of the public message rates $\boldsymbol{r} := (r_i : i \in [3])$, which are given in Table C.3. We will explain briefly how the optimal rates can be obtained from the optimal fractional partitions. By the strong duality theorem [15] for linear programming,

$$\min_{\lambda} \tilde{\beta}(\lambda, P_{\mathsf{X}_V}) = H(\mathsf{X}_{[2]}\mathsf{Y}_{[3]}) - \min_{\boldsymbol{r}} \sum_{i\in[3]} r_i \tag{C.9}$$

where for all $B \in \mathcal{H}_{[2]|\emptyset}$,

$$\sum_{i\in B} r_i \geq H(\mathsf{X}_B\mathsf{Y}_B|\mathsf{X}_{B^c}\mathsf{Y}_{B^c}) = H(\mathsf{Y}_{[3]}|\mathsf{X}_{[2]}) - H(\mathsf{Y}_{B^c}|\mathsf{X}_{B^c}) + H(\mathsf{X}_B) \tag{C.10}$$

By the equality (d) in (C.8), $\lambda^{(1)}$ and $\lambda^{(2)}$ are both optimal solutions to the L.H.S. of (C.9). Applying the complementary slackness theorem [15], we have for any optimal solution $\boldsymbol{r}$ that $\lambda^{(1)}_B > 0$ or $\lambda^{(2)}_B > 0$ implies equality in (C.10) for the particular $B \in \mathcal{H}_{[2]|\emptyset}$. This gives a set of equations, from which we can solve for the optimal rates as given in Table C.3.

C.1.3 Optimal mixed source emulation

The computation for the mixed source emulation approach proceeds in the same way as the pure source emulation approach described in the previous section except that T_1 and T_2 transmit conditionally independent inputs for a chosen public auxiliary component source Q, i.e.

$$P_{\mathsf{Q}^n,\mathsf{X}_1^n,\mathsf{X}_2^n}(q^n,x_1^n,x_2^n)=\prod_{t\in[n]}P_{\mathsf{Q}}(q_t)P_{\mathsf{X}_1|\mathsf{Q}}(x_{1t}|q_t)P_{\mathsf{X}_2|\mathsf{Q}}(x_{2t}|q_t)$$

The maximum key rate is

$$\begin{aligned}C_{\text{mse}}&\overset{\text{(a)}}{=}\max_{\substack{P_{\mathsf{Q},\mathsf{X}_1,\mathsf{X}_2}\\=P_{\mathsf{Q}}P_{\mathsf{X}_1|\mathsf{Q}}P_{\mathsf{X}_2|\mathsf{Q}}}}\min_{\lambda\in\Lambda_{[2]|\emptyset}}E\left[\tilde{\beta}(\lambda,P_{\mathsf{X}_1,\mathsf{X}_2|\mathsf{Q}}(\cdot|\mathsf{Q}))\right]\\&\overset{\text{(b)}}{=}\max_{P_{\mathsf{Q},\mathsf{X}_1,\mathsf{X}_2}}\min_{k\in[2]}\tilde{\beta}(\lambda^{(l)},P_{\mathsf{Q},\mathsf{X}_1,\mathsf{X}_2})\\&\overset{\text{(c)}}{=}\max_{\substack{p_0,p_{iq}\in[0,1]:\\i\in[2],q\in\{0,1\}}}\min\{g'(p_0,p_{10},p_{20},p_{11},p_{21}),\\&\qquad\qquad g'(p_0,1-p_{20},1-p_{10},1-p_{21},1-p_{22})\}\\&\overset{\text{(d)}}{=}\frac{1}{2}(\log 17-3)\approx 0.54\end{aligned}\tag{C.11}$$

where

$$g'(p_0,1-p_{20},1-p_{10},1-p_{21},1-p_{22}):=(1-p_0)g(p_{10},p_{20})+p_0g(p_{11},p_{21})$$

(a) The equality follows from (6.18c).

(b) same reason as (b) of (C.8).

(c) This is by (C.7a) and (C.7b), averaged over a binary auxiliary component source with the following conditional input distributions.

$$\begin{aligned}P_{\mathsf{Q}}&:=\text{Bern}_{p_0}\\P_{\mathsf{X}_i|\mathsf{Q}}(\cdot|q)&:=\text{Bern}_{p_i}\qquad\text{for } i\in[2],q\in\{0,1\}\end{aligned}$$

By the Support Lemma A.3, it does not lose optimality to choose Q binary as there are only two choices for the basic fractional partitions in (b).

(d) The maximum is achieved at

$$p_0 = \frac{1}{2}, \qquad (p_{10}, p_{20}) = \left(0, \frac{2}{17}\right) \qquad \text{and} \qquad (p_{11}, p_{21}) = \left(\frac{15}{17}, 1\right)$$

It can be obtained using a global maximization algorithm. We computed this using the shuffled complex evolution implemented by [17].

We can arrive at the same answer using the alternative form of C_{mse} from (6.18b).

$$\begin{aligned}
C_{\text{mse}} &= \min_{\lambda\in\Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_1,\mathsf{X}_2}=P_{P_{\mathsf{X}_1}P_{\mathsf{X}_2}}} \tilde{\beta}(\lambda, P_{\mathsf{X}_1,\mathsf{X}_2}) \\
&\overset{(a)}{=} \min_{P_{\mathsf{L}}} \max_{P_{\mathsf{X}_1,\mathsf{X}_2}} \tilde{\beta}(E(\lambda^{(\mathsf{L})}), P_{\mathsf{X}_1,\mathsf{X}_2}) \qquad\qquad \text{with } \mathsf{L}\in[3] \\
&\overset{(b)}{=} \min_{P_{\mathsf{L}}} \max_{P_{\mathsf{X}_1,\mathsf{X}_2}} E\left[\tilde{\beta}(\lambda^{(\mathsf{L})}, P_{\mathsf{X}_1,\mathsf{X}_2})\right] \\
&\overset{(c)}{=} \min_{P_{\mathsf{L}}} \max_{P_{\mathsf{X}_1,\mathsf{X}_2}} \left[(P_{\mathsf{L}}(1)+P_{\mathsf{L}}(3))\tilde{\beta}(\lambda^{(1)}, P_{\mathsf{X}_1,\mathsf{X}_2})\right. \\
&\qquad\qquad \left. + (P_{\mathsf{L}}(2)+P_{\mathsf{L}}(3))\tilde{\beta}(\lambda^{(2)}, P_{\mathsf{X}_1,\mathsf{X}_2})\right] \\
&\overset{(d)}{=} \min_{\theta} \max_{P_{\mathsf{X}_1,\mathsf{X}_2}} \left[(1-\theta)\tilde{\beta}(\lambda^{(1)}, P_{\mathsf{X}_1,\mathsf{X}_2}) + \theta\tilde{\beta}(\lambda^{(2)}, P_{\mathsf{X}_1,\mathsf{X}_2})\right] \\
&= \min_{\theta\in[0,1]} \max_{p_1,p_2\in[0,1]} g''(\theta, p_1, p_2) \\
&\overset{(e)}{=} \frac{1}{2}(\log 17 - 3) \approx 0.54
\end{aligned} \tag{C.12}$$

where

$$g''(\theta, p_1, p_2) := \theta g(p_1, p_2) + (1-\theta) g(1-p_2, 1-p_1)$$

(a) because any $\lambda \in \Lambda_{A|D}$ can be written as a convex combination $E(\lambda^{(\mathsf{L})})$ of the basic fractional partitions in (C.1).

(b) by the linearity of expectation and $\beta(\lambda, \tilde{P}_{\mathsf{X}_1\mathsf{X}_2})$ in λ.

(c) by (C.7c).

(d) Since $\tilde{\beta}$ is non-negative, it is optimal to choose $P_{\mathsf{L}}(3) = 0$.

(e) With the help of a global optimization algorithm, it can be shown that the maximum is achieved by choosing $\theta = \frac{1}{2}$. The corresponding optimal choice of (p_1, p_2) for $\theta = \frac{1}{2}$ is $\left(0, \frac{2}{17}\right)$ or $\left(\frac{15}{17}, 1\right)$.

Unlike the previous optimization, the optimal input distribution for the mixed source emulation is not immediately available from the optimal solutions in the current optimization. This is because the operational meanings of the optimal solutions are changed when we apply the minimax-type lemma to obtain the current optimization from the previous one. While the current optimization involves two less parameters than the previous case, it is a minimax problem rather than a pure maximization problem.

C.1.4 Secrecy upper bound

By (6.1), the secrecy upper bound for the coupling channel is

$$\begin{aligned} C_{\mathrm{su}} &= \min_{\lambda \in \Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} \alpha(\lambda, P_{\mathsf{X}_1\mathsf{X}_2}) \\ &\overset{\text{(a)}}{=} \min_{P_{\mathsf{L}}} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} \alpha(E(\lambda^{(\mathsf{L})}), P_{\mathsf{X}_1\mathsf{X}_2}) \\ &\overset{\text{(b)}}{=} \min_{P_{\mathsf{L}}} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} E\left[\alpha(\lambda^{(\mathsf{L})}, P_{\mathsf{X}_1\mathsf{X}_2})\right] \\ &\overset{\text{(c)}}{=} \min_{P_{\mathsf{L}}} \max_{\substack{a,b,c,d \in [0,1]:\\ a+b+c+d=1}} \left[(P_{\mathsf{L}}(1) + P_{\mathsf{L}}(3)) f(a,b,c,d) + (P_{\mathsf{L}}(2) + P_{\mathsf{L}}(3)) f(d,b,c,a)\right] \\ &\overset{\text{(d)}}{=} \frac{1}{2}\log 7 - 2 \approx 0.60 \end{aligned}$$

where

$$f(a,b,c,d) := a - b + (b+d) h\left(\frac{b}{2(b+d)}\right)$$

(a) same reason as (a) in (C.12).

(b) by the linearity of expectation and $\alpha(\lambda, P_{\mathsf{X}_1\mathsf{X}_2})$ in λ.

(c) by (C.3) and (C.4), setting $P_{\mathsf{X}_1\mathsf{X}_2} = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$.

(d) Using [17], we find that $f(a,b,c,d)$ is maximized at

$$(a,b,c,d) = (\frac{3}{7}, \frac{1}{7}, 0, \frac{3}{7}) \qquad \textbf{(C.13)}$$

under the constraint that (a,b,c,d) is stochastic. Since $a = d$ in this case, we have $f(d,b,c,a) = f(a,b,c,d)$, which is also maximized. Thus, (C.13) is indeed the optimal solution for every choice of P_{L}. The optimal choice of P_{L} must have $P_{\mathsf{L}}(3) = 0$. However, $P_{\mathsf{L}}(1)$ and $P_{\mathsf{L}}(2)$ can be arbitrary since $f(d,b,c,a) = f(a,b,c,d)$ optimally.

Alternatively as before, we can turn the minimax problem into a maximization problem by applying the minimax-type lemma,

$$\begin{aligned} C_{\mathrm{su}} &= \max_{P_{\mathsf{Q},\mathsf{X}_1,\mathsf{X}_2}} \min_{\lambda\in\Lambda_{A|D}} E\left[\alpha(\lambda, P_{\mathsf{X}_1\mathsf{X}_2|\mathsf{Q}}(\cdot|\mathsf{Q}))\right] \\ &= \max_{\substack{p,a_0,b_0,c_0,d_0,a_1,b_1,c_1,d_1: \\ a_q+b_q+c_q+d_q=1,\forall q\in\{0,1\}}} \min\left\{(1-p)f(a_0,b_0,c_0,d_0) + pf(a_1,b_1,c_1,d_1),\right. \\ &\qquad\qquad \left.(1-p)f(d_0,b_0,c_0,a_0) + pf(d_1,b_1,c_1,a_1)\right\} \end{aligned}$$

where we have set

$$P_{\mathsf{Q}} = \mathrm{Bern}_p \qquad \text{and} \qquad P_{\mathsf{X}_1\mathsf{X}_2|\mathsf{Q}}(\cdot|q) = \begin{bmatrix} a_q & b_q \\ c_q & d_q \end{bmatrix} \qquad \text{for all } q \in \{0,1\}$$

Solving this with [17] gives the same upper bound.

C.2 Consensus channel

In this section, we will give the detailed computation for the secret key rates of the consensus channel considered in Section 7.3.

C.2.1 Preliminaries

Let the input distribution be

$$P_{\mathsf{X}_1\mathsf{X}_2} := \begin{bmatrix} P_{\mathsf{X}_1\mathsf{X}_2}(0,0) & P_{\mathsf{X}_1\mathsf{X}_2}(0,1) \\ P_{\mathsf{X}_1\mathsf{X}_2}(1,0) & P_{\mathsf{X}_1\mathsf{X}_2}(1,1) \end{bmatrix} := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{C.14}$$

with the constraint that

$$a,b,c,d \in [0,1] : a+b+c+d = 1$$

For the consensus channel $P_{\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2}$ defined in (7.33), we have

$$\begin{aligned}
H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2) &= H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2\mathbb{1}\{\mathsf{X}_1=\mathsf{X}_2\}) \\
&\overset{\text{(i)}}{=} \Pr\{\mathsf{X}_1=\mathsf{X}_2\}\overbrace{H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2,\mathsf{X}_1=\mathsf{X}_2)}^{=0} \\
&\quad + \Pr\{\mathsf{X}_1\neq\mathsf{X}_2\}\underbrace{H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2,\mathsf{X}_1\neq\mathsf{X}_2)}_{=H(\mathsf{N})=1} \\
&= b+c && \textbf{(C.15a)}
\end{aligned}$$

$$\begin{aligned}
H(\mathsf{Y}|\mathsf{X}_1) &= \sum_{x_1\in\{0,1\}} P_{\mathsf{X}_1}(x_1)H(\mathsf{Y}|\mathsf{X}_1=x_1) \\
&\overset{\text{(ii)}}{=} \underbrace{(a+b)h\left(\tfrac{b}{2(a+b)}\right)+(c+d)h\left(\tfrac{c}{2(c+d)}\right)}_{f(a,b,c,d)} && \textbf{(C.15b)} \\
H(\mathsf{Y}|\mathsf{X}_2) &\overset{\text{(iii)}}{=} f(a,c,b,d) && \textbf{(C.15c)}
\end{aligned}$$

(i) This follows from the definition (7.33) of Y.

(ii) Given $\mathsf{X}_1=0$, we have $\mathsf{Y}=1$ iff $\mathsf{X}_2=\mathsf{N}=1$ by (7.33), which occurs with probability

$$P_{\mathsf{X}_2|\mathsf{X}_1}(1|0)P_{\mathsf{N}}(1)=\frac{b}{2(a+b)}$$

by independence. Thus,

$$H(\mathsf{Y}|\mathsf{X}_1=0)=h(\delta_2/2)$$

Similarly, we have

$$H(\mathsf{Y}|\mathsf{X}_1=1)=h((1-\delta_2)/2)$$

(iii) This is by the symmetry of the consensus channel between the two input symbols. i.e. $P_{\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2}(y|x_1,x_2)=P_{\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2}(y|x_2,x_1)$.

Suppose we have X_1 independent of X_2 instead, with

$$P_{\mathsf{X}_1}=\text{Bern}_{\delta_1} \qquad \text{and} \qquad P_{\mathsf{X}_2}=\text{Bern}_{\delta_2} \qquad \textbf{(C.16)}$$

Then, it follows that

$$H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2) = \delta_1(1-\delta_2) + \delta_2(1-\delta_1) \tag{C.17a}$$

$$H(\mathsf{Y}|\mathsf{X}_1) = \underbrace{(1-\delta_1)h\left(\frac{\delta_2}{2}\right) + \delta_1 h\left(\frac{1-\delta_2}{2}\right)}_{g(\delta_1,\delta_2)} \tag{C.17b}$$

$$H(\mathsf{Y}|\mathsf{X}_2) = g(\delta_2, \delta_1) \tag{C.17c}$$

C.2.2 Computation of secrecy bounds

By Theorem 6.3, we have the following secrecy lower bound by the source emulation approach,

$$\begin{aligned}
C_{\text{se}} &:= \min_{\lambda \in \Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_1\mathsf{X}_2} = P_{\mathsf{X}_1}P_{\mathsf{X}_2}} \tilde{\beta}(\lambda, P_{\mathsf{X}_1\mathsf{X}_2}) \\
&\overset{(a)}{=} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} [H(\mathsf{Y}|\mathsf{X}_1) + H(\mathsf{Y}|\mathsf{X}_2) - H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2)] \\
&\overset{(b)}{=} \max_{\delta_1,\delta_2 \in [0,1]} [g(\delta_1,\delta_2) + g(\delta_2,\delta_1) - \delta_1(1-\delta_2) - \delta_2(1-\delta_1)] \\
&\overset{(c)}{=} \frac{7}{2} - \frac{3}{2}\log 3 \approx 1.12
\end{aligned}$$

where

$$g(\delta_1,\delta_2) := (1-\delta_1)h\left(\frac{\delta_2}{2}\right) + \delta_1 h\left(\frac{1-\delta_2}{2}\right)$$

(a) This is because there is only one possible fractional partition.

(b) Let X_1 and X_2 be independent random variables distributed as in (C.16). Then, (b) follows from (C.17).

(c) The maximum is uniquely achieved at $\delta_1 = \delta_2 = \frac{1}{2}$.

n.b. pure and mixed source emulations achieve the same maximum key rate, primarily because the minimization over the fractional partition is trivial with only one possible fractional partition.

By Theorem 6.1, the secrecy upper bound is

$$\begin{aligned}
C_{\mathrm{su}} &:= \min_{\lambda\in\Lambda_{[2]|\emptyset}} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} \alpha(\lambda, P_{\mathsf{X}_1\mathsf{X}_2}) \\
&\overset{\text{(a)}}{=} \max_{P_{\mathsf{X}_1\mathsf{X}_2}} [H(\mathsf{Y}|\mathsf{X}_1) + H(\mathsf{Y}|\mathsf{X}_2) - H(\mathsf{Y}|\mathsf{X}_1\mathsf{X}_2)] \\
&\overset{\text{(b)}}{=} \max_{\substack{a,b,c,d\in[0,1]:\\ a+b+c+d=1}} [f(a,b,c,d) + f(a,c,b,d) - (b+c)] \\
&\overset{\text{(c)}}{=} 2\log 3 - 2 \approx 1.17
\end{aligned}$$

where

$$f(a,b,c,d) := (a+b)h\left(\tfrac{b}{2(a+b)}\right) + (c+d)h\left(\tfrac{c}{2(c+d)}\right)$$

(a) This is because there is only one possible fractional partition.

(b) Let X_1 and X_2 be distributed as in (C.14). Then, (b) follows from (C.15).

(c) The maximum is achieved at $a = d = \frac{1}{6}$ and $b = c = \frac{1}{3}$.

C.3 Computer-assisted tightness test

In this section, we present a computer-assisted test of tightness for the following bound, which is a generalization of the mutual dependence upper bound in (2.13) from entropy function to any supermodular function.

Theorem C.1 *Given any supermodular function $h : \mathcal{F} \mapsto \mathbb{R}$ in* (B.6) *on $\mathcal{F} \in \Phi(A)$, where $\Phi(A)$, $\Lambda(\mathcal{F}, V)$ and $\Pi(\mathcal{F}, V)$ are defined in Definition B.1, we have*

$$\max_{\lambda\in\Lambda(\mathcal{F},V)} \sum_{B\in\mathcal{F}} \lambda_B h(B) \geq \max_{\mathcal{P}\in\Pi(\mathcal{F},V)} \frac{\sum_{C\in\mathcal{P}} h(C^c)}{|\mathcal{P}| - 1} \tag{C.18}$$

This bound is loose for $h(B) := H(\mathsf{Z}_B|\mathsf{Z}_{B^c})$ in Example 2.2, which is a minimal example in the lexicographical order of $(|V|, |A|)$. □

Proof The R.H.S. of (C.18) can be obtained from the L.H.S. with the additional constraint on λ that

$$\lambda_B := \frac{1}{|\mathcal{P}| - 1} \mathbb{1}_{\mathcal{P}}(B^c) \qquad \text{for all } B \in \mathcal{F} \tag{C.19}$$

for some $\mathcal{P} \in \Pi(\mathcal{F}, V)$. In Example 2.2, the L.H.S. is $H(\mathsf{Z}_V) - C_{\mathrm{s}} = \frac{9}{4}$ by Proposition 2.1 while the R.H.S. is $3 - 1 = 2$ because the mutual dependence upper bound is 1. Thus, the bound is loose.

To prove minimality, we will show that the additional condition (C.19) for λ is admissible for all the cases of $(|V|, |A|)$ smaller $(6, 3)$. We will derive a sufficient condition for tightness using the supermodularity assumption of h and test it case-by-case with the help of a computer program.

Note that $\Lambda(\mathcal{F}, V) = \emptyset$ only if $\Pi(\mathcal{F}, V) = \emptyset$ by (C.19). Both sides of (C.18) are $-\infty$ by convention and so the bound is trivially tight. Thus, we can focus on the non-trivial case where $\Lambda(\mathcal{F}, V) \neq \emptyset$.

Consider the linear programming dual of the L.H.S. of (C.18).

$$\text{minimize} \quad \sum_{i \in V} r_i \tag{C.20a}$$

$$\text{subject to} \quad \sum_{i \in B} r_i \geq h(B) \qquad \text{for all } B \in \mathcal{F} \tag{C.20b}$$

This is the same as (B.8) with A replaced by V. Let Γ^* be the set of optimal solutions $\boldsymbol{r} := (r_i : i \in V)$, and $\mathcal{T}$ be the set of tight constraints where

$$\mathcal{T} := \left\{ B \in \mathcal{F} : \forall \boldsymbol{r} \in \Gamma^*, \sum_{i \in B} r_i = h(B) \right\} \tag{C.21}$$

Subclaim C.1A *The bound* (C.18) *is tight if there exists* $\mathcal{P} \in \Pi(\mathcal{F}, V)$ *such that* $C^c \in \mathcal{T}$ *for all* $C \in \mathcal{P}$. ◁

PROOF Suppose the required $\mathcal{P}$ exists. For any optimal solution $\boldsymbol{r} \in \Gamma^*$,

$$\sum_{i \in V} r_i = \sum_{i \in V} \overbrace{\sum_{C \in \mathcal{P} : i \notin C} \frac{1}{|\mathcal{P}| - 1}}^{\overset{\text{(a)}}{=} 1} r_i = \sum_{C \in \mathcal{P}} \frac{1}{|\mathcal{P}| - 1} \overbrace{\sum_{i \in C^c} r_i}^{\overset{\text{(b)}}{=} h(C^c)}$$

where (a) is because $\mathcal{P}$ is a set partition, and (b) is because $C^c \in \mathcal{T}$ corresponds to a tight constraint. The L.H.S. of the above equals that of (C.18) by the strong duality theorem while the R.H.S. of the above equals that of (C.18) as desired. ◂

Let $\Lambda^* \subseteq \Lambda(\mathcal{F}, V)$ be the set of optimal solutions to the primal in (C.18), and

$$\operatorname{supp}(\lambda) := \{B \in \mathcal{F} : \lambda_B > 0\} \qquad \text{for } \lambda \in \Lambda(\mathcal{F}, V)$$

be the support of λ. Then, for any $\lambda \in \Lambda^*$, $\lambda_B > 0$ implies $B \in \mathcal{T}$ by the complementary slackness theorem [16]. Define $\mathcal{T}(\mathcal{S})$ for $\mathcal{S} \subseteq \mathcal{F}$ as follows,

$$B \in \mathcal{S} \implies B \in \mathcal{T}(\mathcal{S}) \tag{C.22a}$$

$$B, B' \in \mathcal{T}(\mathcal{S}) \text{ and } B \cup B' \not\supseteq A \implies B \cap B', B \cup B' \in \mathcal{T}(\mathcal{S}) \tag{C.22b}$$

Using $\mathcal{T}(\mathcal{S})$, we have the following sufficient condition for tightness.

Subclaim C.1B *The bound* (C.18) *is tight if there exists* $\lambda \in \Lambda^*$ *and* $\mathcal{P} \in \Pi(\mathcal{F}, V)$ *such that* $C^c \in \mathcal{T}(\operatorname{supp}(\lambda))$ *for all* $C \in \mathcal{P}$. ◃

PROOF By the definition (B.2) of $\mathcal{F} \in \Phi(A)$, we have $B, B' \in \mathcal{F} : B \cup B' \not\supseteq A$ implies $B \cap B', B \cup B' \in \mathcal{F}$. Thus, by Subclaim B.1A in the proof of Theorem B.1, we have $\mathcal{T}(\operatorname{supp}(\lambda)) \subseteq \mathcal{T}$ for all $\lambda \in \Lambda^*$. Since $C^c \in \mathcal{T}(\operatorname{supp}(\lambda))$ for all $C \in \mathcal{P}$ implies $C^c \in \mathcal{T}$ for all $C \in \mathcal{P}$, the tightness condition here implies that in Subclaim C.1A. ◂

It is impossible to test the condition for every $\lambda \in \Lambda^*$ because Λ^* is not finite. Furthermore, Λ^* depends on the choice of the supermodular function h, which is again impossible to enumerate. We want to further weaken the condition such that it only has a finite number of test cases independent of h. This is possible from the following observation.

Subclaim C.1C Λ^* *must contain some* λ *that satisfies* $|\operatorname{supp}(\lambda)| \leq| V|$.[2] ◃

[2] We can impose the additional constraint that λ is extremal [16], i.e. λ is not a strict convex combination of distinct elements in $\Lambda(\mathcal{F}, V)$. Although this implies the stronger condition that the incidence matrix of $\operatorname{supp}(\lambda)$ has full column rank, the weaker condition suffices here.

Proof Suppose to the contrary that every $\lambda \in \Lambda^*$ has $|\text{supp}(\lambda)| > |V|$. By (B.5),

$$\sum_{B \in \text{supp}(\lambda)} \lambda_B = 1 \qquad \text{for all } i \in V$$

which is a set of $|V|$ linear equations with more than $|V|$ variables $(\lambda_B : B \in \text{supp}(\lambda))$. Thus, there exists a real vector $(e_B : B \in \text{supp}(\lambda)) \neq \mathbf{0}$ satisfying

$$\sum_{B \in \text{supp}(\lambda)} e_B = 0 \qquad \text{for all } i \in V$$

This implies that

$$\sum_{B \in \text{supp}(\lambda)} \overbrace{(\lambda_B + \epsilon e_B)}^{\lambda'_B :=} = 1 \qquad \text{for all } i \in V$$

for any choice of $\epsilon \in \mathbb{R}$. Not only can we guarantee $\lambda'_B \geq 0$ for small enough ϵ, but we can also choose ϵ such that $\text{supp}(\lambda') \subsetneq \text{supp}(\lambda)$ by continuity. It suffices to argue that λ' defined this way is in Λ^*. This indeed follows immediately from the primal/dual optimality criteria in [16] since

$$\lambda'_B > 0 \implies \lambda_B > 0 \implies \sum_{i \in B} r_i = h(B)$$

for all $B \in \mathcal{F}$ and $\boldsymbol{r} \in \Gamma^*$. ◀

Using this, we can remove the dependence on Λ^* and therefore h from the tightness condition.

Subclaim C.1D *The bound* (C.18) *is tight if, for all* $\lambda \in \Lambda(\mathcal{F}, V) : |\text{supp}(\lambda)| \leq |V|$, *there exists* $\mathcal{P} \in \Pi(\mathcal{F}, V)$ *such that* $C^c \in \mathcal{T}(\text{supp}(\lambda))$ *for all* $C \in \mathcal{P}$. ◁

Proof Suppose this condition is satisfied. The desired $\mathcal{P}$ for the tightness condition in Subclaim C.1B also exists for some $\lambda \in \Lambda^*$ since at least one must satisfy $|\text{supp}(\lambda)| \leq| V|$ by Subclaim C.1C. ◀

Instead of enumerating the infinitely-valued λ, we can enumerate the finitely-valued $\text{supp}(\lambda)$. Define $\boldsymbol{M}(\text{supp}(\lambda))$ as the boolean matrix whose entry at row $B \in \text{supp}(\lambda)$ and column $i \in V$ is $\mathbb{1}_i(B)$. It must satisfy the following property.

Subclaim C.1E *Columns of $\boldsymbol{M}(\operatorname{supp}(\lambda))$ indicate sets that are not proper subsets of each other. We denote this by $\boldsymbol{M}(\operatorname{supp}(\lambda)) \in \mathrm{NPS}$ and call the matrix NPS.* $\lhd$

PROOF Suppose to the contrary that there exists $i, j \in V$ with

$$\{B \in \operatorname{supp}(\lambda) : i \in B\} \subsetneq \{B \in \operatorname{supp}(\lambda) : j \in B\}$$

Then, by (B.5),

$$1 = \sum_{B \ni i} \lambda_B = \sum_{B} \lambda_B \mathbb{1}_B(i) < \sum_{B} \lambda_B \mathbb{1}_B(i) = \sum_{B \ni j} \lambda_B = 1$$

which is a contradiction. $\blacktriangleleft$

We call an NPS matrix $\boldsymbol{M}$ basic, denoted as $\boldsymbol{M} \in \mathrm{BNPS}$, if $\boldsymbol{M}$ is not NPS after removing any row. Using this property, we have the following tightness condition.

Subclaim C.1F *The bound* (C.18) *is tight if, for all $\mathcal{S} \subseteq \mathscr{F}(A)$ in* (B.1a) *satisfying $|\mathcal{S}| \leq |V|$ and $\boldsymbol{M}(\mathcal{S}) \in \mathrm{BNPS}$, there exists $\mathcal{P} \in \Pi(\mathscr{F}(A), V)$ such that $C^c \in \mathcal{T}(\mathcal{S})$ for all $C \in \mathcal{P}$.* $\lhd$

PROOF Suppose the condition is satisfied. This implies the same condition with the term "BNPS" replaced by "NPS" since every NPS matrix contains a basic NPS submatrix.[3] Consider any $\mathcal{F} \in \Phi(A)$ and $\lambda \in \Lambda(\mathcal{F}, V) : |\operatorname{supp}(\lambda)| \leq | V|$. Since we have $\operatorname{supp}(\lambda) \subseteq \mathcal{F} \subseteq \mathscr{F}(A)$ from their definitions[4] and $\boldsymbol{M}(\operatorname{supp}(\lambda)) \in \mathrm{NPS}$ by Subclaim C.1E, the tightness condition here implies the tightness condition in Subclaim C.1D. More precisely, the condition implies existence of $\mathcal{P} \in \Pi(\mathscr{F}(A), V)$ with $C^c \in \mathcal{T}(\operatorname{supp}(\lambda))$ for all $C \in \mathcal{P}$. Since $\mathcal{T}(\operatorname{supp}(\lambda)) \subseteq \mathcal{F}$ by the definitions (C.22) and (B.2), we also have $\mathcal{P} \in \Pi(\mathcal{F}, V)$ by the definition (B.3). $\blacktriangleleft$

This tightness condition can be tested by enumerating BNPS matrices, which is finitely-valued for a given finite dimension, and independent of the choice of h and

[3] If a matrix is NPS but not BNPS, there is a row one can remove without loosing the NPS property. Repeating this eventually gives a BNPS submatrix.

[4] $\mathscr{F}(A)$ contains all subsets of V that are not supersets of A. Thus, $\mathcal{F} \subseteq \mathscr{F}(A)$ because it does not contain supersets of A by (B.2).

$\mathcal{F}$. It is implemented as follows using the IT++ library [46] and tested for the cases of $(|V|, |A|)$ equal to $(3, 2)$, $(4, 2)$, $(4, 3)$, $(5, 2)$, $(5, 3)$ and $(6, 2)$. The remaining cases $(3, 3)$, $(4, 4)$ and $(5, 5)$ with $A = V$ follows from Theorem B.1. ■

sklib.h:

```
#ifndef SKLIB_H
#define SKLIB_H
#endif

#include <itpp/itbase.h>

using namespace itpp;

namespace sk {
  bool NPS(bmat A);
  bvec bitget(int x,ivec b);
  bmat SWM(int m,int a);
  void newBNPS(bmat A,ivec S,Array<ivec> C,
               ivec& newS,Array<ivec>& newC);
  Array<ivec> genBNPS(bmat A);
  typedef Array<Array<ivec> > Dstruct;
  bvec Union(bmat A);
  bool TC(int a, bmat tA, Dstruct D);
  void TMD(int a, bmat A,Array<ivec> C,
           Array<Dstruct> & Ds,ivec& P);
}
```

sklib.cpp:

```
#include "sklib.h"

using namespace itpp; using namespace std;

namespace sk {
  /**
   * enumerate 𝓕(A).
   * @param m is the number of terminals. V := [m]
   * @param a is the number of active users. A := [a]
   * @return the incidence matrix M(𝓕(A) \ {∅}).
   */
  bmat SWM(int m,int a) {
```

```
  int i; bmat A1,A2;
  // enumerate proper subsets of A.
  for(i=0;i<pow2(a)-1;i++)
    A1.append_row(dec2bin(a,i));
  if(m<=a)
    // the empty set is not needed for the test.
    return A1.get_rows(1,A1.rows()-1);
  // enumerate subsets of helpers A^c.
  for(i=0;i<pow2(m-a);i++)
    A2.append_row(dec2bin(m-a,i));
  return concat_horizontal(kron(A1,ones_b(A2.rows(),1)),
                           kron(ones_b(A1.rows(),1),A2))
    .get_rows(1,A1.rows()*A2.rows()-1);
}

/** test a matrix the NPS property.
 * @param tA is an incidence matrix.
 * @return true if tA is NPS.
 */
bool NPS(bmat tA) {
  for(int i=0;i<tA.cols();i++)
    for(int j=i+1;j<tA.cols();j++) {
      ivec cmp=to_ivec(tA.get_col(i))
        -to_ivec(tA.get_col(j));
      // NPS implies that non-zero differences of any
      // two columns cannot share the same sign.
      if((bool) sum(to_ivec(cmp==1))
         != (bool) sum(to_ivec(cmp==-1))) return false;
    }
  return true;
}

/**
 * find new BNPS submatrices using NPS().
 * @param A is an incidence matrix.
 * @param S is a vector of (increasing) row indices of A
 *          previously selected for the submatrices.
 * @param C is an array of row selections for the
 *          previously tested NPS submatrices.
 * @param newS will be assigned a vector of newly selected
 *          rows needed for further testing.
 * @param newC will be assigned an array of row selections
 *          for the newly tested NPS submatrices.
 */
```

```
void newBNPS(bmat A,ivec S,Array<ivec> C,
             ivec& newS,Array<ivec>& newC) {
  newS.set_length(0);newC.set_length(0);ivec I=concat(S,-1);
  for(int s=S(S.length()-1)+1; s<A.rows();s++) {
    // add a new row from A and test if the resulting matrix
    // contains any NPS submatrix.
    I(S.length())=s; int pos=0;
    while(pos<C.length()){
      int i=0,j=0; bool submatrix=true;
      while (submatrix && i<C(pos).length() && j<I.length())
        if(I(j)==C(pos)(i)) { i++; j++; }
        else if(I(j)<C(pos)(i) && j<I.length()-1) j++;
        else submatrix=false;
      if(submatrix && i==C(pos).length()) break;
      else pos++;
    }
    // BNPS matrix is NPS but without any NPS submatrix.
    if(pos==C.length()) {
      if(NPS(A.get_rows(I))) newC= concat(newC,I);
      else newS= concat(newS,s);
    }}}

/**
 * enumerate BNPS submatrices using newBNPS().
 * @param A is an incidence matrix.
 * @return an array of row selections of A that
 *         correspond to BNPS submatrices.
 */
Array<ivec> genBNPS(bmat A) {
  imat SM("0:"+to_str(A.rows()-1)); Array<ivec> C;
  while(SM.cols()) {
    // any column of SM, if exists, requires further testing for BNPS.
    imat newSM; Array<ivec> newC;
    for(int i=0; i<SM.cols(); i++) {
      Array<ivec> newC1; ivec newS;
      newBNPS(A,SM.get_col(i),C,newS,newC1);
      // no need to enumerate BNPS with more that m rows.
      if(SM.rows()<A.cols()-1 && newS.length()) {
        // enumerate new submatrices for further testing.
        imat tmp=repmat(SM.get_col(i),1,newS.length());
        tmp.append_row(newS);
        newSM=newSM.cols()?
            concat_horizontal(newSM,tmp):tmp;
      }
      if(newC1.length()) newC=concat(newC,newC1);
    }
    C=concat(C,newC); SM=newSM;
    // report the number of BNPS submatrices found and
    // the number of cases for further testing.
    cout << "%genBNPS␣Level:" << SM.rows()+1
         << "\t#BNPS:" << C.length()
         << "\t#tests␣next:" << SM.cols() << endl;
  }
  return C;
}

/**
 * carry out the union operation.
 * @param sA is an incidence matrix.
 * @return the elementwise 'or' of rows of sA.
 */
bvec Union(bmat sA) {
  bvec out; out.set_length(sA.cols());
  for(int i=0;i<sA.cols();i++)
    out(i)=bin2dec(sA.get_col(i))>0;
  return out;
}

/**
 * Test the tightness condition for a given choice of $\mathcal{P}$
 * and $\mathcal{S}$.
 * @param a is the number of active users.
 * @param tA is the incidence matrix $M(\mathcal{S})$.
 * @param D defines $\mathcal{P}$. It has type Dstruct defined as
 *            typedef Array<Array<ivec> > Dstruct;
 *        $C^c$ for $C \in \mathcal{P}$ are specified as unions and
 *        intersections of rows of tA that are in $\mathcal{T}(\mathcal{S})$.
 *        e.g. D="{{[0 1] [1 2]} {[2 3]}}" means taking
 *        unions of first two and second two rows,
 *        and the intersection of the next two rows.
 *        n.b. 0 indices the first row.
 */
bool TC(int a, bmat tA, Dstruct D) {
  bmat B;
  // take the unions specified in D
  for(int i=0;i<D(0).length();i++) {
    if(!D(0)(i).length() || max(D(0)(i))>=tA.rows() ||
       min(D(0)(i))<0) return false; // invalid row selection
    bvec tmp=Union(tA.get_rows(D(0)(i)));
    if(prod(tmp.left(a))) return false; // not in 𝓕(A)
    B.append_row(tmp);
  }
  // take the intersections specified in D
  for(int i=0;i<D(1).length();i++) {
    if(!D(1)(i).length() || max(D(1)(i))>=tA.rows() ||
       min(D(1)(i))<0) return false;
    // make sure the union is in 𝓕(A)
    bvec tmp=Union(tA.get_rows(D(1)(i)));
    if(prod(tmp.left(a))) return false; // not in 𝓕(A)
    tmp=prod(tA.get_rows(D(1)(i)),1);
    if(!bin2dec(tmp)) return false; // empty set ignored
    B.append_row(tmp);
  }
  // requires 2 ≤ |𝒫| ≤ |A|.
  if(B.rows()>a||B.rows()<2) return false;
  // 𝒫 must be a partition of V. i.e. every column of B has one 0.
  return prod(B.rows()-sum(to_imat(B),1))==1;
}


/**
 * test for tightness using TC().
 * @param a is the number of active users.
 * @param A is an incidence matrix.
 * @param C is the set of BNPS submatrices sufficient.
 * @param Ds is an array of candidates of D for TC().
 * @param P is a vector of the choices of D in Ds that
 *          satisfy TC() for the corresponding C.
 */
void TMD(int a, bmat A,
         Array<ivec> C,Array<Dstruct> & Ds,ivec& P) {
  P.set_length(C.length());
  for(int i=0;i<C.length();i++) {
    bmat tA = A.get_rows(C(i));
    int j=0; while(j<Ds.length() && !TC(a,tA,Ds(j))) j++;
    if(j<Ds.length()) P(i)=j;
    else { // ask when no specified D satisfies TC().
      Dstruct tmp; char Input[100];
      cout << tA << endl << C.length()-i << "␣left.␣D=?";
      do {
        cin.getline(Input,100);
        if(*Input=='\0'){cout<<Ds<<endl<<"D=?"; continue;}
        tmp=Input; if(TC(a,tA,tmp)) break;
```

```
          cout << "Invalid.␣D=?";
        } while(true);
        P(i)=Ds.length(); Ds=concat(Ds,tmp);
      }}}
}
```

main.cpp:

```
#include <itpp/itcomm.h>
#include "sklib.h"
#include <stdio.h>

using namespace sk;
using namespace itpp; using namespace std;

/**
 * Run the tightness test.
 * The first two command line arguments are the numbers of
 * terminals and active users respectively. The third
 * argument is the result filename (default: "results.it").
 * e.g. the following command test the tightness for |V| = 3,
 * |A| = 2 and record the result in results.it
 *        ./main 3 2 results.it
 * Successful termination implies tightness.
 * The file contains details that can be loaded into Matlab
 *  with itload.m from the IT++ library.
 */
int main(int argc, char *argv[]) {
  int m,a;
  if (argc<3) {
    cerr << "Missing␣arguments:␣m␣and␣a" << endl; return 1;
  }
  m=atoi(argv[1]); a=atoi(argv[2]);
  if(m<a) {
    cerr << "m<a" << endl; return 1;
  }
  string fn="results.it";
  if (argc>3 && *argv[3]!='\0') fn=argv[3];
  string s=to_str(m)+"_"+to_str(a);
  bmat A; Array<ivec> C; ivec P; Array<Dstruct> Ds;
  it_file ff(fn);
  // read the SW matrix and the set of NPS submatrices for testing
  if(ff.seek("A"+s)) ff >> Name("A"+s) >> A;
  else A=SWM(m,a);
  if(ff.seek("C"+s)) ff >> Name("C"+s) >> C;
  else C=genBNPS(A);
  cout << "A"+s+"=" << A << endl << "C"+s+"=" << C << endl;
  ff << Name("A"+s) << A; ff << Name("C"+s) << C;
  if (ff.seek("Ds"+s)) { // load Ds from file if available
    string tmp; ff >> Name("Ds"+s) >> tmp; set_array(Ds,tmp);
  }
  TMD(a,A,C,Ds,P); // run the test
  ff << Name("Ds"+s) << to_str(Ds); ff << Name("P"+s) << P;
  ff.flush(); ff.close();
  cout << "Ds"+s+"=" << Ds << endl << "P"+s+"=" << P << endl;
  cout << "%␣MD␣upper␣bound␣is␣tight␣for␣" << m <<
    "␣terminals␣with␣" << a << "␣active␣users." << endl;
  return 0;
}
```

initD.cpp:

```
#include <itpp/itcomm.h>

using namespace itpp;
using std::string;

/**
 * Initialize Ds for the tightness test.
 * The argument is the result filename (default: "results.it").
 * Run this before the main program by the command
 *     ./initD results.it
 */
int main(int argc, char *argv[]) {
  string fn="results.it";
  if (argc>1 && *argv[1]!='\0') {
    fn=argv[1];
  }
  it_file ff(fn);
  ff << Name("Ds3_2")
     << string("{{{[0]␣[1]}␣{}}␣{{[0␣1]␣[2]}␣{}}}");
  ff << Name("Ds4_2")
     << string("{{{[0]␣[1]}␣{}}␣{{[0␣1]␣[2]}␣{}}␣{{[1]␣[0␣2]}␣{}}␣{{[0]␣[2␣3]}␣{}}"
               "␣{{[0␣1␣2]␣[3]}␣{}}}");
  ff << Name("Ds4_3")
     << string("{{{[0]␣[1]}␣{}}␣{{[1]␣[0␣2]}␣{}}␣{{[0]␣[1]␣[2]}␣{}}"
               "␣{{[1]␣[0␣2]␣[0␣3]}␣{}}␣{{[1]␣[0␣2]␣[3]}␣{}}␣{{[0]␣[1␣2]}␣{}}"
               "␣{{[0]␣[1␣3]}␣{}}␣{{[0␣1]␣[2]␣[3]}␣{}}␣{{[0]␣[1␣2]␣[3]}␣{}}"
               "␣{{[0␣1␣2]␣[3]}␣{}}}");
ff << Name("Ds5_2")
   << string("{{{[0]␣[1]}␣{}}␣{{[0␣1]␣[2]}␣{}}␣{{[1]␣[0␣2]}␣{}}␣{{[0]␣[2␣3]}␣{}}"
               "␣{{[2␣3]}␣{[0␣1]}}␣{{[0]␣[2␣3␣4]}␣{}}␣{{[1]␣[2␣3]}␣{}}"
               "␣{{[0␣1␣2]␣[3]}␣{}}␣{{[0␣1␣2␣3]␣[4]}␣{}}␣{{[0␣1␣2]}␣{[3␣4]}}}");
ff << Name("Ds5_3")
   << string("{{{[0]␣[1]}␣{}}␣{{[0␣1]␣[2]}␣{}}␣{{[0]␣[1]␣[2]}␣{}}"
               "␣{{[1]␣[0␣2]␣[0␣3]}␣{}}␣{{[1]␣[0␣2]␣[3]}␣{}}␣{{[1]␣[2␣3]}␣{}}"
               "␣{{[1]␣[2]␣[0␣3]}␣{}}␣{{[0]␣[2␣3]}␣{}}␣{{[0]␣[1␣3]}␣{}}"
               "␣{{[1]␣[0␣2]}␣{}}␣{{[0␣1]␣[2]␣[3]}␣{}}␣{{[1␣3]}␣{[0␣2]}}"
               "␣{{[2␣3]}␣{[0␣1]}}␣{{[1]␣[0␣2␣3]␣[0␣4]}␣{}}␣{{[0]␣[1␣2␣4]}␣{}}"
               "␣{{[0]␣[1␣4]}␣{}}␣{{[0]␣[3␣4]}␣{}}␣{{[0␣3]␣[1]}␣{}}␣{{[4]}␣{[0␣1]}}"
               "␣{{[0␣1]}␣{[2␣4]}}␣{{[1]␣[2]␣[3␣4]}␣{}}␣{{[0]␣[2]␣[3␣4]}␣{}}"
               "␣{{[0]␣[1]␣[3␣4]}␣{}}␣{{[0]␣[1␣2]␣[3]}␣{}}␣{{[0␣1]␣[3]␣[4]}␣{}}"
               "␣{{[0␣1]␣[3]␣[2␣4]}␣{}}␣{{[0␣1]␣[2]␣[3␣4]}␣{}}"
               "␣{{[1]␣[0␣2]␣[3␣4]}␣{}}␣{{[0␣1]␣[2␣3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]}␣{}}"
               "␣{{[0␣1␣2␣3]␣[4]}␣{}}␣{{[2]␣[0␣1␣3]␣[0␣1␣4]}␣{}}"
               "␣{{[0␣1␣2]␣[0␣1␣3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]␣[0␣1␣4]}␣{}}"
               "␣{{[0␣1␣2]}␣{[3␣4]}}␣{{[0␣1␣2]␣[3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]␣[0␣4]}␣{}}"
               "␣{{[1]␣[3]␣[0␣4]}␣{}}␣{{[1]␣[2]␣[0␣4]}␣{}}␣{{[0␣1]}␣{[2␣3]}}"
               "␣{{[0]␣[2␣3]␣[4]}␣{}}␣{{[0␣1]␣[2]␣[4]}␣{}}␣{{[0]}␣{[1␣2]}}}");
ff << Name("Ds5_4")
   << string("{{{[0]␣[1]}␣{}}␣{{[0␣2]␣[1]}␣{}}␣{{[0]␣[1]␣[2]}␣{}}"
               "␣{{[1]␣[0␣2]␣[0␣3]}␣{}}␣{{[1]␣[0␣2]␣[3]}␣{}}␣{{[0]␣[1␣2␣3]}␣{}}"
               "␣{{[0]␣[1␣3]}␣{}}␣{{[2]}␣{[0␣1]}}␣{{[0]␣[1]␣[2]␣[3]}␣{}}"
               "␣{{[0␣1]}␣{[2␣3]}}␣{{[1]␣[0␣2]␣[0␣3]␣[0␣4]}␣{}}"
               "␣{{[1]␣[0␣2]␣[0␣3]␣[4]}␣{}}␣{{[1]␣[0␣2]␣[3]␣[4]}␣{}}"
               "␣{{[0]␣[1␣2]}␣{}}␣{{[0]␣[1␣4]}␣{}}␣{{[1␣2]}␣{[0␣3]}}"
               "␣{{[0␣1]␣[0␣2]␣[3]␣[4]}␣{}}␣{{[0␣1]␣[2]␣[3]␣[4]}␣{}}␣{{[4]}␣{[0␣2]}}"
               "␣{{[2]}␣{[1␣3]}}␣{{[4]}␣{[0␣1]}}␣{{[1]␣[2]␣[3␣0]␣[4]}␣{}}"
               "␣{{[0]␣[2]␣[3␣1]␣[4␣1]}␣{}}␣{{[0]}␣{[1␣3]}}␣{{[0]}␣{[1␣2]}}"
               "␣{{[0]␣[2]␣[3␣1]␣[4]}␣{}}␣{{[0␣1]␣[0␣2]␣[3]}␣{}}␣{{[0␣1]␣[2]␣[3]}␣{}}"
               "␣{{[0␣1]␣[2]␣[1␣3]}␣{}}␣{{[0␣1]␣[3]␣[4]}␣{}}␣{{[0]␣[2]␣[3]␣[1␣4]}␣{}}"
               "␣{{[0]␣[1]␣[2␣3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]}␣{}}␣{{[0␣1␣2␣3]␣[4]}␣{}}"
               "␣{{[0␣1␣2]␣[0␣1␣3]␣[4]}␣{}}␣{{[1]␣[2␣3]}␣{}}␣{{[1]␣[2]␣[0␣3]}␣{}}"
               "␣{{[0]␣[2␣3]}␣{}}␣{{[0]␣[1␣3]}␣{}}␣{{[1]␣[0␣2]}␣{}}"
               "␣{{[0␣1]␣[2]␣[3]}␣{}}␣{{[1␣3]}␣{[0␣2]}}␣{{[2␣3]}␣{[0␣1]}}"
               "␣{{[1]␣[0␣2␣3]␣[0␣4]}␣{}}␣{{[0]␣[1␣2␣4]}␣{}}␣{{[0]␣[1␣4]}␣{}}"
               "␣{{[0]␣[3␣4]}␣{}}␣{{[0␣3]␣[1]}␣{}}␣{{[4]}␣{[0␣1]}}␣{{[0␣1]}␣{[2␣4]}}"
               "␣{{[1]␣[2]␣[3␣4]}␣{}}␣{{[0]␣[2]␣[3␣4]}␣{}}␣{{[0]␣[1]␣[3␣4]}␣{}}"
               "␣{{[0]␣[1␣2]␣[3]}␣{}}␣{{[0␣1]␣[3]␣[4]}␣{}}␣{{[0␣1]␣[3]␣[2␣4]}␣{}}"
               "␣{{[0␣1]␣[2]␣[3␣4]}␣{}}␣{{[1]␣[0␣2]␣[3␣4]}␣{}}"
               "␣{{[0␣1]␣[2␣3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]}␣{}}␣{{[0␣1␣2␣3]␣[4]}␣{}}"
               "␣{{[2]␣[0␣1␣3]␣[0␣1␣4]}␣{}}␣{{[0␣1␣2]␣[0␣1␣3]␣[4]}␣{}}"
               "␣{{[0␣1␣2]␣[3]␣[0␣1␣4]}␣{}}␣{{[0␣1␣2]}␣{[3␣4]}}"
               "␣{{[0␣1␣2]␣[3]␣[4]}␣{}}␣{{[0␣1␣2]␣[3]␣[0␣4]}␣{}}"
               "␣{{[1]␣[3]␣[0␣4]}␣{}}␣{{[1]␣[2]␣[0␣4]}␣{}}␣{{[0␣1]}␣{[2␣3]}}"
               "␣{{[0]␣[2␣3]␣[4]}␣{}}␣{{[0␣1]␣[2]␣[4]}␣{}}␣{{[0]}␣{[1␣2]}}}");
  ff << Name("Ds6_2")
     << string("{{{[0]␣[1]}␣{}}␣{{[0␣1]␣[2]}␣{}}␣{{[1]␣[2␣3]}␣{}}␣{{[0]␣[2␣3]}␣{}}"
               "␣{{[2␣3]}␣{[0␣1]}}␣{{[1]␣[0␣2]}␣{}}␣{{[0]␣[2␣3␣4]}␣{}}"
               "␣{{[2␣3␣4]}␣{[0␣1]}}␣{{[0␣1␣2]␣[3]}␣{}}␣{{[0␣1␣2␣3]␣[4]}␣{}}"
               "␣{{[2]␣[0␣3␣4]}␣{}}␣{{[1]␣[0␣3␣4]}␣{}}␣{{[1]␣[3␣4]}␣{}}"
               "␣{{[1␣2]␣[3]}␣{}}␣{{[0␣1␣2]}␣{[3␣4]}}␣{{[0␣1␣2␣3␣4]␣[5]}␣{}}"
               "␣{{[0␣1␣2␣3]}␣{[4␣5]}}␣{{[0␣1␣2]}␣{[3␣4␣5]}}␣{{[0␣1]}␣{[2␣3]}}"
               "␣{{[0␣1]}␣{[2␣3␣4]}}}");
  ff.flush(); ff.close();
  return 0;
}
```

Bibliography

[1] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography—Part I: Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, Jul 1993.

[2] Amir Salman Avestimehr, Suhas N. Diggavi, and David N. C. Tse. Wireless network information flow: A deterministic approach. *CoRR*, abs/cs/0906.5394, 2009.

[3] Jorgen Bang-Jensen and Stephan Thomassé. Decompositions and orientations of hypergraphs. Preprint no. 10, Department of Mathematics and Computer Science, University of Southern Denmark, May 2001.

[4] Anthony J. Bell. The co-information lattice. In *ICA2003*, Nara, Japan, April 2003.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[6] Chung Chan and Lizhong Zheng. Mutual dependence for secret key agreement. In *Proceedings of 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010.

[7] Claude Crepeau Charles H. Bennett, Gilles Brassard and Ueli M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, Nov 1995.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience Publication, 1991.

[9] Imre Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):48–57, 1996.

[10] Imre Csiszár and János Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, May 1978.

[11] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Akadémiai Kiadó, Budapest, 1981.

[12] Imre Csiszár and Prakash Narayan. Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory*, 50(12), Dec 2004.

[13] Imre Csiszár and Prakash Narayan. Secrecy capacities for multiterminal channel models. *IEEE Transactions on Information Theory*, 54(6):2437–2452, June 2008.

[14] Paul Cuff. Communication requirements for generating correlated random variables. In *Proceedings of 2008 IEEE International Symposium on Information Theory*, pages 1393–1397, 2008.

[15] George B. Dantzig and Mukund N. Thapa. *Linear Programming. 1: Introduction.* Springer-Verlag New York, 1997-2003.

[16] George B. Dantzig and Mukund N. Thapa. *Linear Programming. 2: Theory and Extensions.* Springer-Verlag New York, 1997-2003.

[17] Brecht Donckels. Matlab implementation of shuffled complex evolution. `http://biomath.ugent.be/~brecht/download/SCE.zip`, 2006.

[18] Randall Dougherty, Chris Freiling, and Kenneth Zeger. Networks, matroids, and non-Shannon information inequalities. *IEEE Transactions on Information Theory*, 53(6):1949–1969, Jun 2007.

[19] Harold G. Eggleston. *Convexity.* Cambridge University Press, 1966.

[20] András Frank, Tamás Király, and Matthias Kriesell. On decomposing a hypergraph into k-connected sub-hypergraphs. *Discrete Applied Mathematics*, 131(2):373–383, September 2003.

[21] Satoru Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39(1):55–72, 1978.

[22] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, Feb 1972.

[23] Amin A. Gohari and Venkat Anantharam. Communication for omniscience by a neutral observer and information-theoretic key agreement of multiple terminals. In *Proceeding of 2007 IEEE International Symposium of Information Theory*, pages 2056–2060, Nice, France, Jun 2007.

[24] Amin A. Gohari and Venkat Anantharam. New bounds on the information-theoretic key agreement of multiple terminals. In *Proceeding of 2008 IEEE International Symposium of Information Theory*, Toronto, Ontario, Canada, Jun 2008.

[25] Te S. Han. Linear dependence structure of the entropy space. *Information and Control*, 29:337–368, 1975.

[26] Te S. Han. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control*, 36:133–156, 1978.

[27] Te S. Han. Multiple mutual informations and multiple interactions in frequency data. *Information and Control*, 46:26–45, 1980.

[28] Tracey Ho and Desmond S. Lun. *Network Coding: An Introduction.* Cambridge University Press, 2008.

[29] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[30] Aleks Jakulin and Ivan Bratko. Quantifying and visualizing attribute interactions: An approach based on entropy. *CoRR*, abs/cs/0308002, 2004.

[31] Edwin T. Jaynes. Information theory and statistical mechanics. In K. Ford, editor, *Statistical Physics*, volume 3, pages 181–218, New York, 1963. W. A. Benjamin Inc.

[32] Frank Jones. *Lebesgue Integration on Euclidean Spaces.* Jones and Barlett Publishers, revised edition, 2000.

[33] A. Romashchenko K. Makarychev, Y. Makarychev and N. Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, Dec 2002.

[34] Ralf Koetter and Muriel Médard. An algebraic approach to network coding. *IEEE/ACM Transactions on Networking*, 11(5), October 2003.

[35] Hidetoshi Komiya. Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.

[36] János Körner and K. Marton. The comparison of two noisy channels. In I. Csiszár and P. Elias, editors, *Topics in Information Theory*, pages 411–423. Colloquia Mathematica Societatis János Bolyai, North Holland, Amsterdam, 1975.

[37] Zongpeng Li and Baochun Li. Network coding in undirected networks. In *Proceedings of 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004.

[38] Mokshay Madiman and Prasad Tetali. Information inequalities for joint distributions, with interpr etations and applications. *IEEE ransactions of Information Theory*, 2008. to appear.

[39] Frantisek Matúš. Infinitely many information inequalities. In *Proceeding of 2007 IEEE International Symposium of Information Theory*, pages 41–44, Nice, France, Jun 2007.

[40] Ueli Maurer and Stefan Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. *Lecture Notes in Computer Science*, 1807:351+, 2000.

[41] Ueli M. Maurer. Perfect cryptographic security from partially independent channels. In *Proc. 23rd ACM Symposium on Theory of Computing*, pages 561–571, 1991.

[42] Ueli M. Maurer. Secret key agreement by public discussion from common informat ion. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[43] Ueli M. Maurer and Stefan Wolf. Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Transaction on Information Theory*, 45(2):499–514, Mar 1999.

[44] William J. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, June 1954.

[45] Chandra Nair, Balaji Prabhakar, and Devavrat Shah. On entropy for mixtures of discrete and continuous variables. *CoRR*, abs/cs/0607075, 2006.

[46] T. Ottosson and A. Piątyszek. IT++—C++ library of mathematical, signal processing, speech processing, and communications classes and functions, 2006.

[47] Dimitris N. Politis. On the entropy of a mixture distribution. Technical Report 91-67, Department of Statistics, Purdue University, November 1991.

[48] Renato Renner and Stefan Wolf. New bounds in secret-key agreement: The gap between formation and secrecy extraction. In *Advances in Cryptology–EUROCRYPT 2003*, volume 2656 of *Lecture Notes in Computer Science*, page 643. Springer Berlin / Heidelberg, 2003.

[49] Edward R. Scheinerman and Daniel H. Ullman. *Fractional Graph Theory.* A Wiley-Interscience Publication, 1997.

[50] Alexander Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency.* Springer, 2002.

[51] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[52] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, October 1984.

[53] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[54] Jural Skripsky. The gap between intrinsic information and the secret-key rate. Diploma thesis, Institute for Theoretical Computer Science, ETH Zürich, Switzerland, Oct 2002.

[55] İ. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 6(11):585–595, 1999.

[56] John von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, third edition, 1953.

[57] Satosi Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1):66–82, 1960.

[58] Hans S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal of Applied Mathematics*, 28(1):100–112, Jan 1975.

[59] Hans S. Witsenhausen. Values and bounds for the common information of two discrete random variables. *SIAM Journal of Applied Mathematics*, 31(2):313–333, September 1976.

[60] Aaron D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, Mar 1975.

[61] Aaron D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387, 1975.

[62] Chunxuan Ye and Alex Reznik. Group secret key generation algorithms. In *IEEE International Symposium on Information Theory, 2007. ISIT 2008.*, pages 2596–2600, June 2007.

[63] Raymond W. Yeung. A new outlook on Shannon's information measures. *IEEE Transactions of Information Theory*, 37(3):466–474, May 1991.

[64] Raymond W. Yeung. A framework for linear information inequalities. *IEEE Transactions on Information Theory*, 43(6):1924–1934, Nov 1997.

[65] Zhen Zhang and Raymond W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory*, 43(6):1982–1986, Nov 1997.

[66] Zhen Zhang and Raymond W. Yeung. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory*, 44(4), Jul 1998.

Index