AN OPERATIONAL APPROACH TO INFORMATION LEAKAGE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ibrahim Issa

August 2017

AN OPERATIONAL APPROACH TO INFORMATION LEAKAGE

Ibrahim Issa, Ph.D.

Cornell University 2017

Given two random variables X and Y, how much information does Y "leak" about X? An operational approach is undertaken to answer this question, which is fundamental to the study of communication security. The resulting measure $\mathcal{L}(X \to Y)$ is called *maximal leakage*, and is defined as the multiplicative increase, upon observing Y, of the probability of correctly guessing a randomized function of X, maximized over all such randomized functions. A closed form expression for $\mathcal{L}(X \to Y)$ is given for discrete X and Y, and it is subsequently generalized to handle a large class of random variables. The resulting properties are shown to be consistent with an axiomatic view of a leakage measure, and the definition is shown to be robust to variations in the setup.

Moreover, the *guessing* framework is used to give operational definitions to commonly used leakage measures, such as Shannon capacity, maximal correlation, and local differential privacy. Counter-intuitively, it is shown that Shannon capacity *underestimates* leakage. Furthermore, a variant of the Shannon cipher system is studied, in which performance of an encryption scheme is measured using maximal leakage. A single-letter characterization of the optimal limit of (normalized) maximal leakage is derived and asymptotically-optimal encryption schemes are demonstrated.

The Shannon cipher system is also studied when there is a known distortion function up to which the adversary is interested in X, in which case the relevant metric is the (exponent of the) probability of a successful guess. A

single-letter characterization of the highest achievable exponent is provided, and asymptotically-optimal strategies for both the primary user and the adversary are demonstrated.

Finally, the sample complexity of estimating maximal leakage from data is studied. It is shown that the task is possibly only if the minimum strictly positive probability of a source symbol, θ, is known. In that case, $O\left(\frac{|\mathcal{Y}|\log|\mathcal{X}|}{\theta}\right)$ samples are sufficient, and $\Omega(|\mathcal{Y}|^{1-\eta}/\theta)$ samples, for any $\eta > 0$, are necessary.

BIOGRAPHICAL SKETCH

Ibrahim Issa received his B.E. degree in Computer and Communications Engineering from the American University of Beirut, Lebanon, in 2012. He pursued his Ph.D in Electrical and Computer Engineering at Cornell, Ithaca, NY, where he was also awarded the M.S. degree. His research interests include information-theoretic security and quantum information theory. At Cornell, he was the recipient of the ECE Outstanding Thesis Research Award (2017), and of the Jacobs fellowship (2012-2013).

To my loving family,

whose endless support made this achievement possible.

ACKNOWLEDGEMENTS

I have been fortunate to have had a lot of support during my time at Cornell. Firstly, I would like to express my sincerest gratitude to my advisor, Prof. Aaron Wagner. My work has greatly benefited from his continuous support, his unique insights, and his great patience for teaching. His genuine enthusiasm for research, his attention to detail while maintaining an impressively clear view of the broader picture, his commitment to scientific communication, and his outstanding abilities as a lecturer are a source of inspiration to me. In addition to the academic training, Aaron was very patient and supportive during difficult times I had during my PhD. Being his student was a privilege and a pleasure, and I will always be grateful for his mentorship and friendship.

I would also like to thank my committee members, Prof. Suh, Prof. Wicker, and Prof. Acharya. They provided valuable guidance at various stages of my research, and our discussions have significantly improved my work.

I think most people who know me at Cornell (or on Facebook) know that I am part of a trio, consisting of me and two other wonderful people, Barbara and Raphael. Barbara reminded us that there is life outside engineering! Her support was and is invaluable and her presence brightened my time at Cornell. Raphael was my partner in crime from day one, and is my go-to person for academic and non-academic stuff alike. Mostly, I would like to thank them for letting me nag.

I would also like to thank my office mates, Nirmal, Omer, Yi, and Yuguang, for putting up with my singing. Sinem, Xiaoqing, Yiting, Sam, and Kia have made my time at Cornell much more enjoyable. I would be remiss not to mention the dedicated people I met while working with Cornell Graduate Students United, especially Alex, Maggie, and Michaela. I wish them luck!

My friends from Lebanon were also very supportive of me and kept me connected to home. I would like to especially thank Yasmine, Rafah, Abbas, Lara, Ali, and Mayssa.

Finally, and most importantly, I am forever indebted to my beautiful family: my parents, Hassan and Randa, my brothers, Ahmad and Adam, and my sister-in-law (and in-spirit), Zeinab. Your love, dedication, and support are the constant bright spot in my life. You have sacrificed so much for me and for my education. My time away from you has been difficult; but it would be worth it if it only makes you proud. This work is the fruit of your labor first, and mine second.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Any modern communication system has to satisfy strict security guarantees. Indeed, as communication networks proliferate and are frequently used to transmit sensitive information (such as banking information, medical data, private correspondences, etc.), security concerns become more central to the design of such systems, especially as attacks become more and more sophisticated. These concerns have traditionally been addressed by the cryptography community that made significant advances with cryptographic protocols such as the Advanced Encryption Standard (AES), and the RSA algorithm.

However, cryptography misses a salient feature of communication systems that proves highly compromising of the purported security guarantees: the existence of *side-channels*. A side-channel is an unconventional type of communication channel, which is defined to be any process that unintentionally and inevitably leaks information to an unauthorized user. Examples of side-channels are:

- When using the Secure Shell (SSH), after the initial handshake, each keystroke is sent immediately to the remote machine, as shown in Figure 1.1. When communicating over a wireless network, an eavesdropper can observe the timing of the packets and hence also the timing of the keystrokes, which are correlated with the input of the user (e.g., the inter-keystroke interval when typing 'a' followed by 'k' is significantly smaller

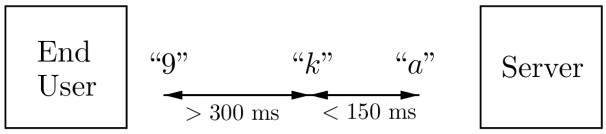

Figure 1.1: The Secure Shell: each keystroke is sent immediately to the remote machine

than when typing ‘k’ followed by ‘9’).

- Consider an on-chip network that has several processes running simultaneously, one of which is malicious. Because resources such as memory and buses are shared on the chip, the timing characteristics observed by the malicious application are affected by the behavior of the remaining applications. Similar phenomena occur when users share links or buffers in a communication network.
- In some implementations of RSA, the time needed to perform private key operations depends on the value of the key. A receiver that can observe the completion time of the algorithm can therefore glean some information about the key.
- The power consumption in some cryptographic devices depends on the value of the key. This is also true for the device’s electromagnetic emissions. A receiver that can observe either of these can therefore acquire information about the key.
- Wiretap channels.

Although at first glance such side-channels may seem innocuous, many works have shown that they in fact pose a significant security threat [9,24–27,46,52, 59]. For instance, Zhang and Wang [59] show how to use the keystroke timing (in SSH) to reduce the search space for passwords by a factor of at least 250.

Kocher [26] shows how to break (early implementations of) the RSA encryption using timing information. Ristenpart *et al.* [37] show how secret keys can be extracted from co-resident virtual machines on production Amazon EC2 servers through microarchitectural timing channels.

As these vulnerabilities exist at the physical layer level, information theory, as opposed to cryptography, is best suited to address them. For example, SSH is considered secure from a cryptography point of view (in fact, an improvement over FTP), which is oblivious to the existing side-channel. However, information theory has offered relatively little prescriptive advice for minimizing information leakage in side-channels, especially when compared with the success that the field has experienced in suggesting practical schemes for conventional communication systems. These successes include the use of multiple-antennas (MIMO), low-density parity-check codes (LDPCs), polar codes, orthogonal-frequency division multiplexing (OFDM), multiuser interference, and opportunistic communication.

The main difficulty in side-channel analysis is summarized by the following question: *how does one quantify how much information is "leaked" through a side-channel?* In mathematical terms, given two random variables X and Y, where X represents sensitive information and Y represents information available to an adversary:

How much information does Y leak about X?

If X and Y are independent, then the only reasonable answer is zero. Often in practice, Y cannot be made independent of X [28,30,31,35,46,47,49,52,57], in which case the answer is not obvious. One might be tempted to use mutual in-

formation as a leakage measure, as Shannon and many subsequent researchers have indeed done [10,18,20,28,39,43,57]. This choice, however, overlooks the context in which mutual information arises. In particular, the goal in security problems is different from compression and transmission problems, in which mutual information arises. It is also worth noting that in the latter contexts, mutual information arises as part of a computable characterization, not as part of the operational formulation of the problem. That is, the rate of transmission (or compression) is defined in terms of an operational engineering problem, and its computation involves mutual information. To adopt the latter as a leakage measure is to include it in the very formulation of the problem. As such, even though mutual information is not an unreasonable choice, there is no cogent, operational justification for its adoption in this context either.

1.2 Contributions

We describe a specific threat model and give an operational definition of leakage that is motivated by the setup of a *guessing* adversary. More specifically, upon observing Y, the adversary tries to guess a (possibly randomized) function of X. Leakage for a specific function is considered to be the logarithm of the ratio of the probability of a correct guess when Y is observed, to the probability of a correct guess when it is not (i.e., a blind guess). Maximal leakage, which we denote by $\mathcal{L}(X \to Y)$, is then defined as the maximum leakage over all such randomized functions (cf. Definition 1). This maximization, which is formally over discrete random variables U for which the Markov chain $U - X - Y$ holds, represents a worst-case analysis on the function of interest U, and models scenarios in which the conditional distribution $P_{U|X}$ is unknown. It is also inspired

by the strong data processing constant [4].

Although the maximization is an infinite-dimensional problem, $\mathcal{L}(X \to Y)$ admits a simple form (cf. Theorem 1). It turns out to equal the Sibson mutual information of order infinity $I_\infty(X;Y)$ [44,50] (cf. Corollary 1), endowing it with an operational significance. The resulting properties (cf. Corollary 2) are consistent with an axiomatic view of a leakage measure: it is zero if and only if X and Y are independent, it is not symmetric, it satisfies the data processing inequality, and it is additive over independent pairs $\{(X_i, Y_i)\}$. Significantly, it is shown that $\mathcal{L}(X \to Y) \geq I(X;Y)$, signifying that mutual information *underestimates* leakage (we further explore this fact in Chapter 3). Moreover, $\mathcal{L}(X \to Y)$ is *convex* in $P_{Y|X}$ for a fixed P_X, and depends on P_X only through its support.

Furthermore, we show that the definition of maximal leakage is robust in several respects. In the definition of $\mathcal{L}(X \to Y)$, we allow the adversary *one* guess only. A natural extension would be to allow for, say, k guesses for some integer k. This is particularly relevant for privacy problems. For example, if U is a password to some system, then an adversary is typically allowed several incorrect guesses before he/she is possibly locked out. We call the modified measure k-maximal leakage, and denote it by $\mathcal{L}^{(k)}(X \to Y)$. We show that, in fact, the two definitions are equivalent for all k (cf. Theorem 2). We also consider the case in which the adversary only needs the guess to be within a certain distance of the true function value, according to an arbitrary distance metric. We call this modified measure maximal locational leakage, and we denote it by $\mathcal{L}_{\mathcal{U}}(X \to Y)$. We show that $\mathcal{L}_{\mathcal{U}}(X \to Y) \leq \mathcal{L}(X \to Y)$, and equality holds under an unboundedness condition on the metric space $\mathcal{U}$ (cf. Theorem 3).

We further extend the notion of maximal leakage in two directions. We gen-

eralize the formula for maximal leakage to cover a large class of random variables (cf. Theorem 4), which includes point processes. Moreover, we propose a conditional form of maximal leakage, which attempts to answer the question: how much does Y leak about X when Z is given? We again provide an operational definition in the guessing framework (cf. Definition 4), and derive a simple form for $\mathcal{L}(X \to Y|Z)$ (cf. Theorem 5). Both the general and the conditional form retain the axiomatic properties of a leakage measure, and are lower-bounded by mutual information and conditional mutual information, respectively.

To summarize, we develop a measure of information leakage that is operationally motivated, simple to compute, robust to variations in the threat model, and lower-bounded by mutual information. Proofs and discussions of these results are the subject of Chapter 2.

Leakage Metrics in the Guessing Framework

Chapter 3 explores the connection between maximal leakage and existing metrics, with particular emphasis on mutual information, capacity, maximal correlation, local differential privacy, as well as rate-distortion-based metrics. We show that the threat model we consider illuminates important features of the given metrics.

We show that Shannon capacity corresponds to an adversary that is interested in functions of X that can be *reliably* recovered (cf. Definition 5 and Theorem 6). As such, capacity is *upper-bounded* by maximal leakage. As for maximal correlation, we show that it captures the change in the *variance* of functions of

X, after observing Y, as opposed to probabilities of correct guessing (cf. Definition 6 and Theorem 7). Local differential privacy captures the multiplicative increase of the guessing probability of functions of X, maximized over *realizations* of Y and over distributions P_X (cf. Theorem 9), hence it upper-bounds maximal leakage. Moreover, maximizing over realizations of Y for a fixed P_X yields a valid leakage measure, which is equal to the maximum information rate (cf. Theorem 8). We also propose a "dual" notion of maximal leakage in which an adversary attempts to minimize a (positive) cost function, rather than maximize a (positive) gain functions (cf. Definition 10 and Theorem 11). Finally, we reveal inadequacies of rate-distortion based approaches by considering them in an operational framework.

Application: Shannon Cipher System

Having established maximal leakage as the proper information leakage metric, we study an instance of a secrecy system using maximal leakage as performance metric. In particular, we consider the Shannon cipher system, shown in Figure 1.2. It consists of a transmitter and a legitimate receiver that are linked by a public noiseless channel and share a common key, and an eavesdropper who has access to the public channel and is aware of the source statistics and the used encryption schemes. The encryption schemes must allow the legitimate receiver to reconstruct the source sequence up to a fidelity constraint.

The objective, in this case, is to minimize the (normalized) maximal leakage between the source X^n and the public message M. We also introduce the information blurring system, which considers the case in which R is large and $r = 0$ and thus represents a stylized model of a side-channel. It is shown that rate-

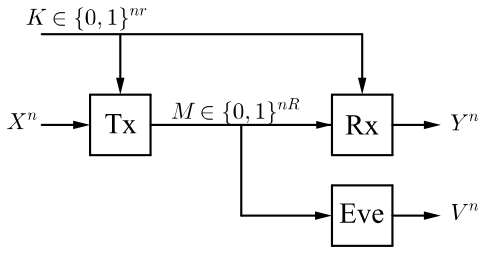

Figure 1.2: The Shannon cipher system with lossy communication.

distortion codes are asymptotically optimal, and the optimal limit is derived (cf. Theorems 13 and 14). Moreover, memoryless schemes are proven to be strictly suboptimal (cf. Lemma 6). One can interpret this result as: "maximal leakage favors quantization over adding noise". This is noteworthy since, in practice, many schemes resort to adding independent noise to guarantee privacy.

Finally, we consider a more "traditional" setup. That is, we assume the eavesdropper is interested in the source up to a known distortion function. We study this setup when the figure of merit is the (exponent) of the probability of a successful guess (i.e., a guess satisfying the distortion constraint). We show that the problem is related to source coding with side information, in which the transmitter's goal is to provide the eavesdropper with the "worst" side information. For a discrete memoryless source, we derive a single-letter characterization of the optimal exponent (cf. Theorem 16), and demonstrate asymptotically-optimal strategies for both the primary user and the eavesdropper.

Learning Complexity

In the final chapter of this dissertation, we consider the following natural question: can we estimate maximal leakage from data? In particular, how many samples do we need to estimate $\mathcal{L}(X \to Y)$ up to ϵ-accuracy, for a given $\epsilon > 0$?

We show that this task is only possible if we know the minimum strictly positive probability of a symbol $x \in \mathcal{X}$. That is, let $\theta = \min_{x \in \mathcal{X}: P_X(x) > 0} P_X(x)$. Then, we show that the number of needed samples, n, satisfies $n \geq \Omega(|\mathcal{Y}|^{1-\eta}/\theta)$, for any $\eta > 0$ (cf. Theorem 18). In particular, the lower bounds goes to infinity if θ goes to zero. The techniques to prove the lower bound draw heavily on the work of Renyi entropy estimation [1]. On the other hand, we show that $O\left(\frac{|\mathcal{Y}| \log |\mathcal{X}|}{\theta}\right)$ samples are sufficient (cf. Theorem 17).

1.3 Literature Overview

The literature on leakage and privacy measures is vast, spanning the fields of information theory, computer science, and computer security. The closest to our work comes from computer security [2,3,7,16,45]. In particular, Smith [45] defines leakage from X to Y as the logarithm of the multiplicative increase, upon observing Y, of the probability of guessing X *itself* correctly, neglecting that the adversary might be interested in certain functions of X. Braun *et al.* [7] consider a *worst case* modification of the metric, and maximize the previous quantity over all distributions on the alphabet of X (while $P_{Y|X}$ is fixed). The resulting quantity turns out to equal $\mathcal{L}(X \to Y)$. In the computer security literature, it is denoted by $ML(P_{Y|X})$, and its properties were further studied by Alvim *et al.* [3] and Espinoza and Smith [16]. The formula for maximal leakage is also derived in a different work [2]. Instead of looking at the (normalized) probability of guessing X, a gain function $g : \mathcal{X} \times \hat{\mathcal{X}} \to [0, 1]$ is introduced, and the normalized maximal gain is considered. Maximizing over g, for a fixed P_X, is shown to yield maximal leakage [2]. However, this threat model still focuses on X itself. In fact, this result was included merely as an additional computable formula,

and was de-emphasized by the authors [2].

Another connected line of work stems from cryptography, and in particular from the notion of *semantic security* [17] which considers the security of encryption schemes. First, Goldwasser and Micali [17] introduce the notion of "advantage" for a given function of the messages. It is the *additive* increase, upon observing the encrypted message (i.e., the ciphertext), of the probability of correctly guessing the value of the function. Semantic security then requires that, for an adversary that can work only for a polynomial (in the length of the message) amount of time, the advantage is negligible for all *deterministic* functions that are computable in polynomial time, and for all input distributions. Note that, in our framework, "advantage" is defined as the multiplicative increase. Since one is typically interested in securing hard-to-guess functions for which the probability of a correct guess is small, the multiplicative increase is arguably more descriptive of the change. It is also the more natural choice when viewing leakage in terms of leaked *bits*.

There are several variants of semantic security. In particular, entropic security [12,38] drops the computational bounds (on the adversary and the considered functions), but restricts its attention to input distributions with high min-entropy. Bellare *et al.* [6] introduce semantic security to the wiretap channel, and do not restrict it to computationally bounded adversaries, nor deterministic polynomial-time computable functions. For a given encryption scheme, they then upper and lower-bound the advantage of semantic security in terms of—what the authors call—mutual information security advantage, which is defined as the maximum, over all input distributions, of the mutual information between the message and the output of the channel whose input is the encryp-

tion of the message. Moreover, for discrete random variables X and Y, Calmon *et al.* [8] upper-bound the advantage over all deterministic functions in terms of their maximal correlation, which inspired Li and El Gamal [29] to use the latter quantity as a secrecy metric. Calmon *et al.* [8], inspired by the correspondence analysis literature [19], also generalized maximal correlation to k-correlation, which is defined as the sum of the k largest principal inertial components of the joint distribution P_{XY}.

Finally, other approaches to leakage metrics can be found in the information-theoretic literature. However, similarly to the choice of mutual information, they merely "borrow" information-theoretic metrics developed in other contexts, such as rate-distortion theory. As such, they lack a clear operational motivational, and in some cases, label obviously insecure systems as secure (see Section 3.6 for more details). These include expected distortion [58] incurred at the eavesdropper, expected minimum distortion in a list [41], expected number of guesses needed to satisfy a distortion constraint [33], the probability of satisfying the constraint [22,54], etc. Although the particular metrics differ among those works, they all assume that there is a priori known distortion function up to which the adversary is interested in the sensitive information X. For further discussion of privacy metrics, we refer the reader to Wagner and Eckhoff's work [51], which categorizes over eighty such metrics.

CHAPTER 2

MAXIMAL LEAKAGE

2.1 Threat Model and Definition

We give an operational definition of leakage that is motivated by the setup of a *guessing* adversary. More specifically, the adversary is interested in a (possibly randomized) function of X, called U. The distribution $P_{U|X}$ is unknown to us, and upon observing Y, the eavesdropper wants to guess U. To further illustrate this, consider the SSH example. Let X represent the actual keystroke timings, and Y represent the (perturbed) timings observed by the eavesdropper. The question we aim to answer is: how much information does Y leak about X? We will assume that the adversary is not interested in the keystroke timings themselves, but rather is interested in the input, which might represent a password, that generated those timings. Let U denote the password, and assume that the Markov chain $U - X - Y$ holds. It is reasonable to assume that the adversary has a verification mechanism for his/her guess $\hat{U}$ of the password, e.g., s/he could try logging in as the user. The quantity of interest is then the chance that the adversary guesses the password correctly after observing Y, namely $\mathbf{Pr}(U = \hat{U}(Y))$, where $\hat{U}(Y)$ is the best estimator of U given Y. The operational meaning of this quantity is clear: it is the chance that the adversary can break into the account (with one guess; the case of multiple guesses will be discussed later). Then, to understand the leakage due to Y, we should compare this quantity before and after observing Y, i.e., $\mathbf{Pr}(U = \hat{U}(Y))$ and $\max_u \mathbf{Pr}(U = u)$. Since for U's of interest (such as passwords) $\max_u \mathbf{Pr}(U = u)$ is small, we consider the *ratio* of these two quantities to be the appropriate description of the change (i.e., we consider

geometric gain instead of additive gain). Taking $\log_2$ of the ratio then yields an answer in bits: a leak of ℓ bits corresponds to a multiplicative increase in the correct guessing probability of 2^ℓ. The final, and key, step is to realize that the conditional distribution $P_{U|X}$ (i.e., the distribution of the passwords given the timings in this example) is not known. Therefore, we define maximal leakage as the mentioned ratio *maximized* over all $P_{U|X}$. This maximization can also be viewed as modeling the scenario in which we do not know which variable U is of interest to the adversary, and is inspired by the strong data processing inequality [4].

Definition 1 (Maximal Leakage) *Given a joint distribution P_{XY} on alphabets $\mathcal{X}$ and $\mathcal{Y}$, the* maximal leakage *from X to Y is defined as*

$$\mathcal{L}(X \to Y) = \sup_{U-X-Y-\hat{U}} \log \frac{\mathbf{Pr}\left(U = \hat{U}\right)}{\max_{u \in \mathcal{U}} P_U(u)}, \tag{2.1}$$

where the supremum is over all U and $\hat{U}$ taking values in the same finite, but arbitrary, alphabet.

2.2 Main Result

The optimization problem in (2.1) is infinite-dimensional, and it is not clear a priori that it is computable. In fact, one can show that it is impossible to bound the cardinality of the alphabet $\mathcal{U}$ in terms of the cardinalities of the alphabets $\mathcal{X}$ and $\mathcal{Y}$. Nonetheless, we can show that, maximal leakage is indeed computable and actually takes a simple form.

Theorem 1 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal leakage from X to Y is given by*

$$\mathcal{L}(X\to Y) = \log \sum_{y\in\mathcal{Y}} \max_{\substack{x\in\mathcal{X}:\\ P_X(x)>0}} P_{Y|X}(y|x).$$

Before proving the theorem, we investigate some of its consequences. First, it reveals two of the more useful aspects of maximal leakage from an engineering perspective: minimizing $\mathcal{L}(X\to Y)$ over $P_{Y|X}$, for a fixed support of P_X, amounts to minimizing a convex function, and $\mathcal{L}(X\to Y)$ depends on P_X only through its support. The latter fact is very useful because in practice P_X is typically complicated and outside our control. P_X is also typically used to model the adversary's prior knowledge of X, which is not necessarily known to us.

Moreover, the right-hand side of Theorem 1 is $I_\infty(X;Y)$ [44,50], The Sibson mutual information of order infinity. Sibson's $I_\alpha(X;Y)$ ($\alpha \geq 0$) is an extension of the concept of Renyi entropy $H_\alpha(X)$ (itself an extension of entropy) and Renyi divergence $D_\alpha(P\|Q)$. Although there are other possible extensions, Verdu [50] argues for the adoption of Sibson's definition. By endowing $I_\infty(X;Y)$ with an operational meaning, our result could be seen as also supporting that claim (more recently, $I_\infty(X;Y)$ has been used as a complexity measure in the study of communication complexity [34]). This equivalence is quite useful, so we state it as a separate Corollary.

Corollary 1 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$,*

$$\mathcal{L}(X\to Y) = I_\infty(X;Y),$$

where $I_\infty(X;Y)$ is the Sibson mutual information of order infinity.

For binary-valued X, say $\mathcal{X} = \{0, 1\}$, Sibson [44] showed that

$$I_\infty(X;Y) = \log_2\left(1 + \frac{1}{2}\|P_{Y|X}(\cdot|1) - P_{Y|X}(\cdot|0)\|\right) = 1 + \log_2\left(\frac{1}{2} + \frac{1}{4}\|P_{Y|X}(\cdot|1) - P_{Y|X}(\cdot|0)\|\right),$$

where $\|.\|_1$ is the L-1 distance. The term inside the $\log_2$ is the probability of success in binary hypothesis testing, which sheds light on why $I_\infty(X;Y)$ arises as maximal leakage. The following corollary summarizes some useful properties of $\mathcal{L}(X{\to}Y)$.

Corollary 2 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$,*

1. *(*Data Processing Inequality*) If the Markov chain $X - Y - Z$ holds for a discrete random variable Z, then $\mathcal{L}(X{\to}Z) \leq \min\{\mathcal{L}(X{\to}Y), \mathcal{L}(Y{\to}Z)\}$.*
2. $\mathcal{L}(X{\to}X) = H_0(X) = \log|\{x : P_X(x) > 0\}|$.
3. $\mathcal{L}(X{\to}Y) \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$.
4. $\mathcal{L}(X{\to}Y) \geq I(X;Y)$.
5. $\mathcal{L}(X{\to}Y) = 0$ *iff X and Y are independent.*
6. $\mathcal{L}(X{\to}Y)$ *is not symmetric in X and Y.*
7. *(*Additivity*) If $\{(X_i, Y_i)\}_{i=1}^{\ell}$ are mutually independent, then*

$$\mathcal{L}(X_1^\ell{\to}Y_1^\ell) = \sum_{i=1}^{\ell} \mathcal{L}(X_i{\to}Y_i).$$

8. $\exp\{\mathcal{L}(X{\to}Y)\}$ *is convex in $P_{Y|X}$ for fixed support of P_X.*

Proof: Properties 1) through 4), and 7) are shown for $I_\infty(X;Y)$ [44,50]. 5) follows from the definition and 4). That is, if X and Y are independent, $\mathcal{L}(X{\to}Y) = 0$ follows straightforwardly from the definition. Otherwise, we note that $\mathcal{L}(X{\to}Y) \geq$

$I(X;Y) > 0$. 6) is clear and is illustrated in Example 9 below. 8) follows from the fact that, for each $y \in \mathcal{Y}$, $\max_x P_{Y|X}(y|x)$ is convex in $P_{Y|X}$. ■

Note that properties 1), 5), and 7) can be regarded as axiomatic for a leakage measure, reinforcing the choice of maximal leakage. Moreover, Property 6) reveals a "weakness" in some suggested leakage metrics, including mutual information. In particular, there is no reason to expect a priori that X leaks about Y as much as Y leaks about X. Therefore, metrics that are symmetric by design miss that fact.

Property 4) is crucial and shows that a small maximal leakage is a more stringent requirement than a small mutual information. Since $\mathcal{L}(X \to Y)$ depends on P_X only through its support, it follows that maximal leakage is at least the Shannon capacity of the channel $P_{Y|X}$ when X has full support, and this inequality can be strict. This justifies the claim in the introduction that the Shannon capacity of a side-channel does not necessarily upper-bound its leakage. The maximization in the definition of maximal leakage hints at the reason why. In particular, Shannon capacity is concerned with (the size of) message sets that can be *reliably* reconstructed at the receiver, i.e., $\mathbf{Pr}(U = \hat{U}(Y)) \geq 1 - \epsilon$ for some small ϵ. Leakage, on the other hand, is concerned with the advantage in guessing, without any notion of reliability. This observation is made mathematically precise in Section 3.1. It is also worth noting that maximal leakage is upper bounded by local differential privacy [13], which is known to be too pessimistic (e.g., [13]). This is further discussed in Section 3.4.

Property 8) shows that minimizing maximal leakage, for a fixed support of P_X, amounts to minimizing a convex function. That is, one can efficiently solve the problem of finding the randomization mechanism $P_{Y|X}$ that minimizes max-

imal leakage, subject to a convex constraint.

We evaluate $\mathcal{L}(X \to Y)$ for some special cases.

Example 1 *If $X \sim Ber(q)$, $0 < q < 1$, and Y is the output of a BSC with parameter p, $0 \leq p \leq 1/2$, then $\mathcal{L}(X \to Y) = \log(2(1-p))$.*

Example 2 *If $X \sim Ber(q)$, $0 < q < 1$, and Y is the output of a BEC with parameter ϵ, $0 \leq \epsilon < 1$, then $\mathcal{L}(X \to Y) = \log(2-\epsilon)$, and $\mathcal{L}(Y \to X) = \log 2$.*

Example 3 *For any deterministic law $P_{Y|X}$, $\mathcal{L}(X \to Y) = \log |\{y : P_Y(y) > 0\}|$.*

Finally, as will be seen in the proof, it is worth noting that the conditional distribution $P_{U|X}$ that achieves the supremum in (2.1) depends on P_{XY} only through its $X-$marginal, P_X. In particular, $P_{U|X}$ is such that: for distinct x's, the supports of $P_{U|X=x}$'s are disjoint, and each $P_{U|X=x}$ effectively "shatters" the atom x into (almost) uniformly distributed u's to get an (almost) uniform marginal P_U. A special case to consider is the uniform P_X, in which case $P_{U|X}$ is simply the identity map. In light of this, one might wonder if there is always a deterministic map $P_{U|X}$ that achieves $\mathcal{L}(X \to Y)$. This is, however, not true in general. Suppose P_{XY} satisfies the following condition: there exists $x^\star \in \mathcal{X}$ such that for all $y \in \mathcal{Y}$, $P_{X|Y}(x^\star|y) \geq 1/2$. Then, for any deterministic function f, $f(x^\star)$ is always the optimal choice for the adversary, with and without the observation of Y. The above condition, however, is not sufficient for X and Y to be independent. That is, we can construct P_{XY} such that $\mathcal{L}(X \to Y) > 0$, whereas observing Y does not affect the probability of guessing any deterministic function of X.

2.2.1 Proof of Theorem 1

Assume, without loss of generality, that $P_X(x) > 0$ for all $x \in \mathcal{X}$. To show that $\mathcal{L}(X \to Y) \leq I_\infty(X;Y)$, consider any U satisfying $U - X - Y$. Let

$$\mathcal{L}(X \to Y)[U] = \log \frac{\sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} P_{UY}(u,y)}{\max_{u \in \mathcal{U}} P_U(u)}, \tag{2.2}$$

so that $\mathcal{L}(X \to Y) = \sup_{U : U - X - Y} \mathcal{L}(X \to Y)[U]$. Then,

$$\begin{aligned}
\sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} P_{UY}(u,y) &= \sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_X(x) P_{U|X}(u|x) P_{Y|X}(y|x) \\
&\leq \sum_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_X(x) P_{U|X}(u|x) \max_{x' \in \mathcal{X}} P_{Y|X}(y|x') \\
&= \sum_{y \in \mathcal{Y}} \left(\max_{x' \in \mathcal{X}} P_{Y|X}(y|x') \right) \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_X(x) P_{U|X}(u|x) \\
&= \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x) \max_{u \in \mathcal{U}} P_U(u).
\end{aligned}$$

Therefore, $\mathcal{L}(X \to Y)[U] \leq I_\infty(X;Y)$ for all $P_{U|X}$, hence $\mathcal{L}(X \to Y) \leq I_\infty(X;Y)$.

For the reverse inequality, we construct a $P_{U|X}$ for which $\mathcal{L}(X \to Y)[U] = I_\infty(X;Y)$, which we will call the "shattering" $P_{U|X}$. To that end, let $p^\star = \min_{x \in \mathcal{X}} P_X(x)$. For each $x \in \mathcal{X}$, let $k(x) = P_X(x)/p^\star$, and let $\mathcal{U} = \bigcup_{x \in \mathcal{X}} \{(x,1), (x,2), \ldots, (x, \lceil k(x) \rceil)\}$. For each $u = (i_u, j_u) \in \mathcal{U}$ and $x \in \mathcal{X}$, let $P_{U|X}(u|x)$ be:

$$P_{U|X}((i_u,j_u)|x) = \begin{cases} \frac{p^\star}{P_X(x)}, & i_u = x, \;\; 1 \leq j_u \leq \lfloor k(x) \rfloor, \\ 1 - \frac{(\lceil k(x) \rceil - 1) p^\star}{P_X(x)}, & i_u = x, \;\; j_u = \lceil k(x) \rceil, \\ 0, & i_u \neq x, \;\; 1 \leq j_u \leq \lceil k(i_u) \rceil. \end{cases} \tag{2.3}$$

Remark 1 *It is easy to check that if $\lfloor k(x) \rfloor = \lceil k(x) \rceil$, then the corresponding formulas are equal.*

Then, for each $((i_u, j_u), x) \in \mathcal{U} \times \mathcal{X}$,

$$P_{UX}((i_u, j_u), x) = \begin{cases} p^\star, & i_u = x, 1 \le j_u \le \lfloor k(x) \rfloor, \\ P_X(x) - (\lceil k(x) \rceil - 1)p^\star, & i_u = x, j_u = \lceil k(x) \rceil, \\ 0, & i_u \ne x, 1 \le j_u \le \lceil k(i_u) \rceil. \end{cases} \tag{2.4}$$

As mentioned earlier, the supports of $P_{U|X=x}$ are disjoint for distinct x's, and each x is effectively shattered into shards of probability $p^\star$. Now, note that

$$\max_{u \in \mathcal{U}} P_U(u) = \max_{(i_u, j_u) \in \mathcal{U}} P_{UX}((i_u, j_u), i_u) = p^\star. \tag{2.5}$$

Now, consider any $(u, y) \in \mathcal{U} \times \mathcal{Y}$. We have

$$P_{UY}((i_u, j_u), y) = \sum_{x \in \mathcal{X}} P_X(x) P_{U|X}((i_u, j_u)|x) P_{Y|X}(y|x) \tag{2.6}$$

$$= P_X(i_u) P_{U|X}((i_u, j_u)|i_u) P_{Y|X}(y|i_u) \tag{2.7}$$

$$= \begin{cases} p^\star P_{Y|X}(y|i_u), & 1 \le j_u \le \lfloor k(i_u) \rfloor, \\ (P_X(x) - (\lceil k(x) \rceil - 1)p^\star) P_{Y|X}(y|i_u), & j_u = \lceil k(i_u) \rceil. \end{cases} \tag{2.8}$$

Then, for a given $y \in \mathcal{Y}$,

$$\max_{(i_u, j_u) \in \mathcal{U}} P_{UY}((i_u, j_u), y) = \max_{(i_u, 1) \in \mathcal{U}} p^\star P_{Y|X}(y|i_u) = \max_{x \in \mathcal{X}} p^\star P_{Y|X}(y|x). \tag{2.9}$$

Finally, we get

$$\mathcal{L}(X \to Y) \ge \mathcal{L}(X \to Y)[U] = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{Y|X}(y|x),$$

where the inequality follows from the definition, and the equality follows from equations (2.2), (2.5), and (2.9). ■

2.3 Robustness of Maximal Leakage

We consider two natural variations on the definition of maximal leakage. The first allows the adversary multiple guesses, and the second allows for U's that

are continuous, in which case the adversary wants only to approximate U. In both cases, the resulting metric is unchanged.

2.3.1 Multiple Guesses

The definition of maximal leakage (Definition 1) allowed the adversary one guess. However, an adversary might be able to make several guesses in some practical scenarios. For example, if the adversary is trying to guess a password U of some system, s/he can typically try several passwords before s/he is locked out. We can modify the definition to allow for k guesses, for some integer k, as follows.

Definition 2 (k-Maximal Leakage) *Given a joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and a positive integer k, the k-maximal leakage from X to Y is defined as*

$$\mathcal{L}^{(k)}(X\to Y) = \sup_{U-X-Y-(\hat{U}_i)_{i=1}^k} \log \frac{\mathbf{Pr}\left(\bigvee_{i=1}^k U = \hat{U}_i\right)}{\max_{\substack{S\subseteq\mathcal{U}\\ |S|\leq k}} P_U(S)}.$$

Theorem 2 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and any $k \in \mathbb{N}$,*

$$\mathcal{L}^{(k)}(X\to Y) = \mathcal{L}(X\to Y).$$

Proof: To show $\mathcal{L}^{(k)}(X \to Y) \geq \mathcal{L}(X \to Y)$, we consider an arbitrary $P_{U|X}$ and construct $P_{V|X}$ such that $\mathcal{L}^{(k)}(X\to Y)[V] = \mathcal{L}(X\to Y)[U]$. In particular, for a given $P_{U|X}$ and associated alphabet $\mathcal{U}$, let

$$\mathcal{V} = \bigcup_{u\in\mathcal{U}} \{(u,1),(u,2),\ldots,(u,k)\}, \text{ and } P_{V|X}(v|x) = P_{V|X}((a_v,b_v)|x) = P_{U|X}(a_v|x)/k.$$

Then, the probability of correctly guessing V with k guesses after observing Y is:

$$\begin{aligned}\sup_{X-Y-(\hat{V}_i)_{i=1}^k} \mathbf{Pr}(V = \hat{V}_1 \vee \cdots \vee V = \hat{V}_k) &= \sum_{y\in\mathcal{Y}} \max_{\substack{v_1,v_2,\ldots,v_k \\ v_i\neq v_j, i\neq j}} \sum_{i=1}^{k} \sum_{x\in\mathcal{X}} P_X(x)P_{V|X}(v_i|x)P_{Y|X}(y|x) \\ &= \sum_{y\in\mathcal{Y}} \sum_{i=1}^{k} \max_{v_i\neq v_1,\ldots,v_{i-1}} \sum_{x\in\mathcal{X}} P_X(x)P_{V|X}(v_i|x)P_{Y|X}(y|x) \\ &\overset{(a)}{=} \sum_{y\in\mathcal{Y}} \max_{u} \sum_{x\in\mathcal{X}} P_X(x)P_{U|X}(u|x)P_{Y|X}(y|x), \quad (2.10)\end{aligned}$$

where (a) follows by setting $v_i = (u^\star, i)$, where

$$u^\star = \operatorname*{argmax}_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_X(x)P_{U|X}(u|x)P_{Y|X}(y|x).$$

Now, note that (2.10) is simply the probability of guessing U correctly with a single guess after observing Y. A similar argument shows that, with no Y observation, the probability of guessing V correctly with k guesses is equal to the probability of guessing U correctly with a single guess, hence $\mathcal{L}^{(k)}(X\to Y)[V] = \mathcal{L}(X\to Y)[U]$, which establishes $\mathcal{L}^{(k)}(X\to Y) \geq \mathcal{L}(X\to Y)$.

It remains to show $\mathcal{L}(X\to Y) \geq \mathcal{L}^{(k)}(X\to Y)$. For any $P_{V|X}$, we construct $P_{U|X}$ such that $\mathcal{L}(X\to Y)[U] = \mathcal{L}^{(k)}(X\to Y)[V]$. So let $P_{V|X}$ be given, with associated alphabet $\mathcal{V}$, and let $\ell \triangleq |\mathcal{V}| \geq k$. Now, let

$$\mathcal{U} = \{S \subset \mathcal{V} : |S| = k\}, \text{ and } P_{U|X}(u|x) = c \sum_{v\in u} P_{V|X}(v|x),$$

where $c = 1/\binom{\ell-1}{k-1}$. Then, observing Y, the probability of guessing U correctly with a single guess is

$$\begin{aligned}\sup_{X-Y-\hat{U}} \mathbf{Pr}(U = \hat{U}) &= \sum_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_X(x)P_{U|X}(u|x)P_{Y|X}(y|x) \\ &= \sum_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_X(x) \sum_{v\in u} P_{V|X}(v|x)P_{Y|X}(y|x)c \\ &= c\sum_{y\in\mathcal{Y}} \max_{\substack{v_1,v_2,\ldots,v_k \\ v_i\neq v_j, i\neq j}} \sum_{x\in\mathcal{X}} \sum_{i=1}^{k} P_X(x)P_{V|X}(v_i|x)P_{Y|X}(y|x),\end{aligned}$$

which is the probability, normalized by c, of guessing V correctly with k guesses after observing Y. A similar argument shows that, with no Y observation, the probability of guessing U correctly with a single guess is equal to the probability, normalized by c, of guessing V correctly with k guesses, hence $\mathcal{L}(X \to Y)[V] = \mathcal{L}^{(k)}(X \to Y)[U]$, which establishes $\mathcal{L}(X \to Y) \geq \mathcal{L}^{(k)}(X \to Y)$. ■

2.3.2 Approximate Guessing

We consider the case in which the adversary only needs the guess to be within a certain distance of the true function value, according to a given distance metric. As such, the random variable U, over which we are optimizing, now lives in a given metric space $\mathcal{U}$ and is no longer restricted to be discrete. We call this modified measure maximal locational leakage. The term "locational" is motivated by the scenario in which the variable of interest U is a geographical location.

Definition 3 (Maximal Locational Leakage) *Given a joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and a metric space $\mathcal{U}$ (with its associated Borel σ-field), the maximal locational leakage from X to Y is defined as*

$$\mathcal{L}_{\mathcal{U}}(X \to Y) = \sup_{\substack{U: U - X - Y \\ \exists u: \mathbf{Pr}(U \in B(u)) > 0}} \log \frac{\sup_{\hat{u}(.)} \mathbf{Pr}(U \in B(\hat{u}(Y)))}{\sup_{\hat{u}} \mathbf{Pr}(U \in B(\hat{u}))}, \tag{2.11}$$

where $B(u)$ is the closed unit ball centered at $u \in \mathcal{U}$.

Theorem 3 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and any metric space $\mathcal{U}$,*

$$\mathcal{L}_{\mathcal{U}}(X \to Y) \leq \mathcal{L}(X \to Y),$$

with equality if $\mathcal{U}$ has a countably infinite subset S such that no pair of its elements can be contained in a single unit ball.

Proof: Consider any U and $\hat{u}(Y)$ in the maximization of (2.11):

$$\begin{aligned}
\mathbf{Pr}(U \in B(\hat{u}(Y)) &\leq \sum_{y\in\mathcal{Y}} \sup_{u\in\mathcal{U}} P(U \in B(u), Y = y) \\
&= \sum_{y\in\mathcal{Y}} \sup_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P(U \in B(u), X = x, Y = y) \\
&= \sum_{y\in\mathcal{Y}} \sup_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P(U \in B(u))P(X = x|U \in B(u))P_{Y|X}(y|x) \\
&\leq \sum_{y\in\mathcal{Y}} \sup_{u\in\mathcal{U}} P(U \in B(u)) \sup_{x\in\mathcal{X}} p_{Y|X}(y|x) \\
&= \left[\sum_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} p_{Y|X}(y|x)\right] \sup_{u\in\mathcal{U}} P(U \in B(u)).
\end{aligned}$$

Therefore,

$$\mathcal{L}_{\mathcal{U}}(X{\to}Y) \leq \log \sum_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} P_{Y|X}(y|x) = \mathcal{L}(X{\to}Y).$$

If $\mathcal{U}$ satisfies the given condition (e.g., $\mathcal{U}$ is unbounded), then exact guessing of discrete functions can be simulated by choosing S to be the support of U. Hence $\mathcal{L}_{\mathcal{U}}(X{\to}Y) \geq \mathcal{L}(X{\to}Y)$, which implies the equality. ■

2.4 Extensions

We extend the notion of maximal leakage in two ways. We generalize the formula for maximal leakage, beyond the discrete case, to cover a larger class of random variables, including point processes. Moreover, we investigate a conditional version of the problem. That is, we consider the question: how much does Y leak about X when Z is known?

2.4.1 General Formula

Note that the definition of maximal leakage (Definition 1) is not restricted to discrete X and Y, but Theorem 1 is. Moreover, the input and output of side-channels is not necessarily discrete, as in the SSH and microarchitectural timing channels in which the input is better modeled as a point process. Hence, a more general form is needed.

Before stating the theorem, we introduce the following notation. For a given probability distribution P_X, and a measurable function $f : \mathcal{X} \to \mathbb{R}$, we define the *essential supremum* of f with respect to P_X as follows:

$$\text{ess-sup}_{P_X} f(x) = \inf\{\alpha : P_X(\{x : f(x) > \alpha\}) = 0\}. \tag{2.12}$$

Equivalently,

$$\text{ess-sup}_{P_X} f(x) = \sup\{\beta : P_X(\{x : f(x) > \beta\}) > 0\}. \tag{2.13}$$

To verify the equivalence, let A be the set on the right-hand side of (2.12) with $a = \inf A$, and B be the set on the right-hand side of (2.13) with $b = \sup B$. It follows immediately from the definitions of the sets that $a \geq b$. Now consider $r > b$. Then, $r \notin B \Rightarrow P_X(\{x : f(x) > r\}) = 0 \Rightarrow r \in A \Rightarrow r \geq a$. As such, $r > b \Rightarrow r \geq a$, which implies $b \geq a$. Hence, $a = b$.

Theorem 4 *Let $(\mathcal{X} \times \mathcal{Y}, \sigma_{XY}, P_{XY})$ be a probability space with associated probability spaces $(\mathcal{X}, \sigma_X, P_X)$ and $(\mathcal{Y}, \sigma_Y, P_Y)$.*

1. *If $P_{XY} \ll P_X \times P_Y$ and σ_X is generated by a countable set, then*

$$\mathcal{L}(X \to Y) = \log \int_{\mathcal{Y}} \text{ess-sup}_{P_X} f(x, y) P_Y(dy), \tag{2.14}$$

where $f(x, y) = \frac{dP_{XY}}{d(P_X \times P_Y)}(x, y)$.

2. *If absolute continuity fails, then* $\mathcal{L}(X \to Y) = +\infty$.

We defer the proof to the end of the section and discuss some implications and examples of the theorem. As in the discrete case, the general formula satisfies the desirable properties of a leakage measure, and is lower-bounded by $I(X;Y)$.

Corollary 3 *Let* $(\mathcal{X} \times \mathcal{Y}, \sigma_{XY}, P_{XY})$ *be a probability space with associated probability spaces* $(\mathcal{X}, \sigma_X, P_X)$ *and* $(\mathcal{Y}, \sigma_Y, P_Y)$. *Assume* $P_{XY} \ll P_X \times P_Y$ *and* σ_X *is generated by a countable set. Then,*

1. *(*Data Processing Inequality*) If the Markov chain* $X - Y - Z$ *holds for a random variable* Z, *then* $\mathcal{L}(X \to Z) \leq \min\{\mathcal{L}(X \to Y), \mathcal{L}(Y \to Z)\}$.
2. $\mathcal{L}(X \to Y) \geq I(X;Y)$.
3. $\mathcal{L}(X \to Y) = 0$ *iff* X *and* Y *are independent.*
4. *(*Additivity*) If* $\{(X_i, Y_i)\}_{i=1}^n$ *are mutually independent, then*

$$\mathcal{L}(X_1^n \to Y_1^n) = \sum_{i=1}^n \mathcal{L}(X_i \to Y_i).$$

Proof: The data processing inequality follows directly from the definition (as shown in Lemma 1 below). To verify 2), consider the following.

$$\begin{aligned} I(X;Y) = \mathbf{E}[\log f(X,Y)] &\overset{(a)}{\leq} \log \mathbf{E}[f(X,Y)] \\ &= \log \int_{\mathcal{Y}} \int_{\mathcal{X}} f^2(x,y) P_X(dx) P_Y(dy) \\ &\leq \log \int_{\mathcal{Y}} (\text{ess-sup}_{P_X} f(x,y)) \int_{\mathcal{X}} f(x,y) P_X(dx) P_Y(dy) \\ &\overset{(b)}{=} \mathcal{L}(X \to Y), \end{aligned}$$

where (a) follows from Jensen's inequality, and (b) follows from the fact $\int_{\mathcal{X}} f(x,y)P_X(dx) = 1$. 3) follows from the definition and 2). 4) follows from the fact that, if (X_1, Y_1) is independent of (X_2, Y_2), then $f(x_1, x_2, y_1, y_2) = f(x_1, y_1)f(x_2, y_2)$. ■

Note that Theorem 4 cover all the combinations of discrete, countable, or continuous random variables X and Y.

Corollary 4 *If X and Y are jointly continuous real random variables,*

$$\mathcal{L}(X\to Y) = \log \int_{\mathbb{R}} \sup_{x:p_X(x)>0} p_{Y|X}(y|x)dy, \tag{2.15}$$

where p_X and $p_{Y|X}(.|.)$ are the marginal pdf of X and the conditional pdf of Y given X, respectively.

Example 4 *If X and Y are jointly Gaussian, then*

$$\mathcal{L}(X\to Y) = \begin{cases} 0, & \text{if } X \text{ and } Y \text{ are independent,} \\ +\infty, & \text{otherwise.} \end{cases}$$

Example 5 *Suppose X is real and its pdf satisfies $p_X(x) > 0$ for all $x \in \mathbb{R}$. Let $Y = X+Z$, where Z is a continuous real random variable independent of X. Let $z_0 = \operatorname{argmax} p_Z(z)$. Then,*

$$\mathcal{L}(X\to Y) = \log \int_{\mathbb{R}} \sup_{x\in\mathbb{R}} p_{Y|X}(y|x)dy = \log \int_{\mathbb{R}} \sup_{x} p_Z(y-x)dy = \log \int_{\mathbb{R}} p_Z(z_0)dy = +\infty.$$

The above examples suggest that "adding independent noise" is not necessarily secure in the maximal leakage sense. A similar phenomenon will be seen in the context of the Shannon cipher system (cf. Section 4.5), in which quantization arises as an optimal mechanism, whereas memorylessly adding noise is shown to be strictly suboptimal. The following example further illustrates the point.

Example 6 *Fix $T \in \mathbb{R}_+$. Let Ω_T be the set of all counting functions on $[0,T]$ and let $\{\mathcal{F}_t\}_{t=0}^T$ be the filtration over Ω_T generated by the mapping $\omega \mapsto \omega_t$. Let X_0^T be a Poisson process of rate λ. Let Y_0^T be the output of an (initially empty) exponential server queue with rate μ and input X_0^T. Then,*

$$\frac{1}{T}\mathcal{L}(X_0^T \to Y_0^T) = \mu.$$

Proof: *Let P_0 be the probability measure on $(\Omega_T, \mathcal{F}_T)$ under which the output is distributed as a Poisson process of rate one. It is known that, for $(x,y) \in \Omega_T \times \Omega_T$,*

$$\frac{dP_{XY}}{dP_X \times P_0}(x,y) = \exp\left[\int_0^T \log(\mu \mathbb{I}(x_t > y_{t-}))dy_t + \int_0^T (1 - \mu\mathbb{I}(x_t > y_t))dt\right] =: L(x,y).$$

Now, note that,

$$\frac{dP_{XY}}{dP_X \times P_Y} dP_Y = \frac{dP_{XY}}{dP_X \times P_Y}\frac{dP_X \times P_0}{dP_X \times P_0}\frac{dP_Y}{dP_0}dP_0 = L dP_0,$$

so that

$$\mathcal{L}(X_0^T \to Y_0^T) = \int_{\Omega_T} \textit{ess-sup}_{P_X} L(x,y) P_0(dy).$$

It is easy to check that the first term in $L(x,y)$ is equal to $y_T \log\mu$, and the second term can be made arbitrarily close to T under the sup. By noting that y_T is distributed as Poi(T) *under P_0, we get*

$$\frac{1}{T}\mathcal{L}(X_0^T \to Y_0^T) = \frac{1}{T}\log\int_{\Omega_T} \exp\left[y_T \log\mu + T\right] P_0(dy) = \frac{1}{T}\log\exp[T(e^{\log\mu} - 1) + T] = \mu.$$

■

Proof of Theorem 4: Proof of 1): To show that the left-hand side upper-

bounds the right-hand side, fix any $P_{U|X}$, and consider the following

$$\begin{aligned}
\sup_{\hat{U}(Y)} \mathbf{Pr}(U = \hat{U}(Y)) &= \int_{\mathcal{Y}} \max_{u\in\mathcal{U}} \int_{\mathcal{X}} P_{U|X}(u|x) P_{XY}(dxdy) \\
&= \int_{\mathcal{Y}} \max_{u\in\mathcal{U}} \int_{\mathcal{X}} P_{U|X}(u|x) f(x,y) P_X(dx) P_Y(dy) \\
&\leq \int_{\mathcal{Y}} \max_{u\in\mathcal{U}} \int_{\mathcal{X}} P_{U|X}(u|x) (\sup_{P_X} f(x,y)) P_X(dx) P_Y(dy) \\
&= \int_{\mathcal{Y}} (\sup_{P_X} f(x,y)) \left(\max_{u\in\mathcal{U}} \int_{\mathcal{X}} P_{U|X}(u|x) P_X(dx) \right) P_Y(dy) \\
&= (\max_{u\in\mathcal{U}} P_U(u)) \int_{\mathcal{Y}} (\sup_{P_X} f(x,y)) P_Y(dy).
\end{aligned}$$

To show the reverse direction, we will show it first for discrete X, and extend the result by discretizing more general X's. The argument uses the data processing inequality, which we need to prove at the outset.

Lemma 1 *If the Markov chain $X - Y - Z$ holds, then $\mathcal{L}(X \to Z) \leq \min\{\mathcal{L}(X\to Y), \mathcal{L}(Y\to Z)\}$.*

Proof: Consider any U such that $U - X - Z$ holds. Then, by marginalizing over X, we can construct V such that $V - Y - Z$ holds and $P_{UZ} = P_{VZ}$. Hence, $\mathcal{L}(X\to Z) \leq \mathcal{L}(Y\to Z)$. Moreover, we can construct W such that $W - X - Y - Z$ holds and $P_{WXZ} = P_{UXZ}$. Clearly, $\mathbf{Pr}(W = \hat{W}(Y)) \geq \mathbf{Pr}(W = \hat{W}(Z))$. Since $\mathbf{Pr}(W = \hat{W}(Z)) = \mathbf{Pr}(U = \hat{U}(Z))$, we get $\mathcal{L}(X\to Z) \leq \mathcal{L}(X\to Y)$. ■

Now, suppose X has finite alphabet. Note that, in this case, $\sigma(X)$ is generated by a finite set, and $P_{XY} \ll P_X \times P_Y$ since $I(X;Y) \leq H(X) < \infty$. Without loss of generality, suppose X has full support. Consider the "shattering" $P_{U|X}$. Recall: $p^\star = \min_{x\in\mathcal{X}} P_X(x)$. For each $x \in \mathcal{X}$, let $k(x) = P_X(x)/p^\star$, and let $\mathcal{U} = \bigcup_{x\in\mathcal{X}} \{(x,1),(x,2),\ldots,(x,\lceil k(x)\rceil)\}$. For each $u = (i_u, j_u) \in \mathcal{U}$ and $x \in \mathcal{X}$, let

$P_{U|X}(u|x)$ be:

$$P_{U|X}((i_u, j_u)|x) = \begin{cases} \frac{p^\star}{P_X(x)}, & i_u = x, \;\; 1 \le j_u \le \lfloor k(x) \rfloor, \\ 1 - \frac{(\lceil k(x) \rceil - 1)p^\star}{P_X(x)}, & i_u = x, \;\; j_u = \lceil k(x) \rceil, \\ 0, & i_u \neq x, \;\; 1 \le j_u \le \lceil k(i_u) \rceil. \end{cases}$$

Then,

$$\begin{aligned} \sup_{\hat{U}(Y)} \mathbf{Pr}(U = \hat{U}(Y)) &= \int_{\mathcal{Y}} \max_{(i_u, j_u) \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{U|X}((i_u, j_u)|x) f(x, y) P_X(x) P_Y(dy) \\ &= \int_{\mathcal{Y}} \max_{(i_u, 1) \in \mathcal{U}} p^\star f(i_u, y) P_Y(dy) \\ &= p^\star \int_{\mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y) P_Y(dy). \end{aligned}$$

The proof for the discrete case is completed by noticing that $p^\star = \max_u P_U(u)$.

Now, consider the more general case. Let $\{A_n\}_{n=1}^\infty$ be a countable collection of sets generating $\sigma(X)$. We will prove the result by considering a series of discretizations of X, each of which is a refinement of the previous one. To that end, let $\mathcal{S}_n$ be a collection of sets such that

$$\sigma(\mathcal{S}_n) = \sigma\left(\bigcup_{i=1}^n A_i\right), \text{ and for all } S_i, S_j \in \mathcal{S}_n, i \neq j, S_i \cap S_j = \varnothing.$$

In particular, every finite sigma algebra is generated by a finite partition. So let $\mathcal{S}_n$ be the finite partition generating $\sigma(\cup_{i=1}^n A_i)$. It can be readily verified that $\mathcal{S}_{n+1}$ is a refinement of $\mathcal{S}_n$. Let $N_n = |\mathcal{S}_n|$, $\mathcal{S}_n = \{S_{n,1}, S_{n,2}, \cdots, S_{n,N_n}\}$, and define

$$U_n(X) = \sum_{i=1}^{N_n} i \, \mathbb{I}\{X \in S_{n,i}\}.$$

Then, we get $\mathcal{L}(X \to Y) \ge \mathcal{L}(U_n \to Y)$ since $U_n - X - Y$ is a Markov chain, and the data processing inequality holds by Lemma 1. By the previous result for finite X, we get

$$\mathcal{L}(U_n \to Y) = \log \int_{\mathcal{Y}} \sup_{u : P_{U_n}(u_n) > 0} f_n(u_n, y) P_Y(dy),$$

where $f_n(u_n, y) = \frac{dP_{U_n Y}}{d(P_{U_n} \times P_Y)}$. We need to find $f_n(u_n, y)$. Let $A \subseteq \mathcal{U}_n \times \mathcal{Y}$, then

$$\begin{aligned} P_{U_n,Y}(A) &= \int_{\mathcal{Y}} \sum_{u_n} \mathbb{I}\{(u_n, y) \in A\} \int_{\mathcal{X}} P_{U_n|X}(u_n|x) f(x, y) P_X(dx) P_Y(dy) \\ &= \int_{\mathcal{Y}} \sum_{u_n} \mathbb{I}\{(u_n, y) \in A\} \left(\int_{S_{n,u_n}} f(x, y) P_X(dx) \right) P_Y(dy) \\ &= \int_{\mathcal{Y}} \sum_{u_n : P_{U_n}(u_n) > 0} \mathbb{I}\{(u_n, y) \in A\} \left(\frac{\int_{S_{n,u_n}} f(x, y) P_X(dx)}{\int_{S_{n,u_n}} P_X(dx)} \right) P_{U_n}(u_n) P_Y(dy), \end{aligned}$$

so that

$$f_n(u_n, y) = \frac{\int_{S_{n,u_n}} f(x, y) P_X(dx)}{\int_{S_{n,u_n}} P_X(dx)}.$$

Let $S_n(x)$ be the set in $\mathcal{S}_n$ containing x. Then, we can view $f_n(u_n, y)$ as a function of (x, y):

$$f_n(x, y) = \frac{\int_{S_n(x)} f(x, y) P_X(dx)}{\int_{S_n(x)} P_X(dx)}.$$

We can rewrite $f_n(x, y) = \mathbf{E}[f(X, y)|X \in S_n(x)]$, so that

$$f_n(X, y) = \mathbf{E}[f(X, y)|\sigma(\mathcal{S}_n)]. \tag{2.16}$$

Since $\mathcal{S}_n$'s are refinements, then $f_n(X, y)$ is a martingale process, and it follows by Levy's upward Theorem [56, Theorem 14.2] that

$$f_n(X, y) \overset{\text{a.s.}}{\to} \mathbf{E}\left[f(X, y)|\sigma\left(\cup_{i=1}^{\infty} \mathcal{S}_i\right)\right]. \tag{2.17}$$

Then,

$$\mathbf{E}\left[f(X, y)|\sigma\left(\cup_{i=1}^{\infty} \mathcal{S}_i\right)\right] = \mathbf{E}\left[f(X, y)|\sigma\left(\cup_{i=1}^{\infty} A_i\right)\right] = \mathbf{E}\left[f(X, y)|\sigma\left(X\right)\right] \overset{\text{a.s.}}{=} f(X, y). \tag{2.18}$$

Moreover,

$$\mathcal{L}(X \to Y) \geq \lim_{n \to \infty} \mathcal{L}(U_n \to Y) = \lim_{n \to \infty} \log \int_{\mathcal{Y}} \sup_{u : P_{U_n}(u_n) > 0} f_n(u_n, y) P_Y(dy) \tag{2.19}$$

$$= \lim_{n \to \infty} \log \int_{\mathcal{Y}} \sup_{x : P_X(S_n(x)) > 0} f_n(x, y) P_Y(dy). \tag{2.20}$$

Since $\mathcal{S}_{n+1}$ is a refinement of $\mathcal{S}_n$, the integrand is increasing. Therefore, by the monotone convergence theorem,

$$\mathcal{L}(X \to Y) \geq \log \int_{\mathcal{Y}} \lim_{n\to\infty} \sup_{x:P_X(S_n(x))>0} f_n(x,y) P_Y(dy). \tag{2.21}$$

Then, it remains to show that

$$\lim_{n\to\infty} \sup_{x:P_X(S_n(x))>0} f_n(x,y) \geq \text{ess-sup}_{P_X} f(x,y). \tag{2.22}$$

To that end, let $B = \{N : P_X(f(X,y) > N) > 0\}$. Consider $r \in B$ and let $E_r = \{x : f(x,y) > r\}$. Then, $P_X(E_r) > 0$. Therefore, by (2.17) and (2.18), $f_n(X,y)$ converges almost everywhere to $f(X,y)$ on E_r. By Egoroff's Theorem [5, Theorem 7.12], for every $\delta > 0$, there exists E'_δ such that f_n converges uniformly to f on $E_r \backslash E'_\delta$. Call the latter set $E_{r\backslash\delta}$. So fix $\delta > 0$ small such that $P_X(E_{r\backslash\delta}) > 0$. For each n, let $\mathcal{S}_n(E_{r\backslash\delta})$ be a collection of sets in $\mathcal{S}_n$ satisfying: $\cup_{S\in\mathcal{S}_n(E_{r\backslash\delta})} \supseteq E_{r\backslash\delta}$ and $S \in \mathcal{S}_n(E_{r\backslash\delta}) \Rightarrow S \cap E_{r\backslash\delta} \neq \varnothing$. Then there must exist $S \in \mathcal{S}_n(E_{r\backslash\delta})$ satisfying $P(S) > 0$. Denote the latter set by $S_n(E_{r\backslash\delta})$. Hence,

$$\begin{aligned}
\lim_{n\to\infty} \sup_{x:P_X(S_n(x))>0} f_n(x,y) &\overset{(a)}{\geq} \lim_{n\to\infty} \sup_{x\in S_n(E_{r\backslash\delta})} f_n(x,y) && (2.23)\\
&\overset{(b)}{\geq} \lim_{n\to\infty} \inf_{x\in E_{r\backslash\delta}} f_n(x,y) && (2.24)\\
&\overset{(c)}{=} \inf_{x\in E_{r\backslash\delta}} f(x,y) && (2.25)\\
&\geq r, && (2.26)
\end{aligned}$$

where (a) follows from the fact that $P_X(S_n(E_{r\backslash\delta}) > 0$, (b) follows from the fact that $S_n(E_{r\backslash\delta}) \cap E_{r\backslash\delta} \neq \varnothing$, and (c) follows from the fact that $f_n(x,y)$ converges uniformly to f on $E_{r\backslash\delta}$. Finally, since r was chosen arbitrarily from B, we get

$$\lim_{n\to\infty} \sup_{x:P_X(S_n(x))>0} f_n(x,y) \geq \sup B = \text{ess-sup}_{P_X} f(x,y), \tag{2.27}$$

as desired. ■

Proof of 2): Note that if absolute continuity does not hold, $I(X;Y) = +\infty$. Then, there exists a sequence of discretizations (X_n, Y_n) such that $I(X_n;Y_n) \to +\infty$. The result then follows by noting that $\mathcal{L}(X \to Y) \geq \mathcal{L}(X_n \to Y_n) \geq I(X_n;Y_n)$. ■

2.4.2 Conditional Maximal Leakage

We consider a natural extension of maximal leakage. In particular, how much does Y leak about X when Z is known? We define conditional maximal leakage analogously to Definition 1.

Definition 4 (Conditional Maximal Leakage) *Given a joint distribution P_{XYZ} on alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, the* conditional maximal leakage *from X to Y given Z is defined as*

$$\mathcal{L}(X \to Y|Z) = \sup_{U:U-X-Y|Z} \log \frac{\mathbf{Pr}(U = \hat{U}(Y,Z))}{\mathbf{Pr}(U = \tilde{U}(Z))}, \tag{2.28}$$

where U takes values in a finite, but arbitrary, alphabet, and $\hat{U}(Y,Z)$ and $\tilde{U}(Z)$ are the optimal (i.e., MAP) estimators of U given (Y,Z) and Z, respectively.

Remark 2 *The Markov chain $U - X - Y|Z$ is equivalent to $U - (X,Z) - Y$.*

In the remainder, assume, without loss of generality, that X and Z have full support.

Theorem 5 *Given a joint distribution P_{XYZ} on finite alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, the* conditional maximal leakage *from X to Y given Z is given by*

$$\mathcal{L}(X \to Y|Z) = \log \left(\max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z) \right). \tag{2.29}$$

In other terms, $\mathcal{L}(X \to Y|Z) = \max_z \mathcal{L}(X \to Y|Z = z)$, where the latter term is interpreted as the unconditional maximal leakage evaluated with respect to the joint $P_{XY|Z=z}$.

Proof: To show that the left-hand side is upper-bounded by the right-hand side, fix $P_{U|XZ}$ and consider the following.

$$\frac{\mathbf{Pr}(U = \hat{U}(Y,Z))}{\mathbf{Pr}(U = \tilde{U}(Z))} = \frac{\sum_z p(z) \sum_y p(y|z) \max_u p(u|y,z)}{\sum_z p(z) \max_u p(u|z)} \leq \max_z \frac{\sum_y p(y|z) \max_u p(u|y,z)}{\max_u p(u|z)}.$$

Then, by noting that term being maximized is $\exp\{\mathcal{L}(X \to Y|Z = z)\}$, we get

$$\sup_{U:U-(X,Z)-Y} \log \frac{\mathbf{Pr}(U = \hat{U}(Y,Z))}{\mathbf{Pr}(U = \tilde{U}(Z))} \leq \max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z).$$

To get the lower bound, let $\epsilon_n = 1/n$ for $n \in \mathbb{N}$, $z^\star \in \operatorname{argmax} \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z)$, and $p^\star = \min_{x:p(x|z^\star)>0} p(x|z^\star)$. Construct $P_{U|XZ}$ as follows. If $Z = z^\star$, then $P_{U|X,Z=z^\star}$ is the "shattering" conditional with respect to the distribution $P_{X|Z=z^\star}$. If $Z \neq z^\star$, then $U \sim Unif([n])$, independently of X. We get

$$\begin{aligned}
\frac{\mathbf{Pr}(U = \hat{U}(Y,Z))}{\mathbf{Pr}(U = \tilde{U}(Z))} &= \frac{\sum_{z\neq z^\star} p(z) \sum_y p(y|z) \max_u p(u|y,z) + p(z^\star) \sum_y p(y|z^\star) \max_u p(u|y,z^\star)}{\sum_{z\neq z^\star} p(z) \max_u p(u|z) + p(z^\star) \max_u p(u|z^\star)} \\
&= \frac{\sum_{z\neq z^\star} p(z)\epsilon_n + p(z^\star)p^\star \sum_y \max_{x:P_{X|Z}(x|z^\star)>0} P_{Y|XZ}(y|x,z^\star)}{\sum_{z\neq z^\star} p(z)\epsilon_n + p(z^\star)p^\star} \\
&= \frac{(1-p(z^\star))\epsilon_n + p(z^\star)p^\star \sum_y \max_{x:P_{X|Z}(x|z^\star)>0} P_{Y|XZ}(y|x,z^\star)}{(1-p(z^\star))\epsilon_n + p(z^\star)p^\star}
\end{aligned}$$

Letting $n \to \infty$ (i.e., $\epsilon_n \to 0$) yields our lower bound. ■

The following corollary summarizes important properties of conditional maximal leakage.

Corollary 5 *Given a joint distribution P_{XYZ} on finite alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$,*

1. *(*Data Processing Inequality*) If the Markov chain $X-Y-V|Z$ holds for a discrete random variable V, then $\mathcal{L}(X \to V|Z) \leq \min\{\mathcal{L}(X \to Y|Z), \mathcal{L}(Y \to V|Z)\}$.*

2. $\mathcal{L}(X{\to}Y|Z) \leq \min\{\log|\mathcal{X}|, \log|\mathcal{Y}|\}$.

3. $\mathcal{L}(X{\to}Y|Z) \geq I(X;Y|Z)$.

4. $\mathcal{L}(X{\to}Y|Z) = 0$ *iff* $X - Z - Y$ *holds.*

5. $\mathcal{L}(X{\to}Y|Z)$ *is not symmetric in* X *and* Y.

6. (Additivity) *If* $\{(X_i, Y_i, Z_i)\}_{i=1}^{\ell}$ *are mutually independent, then*

$$\mathcal{L}(X_1^{\ell}{\to}Y_1^{\ell}|Z^{\ell}) = \sum_{i=1}^{\ell} \mathcal{L}(X_i{\to}Y_i|Z_i).$$

7. *If* $Z - X - Y$ *holds, then*

$$\mathcal{L}(X{\to}Y|Z) \leq \mathcal{L}(X{\to}Y),$$

with equality if, for some z*, the support of* $P_{X|Z=z}$ *is the same as the support of* P_X.

8. $\mathcal{L}(X{\to}(Y,Z)) \leq \mathcal{L}(X{\to}Z) + \mathcal{L}(X{\to}Y|Z)$.

Similarly to maximal leakage, properties 1), 2), 4) and 6) can be seen as axiomatic for a conditional leakage metric. Property 3) is analogous to the relationship between maximal leakage and mutual information. Property 7) is interesting in that it exhibits a behavior similar to mutual information. Indeed, if $Z - X - Y$ holds, then $I(X;Y|Z) \leq I(X;Y)$. Property 8) is a form of a chain rule. An interesting consequence of 7) and 8) is that: if $Z - X - Y$ holds, then $\mathcal{L}(X \to (Y,Z)) \leq \mathcal{L}(X \to Z) + \mathcal{L}(X \to Y)$. This can be interpreted as follows. If an adversary has access to side information Z, which is not known to us (which is the case in practice), then minimizing $\mathcal{L}(X \to Y)$ (irrespective of Z) is the right objective.

Proofs: Property 1) follow directly from the definition. Property 2) follows from Theorems 1 and 5. Property 4) and 6) also follow straightforwardly from

the theorem.

Proof of 3): $I(X;Y|Z) \leq \max_z I(X;Y|Z=z) \leq \max_z \mathcal{L}(X{\to}Y|Z=z) = \mathcal{L}(X{\to}Y|Z)$. ■

Proof of 7):

$$\begin{aligned}\mathcal{L}(X{\to}Y|Z) &= \log\left(\max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z)\right)\\ &= \log\left(\max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|X}(y|x)\right)\\ &\leq \log\left(\sum_y \max_{x:P_X(x)>0} P_{Y|X}(y|x)\right) = \mathcal{L}(X{\to}Y),\end{aligned}$$

where the last inequality follows from the fact that $supp(X) \supseteq supp(X|Z=z)$ for any z. ■

Proof of 8):

$$\begin{aligned}\mathcal{L}(X{\to}(Y,Z)) - \mathcal{L}(X{\to}Z) &= \log \frac{\sum_{z,y} \max_{x:P_X(x)>0} P_{YZ|X}(y,z|x)}{\sum_z \max_{x:P_X(x)>0} P_{Z|X}(z|x)}\\ &\leq \log \max_z \frac{\sum_y \max_{x:P_X(x)>0} P_{YZ|X}(y,z|x)}{\max_{x:P_X(x)>0} P_{Z|X}(z|x)}\\ &= \log \max_z \frac{\sum_y \max_{x:P_X(x)>0} P_{Z|X}(z|x)P_{Y|XZ}(y|x,z)}{\max_{x:P_X(x)>0} P_{Z|X}(z|x)}\\ &\overset{(a)}{=} \log \max_z \frac{\sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Z|X}(z|x)P_{Y|XZ}(y|x,z)}{\max_{x:P_X(x)>0} P_{Z|X}(z|x)}\\ &= \log \max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z)\frac{P_{Z|X}(z|x)}{\max_{x':P_X(x')>0} P_{Z|X}(z|x')}\\ &\leq \log \max_z \sum_y \max_{x:P_{X|Z}(x|z)>0} P_{Y|XZ}(y|x,z) = \mathcal{L}(X{\to}Y|Z),\end{aligned}$$

where (a) follows from the fact that $P_{X|Z}(x|z) = 0$ and $P_Z(z) > 0$ implies that $P_{Z|X}(z|x) = 0$, so that the maximum is achieved outside this set. ■

CHAPTER 3

LEAKAGE METRICS IN THE GUESSING FRAMEWORK

In the following, we investigate the connections between maximal leakage and mutual information, maximal correlation, and local differential privacy in the guessing framework. The analysis will naturally lead us to define an information metric that is intermediate between maximal leakage and local differential privacy, which we call maximal realizable leakage (cf. Section 3.3). Moreover, we introduce a "cost"-based notion of leakage in Section 3.5. Finally, we analyze rate-distortion based metrics in the guessing framework.

3.1 Mutual Information and Capacity

Shannon justifies the choice of mutual information by arguing that "From the point of view of the cryptanalyst [i.e., the adversary], a secrecy system is almost identical with a noisy communication system" [43]. This argument is not persuasive, however, because a noisy communication system (the rate of which is governed by mutual information) relies on coding, of which there is generally none in the leakage setting. One could argue that Shannon is simply taking a "pessimistic" view by upper-bounding leakage by assuming that the transmitter is a cooperative participant and thus willing to code. However, as mentioned before, the key observation is that coding is concerned with (the size of) message sets that can be *reliably* reconstructed at the receiver, whereas leakage does not impose such a constraint. This inspires the following definition.

Definition 5 (Recoverable Leakage) *Given $\epsilon > 0$ and a joint distribution P_{XY} on*

finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the recoverable leakage *from X to Y is defined as*

$$\mathcal{L}_\epsilon^C(X{\to}Y) = \sup_{\substack{(U,X):U-X-Y \\ \mathbf{Pr}(U=\hat{U}(Y))\geq 1-\epsilon}} \log \frac{\mathbf{Pr}(U=\hat{U}(Y))}{\max_u P_U(u)}, \tag{3.1}$$

where the support of U is finite but of arbitrary size, and $\hat{U}(Y)$ is the MAP estimator.

Remark 3 *$\mathcal{L}_\epsilon^C(X{\to}Y)$ depends on P_{XY} only through $P_{Y|X}$.*

Theorem 6 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$,*

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathcal{L}_\epsilon^C(X^n{\to}Y^n) = C(P_{Y|X}), \tag{3.2}$$

where (X^n, Y^n) is distributed i.i.d according to P_{XY}, and $C(P_{Y|X})$ is the capacity of the channel $P_{Y|X}$.

To compare with maximal leakage, say X has full support. Then,

$$\begin{aligned}
\mathcal{L}(X{\to}Y) &\overset{(a)}{=} \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathcal{L}(X^n{\to}Y^n) \\
&\overset{(b)}{=} \lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{\substack{(U,X^n):\\ U-X^n-Y^n}}\log\frac{\mathbf{Pr}(U=\hat{U}(Y))}{\max_u P_U(u)} \\
&\geq \lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{\substack{(U,X^n):\\ U-X^n-Y^n \\ \mathbf{Pr}(U=\hat{U}(Y))\geq 1-\epsilon}}\log\frac{\mathbf{Pr}(U=\hat{U}(Y))}{\max_u P_U(u)} \\
&= C(P_{Y|X}),
\end{aligned}$$

where (a) follows from the additivity of maximal leakage, and (b) follows from the fact that $\mathcal{L}(X{\to}Y)$ depends on P_X only through its support. One can readily verify that the inequality can be strict.

Example 7 *Say $X \sim Ber(p)$, $p \in (0, \frac{1}{2})$. If Y is the output of a BEC(ϵ) ($\epsilon \in (0,1)$) with input X, then $\mathcal{L}(X{\to}Y) = \log(2-\epsilon) > (1-\epsilon)\log 2 = C(P_{Y|X})$.*

Proof of Theorem 6: To show that the left-hand side upper-bounds the right-hand side, consider

$$\begin{aligned}\mathcal{L}_\epsilon^C(X^n{\to}Y^n) &= \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} \log \frac{\mathbf{Pr}(U=\hat{U}(Y))}{\max_u P_U(u)} \\ &\geq \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon\\ U\sim\text{uniform}}} \log|\mathcal{U}|+\log(1-\epsilon). \end{aligned} \tag{3.3}$$

Note that the right-hand side of the above equation is exactly the channel coding setup: U is the uniform message, $P_{X^n|U}$ is the (stochastic) encoding map, $P_{Y|X}$ is the memoryless channel, and ϵ is the allowed average probability of decoding error. Therefore, for any $\delta > 0$, any U with $|\mathcal{U}| < 2^{n(C-\delta)}$ is feasible for large enough n, yielding the lower bound. For the reverse direction, consider the following.

$$\begin{aligned}\mathcal{L}_\epsilon^C(X^n{\to}Y^n) &= \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} \log \frac{\mathbf{Pr}(U=\hat{U}(Y))}{\max_u P_U(u)} \\ &\leq \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} \log \frac{1}{2^{-H_\infty(P_U)}} \\ &= \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} H_\infty(P_U) \\ &\overset{\text{(a)}}{\leq} \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} \frac{I(U;\hat{U})+1}{\mathbf{Pr}(U=\hat{U})} \\ &\leq \sup_{\substack{(U,X^n):U-X^n-Y^n\\ \mathbf{Pr}(U=\hat{U}(Y^n))\geq 1-\epsilon}} \frac{nC(P_{Y|X})+1}{1-\epsilon}, \end{aligned}$$

where (a) follows from [21, Theorem 5]. Taking the limit as $n \to \infty$ and $\epsilon \to 0$ yields the upper bound. ■

Clearly, if $C(P_{Y|X})$ underestimates leakage, then so does $I(X;Y)$. Nevertheless, one might ask: under what conditions are maximal leakage and mutual information equal? The answer is given by the following lemma.

Lemma 2 *$\mathcal{L}(X\to Y) = I(X;Y)$ if and only if*

1. *If $P_{XY}(x,y) > 0$ and $P_{XY}(x',y) > 0$, then $P_{Y|X}(y|x) = P_{Y|X}(y|x')$.*
2. *For all $y, y' \in supp(Y)$,*

$$\sum_{x:P_{XY}(x,y)>0} P_X(x) = \sum_{x':P_{XY}(x',y')>0} P_X(x').$$

Remark 4 *If X has full support, then $\mathcal{L}(X\to Y) = I(X;Y) \Rightarrow \mathcal{L}(X\to Y) = C(P_{Y|X})$.*

Proof: Consider the following chain of inequalities.

$$\begin{aligned}
&I(X;Y)\\
&= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{XY}(x,y)\log\frac{P_{Y|X}(y|x)}{P_Y(y)}\\
&= \sum_{\substack{x\in\mathcal{X},y\in\mathcal{Y}:\\ P_X(x)P_Y(y)>0}} P_{XY}(x,y)\log\frac{P_{Y|X}(y|x)}{P_Y(y)}\\
&\overset{(a)}{\le} \log \sum_{\substack{x\in\mathcal{X},y\in\mathcal{Y}:\\ P_X(x)P_Y(y)>0}} P_{XY}(x,y)\frac{P_{Y|X}(y|x)}{P_Y(y)}\\
&= \log \sum_{\substack{x\in\mathcal{X},y\in\mathcal{Y}:\\ P_X(x)P_Y(y)>0}} P_{X|Y}(x|y)P_{Y|X}(y|x)\\
&\overset{(b)}{\le} \log \sum_{\substack{x\in\mathcal{X},y\in\mathcal{Y}:\\ P_X(x)P_Y(y)>0}} P_{X|Y}(x|y)\max_{\substack{x'\in\mathcal{X}:\\ P_X(x')>0}} P_{Y|X}(y|x')\\
&= \log \sum_{\substack{y\in\mathcal{Y}:\\ P_Y(y)>0}} \max_{x\in\mathcal{X}:P_X(x)>0} P_{Y|X}(y|x)\\
&\overset{(c)}{=} \log \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}:P_X(x)>0} P_{Y|X}(y|x) = \mathcal{L}(X\to Y),
\end{aligned}$$

where (a) follows from Jensen's inequality, and (c) follows from the fact that $P_Y(y) = 0$ implies that $\max_{x\in\mathcal{X}:P_X(x)>0} P_{Y|X}(y|x) = 0$. Now, note that (b) can be turned into equality if and only if condition 1) holds. Given condition 1), it can be

seen that condition 2) is necessary and sufficient for (a) to become equality (by expanding $P_Y(y) = \sum_{x:P_{XY}(x,y)>0} P_X(x)P_{Y|X}(y|x)$).

3.2 Maximal Correlation

Recall that, given a joint distribution P_{XY}, maximal correlation $\rho_m(X;Y)$ is defined as

$$\rho_m(X;Y) = \sup_{\substack{f,g: \\ \mathbf{E}[f]=\mathbf{E}[g]=0 \\ \mathbf{E}[f^2]=\mathbf{E}[g^2]=1}} \mathbf{E}[f(X)g(Y)]. \tag{3.4}$$

Li and El-Gamal [29] proposed maximal correlation as a secrecy metric. A main motivation for such choice is the result of Calmon *et al.* [8], which bounds in terms of $\rho_m(X;Y)$ the additive increase in the probability of correctly guessing any deterministic function of X, assuming X is uniformly distributed. We posit that maximal correlation is actually capturing the change in variance. That is, we define *variance leakage* as follows.

Definition 6 (Variance Leakage) *Given a joint distribution P_{XY} on alphabets $\mathcal{X}$ and $\mathcal{Y}$, the* variance leakage *from X to Y is defined as*

$$\mathcal{L}^v(X\to Y) = \sup_{\substack{U:U-X-Y \\ \mathbf{var}(U)>0}} \log \frac{\mathbf{var}(U)}{\mathbf{E}[(U-\mathbf{E}[U|Y])^2]}. \tag{3.5}$$

Theorem 7 *For any joint distribution P_{XY} on alphabets $\mathcal{X}$ and $\mathcal{Y}$, the variance leakage from X to Y is given by*

$$\mathcal{L}^v(X\to Y) = -\log(1-\rho_m^2(X;Y)), \tag{3.6}$$

where $\rho_m(X;Y)$ is the maximal correlation.

As such, maximal correlation is capturing the multiplicative decrease in variance. If the U of interest is discrete, which is often the case in practice (e.g., U is a password, a social security number, etc.), the probability of correct guessing is arguably the more relevant quantity.

Proof: Without loss of generality, we can restrict the optimization in (3.5) to U's that satisfy $\mathbf{E}[U] = 0$, and $\mathbf{E}[U^2] = 1$. So, we rewrite

$$\begin{aligned}\mathcal{L}^v(X \to Y) &= \sup_{\substack{U: U-X-Y \\ \mathbf{E}[U]=0,\ \mathbf{E}[U^2]=1}} \log \frac{1}{\mathbf{E}[U^2] - \mathbf{E}[\mathbf{E}[U|Y]^2]} \\ &= \sup_{\substack{U: U-X-Y \\ \mathbf{E}[U]=0,\ \mathbf{E}[U^2]=1}} -\log\left(1 - \mathbf{E}\left[\mathbf{E}[U|Y]^2\right]\right). \end{aligned} \tag{3.7}$$

Also, we can rewrite maximal correlation using Renyi's equivalent characterization [36]:

$$\rho_m(X;Y) = \sup_{\substack{f: \mathbf{E}[f(X)]=0 \\ \mathbf{E}[f^2(X)]=1}} \sqrt{\mathbf{E}\left[\mathbf{E}[f(X)|Y]^2\right]}. \tag{3.8}$$

Now, note that

$$\begin{aligned}\rho_m^2(X;Y) &= \sup_{f:\ \mathbf{E}[f]=0,\ \mathbf{E}[f^2]=1} \mathbf{E}\left[\mathbf{E}[f(X)|Y]^2\right] \\ &\overset{(a)}{\leq} \sup_{\substack{U: U-X-Y \\ \mathbf{E}[U]=0,\ \mathbf{E}[U^2]=1}} \mathbf{E}\left[\mathbf{E}[U|Y]^2\right] \\ &\leq \sup_{\substack{U: U-X-Y \\ \mathbf{E}[U]=0,\ \mathbf{E}[U^2]=1}} \sup_{\substack{h:\ \mathbf{E}[h(U)]=0, \\ \mathbf{E}[h^2(U)]=1}} \mathbf{E}\left[\mathbf{E}[h(U)|Y]^2\right] \\ &= \sup_{\substack{U: U-X-Y \\ \mathbf{E}[U]=0,\ \mathbf{E}[U^2]=1}} \rho_m^2(U;Y) \overset{(b)}{\leq} \rho_m^2(X;Y), \end{aligned} \tag{3.9}$$

where (b) follows from the fact that maximal correlation obeys the data processing inequality [8, Theorem 2]. Therefore (a) is in fact an equality. Plugging it in (3.7) yields our desired result. ■

3.3 Maximal Realizable Leakage

We now consider a variation of the definition of maximal leakage, which captures a different scenario of interest. It will be also useful for interpreting local differential privacy in the guessing framework (cf. Section 3.4).

In particular, maximal leakage considers the *average* guessing performance of the adversary, $\mathbf{Pr}(U = \hat{U}(Y))$, for each U satisfying $U - X - Y$. As such, realizations y of Y that lead to a high probability of correct guessing are "tolerable" if the corresponding probabilities $P_Y(y)$'s are very small. For scenarios in which such small probability events are still unacceptable (e.g., U is highly classified data), we need to consider the *maximum* instead of the average performance. This leads to the following definition.

Definition 7 (Maximal Realizable Leakage) *Given a joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the* maximal realizable leakage *from X to Y is defined as*

$$\mathcal{L}^r(X \to Y) = \sup_{U - X - Y} \log \frac{\max_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} P_{U|Y}(u|y)}{\max_{u \in \mathcal{U}} P_U(u)}, \tag{3.10}$$

where the support U is finite but of arbitrary size.

Theorem 8 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal realizable leakage from X to Y is given by*

$$\mathcal{L}^r(X \to Y) = \max_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ P_{XY}(x,y) > 0}} \log \frac{P_{Y|X}(y|x)}{P_Y(y)}. \tag{3.11}$$

As opposed to maximal leakage, maximal realizable leakage is symmetric in X and Y (since $P_{Y|X}(y|x)/P_Y(y) = P_{X|Y}(x|y)/P_X(x)$). It is equal to the maximum *information rate* (the latter is the random variable the expectation of which

is mutual information). It follows straightforwardly from the definitions that $\mathcal{L}^r(X{\to}Y) \geq \mathcal{L}(X{\to}Y)$. Moreover, $\mathcal{L}^r(X{\to}Y)$ cannot be bounded in terms of $|\mathcal{X}|$ and $|\mathcal{Y}|$: in Example 7, $\mathcal{L}^r(X{\to}Y) = \log(1/p) \xrightarrow{p\to 0} \infty$.

Furthermore, $\mathcal{L}^r(X \to Y)$ exhibits desirable properties of a leakage metric: it satisfies the data processing inequality, it is zero if and only X and Y are independent, and it is additive over independent pairs $\{(X_i, Y_i)\}$. The proofs of these facts are similar to previous derivations and are omitted.

Remark 5 *The fact that using the max in* (3.10) *and the average in* (2.1) *both lead to quantities with desirable properties suggests that we could also consider weighted averages, i.e., replace the numerator by* $\left(\sum_y P_Y(y) \max_u P_{U|Y}^{\alpha}(u|y)\right)^{1/\alpha}$*, for some* $\alpha > 0$*.*

Proof of Theorem 8: Without loss of generality, assume X and Y have full support. To show the upper bound, consider the following.

$$\begin{aligned}
\max_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} P_{U|Y}(u|y) &= \max_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_{U|X}(u|x) P_{X|Y}(x|y) \\
&= \max_{y\in\mathcal{Y}} \max_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_{U|X}(u|x) P_X(x) \frac{P_{Y|X}(y|x)}{P_Y(y)} \\
&\leq \max_{y\in\mathcal{Y}} \max_{x'\in\mathcal{X}} \frac{P_{Y|X}(y|x')}{P_Y(y)} \max_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_{UX}(u,x) \\
&= \max_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} \frac{P_{Y|X}(y|x)}{P_Y(y)} \max_{u\in\mathcal{U}} P_U(u).
\end{aligned}$$

For the reverse direction, we again consider the shattering $P_{U|X}$ (cf. equation 2.3). It is a simple exercise to check that this choice yields the desired lower bound. ■

3.4 Local Differential Privacy

Local differential privacy [13] adapts that notion to the setting of a given conditional distribution $P_{Y|X}$. It is defined as:

$$L^{dp}(X\to Y) = \max_{\substack{y\in\mathcal{Y},\\ x,x'\in\mathcal{X}}} \log \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')}. \tag{3.12}$$

Local differential privacy is known to be pessimistic [13]. It is indeed very strict: in Example 7, $L^{dp}(X\to Y) = \infty$. Interestingly, we also noted in the previous section that $\lim_{p\to 0} \mathcal{L}^r(X\to Y) = \infty$.

So what operational problem is local differential privacy solving? Similarly to maximal realizable leakage, local differential privacy is concerned with worst-case analysis over the *realizations* of Y. Moreover, being a function of $P_{Y|X}$, it is robust against the worst-case distribution P_X. This yields the following definition.

Definition 8 *Given a conditional distribution $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets, let*

$$\begin{aligned}\mathcal{L}^{dp}(X\to Y) &= \sup_{P_X} \sup_{U-X-Y} \log \frac{\max_y \max_u P_{U|Y}(u|y)}{\max_u P_U(u)} \\ &= \sup_{P_X} \mathcal{L}^r(X\to Y).\end{aligned} \tag{3.13}$$

Theorem 9 *For any conditional distribution $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets,*

$$\mathcal{L}^{dp}(X\to Y) = L^{dp}(X\to Y). \tag{3.14}$$

Clearly, $\mathcal{L}^{dp}(X \to Y) \geq \mathcal{L}^r(X \to Y) \geq \mathcal{L}(X\to Y)$. Theorems 8 and 9 imply that $\mathcal{L}^{dp}(X \to Y) = \mathcal{L}^r(X\to Y)$ if and only if X and Y are independent. Thus,

$\mathcal{L}^{dp}(X \to Y) = \mathcal{L}(X \to Y)$ if and only if X and Y are independent. Moreover, an interesting implication of (3.13) is that one could incorporate information about the marginal P_X by restricting the optimization set of the sup.

Proof: By Theorem 8, we can rewrite (3.13) as

$$\mathcal{L}^{dp}(X \to Y) = \sup_{P_X} \max_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ P_{XY}(x,y) > 0}} \log \frac{P_{Y|X}(y|x)}{P_Y(y)}.$$

The upper bound thus follows from the fact that $P_Y(y) \geq \min_x P_{Y|X}(y|x)$. For the lower bound, consider the following. Let $y^\star$ be an element achieving the max in (3.12). Let $x_0 \in \operatorname{argmin}_x P_{Y|X}(y^\star|x)$ and $x_1 \in \operatorname{argmax}_x P_{Y|X}(y^\star|x)$. Finally, for a given $\alpha > 0$, let $P_X(x_0) = 1 - \alpha$ and $P_X(x_1) = \alpha$. Then,

$$\begin{aligned}
\mathcal{L}^{dp}(X \to Y) &\geq \log \frac{\max_x P_{Y|X}(y^\star|x)}{P_Y(y^\star)} \\
&= \log \frac{P_{Y|X}(y^\star|x_1)}{(1-\alpha) P_{Y|X}(y^\star|x_0) + \alpha P_{Y|X}(y^\star|x_1)} \\
&\xrightarrow{\alpha \to 0} \log \frac{P_{Y|X}(y^\star|x_1)}{P_{Y|X}(y^\star|x_0)} = L^{dp}(X \to Y).
\end{aligned}$$

■

We also provide another equivalent definition of local differential privacy. More specifically, for any given $U : U - X - Y$, maximal leakage compares the *average* probabilities of correct guessing before and after observing Y. Another approach would be to compare, for each y, the performance of an adversary who sees y and thus acts according to $P_{U|Y}(u|y)$, and another who doesn't see y and thus acts according to $P_U(u)$. Let $u^\star = \operatorname{argmax}_u P_U(u)$ (suppose it is unique, for now). So for each y, we are interested in

$$\frac{\max_u P_{U|Y}(u|y)}{P_{U|Y}(u^\star|y)}. \tag{3.15}$$

If $u^\star$ is not unique, then we reformulate as follows. Let $U^\star = \operatorname{argmax}_u P_U(u)$. We give the informed adversary maximum advantage by considering

$$\frac{\max_u P_{U|Y}(u|y)}{\min_{u \in U^\star} P_{U|Y}(u|y)}. \tag{3.16}$$

We also consider the best case for the informed adversary over y, yielding the following definition.

Definition 9 *For any joint distribution P_{XY}, define*

$$\mathcal{L}^{diff}(X \to Y) = \sup_{U:U-X-Y} \log \max_y \frac{\max_u P_{U|Y}(u|y)}{\min_{u \in U^\star} P_{U|Y}(u|y)}. \tag{3.17}$$

Theorem 10 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$,*

$$\mathcal{L}^{diff}(X \to Y) = \log \max_{y,x,x'} \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')}. \tag{3.18}$$

Proof:

To show that the left-hand side is upper-bounded by by the right-hand side, fix any U satisfying $U - X - Y$, and fix $y \in \mathcal{Y}$. Then,

$$\begin{aligned}
\frac{\max_u P_{U|Y}(u|y)}{\min_{u \in U^\star} P_{U|Y}(u|y)} &= \frac{\max_u \sum_x P_{U|X}(u|x) P_{X|Y}(x|y)}{\min_{u \in U^\star} \sum_x P_{U|X}(u|x) P_{X|Y}(x|y)} = \frac{\max_u \sum_x P_{U|X}(u|x) P_{XY}(x,y)}{\min_{u \in U^\star} \sum_x P_{U|X}(u|x) P_{XY}(x,y)} \\
&= \frac{\max_u \sum_x P_{UX}(u,x) P_{Y|X}(y|x)}{\min_{u \in U^\star} \sum_x P_{UX}(u,x) P_{Y|X}(y|x)} \\
&\leq \frac{(\max_x P_{Y|X}(y|x)) \max_u \sum_x P_{UX}(u,x)}{(\min_x P_{Y|X}(y|x)) \min_{u \in U^\star} \sum_x P_{UX}(u,x)} = \max_{x,x'} \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')}.
\end{aligned}$$

For the reverse direction, consider the "shattering" $P_{U|X}$, given by (2.3). Now, note that, for a given y

$$\begin{aligned}
P_{U|Y}((i_u, j_u)|y) &= \sum_x P_{X|Y}(x|y) P_{U|X}((i_u, j_u)|x) \\
&= P_{X|Y}(i_u|y) P_{U|X}((i_u, j_u)|i_u) \\
&= P_{X|Y}(i_u|y). \begin{cases} \frac{p^\star}{P_X(i_u)}, & 1 \leq j_u \leq \lfloor k(i_u) \rfloor, \\ 1 - \frac{(\lceil k(i_u) \rceil - 1) p^\star}{P_X(i_u)}, & j_u = \lceil k(i_u) \rceil. \end{cases}
\end{aligned}$$

Therefore,

$$\max_u P_{U|Y}((i_u, j_u)|y) = \max_{(i_u,1)} P_{X|Y}(i_u|y)\frac{p^\star}{P_X(i_u)} = \max_x \frac{p^\star P_{X|Y}(x|y)}{P_X(x)} = \max_x \frac{p^\star P_{Y|X}(y|x)}{P_Y(y)}. \tag{3.19}$$

Furthermore, $U^\star = \bigcup_{x\in\mathcal{X}}\{(x,1),(x,2),\ldots,(x,\lfloor k(x)\rfloor)\}$. Therefore,

$$\min_{u\in U^\star} P_{U|Y}((i_u, j_u)|y) = \min_{(i_u,1\le j\le\lfloor k(i_u)\rfloor)} P_{X|Y}(i_u|y)\frac{p^\star}{P_X(i_u)} = \min_x \frac{p^\star P_{X|Y}(x|y)}{P_X(x)} = \min_x \frac{p^\star P_{Y|X}(y|x)}{P_Y(y)}. \tag{3.20}$$

Combining (3.19) and (3.20), we get, for a given y,

$$\frac{\max_u P_{U|Y}(u|y)}{\min_{u\in U^\star} P_{U|Y}(u|y)} = \frac{\max_x P_{Y|X}(y|x)}{\min_x P_{Y|X}(y|x)} = \max_{x,x'} \frac{P_{Y|X}(y|x)}{P_{Y|X}(y|x')}, \tag{3.21}$$

as desired. ■

3.5 Maximal Cost Leakage

In this section, we introduce a leakage metric that is *dual* to maximal leakage. Whereas maximal leakage considers the maximum *gain* that the adversary achieves, we could alternatively consider the maximum reduction in *cost* s/he incurs.

Definition 10 (Maximal Cost Leakage) *Given a joint distribution P_{XY} on alphabets $\mathcal{X}$ and $\mathcal{Y}$, the* maximal cost leakage *from X to Y is defined as*

$$\mathcal{L}^c(X\to Y) = \sup_{\substack{U:U-X-Y\\ \hat{\mathcal{U}},\ d:\hat{\mathcal{U}}\times\mathcal{U}\to\mathbb{R}_+}} \log \frac{\inf_{\hat{u}\in\hat{\mathcal{U}}} \mathbf{E}[d(U,\hat{u})]}{\inf_{\hat{U}:X-Y-\hat{U}} \mathbf{E}[d(U,\hat{U})]}, \tag{3.22}$$

where $\hat{\mathcal{U}}$ is a finite alphabet, and $\frac{0}{0} = 1$ by convention.

Maximal cost leakage also admits a simple form for discrete X and Y, given in the following theorem. We defer the proof to Appendix A.1.

Theorem 11 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal cost leakage from X to Y is given by*

$$\mathcal{L}^c(X{\to}Y) = -\log \sum_{y\in\mathcal{Y}} \min_{\substack{x\in\mathcal{X}:\\ P_X(x)>0}} P_{Y|X}(y|x). \tag{3.23}$$

The following corollary, the proof of which is given in Appendix A.2, summarize useful properties of $\mathcal{L}^c(X{\to}Y)$.

Corollary 6 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$,*

1. *(Data Processing Inequality) If the Markov chain $X - Y - Z$ holds for a discrete random variable Z, then $\mathcal{L}^c(X{\to}Z) \le \min\{\mathcal{L}^c(X{\to}Y), \mathcal{L}^c(Y{\to}Z)\}$.*
2. *For any non-trivial deterministic law $P_{Y|X}$ (i.e., $|\{y : P_Y(y) > 0\}| > 1$), $\mathcal{L}^c(Y{\to}X) = +\infty$.*
3. *$\mathcal{L}^c(X{\to}Y) = 0$ iff X and Y are independent.*
4. *$\mathcal{L}^c(X{\to}Y) \le L^{dp}(X{\to}Y)$.*
5. *$\mathcal{L}^c(X{\to}Y)$ is not symmetric in X and Y.*
6. *(*Additivity*) If $\{(X_i, Y_i)\}_{i=1}^{\ell}$ are mutually independent, then*

$$\mathcal{L}^c(X_1^\ell{\to}Y_1^\ell) = \sum_{i=1}^{\ell} \mathcal{L}^c(X_i{\to}Y_i).$$

7. *$\mathcal{L}^c(X{\to}Y)$ is convex in $P_{Y|X}$ for fixed P_X.*

We evaluate $\mathcal{L}^c(X{\to}Y)$ for some examples.

Example 8 *If $X \sim Ber(q)$, $0 < q < 1$, and Y is the output of a BSC with input X and parameter p, $0 \leq p \leq 1/2$, then $\mathcal{L}^c(X \to Y) = -\log(2p)$.*

Example 9 *If $X \sim Ber(q)$, $0 < q < 1$, and Y is the output of a BEC with input X and parameter ϵ, $0 \leq \epsilon < 1$, then $\mathcal{L}^c(X \to Y) = -\log(\epsilon)$, and $\mathcal{L}^c(Y \to X) = +\infty$.*

Example 10 *If X is not deterministic, $\mathcal{L}^c(X \to X) = +\infty$.*

Remark 6 *One can note that in each of the examples above, $\mathcal{L}^c(X \to Y) \geq \mathcal{L}(X \to Y)$. This is always true when $|\mathcal{X}| = |\mathcal{Y}| = 2$, but it is not necessarily true in general. As a counter example, say X has full support and $P_{Y|X} = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$. Then $\exp\{\mathcal{L}(X \to Y)\} = 1.2$ and $\exp\{\mathcal{L}(X \to Y)\} = 1/0.9 = 1.\bar{1}$.*

Maximal Realizable Cost

Similarly to the modification of maximal leakage to maximal realizable leakage, we could consider the minimum cost incurred at the adversary, instead of the average cost. We show next that this yields the maximum of the *negative* of the information rate. Maximizing it over the input distribution also yields local differential privacy.

Definition 11 (Maximal Realizable Cost) *Given a joint distribution P_{XY} on alphabets $\mathcal{X}$ and $\mathcal{Y}$, the* maximal realizable cost *from X to Y is defined as*

$$\mathcal{L}^{rc}(X \to Y) = \sup_{\substack{U : U - X - Y \\ \hat{\mathcal{U}},\, d : \hat{\mathcal{U}} \times \mathcal{U} \to \mathbb{R}_+}} \log \frac{\inf_{\hat{u} \in \hat{\mathcal{U}}} \mathbf{E}[d(U, \hat{u})]}{\min_{y \in \text{supp}(Y)} \inf_{\hat{u} \in \hat{\mathcal{U}}} \mathbf{E}[d(U, \hat{u}) | Y = y]}, \tag{3.24}$$

where $\hat{\mathcal{U}}$ is a finite alphabet, and $\frac{0}{0} = 1$ by convention.

Theorem 12 *For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal realizable cost from X to Y is given by*

$$\mathcal{L}^{rc}(X \to Y) = -\log \min_{\substack{x,y: \\ P_{XY}(x,y)>0}} \frac{P_{Y|X}(y|x)}{P_Y(y)} = \log \max_{\substack{x,y: \\ P_{XY}(x,y)>0}} \frac{P_Y(y)}{P_{Y|X}(y|x)}. \tag{3.25}$$

Corollary 7 *For any conditional distribution $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets,*

$$\max_{P_X} \mathcal{L}^{rc}(X \to Y) = L^{dp}(X \to Y). \tag{3.26}$$

The proofs are given in Appendices A.3 and A.4, respectively.

3.6 Rate-Distortion Based Approaches

As mentioned in the overview, all rate-distortion based approaches basically assume that the adversary is only interested in a deterministic function of X and that we know what that function is. Nevertheless, it is worth examining some common approaches.

For ease of comparison, we will consider the setup of the Shannon cipher system with lossy communication, shown in Figure 3.1. It consists of a transmitter and a legitimate receiver that are linked by a public noiseless channel and share a common key, and an eavesdropper who has access to the public channel and is aware of the source statistics and the used encryption schemes. The encryption schemes must allow the legitimate receiver to reconstruct the source sequence up to a fidelity constraint.

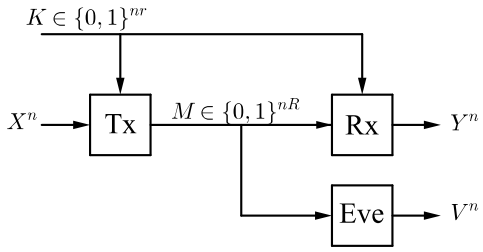

Figure 3.1: The Shannon cipher system with lossy communication.

3.6.1 Expected Distortion

Suppose we use the expected distortion of the eavesdropper's best reconstruction as the leakage metric [58]. A standard example, discussed and generalized in [40], shows why this metric is inadequate: Suppose X^n is a sequence of independent and identically distributed bits with $X_i \sim \text{Ber}(1/2)$, the transmitter and the legitimate receiver have access to one common bit $K \sim \text{Ber}(1/2)$, and the distortion function is the Hamming distance. The transmitter then sends the sequence X^n as is if $K = 0$, and flips all its bits if $K = 1$. The induced expected distortion at the eavesdropper is then equal to $1/2$, which is also the maximum expected distortion that the eavesdropper can possibly incur, since it is achievable even if the public message is not observed. However, this "optimal" scheme in fact reveals a lot about the true source sequence, namely, it is one of only two possible candidates. Indeed, a simple computation shows that, $\mathcal{L}(X^n \to M) = n \log |\mathcal{X}| - 1$, which is exactly describing that X^n is completely revealed except for one bit.

3.6.2 Expected Distortion in a List/with Feedforward

To overcome this limitation of expected distortion, Schieler and Cuff [41,42] allow the eavesdropper to generate an exponentially-sized list of estimates and propose the expected minimum distortion over the list as a secrecy metric. It is not clear, however, how to operationally interpret a list of *exponential* size. Moreover, this metric leads to a degenerate trade-off between the key rate r, the allowed list exponent R_L, and the expected minimum distortion in the list D_e. For example, if the legitimate receiver must reconstruct X^n losslessly, one of two cases must occur (see [41, Theorem 1]): either the public message M is made completely useless to the eavesdropper when $r > R_L$, and D_e is then given by the distortion-rate function at R_L; or, when $r \leq R_L$, the eavesdropper can trivially find the exact sequence by listing all the possible keys.

Another approach to overcome this limitation is to consider expected distortion with feedforward [40]. That is, at time i, the adversary has access to the public message M as well as the first $i-1$ symbols of the source, X^{i-1}. Again, it is not clear what is the operational motivation of such a setup.

3.6.3 Expected Number of Guesses to Satisfy the Constraint

Merhav and Arikan [33] proposed a more direct approach: they consider an i.i.d. source and they measure secrecy by the expected number of guesses that the eavesdropper needs to make before finding the correct source sequence, which they denote by $\mathbf{E}[G(X^n|M)]$, where $G(.|m)$ is a "guessing" function defined for each possible public message m (this can be modified to allow a certain level of distortion instead of exact guessing). This is intended to capture the sce-

nario in which the eavesdropper has a testing mechanism to check whether or not his/her guess is correct. Such mechanism exists, for example, if the source message is a password to a computer account. When the source is discrete and memoryless, and the transmitter and the legitimate receiver have access to nr purely random common bits (where r is the key rate), the optimal exponent of $\mathbf{E}[G(X^n|M)]$ is found to be [33, Theorem 1]:

$$E(P,r) \triangleq \lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}[G(X^n|M)] = \max_{Q} \{\min\{H(Q), r\} - D(Q\|P)\}, \tag{3.27}$$

where P is the source distribution and $D(\cdot\|\cdot)$ is the Kullback-Leibler (KL) divergence. Two issues arise with this metric. First, even if a testing mechanism exists, any practical system would only allow a small number of incorrect inputs. Thus, it is not clear how to interpret an exponentially large number of guesses. Second, and more importantly, because expected behavior can be quite different from typical behavior, it turns out that even highly-insecure systems can appear to be secure under this metric. Indeed, by modifying the asymptotically-optimal scheme proposed in [33], we can construct a scheme for the primary user that allows the eavesdropper to find the source sequence correctly with high probability by the first guess, and yet achieves the optimal exponent in (3.27). The scheme proposed in [33] operates on the source sequences on a type-by-type basis, and it yields:

$$\mathbf{E}\left[G(X^n|M)\middle|X^n \in T_Q\right] \geq 2^{n\min\{r,H(Q)\}-o(n)}, \tag{3.28}$$

where $o(n)/n \to 0$ as $n \to \infty$, and T_Q is the type class of a given type Q, i.e., the set of sequences with empirical distribution Q. Averaging over the probabilities of $\{T_Q\}$ yields the exponent in (3.27) (as a lower bound). However, this means that it is enough to apply the proposed scheme to the type class T_Q that achieves the maximum of $[\min\{H(Q), r\} - D(Q\|P)]$, whereas sequences belonging to other type classes can be sent with no encoding whatsoever with no effect

on the exponent. Therefore, only a set with vanishing probability is encoded, whereas sequences outside that set are immediately known by the eavesdropper[1]. On the other hand, a simple calculation shows that such a scheme yields $\lim_{n\to\infty} \frac{1}{n}\mathcal{L}(X^n \to Y^n) = 1$, which is exactly describing that X^n is completely revealed asymptotically.

[1]Merhav and Arikan actually characterize, for any $\rho > 0$, the exponent of $\mathbf{E}[G^\rho(X^n|M)]$. This more general result can still yield large exponents for systems that are highly insecure, although one could potentially address this issue by requiring schemes that yield large exponents simultaneously over a range of ρ values.

CHAPTER 4

THE SHANNON CIPHER SYSTEM

4.1 Overview

The main goal in quantifying information leakage is to enable the design of mechanisms to mitigate it. As an application, we study a (traditional) secrecy setup known as the Shannon cipher system [43]. The setup consists of a transmitter and a legitimate receiver that are linked by a public noiseless channel and share a common key, and an eavesdropper who has access to the public channel and is aware of the source statistics and the used encryption schemes. The encryption schemes must allow the legitimate receiver to perfectly reconstruct the source sequence.

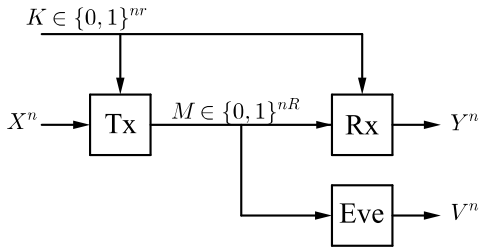

Figure 4.1: The Shannon cipher system with lossy communication: the transmitter and the legitimate receiver have access to a common key K, which consists of nr purely random bits, where r is called the key rate. The transmitter encodes X^n using K, and sends a message M through a noiseless public channel of rate R. Both the legitimate receiver and the eavesdropper are allowed a certain level of distortion. The legitimate receiver generates the reconstruction Y^n based on M and K, whereas the eavesdropper has access to M only to produce an estimate V^n.

Shannon showed that *perfect secrecy* (i.e., making the public message M and the source sequence X^n statistically independent) requires a key rate that is at

least as large as the message rate, which is typically not possible in practice. It is necessary then to quantify *imperfect* or *partial* secrecy.

Therefore, in this chapter, we use maximal leakage to assess the best partial secrecy that can be obtained in the Shannon cipher system, for any encryption scheme. Moreover, we allow for lossy communication by introducing a distortion function d at the legitimate receiver, as shown in Figure 4.1. For a given distortion level D, we require that the probability of violating the distortion constraint decays as $2^{-n\alpha}$, for a given $\alpha > 0$. Then, for a given D and α, we study the asymptotic behavior of the normalized maximal leakage.

For a discrete memoryless source (DMS), we derive the optimal (i.e., minimal) limit of the normalized maximal leakage. The scheme we propose for the primary user (i.e., the transmitter–legitimate receiver pair) operates on a type-by-type basis. With each type, we associate a good rate-distortion code. The codebooks are then divided into bins, and the key is used to randomize, within a bin, the choice of codeword associated with a particular source sequence. However, types with low enough probability are discarded, i.e., a dummy message is associated with all the source sequences belonging to such types. We also derive the optimal limit when the requirement of a decaying probability of violating the distortion constraint is replaced with an expected distortion constraint. We further show that, even in this setup, memoryless schemes are not optimal. Other works have considered the Shannon cipher system allowing for lossy compression [23,41,53,58], although they differ from the present chapter in how they measure partial secrecy.

4.2 Problem Setup and Statement of Result

Let $\mathcal{X}$ and $\mathcal{Y}$ be the alphabets associated with the transmitter and the legitimate receiver, respectively. The transmitter and the legitimate receiver are connected through a noiseless channel of rate R, and share common randomness $K_n \in \mathcal{K}_n = \{0,1\}^{nr}$, where K_n is uniformly distributed over $\mathcal{K}_n$, and $r > 0$ is the rate of the key. The transmitter observes an n-length message $X^n = (X_1, X_2, \cdots, X_n)$, independent of K_n, and wants to transmit a quantized version of it. Let f and h be, respectively, the transmitter's encoding and the receiver's decoding functions. The transmitter then sends a message $M_n = f(X^n, K_n)$, $M_n \in \mathcal{M}_n = \{0,1\}^{nR}$, and the receiver generates a reconstruction $Y^n = h(M_n, K_n)$. Note that we allow the functions f and h to be randomized (beyond the randomness in K_n). For a given distortion function $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$, distortion level D, and distortion excess probability α, we require that $\mathbf{Pr}(d(X^n, Y^n) > D) \leq 2^{-n\alpha}$, where $d(X^n, Y^n) = \frac{1}{n}\sum_{i=1}^{n} d(X_i, Y_i)$.

An eavesdropper intercepts the message M. We assume s/he knows the source statistics as well as the encoding and decoding functions, but does not have access to the key K_n.

The primary user aims to minimize the maximal leakage to the eavesdropper $\mathcal{L}(X^n \to M_n)$. We characterize the asymptotically-optimal normalized maximal leakage under the following assumptions[1]:

(A2) The alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite.

(A1) The source is memoryless and has full support.

[1]Note that it is necessary to have $R \geq \max_{Q:D(Q\|P)\leq\alpha} R(Q, D)$ for the primary user's problem to be feasible.

(A3) The distortion function d is bounded, i.e., there exists $D_{\max}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, $d(x,y) \leq D_{\max}$. Moreover, $D \geq D_{\min}$, where $D_{\min} = \max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} d(x,y)$.

(A4) $R > \max_{Q:D(Q\|P)\leq\alpha} R(Q,D)$, where $R(Q,D)$ is the rate distortion function for source distribution Q.

We denote the optimal limit by $L(P,D,\overrightarrow{R},\alpha)$, where P is the source distribution, and $\overrightarrow{R} = (R,r)$:

$$L(P,D,\overrightarrow{R},\alpha) = \lim_{n\to\infty} \min_{\{f_n\}} \frac{1}{n}\mathcal{L}\left(X^n \to f(X^n, K_n)\right),$$

where $\{f_n\}$ is restricted to the class of functions ensuring the feasibility of the primary user's problem.

Theorem 13 *Under assumptions (A1)-(A4), for any DMS P, and distortion function d with associated distortion level $D \geq D_{\min}$ and distortion excess probability $\alpha > 0$:*

$$L(P,D,\overrightarrow{R},\alpha) = \max_{Q:D(Q\|P)\leq\alpha} [R(Q,D) - r]^+ \ bits, \tag{4.1}$$

where $[a]^+ = \max\{0,a\}$.

Note that the case $\alpha = \infty$ (i.e., when the distortion constraint is imposed almost surely) is included in the theorem. Moreover, in that case, the theorem holds even if the source is not memoryless, as long as the support of X^n is $\mathcal{X}^n$. This follows from the fact that $\mathcal{L}(X^n \to M_n)$ and the constraint, when imposed almost surely, depend on the distribution of X^n only through its support. Therefore, solving for any specific distribution on that support is equivalent to solving for all distributions on the same support.

We set some notation for the remainder of the chapter. In the following, $\mathcal{Z}$ is an arbitrary discrete set, and Z is a random variable over $\mathcal{Z}$.

- For a sequence $z^n \in \mathcal{Z}^n$, Q_{z^n} is the empirical PMF of z^n, also referred to as its type.
- $\mathcal{Q}_{\mathcal{Z}}^n$ is the set of types in $\mathcal{Z}^n$, i.e., the set of rational PMF's with denominator n.
- For $Q_Z \in \mathcal{Q}_{\mathcal{Z}}^n$, the type class of Q_Z is $T_{Q_Z} \triangleq \{z^n \in \mathcal{Z}^n : Q_{z^n} = Q_Z\}$.
- $\mathbf{E}_Q[\cdot]$, $H_Q(\cdot)$, and $I_Q(\cdot;\cdot)$ denote respectively expectation, entropy, and mutual information taken with respect to distribution Q.
- Throughout this chapter, logarithms and exponentials are taken to the base 2.

4.2.1 Special Case: Information Blurring System

Consider the special case in which $R = \log_2 |\mathcal{Y}|$ and $r = 0$, shown in Figure 4.2. We refer to it as the information blurring system (IBS). This is a special case of

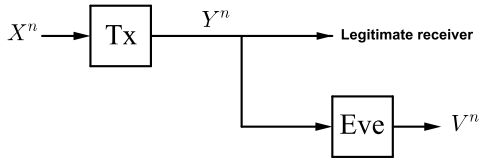

Figure 4.2: Information blurring system: both the legitimate receiver and the eavesdropper are allowed a certain distortion level.

interest as it is intended as a stylized model of a side channel. For example, consider the SSH side channel. Then, X^n corresponds roughly to the original timing vector, and it is mapped to a sequence Y^n that corresponds to the perturbed timings and is observed by both the legitimate receiver and an eavesdropper. The mapping must satisfy a distortion constraint, which corresponds to some quality constraints imposed by the network (e.g., delay constraints). At this point,

we do not require the mapping to be causal as the intent is to provide fundamental limits for a simplified version of the information leakage problem. In broad terms, the transmitter wants to *blur* the information in X^n (hence the name), so that it is no longer useful for the eavesdropper. For example, one approach is to artificially add noise to the input sequence. In that sense, the problem is related to methods for ensuring *differential privacy*, in which a curator wants to publicly release statistical information about a given population without compromising the privacy of its individuals [14,31].

4.3 Achievability Proof of Theorem 13

We will slightly abuse notation and shorten $L(P, D, \overrightarrow{R}, \alpha)$ to L in the following. We now show that the right-hand side of (4.1) upper-bounds L.

Consider any $\epsilon > 0$ and let n be large enough such that we can construct a rate-distortion code $C^n_{Q_X}$, for each type $Q_X \in \mathcal{Q}^n_{\mathcal{X}}$, satisfying the following: each sequence $x^n \in T_{Q_X}$ is covered and $|C^n_{Q_X}| \leq 2^{n(R(Q_X,D)+\epsilon)}$. Such construction is guaranteed by the type covering lemma (Lemma 2.4.1 in [11]). We divide the codebook $C^n_{Q_X}$ into $\left\lceil |C^n_{Q_X}| / 2^{nr} \right\rceil$ bins, each of size 2^{nr}, except for possibly the last one. We denote by $C^n_{Q_X}(i, .)$ the ith partition of the codebook, and by $C^n_{Q_X}(i, j)$ the jth codeword in the ith partition. For each $x^n \in T_{Q_X}$, let i_{x^n} and j_{x^n} denote, respectively, the index of the partition containing the codeword associated with x^n and the index of the codeword within the partition (Note that if more than one codeword can be associated with x^n, we fix any one of them arbitrarily). Finally, let $m(Q_X, i, j)$ be a message consisting of the following:

- $\left\lceil \log_2 |\mathcal{Q}^n_{\mathcal{X}}| \right\rceil$ bits to describe the type Q_X.

- $\left\lceil \log_2 \left\lceil \left|C^n_{Q_X}\right| / 2^{nr} \right\rceil \right\rceil$ bits to describe the index i, where $1 \le i \le \left\lceil \left|C^n_{Q_X}\right| / 2^{nr} \right\rceil$.
- $\left\lceil \log_2 \left|C^n_{Q_X}(i,.)\right| \right\rceil$ bits to describe the index j, where $0 \le j \le \exp_2 \left\lceil \log_2 \left|C^n_{Q_X}(i,.)\right| \right\rceil - 1$.

Now, for any $\delta \in \mathbb{R}$, let $\mathcal{Q}(\alpha,\delta) = \{Q_X : D(Q_X||P) \le \alpha + \delta\}$, $\mathcal{Q}_n(\alpha,\delta) = \{Q_X \in \mathcal{Q}^n_X : D(Q_X||P) \le \alpha + \delta\}$, and consider the following lemma.

Lemma 3

$$\lim_{\delta \to 0} \max_{Q_X \in \mathcal{Q}(\alpha,\delta)} R(Q_X, D) = \max_{Q_X \in \mathcal{Q}(\alpha,0)} R(Q_X, D).$$

Proof: This follows directly from Propositions 22 and 23. □

Now let $\delta > 0$ be such that $\max_{Q_X \in \mathcal{Q}(\alpha,\delta)} R(Q_X, D) < R$ (Such δ exists by Lemma 3 and (A4)). Finally, for each sequence x^n, let $s(x^n) = \left\lceil \log_2 \left|C^n_{Q_X}(i_{x^n},.)\right| \right\rceil$, and let $K_{s(x^n)}$ be the first $s(x^n)$ bits of K_n. The transmitter encodes as follows. Given x^n, if $Q_{x^n} \in \mathcal{Q}_n(\alpha,\delta)$, then

$$f(x^n, K_n) = m\left(Q_{x^n}, i_{x^n}, j_{x^n} \oplus K_{s(x^n)}\right), \tag{4.2}$$

where the XOR-operation is performed bitwise. Note that, in this case, the legitimate receiver can retrieve the type of the transmitted sequence and the index of the bin from the first two parts of the message, and the index of the sequence within the bin using the last part of the message and the key K_n, so that $h(M_n, K_n) = C^n_{Q_{x^n}}(i_{x^n}, j_{x^n})$. Now, consider an $m_0 \in \mathcal{M}_n$ that has not been used by the previous encoding (Assumption (A4) and the choice of δ ensures the existence of such m_0). Then, for all x^n such that $Q_{x^n} \notin \mathcal{Q}_n(\alpha,\delta)$,

$$f(x^n, K_n) = m_0. \tag{4.3}$$

Remark 7 *To verify that the suggested scheme satisfies the excess distortion probability*

constraint, consider the following:

$$\mathbf{Pr}(d(X^n, Y^n) > D) \leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} P(Q) \leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} 2^{-nD(Q_X\|P)} \leq (n+1)^{|\mathcal{X}|} 2^{-n(\alpha+\delta)} < 2^{-n\alpha},$$

where the last inequality holds for large enough n.

Effectively, we are leaking the first two parts of the message Q_{X^n} and i_{X^n}, and *hiding* completely the last part j_{X^n}. Since there are only polynomially many types, the first part does not affect the normalized leakage. The second part, however, consists roughly of $R(Q, D) - r$ bits, whenever $R(Q, D) > r$; otherwise, i.e., when $R(Q, D) \leq r$, there is only one bin and there is no information to be leaked.

For a more rigorous analysis, let P_f be the induced joint probability distribution of (X^n, M_n). Then, for x^n satisfying $Q_{x^n} \in \mathcal{Q}_n(\alpha, \delta)$, we get from (4.2):

$$P_f\left(m(Q_{x^n}, i_{x^n}, j)\middle| x^n\right) = 2^{-s(x^n)}, \quad 0 \leq j \leq 2^{s(x^n)} - 1.$$

Let $S(x^n) = 2^{s(x^n)}$. Note that we can equivalently denote $S(x^n)$ by $S(Q_{x^n}, i_{x^n})$, since the dependence on the sequence is only through the type and the index of the

bin. Therefore, we get

$$\begin{aligned}
\exp_2\{\mathcal{L}(X^n \to M_n)\} &= \sum_{m \in \mathcal{M}_n} \max_{x^n \in \mathcal{X}^n} P_f(m|x^n) \\
&= \max_{x^n \in \mathcal{X}^n} P_f(m_0|x^n) + \sum_{\substack{Q_X \in \\ \mathcal{Q}_n(\alpha,\delta)}} \sum_{i=1}^{\left\lceil |\mathcal{C}^n_{Q_X}|/2^{nr} \right\rceil_{S(Q_X,i)-1}} \sum_{j=0} \max_{x^n \in \mathcal{X}^n} P_f(m(Q_X, i, j)|x^n) \\
&= 1 + \sum_{\substack{Q_X \in \\ \mathcal{Q}_n(\alpha,\delta)}} \sum_{i=1}^{\left\lceil |\mathcal{C}^n_{Q_X}|/2^{nr} \right\rceil_{S(Q_X,i)-1}} \sum_{j=0} S(Q_X, i)^{-1} \\
&\leq 1 + \sum_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} (2^{n(R(Q_X,D)+\epsilon-r)} + 1) \\
&\leq 1 + 2 \sum_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} 2^{n \max\{R(Q_X,D)+\epsilon-r,0\}} \\
&\leq 4(n+1)^{|\mathcal{X}|} \exp_2\{n \max_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} [R(Q_X, D) + \epsilon - r]^+\}. \qquad (4.4)
\end{aligned}$$

Taking the limit as n tends to infinity, and noting that ϵ and δ were arbitrary, we get that

$$L \leq \max_{Q: D(Q\|P) \leq \alpha} [R(Q, D) - r]^+,$$

where the inequality follows from Lemma 3 and the following lemma, the simple of proof of which is omitted.

Lemma 4

$$\lim_{n \to \infty} \max_{Q \in \mathcal{Q}^n_{\mathcal{X}}: D(Q\|P) \leq \alpha} R(Q, D) = \max_{Q: D(Q\|P) \leq \alpha} R(Q, D).$$

4.4 Converse Proof of Theorem 13

We now show that L is lower-bounded by the right-hand side of (4.1). To that end, consider any valid encoding function f. To lower-bound $\mathcal{L}(X^n \to M_n)$, we

consider a specific $P_{U|X^n}$. In particular, we consider the "shattering" $P_{U|X^n}$ given in (2.3). Recall

$$P_{U|X^n}((i_u, j_u)|x^n) = \begin{cases} \frac{p^\star}{P(x^n)}, & i_u = x^n,\ 1 \le j_u \le \lfloor k(x^n) \rfloor, \\ 1 - \frac{(\lceil k(x^n) \rceil - 1)p^\star}{P(x^n)}, & i_u = x^n,\ j_u = \lceil k(x^n) \rceil, \\ 0, & i_u \neq x^n,\ 1 \le j_u \le \lceil k(i_u) \rceil. \end{cases}$$

Remark 8 *The given $P_{U|X^n}$ achieves the supremum in the definition of $\mathcal{L}(X^n \to M_n)$, although this is not needed here.*

Therefore, $\max_{u \in \mathcal{U}} P_U(u) = p^\star$. We will also consider a sub-optimal guessing function for U. The scheme is as follows: the eavesdropper first tries to guess the key K_n by choosing an element uniformly at random from $\{0,1\}^{nr}$. We denote this guess by $\tilde{K}_n$. Then, proceeding by assuming that the key guess was correct, s/he tries to guess the sequence x^n using a guessing function given by Lemma 5 below. We denote this stage by g_1. Finally, again proceeding by assuming that the source sequence guess was correct, the eavesdropper attempts to guess U by using the MAP rule. We denote this stage by g_2, and we get for each $x^n \in \mathcal{X}^n$,

$$g_2(x^n) = (x^n, 1), \text{ and } \mathbf{Pr}(g_2(x^n) = U^n | x^n) = p^\star / P(x^n). \tag{4.5}$$

Lemma 5 *There exists a function $g_1 : \mathcal{Y}^n \to \mathcal{X}^n$ such that, for all (x^n, y^n) satisfying $d(x^n, y^n) \le D$, $\mathbf{Pr}\,(x^n = g(y^n)) \ge c_n 2^{-n(H_{Q_{x^n}}(X) - R(Q_{x^n}, D))}$, where $c_n = (n+1)^{-|\mathcal{X}||\mathcal{Y}|(|\mathcal{X}|+1)}$.*

Proof: This is an application of Lemma 5 in [23]. In particular, we set in Lemma 5 $\mathcal{V}$ to be $\mathcal{X}$, d_e to be the Hamming distortion function, and D_e to be zero. Then, $I_{P_n^\star(Q_{x^n y^n})}(X; V|Y)$ (as defined in [23]) satisfies:

$$\begin{aligned} I_{P_n^\star(Q_{x^n y^n})}(X; V|Y) = H_{Q_{x^n y^n}}(X|Y) &= H_{Q_{x^n}}(X) - H_{Q_{x^n}}(X) + H_{Q_{x^n y^n}}(X|Y) \\ &\le H_{Q_{x^n}}(X) - R(Q_{x^n}, D). \end{aligned}$$

□

To analyze the above scheme, fix $\epsilon > 0$, and let P_f denote the induced joint probability on (X^n, K_n, M_n). Furthermore, without loss of generality, we can assume that the decoding function h is a deterministic function of M_n and K_n. Finally, define

$$\mathcal{M}_D(x^n, k) = \{m \in \mathcal{M}_n : d(x^n, h(m,k)) \le D\}, \quad x^n \in \mathcal{X}^n, k \in \mathcal{K}_n, \tag{4.6}$$

$$\text{and } \mathcal{A} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : d(x^n, y^n) > D\}. \tag{4.7}$$

Letting g be the concatenation of the two stages, we get

$$\begin{aligned}
&\mathbf{Pr}(U = g(M)) \\
&= \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_n} P(x^n) P_{U|X^n}(u|x^n) P_{K_n}(k) P_f(m|x^n, k) P(u = g(m)|x^n, m, k) \\
&\ge \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{U|X^n}(u|x^n) P_{K_n}(k) P_f(m|x^n, k) P(u = g(m)|x^n, m, k) \\
&\ge \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{U|X^n}(u|x^n) P_{K_n}(k) P_f(m|x^n, k) P(\tilde{K}_n = k). \\
&\qquad\qquad P(g_1(h(m,k)) = x^n) P(g_2(x^n) = u|x^n) \\
&\overset{(a)}{\ge} c_n \sum_{x^n \in \mathcal{X}^n} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{K_n}(k) P_f(m|x^n, k) 2^{-nr} 2^{-n(H_{Q_{x^n}}(X) - R(Q_{x^n}, D))} p^\star / P(x^n) \\
&= c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_X^n} \sum_{x^n \in T_{Q_X}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{K_n}(k) P_f(m|x^n, k) 2^{-n(H_{Q_X}(X) - R(Q_X, D))} / P(x^n) \\
&= c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_X^n} \sum_{x^n \in T_{Q_X}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{K_n}(k) P_f(m|x^n, k) 2^{n(R(Q_X, D) + D(Q_X \| P))} \\
&= c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_X^n} 2^{n(R(Q_X, D) + D(Q_X \| P))} P_f(\mathcal{A}^c \cap T_{Q_X}),
\end{aligned} \tag{4.8}$$

where (a) follows from Lemma 5, (4.5), and (4.52). Now, note that for any Q,

$$\begin{aligned}
P_f(\mathcal{A}^c | T_Q) = 1 - P_f(\mathcal{A}|T_Q) &\ge 1 - \min\{1, P_f(\mathcal{A})/P(T_Q)\} \\
&\ge 1 - \min\{1, 2^{-n(\alpha - D(Q\|P) - \frac{|\mathcal{X}|}{n} \log_2(n+1))}\} \\
&= \max\{0, 1 - 2^{-n(\alpha - D(Q\|P) - \frac{|\mathcal{X}|}{n} \log_2(n+1))}\}.
\end{aligned}$$

Then, continuing (4.8), we get

$$
\begin{aligned}
&\mathbf{Pr}(U = g(M)) \\
&\geq c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_X^n} 2^{n(R(Q_X,D)+D(Q_X\|P))} P(T_{Q_X}) \max\{0, 1 - 2^{-n(\alpha - D(Q_X\|P) - \frac{|\mathcal{X}|}{n}\log_2(n+1))}\} \\
&\overset{(a)}{\geq} c'_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_n(\alpha,-\epsilon)} 2^{nR(Q_X,D)} (1 - 2^{-n(\alpha - D(Q_X\|P) - \frac{|\mathcal{X}|}{n}\log_2(n+1))}) \\
&\overset{(b)}{\geq} c'_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}_n(\alpha,-\epsilon)} 2^{nR(Q_X,D)} (1/2) \\
&\geq (c'_n p^\star/2) \max_{Q_X \in \mathcal{Q}_n(\alpha,-\epsilon)} \exp_2\{n(R(Q_X, D) - r)\},
\end{aligned} \tag{4.9}
$$

where (a) and (b) hold for large enough n, and $c'_n = (n+1)^{-|\mathcal{X}|} c_n$. Finally taking the ratio of $\mathbf{Pr}(U = g(M))$ and $\max_u P_U(u)$, and taking the limit as n tends to infinity, and noting that ϵ is arbitrary, we get

$$L \geq \max_{Q: D(Q\|P) \leq \alpha} R(Q, D) - r,$$

where the inequality follows from Lemmas 3 and 4. Since L is positive by definition,

$$L \geq [\max_{Q: D(Q\|P) \leq \alpha} R(Q, D) - r]^+ = \max_{Q: D(Q\|P) \leq \alpha} [R(Q, D) - r]^+.$$

4.5 Expected Distortion Constraint

We discuss a variation of the above problem. Instead of requiring a decaying probability of violating the distortion constraint, we could require that the distortion constraint holds only in expectation—as is common in many works in the literature. In that case, we modify assumption (A4) to be:

(A4′) $R > R(P, D)$.

Then, the following theorem holds.

Theorem 14 *Under assumptions (A1)-(A3) and (A4'), for any DMS P, and distortion function d with associated distortion level $D \geq D_{\min}$:*

$$L(P, D, \overrightarrow{R}) = [R(P, D) - r]^+ \; bits. \tag{4.10}$$

The achievability argument follows by a similar manner as the one given in Section 4.3. However, instead of encoding on a type-by-type basis, we simply use a good rate-distortion code that satisfies the expected distortion requirement. As above, we divide the codebook into bins of size 2^{nr}, except for possibly the last one. Then, an analysis similar to (4.4) yields $L(P, D, \overrightarrow{R}) \leq [R(P, D) - r]^+$. As for the lower bound, we use the fact that $I_\infty(X; Y) \geq I(X; Y)$ [50]. This problem, with mutual information replacing maximal leakage, has already been solved by Schieler and Cuff [40]. More specifically, Corollary 5 of [40] yields that the optimal normalized mutual information is indeed given by $[R(P, D) - r]^+$.

4.5.1 Suboptimality of Memoryless Schemes

With the expected distortion constraint, one might venture that a memoryless scheme, in which the encoder simply passes the source through i.i.d. copies of the optimal test channel, is optimal. Counter to this common intuition, we show that this is not the case when the objective is maximal leakage.

To that end, consider the information blurring system, and suppose the source is binary and the distortion constraint is the Hamming distance. We first derive the "optimal test channel" with respect to maximal leakage. More specifically, let $0 \leq D \leq p \leq 1/2$, and let $X \sim \text{Ber}(p)$. We want to minimize leakage

from X to an output Y subject to a Hamming distortion constraint:

$$\begin{aligned} &\text{minimize } \mathcal{L}(X \to Y) \\ &\text{subject to } \mathbf{Pr}(X \neq Y) \leq D. \end{aligned} \tag{4.11}$$

We show in Appendix B that for the optimal mechanism $P^{\star}_{Y|X}$, we get

$$\mathcal{L}(X \to Y^{\star}) = \log_2(2 - D/p) \text{ bits.} \tag{4.12}$$

Now, consider a source X^n that is i.i.d $Ber(p)$, and let

$$\begin{aligned} L_n = \min_{P_{Y^n|X^n} : P_{Y^n|X^n} = \prod_{i=1}^n P^{(i)}_{Y_i|X_i}} &\frac{1}{n}\mathcal{L}(X^n \to Y^n) \\ \text{subject to } &\mathbf{E}[d(X^n, Y^n)] \leq D. \end{aligned} \tag{4.13}$$

The minimization only considers product distributions, that is, it corresponds to schemes that can be implemented memorylessly.

Lemma 6

$$L_n = 1 - D/p \text{ bits.}$$

On the other hand, Theorem 14 implies that the optimal normalized limit is $R(P, D) = H(p) - H(D)$. Since $H(p) - H(D) < 1 - D/p$ (where the inequality can be checked using elementary calculus), memoryless schemes are suboptimal.

Proof: For any $P_{Y^n|X^n}$ in the minimization, let $D_i = \mathbf{E}[d(X_i, Y_i)]$. Without loss

of generality, we can assume $D_i \leq p$. Then,

$$\begin{aligned}
\mathcal{L}(X^n \to Y^n) = \sum_{i=1}^{n} \mathcal{L}(X_i \to Y_i) &\overset{(a)}{\geq} \sum_{i=1}^{n} \log_2(2 - D_i/p) \\
&= \sum_{i=1}^{n} \log_2\left(2 - \frac{(D_i/p)p + (1 - D_i/p)0}{p}\right) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} (D_i/p)\log_2(1) + (1 - D_i/p)\log_2(2) \\
&= \sum_{i=1}^{n} (1 - D_i/p) \\
&\overset{(c)}{\geq} n(1 - D/p),
\end{aligned}$$

where (a) follows from (4.12), (b) follows from the fact that $\log_2(2 - D/p)$ is concave in D, and (c) follows from the constraint in (4.13). Note that the lower-bound can be achieved by sending $n(1 - D/p)$ bits as is, and sending zeros for the remaining nD/p bits. ■

4.6 Known Distortion Function

4.6.1 Overview

Suppose the eavesdropper wants to estimate the source X^n up to a distortion function that we know a priori, which is the classical setup in the literature. In that case, the relevant metric is the probability that the eavesdropper's guess is successful, i.e., the distortion it incurs is below a given level. The primary user (i.e., the transmitter legitimate-receiver pair) aims then to minimize that probability. Since computing the exact probability is quite difficult, this section will be mainly concerned with asymptotic analysis: we will derive the rate of decay (i.e., the exponent) of the probability of a successful guess.

We first study the information blurring system, as it more closely resembles the side channel mitigation problem (cf. Section 4.2.1). For a discrete memoryless source (DMS), we provide a single-letter characterization of the optimal exponent. We show that the problem is related to source coding with side information. Essentially, the eavesdropper first attempts to guess the joint type of X^n and Y^n. S/he, then, "pretends" that Y^n is received through a memoryless channel the probability law of which is the conditional probability $P(Y|X)$ induced by the joint type. The problem can be viewed at this point as compression with side information, so the eavesdropper picks a codeword from an optimal rate-distortion code. The primary user's objective, therefore, is to supply the "worst" side information. Moreover, we demonstrate asymptotically-optimal universal schemes for both the legitimate receiver and the eavesdropper. The schemes are universal in the sense that they do not depend on the source statistics.

Next, we extend the study to the full setup of the Shannon cipher system. The eavesdropper has full knowledge of the encryption system, except for the realization of the key and the realization of X^n. Since the transmitter is subject to a rate constraint, we allow the primary user to violate the distortion constraint, but restrict the probability of such event to be exponentially decaying. We again derive a single-letter characterization of the optimal exponent (cf. Theorem 16), and demonstrate asymptotically-optimal strategies for both the primary user and the eavesdropper. In particular, similarly to the previous setting, the transmitter operates on a type-by-type basis and associates with each type a rate-distortion code, the construction of which is based on the conditional probability law that provides the worst side information and satisfies the rate constraint (however, types with low enough probability are discarded, by associating a dummy messsage to all the source sequences belonging to such types). To make

use of the shared key, we (randomly) generate many instances of such codes, and use the secret key to randomize the choice of the code selected for encoding X^n. We also investigate conditions under which the resulting codes are optimal rate-distortion codes. As for the eavesdropper, we show that one of the following two schemes is optimal. The first consists of generating a blind guess, i.e., completely ignoring the public message. The second consists of guessing the value of the key to reproduce the reconstruction at the legitimate receiver, and then applying the strategy developed in the first part of the paper.

We note that Theorem 16 subsumes Theorem 15 by setting the key rate to be zero, and the channel rate to be high enough. We nevertheless present them separately for two reasons. We believe the information blurring system to be of independent interest, as it corresponds to problems different from the Shannon cipher system (e.g., the SSH timing attack). As such, Theorem 15 can serve as a baseline for future refinements of this model (say, by requiring the encoding to be causal). Moreover, it significantly simplifies the exposition of the results, by first revealing the connection to source coding with side information and then introducing the key and the rate constraint.

Finally, it should be noted that, concurrently with the derivation of this work, Weinberger and Merhav studied the Shannon cipher system with lossy communication [53] (i.e, the setup of the second part of this section), and independently suggested the same secrecy metric we proposed. Furthermore, they allowed a variable key rate. They derived the optimal exponent in general, as is done here. However, the suggested scheme herein and its subsequent analysis are significantly simpler. In particular, our scheme uses a traditional random coding construction followed by a separate key-based randomization.

4.6.2 The Information Blurring System

In this section, we assume the transmitter and legitimate receiver must satisfy the distortion constraint almost surely. That is, every realization of (X^n, Y^n) satisfies $d(X^n, Y^n) \leq D$. Furthermore, let $\mathcal{V}$ be the alphabet associated with the eavesdropper. The latter, with an associated distortion function $d_e : \mathcal{X} \times \mathcal{V} \to \mathbb{R}_+$, observes Y^n and generates a guess $V^n = g(Y^n)$, aiming to have $d_e(X^n, V^n) = \frac{1}{n}\sum_{i=1}^{n} d_e(X_i, V_i) \leq D_e$ for a given distortion level D_e.

It is assumed that the eavesdropper knows the source statistics and the primary user's encoding function f. The secrecy metric we adopt is the probability that the eavesdropper makes a successful guess, i.e., $\mathbf{Pr}(d_e(X^n, V^n) \leq D_e)$. The primary user's objective is to minimize this probability. So, the problem can be written as:

$$\min_{f_n} \max_{g_n} \mathbf{Pr}\left(d_e\Big(X^n, g_n(f_n(X^n))\Big) \leq D_e\right).$$

We characterize the highest achievable exponent of the probability of a successful guess under assumptions (A1)-(A3) with the following update to (A3):

(A3) The distortion functions d and d_e are bounded, i.e., there exists $D_{\max}$ and $D_{e,\max}$ such that, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $v \in \mathcal{V}$, $d(x, y) \leq D_{\max}$ and $d_e(x, v) \leq D_{e,\max}$. Moreover, $D \geq D_{\min}$, where $D_{\min} = \max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} d(x, y)$. Similarly, $D_e \geq D_{e,\min}$, where $D_{e,\min} = \max_{x\in\mathcal{X}} \min_{v\in\mathcal{V}} d_e(x, v)$.

We denote the optimal exponent by $E(P, D, D_e)$, where P is the source distribution, i.e.,

$$E(P, D, D_e) = \lim_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\left(d_e\Big(X^n, g_n(f_n(X^n))\Big) \leq D_e\right). \tag{4.14}$$

The existence of the limit will be seen later.

We will show that the problem is related to source coding with side information, where Y^n acts as side information for the eavesdropper. Therefore, the primary user's job is to provide the "worst" side information subject to a distortion constraint of his/her own. To this end, we denote the *conditional* rate-distortion function as:

$$R(P_{XY}, D_e) = \min_{\substack{P_{V|X,Y}: \\ \mathbf{E}[d_e(X,V)] \leq D_e}} I(X; V|Y), \tag{4.15}$$

and define the quantity $R(P_X, D, D_e)$ as:

$$R(P_X, D, D_e) = \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \leq D}} R(P_{XY}, D_e). \tag{4.16}$$

Roughly speaking, when the joint type of X^n and Y^n is P_{XY}, the eavesdropper can restrict the guessing space to $2^{nR(P_{XY}, D_e)}$ reconstruction sequences, knowing that at least one of them must satisfy the distortion constraint. The maximization in (4.16) corresponds to the primary user's goal of maximizing that quantity.

We prove the following properties of $R(P_{XY}, D_e)$ and $R(P_X, D, D_e)$ in Appendix C.1.

Proposition 7 *In the following statements, the domains of D and D_e are $[D_{\min}, +\infty)$ and $[D_{e,\min}, +\infty)$, respectively.*

(P1) For fixed P_{XY}, $R(P_{XY}, D_e)$ is a finite-valued, non-increasing convex function of D_e. Furthermore, $R(P_{XY}, D_e)$ is a uniformly continuous function of the pair (P_{XY}, D_e).

(P2) For fixed P_X, $R(P_X, D, D_e)$ is a finite-valued function of (D, D_e). Moreover, for fixed D_e, $R(P_X, D, D_e)$ is a uniformly continuous function of the pair (P_X, D).

(P3) $R_e(P_X, D_e) - R(P_X, D) \leq R(P_X, D, D_e) \leq R_e(P_X, D_e)$, *where* $R(P_X, D)$ *and* $R_e(P_X, D_e)$ *are the rate-distortion functions corresponding to the distortion constraints* d *and* d_e*, respectively.*

Our main result is the characterization of the optimal exponent as follows:

Theorem 15 *Under assumptions (A1)-(A3), for any DMS* P*, and distortion functions* d *and* d_e *with associated distortion levels* $D \geq D_{\min}$ *and* $D_e \geq D_{e,\min}$*, corresponding respectively to the primary user and the eavesdropper:*

$$E(P, D, D_e) = \min_{Q} D(Q||P) + R(Q, D, D_e), \tag{4.17}$$

where Q *ranges over all probability distributions on the source alphabet, and* $R(Q, D, D_e)$ *is as defined in* (4.16)*.*

Remark 9 *We do not require any* ϵ*-backoff for* D *or* D_e *to characterize the associated exponent.*

An interesting feature of Theorem 15 is the emergence of mutual information as part of the solution in (4.17), even though the setup does not include any rate constraints. Moreover, an interesting contrast can be seen between the expression in (3.27) for the expected number of guesses metric and the expression in (4.17) for our metric. Indeed, the former evaluates the performance of a given scheme asymptotically by a weighted *best-case* scenario, whereas the latter evaluates it by a weighted *worst-case* scenario.

As an application of the theorem, we compute the *perfect secrecy* exponent, which we define as the best achievable exponent when the primary user is not

subject to any constraint and denote it by $E_0(P, D_e)$. To this end, we introduce a trivial distortion function: $d(x, y) = 0$, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Then, $R(Q, D) = 0$, for all Q and all $D \geq 0$. It then follows from (P3) of Proposition 7 that $R(Q, D, D_e) = R_e(Q, D_e)$ for all Q. Therefore,

$$E_0(P, D_e) = \min_Q D(Q\|P) + R_e(Q, D_e). \tag{4.18}$$

The next two subsections are devoted to proving Theorem 15. We first propose a scheme for the primary user and show that the induced exponent is lower-bounded by the right-hand side of (4.17). From the eavesdropper's point of view, this is a converse result. Similarly, we propose a scheme for the eavesdropper and show that the induced exponent is upper-bounded by the right-hand side of (4.17), which establishes the desired result.

Achievability

$$\text{Let } E^-(P, D, D_e) = \liminf_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\left(d_e\left(X^n, g_n(f_n(X^n))\right) \leq D_e\right). \tag{4.19}$$

We will show that $E^-(P, D, D_e) \geq \min_Q D(Q\|P) + R(Q, D, D_e)$.

The primary user will operate on the source sequences on a type-by-type basis. For each type $Q_X \in \mathcal{Q}^n_{\mathcal{X}}$, we create a rate distortion code C_{Q_X} to cover each sequence in T_{Q_X} as follows. We associate with Q_X a joint type Q_{XY} from $\mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_X, D)$:[2]

$$\mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_X, D) = \{P_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}} : P_X = Q_X,\ \mathbf{E}_{P_{XY}}[d(X, Y)] \leq D\}. \tag{4.20}$$

The code is then constructed from T_{Q_Y} as given by the following lemma, which bounds the size of the code.

[2] Assumption (A3) guarantees that $\mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_X, D)$ is nonempty for any Q_X.

Lemma 8 *Given $\epsilon > 0$, there exists $n_0(\epsilon, |\mathcal{X}|, |\mathcal{Y}|)$ such that for any $n \geq n_0$, for each joint type $Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}$, there exists a code $(y_1^n, y_2^n, \cdots, y_N^n)$ such that $N \leq 2^{n(I_{Q_{XY}}(X;Y)+\epsilon)}$, and for all $x^n \in T_{Q_X}$, there exists i satisfying $(x^n, y_i^n) \in T_{Q_{XY}}$.*

We later prove a stronger result, Lemma 15, in Appendix C.5.

Remark 10 *One might be tempted to use an optimal rate-distortion code for each type Q_X, presuming that this choice is best at preserving secrecy since it achieves optimal compression, i.e., it only sends the necessary information. However, the problem is more subtle since the "redundancy of information" depends on the eavesdropper's distortion constraint d_e. The optimal choice of Q_{XY} will be revealed when analyzing the eavesdropper's optimal strategy.*

Now, fix $\epsilon > 0$ and let n be at least as large as n_0 in Lemma 8. We will denote by $C^n_{Q_X}$ the rate distortion code associated with type Q_X. Thus, the function f of the primary user is as follows: each sequence x^n is mapped to a sequence $y^n \in C^n_{Q_{x^n}}$ satisfying $Q_{x^n y^n} = Q_{XY}$ (where Q_{XY} is associated with Q_{x^n}) and subsequently $d(x^n, y^n) \leq D$.

To determine the eavesdropper's optimal guess, we define $B_{D_e}(v^n) = \{x^n \in \mathcal{X}^n : d_e(x^n, v^n) \leq D_e\}$. Then, for each observed y^n, the optimal rule is given by

$$g(y^n) = \operatorname*{argmax}_{v^n \in \mathcal{V}^n} \sum_{x^n \in B_{D_e}(v^n)} p(x^n|y^n).$$

This can be understood as the MAP rule, and we denote in the remainder by[3] g_o (where "o" stands for optimal). To upper-bound the probability of a correct

[3]The MAP rule depends on f and thus should be denoted by $g_{o,f}$. Since this is obvious, we drop the subscript f for notational convenience.

guess, we consider a *genie-aided* rule that is aware of the type of the transmitted source sequence. That is, the genie-aided MAP rule yields

$$g_o(y^n, Q_X) = \underset{v^n \in \mathcal{V}^n}{\operatorname{argmax}} \sum_{x^n \in B_{D_e}(v^n) \cap Q_X} p(x^n | y^n, X^n \in T_{Q_X}).$$

Remark 11 *One should not expect the upper bound to be loose since there are only polynomially many types in n, so that the exponent is not affected.*

For a given y^n, let $f_{Q_X}^{-1}(y^n) = \{x^n \in T_{Q_X} : f(x^n) = y^n\}$ be the set of sequences in T_{Q_X} that are mapped to it. Then, the observation of y^n implies that $X^n \in f_{Q_X}^{-1}(y^n)$, and the genie-aided MAP rules makes a successful guess if $X^n \in B_{D_e}(g_o(y^n, Q_X))$. Therefore, we will derive an upper bound on the maximum possible size of the intersection of these two sets. First, note that, $x^n \in T_{Q_X}$ and $f(x^n) = y^n$ implies that $Q_{x^n y^n} = Q_{XY}$, where Q_{XY} is the joint type associated with Q_X. So $f_{Q_X}^{-1}(y^n) \subseteq T_{Q_{X|Y}}(y^n) \triangleq \{x^n \in T_{Q_X} : (x^n, y^n) \in T_{Q_{XY}}\}$. Now, consider any $v^n \in \mathcal{V}^n$,

$$\begin{aligned}
\left| B_{D_e}(v^n) \bigcap f_{Q_X}^{-1}(y^n) \right| &\leq \left| B_{D_e}(v^n) \bigcap T_{Q_{X|Y}}(y^n) \right| \\
&\overset{(a)}{=} \sum_{\substack{P_{XYV} \in \mathcal{Q}^n_{\mathcal{X},\mathcal{Y},\mathcal{V}}: \\ P_{XY} = Q_{XY} \\ \mathbf{E}_{P_{XYV}}[d_e(X,V)] \leq D_e \\ P_{YV} = Q_{y^n v^n}}} \sum_{\substack{x^n: \\ (x^n, y^n, v^n) \in T_{P_{XYV}}}} 1 \\
&\overset{(b)}{=} \sum_{\substack{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e): \\ P_{YV} = Q_{y^n v^n}}} |T_{P_{X|V,Y}}(v^n, y^n)| \\
&\leq (n+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{V}|} \max_{\substack{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e): \\ P_{YV} = Q_{y^n v^n}}} |T_{P_{X|V,Y}}(v^n, y^n)| \\
&\overset{(b)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{V}|} \max_{\substack{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e): \\ P_{YV} = Q_{y^n v^n}}} 2^{n H_{P_{XYV}}(X|V,Y)}, \qquad (4.21)
\end{aligned}$$

where

(a) follows from the fact that $(x^n, y^n, v^n) \in T_{P_{XYV}} \Rightarrow P_{YV} = Q_{y^n v^n}$.

(b) follows from the definition of $\mathcal{Q}^n(Q_{XY}, D_e)$ as:

$$\mathcal{Q}^n(Q_{XY}, D_e) = \{P_{XYV} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}\mathcal{V}} : P_{XY} = Q_{XY},\ \mathbf{E}_{P_{XYV}}[d_e(X, V)] \leq D_e\}. \tag{4.22}$$

(b) follows from Lemma 1.2.5 in [11].

Therefore, for large enough n, we get

$$\max_{v^n \in \mathcal{V}^n} \left| B_{D_e}(v^n) \bigcap f_{Q_X}^{-1}(y^n) \right| \leq \max_{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e)} 2^{n\left(H_{P_{XYV}}(X|V,Y)+\epsilon\right)}. \tag{4.23}$$

Let $P_n^\star(Q_{XY})$ be the joint type achieving the max in (4.23), where the dependence on D_e is suppressed since it is fixed throughout the analysis. We can now upper-bound the probability that the eavesdropper makes a successful guess as follows:

$$\begin{aligned}
\mathbf{Pr}\Big(d_e\Big(X^n, g_o(f(X^n))\Big) \leq D_e\Big) &\leq \mathbf{Pr}\Big(d_e\Big(X^n, g_o\left(f(X^n), Q_{X^n}\right)\Big) \leq D_e\Big). \\
&= \sum_{x^n \in \mathcal{X}^n} P(x^n) \mathbf{1}\left\{x^n \in B_{D_e}\left(g_o(f(x^n), Q_{x^n})\right)\right\} \\
&= \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{y^n \in C^n_{Q_X}} \sum_{\substack{x^n \in T_{Q_X}: \\ f(x^n)=y^n}} P(x^n) \mathbf{1}\left\{x^n \in B_{D_e}\left(g_o(y^n, Q_X)\right)\right\} \\
&\overset{(a)}{\leq} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{y^n \in C^n_{Q_X}} 2^{n\left(-D(Q_X\|P)-H_{Q_X}(X)\right)} 2^{n\left(H_{P_n^\star(Q_{XY})}(X|V,Y)+\epsilon\right)} \\
&\overset{(b)}{\leq} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{n\left(I_{Q_{XY}}(X;Y)+\epsilon-D(Q_X\|P)-H_{Q_X}(X)+H_{P_n^\star(Q_{XY})}(X|V,Y)+\epsilon\right)} \\
&= \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{n\left(-D(Q_X\|P)-H_{Q_{XY}}(X|Y)+H_{P_n^\star(Q_{XY})}(X|V,Y)+2\epsilon\right)} \\
&= \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{-n\left(D(Q_X\|P)+I_{P_n^\star(Q_{XY})}(X;V|Y)-2\epsilon\right)},
\end{aligned} \tag{4.24}$$

where

(a) follows from (4.23).

(b) follows from Lemma 8.

To interpret the exponent in (4.24), note that $P_n^\star(Q_{XY})$ minimizes $I(X;V|Y)$ over $\mathcal{Q}^n(Q_{XY}, D_e)$ (follows readily from (4.23)). Therefore, $I_{P_n^\star(Q_{XY})}(X;V|Y)$ is roughly $R(Q_{XY}, D_e)$. The eavesdropper's scheme can then be seen as picking a codeword from an optimal rate-distortion code that uses side information generated according to $Q_{Y|X}$.

Since Q_{XY} is the choice of the primary user, who is interested in maximizing the exponents in (4.24), we define for each $Q_X \in \mathcal{Q}_{\mathcal{X}}^n$:

$$Q^\star(Q_X) = \operatorname*{argmax}_{Q_{XY} \in \mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X, D)} I_{P_n^\star(Q_{XY})}(X;V|Y),$$

where we have again suppressed the dependence on D and D_e in the notation.

Remark 12 *The maximization does not depend on the source statistics, and consequently neither does the proposed encoding function f.*

With a slight abuse of notation, we rewrite $P_n^\star(Q^\star(Q_X))$ as $P_n^\star(Q_X)$ to get

$$I_{P_n^\star(Q_X)}(X;V|Y) = \max_{\substack{Q_{XY} \in \\ \mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X, D)}} I_{P_n^\star(Q_{XY})}(X;V|Y) = \max_{\substack{Q_{XY} \in \\ \mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X, D)}} \min_{\substack{Q_{XYV} \in \\ \mathcal{Q}^n(Q_{XY}, D_e)}} I_{Q_{XYV}}(X;V|Y). \quad (4.25)$$

We can now rewrite (4.24) as

$$\begin{aligned}
&\mathbf{Pr}\Big(d_e\Big(X^n, g_o(f(X^n))\Big) \\
&\leq D_e\Big) \leq \sum_{Q_X \in \mathcal{Q}_{\mathcal{X}}^n} 2^{-n\left(D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y) - 2\epsilon\right)} \\
&\leq (n+1)^{|\mathcal{X}|} \max_{Q_X \in \mathcal{Q}_{\mathcal{X}}^n} 2^{-n\left(D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y) - 2\epsilon\right)} \\
&= (n+1)^{|\mathcal{X}|} \exp\left\{-n\left(-2\epsilon + \min_{Q_X \in \mathcal{Q}_{\mathcal{X}}^n} [D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y)]\right)\right\}. \quad (4.26)
\end{aligned}$$

Taking the limit as n goes to infinity, and noting that ϵ is arbitrary, we get

$$E^{-}(P, D, D_e) = \liminf_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\Big(d_e\Big(X^n, g_n(f_n(X^n))\Big) \le D_e\Big)$$
$$\ge \min_{Q} D(Q\|P) + R(Q, D, D_e), \tag{4.27}$$

where the last inequality follows from the following proposition, the proof of which is given in Appendix C.2.

Proposition 9

$$\lim_{n\to\infty} \min_{Q_X \in \mathcal{Q}_{\mathcal{X}}^n} [D(Q_X\|P) + I_{P_n^\star(Q_X)}(X; V|Y)] = \min_{Q} D(Q\|P) + R(Q, D, D_e).$$

Converse

$$\text{Let } E^{+}(P, D, D_e) = \limsup_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\Big(d_e\Big(X^n, g_n(f_n(X^n))\Big) \le D_e\Big). \tag{4.28}$$

We will now show that $E^{+}(P, D, D_e) \le \min_Q D(Q\|P) + R(Q, D, D_e)$. This means that the eavesdropper can achieve the exponent in (4.17) for any function f the primary user implements.

We propose a two-stage scheme for the eavesdropper. In the first stage, observing y^n, s/he tries to guess the joint type of x^n and y^n by choosing an element uniformly at random from the set $\mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_{y^n}, D)$, where

$$\mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_Y, D) = \{P_{XY} \in \mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n : P_Y = Q_Y,\ \mathbf{E}_{P_{XY}}[d(X, Y)] \le D\}. \tag{4.29}$$

The correct joint type must fall in this set since the restriction $d(X^n, Y^n) \le D$ is imposed on each realization of (X^n, Y^n). We denote the function corresponding to this stage by $g_1 : \mathcal{Y}^n \to \mathcal{Q}_{XY}^n$.

Table 4.1: Summary of the defined sets.

Set Notation	Description
$\mathcal{Q}^n_{\mathcal{XY}}(Q_X, D)$	$P_{XY} \in \mathcal{Q}^n_{\mathcal{XY}} : P_X = Q_X,\ \mathbf{E}_{P_{XY}}[d(X,Y)] \leq D.$
$\mathcal{Q}^n_{\mathcal{XY}}(Q_Y, D)$	$P_{XY} \in \mathcal{Q}^n_{\mathcal{XY}} : P_Y = Q_Y,\ \mathbf{E}_{P_{XY}}[d(X,Y)] \leq D.$
$\mathcal{Q}^n(Q_{XY}, D_e)$	$P_{XYV} \in \mathcal{Q}^n_{\mathcal{XYV}} : P_{XY} = Q_{XY},\ \mathbf{E}_{P_{XYV}}[d_e(X,V)] \leq D_e.$

Remark 13 *We differentiate between $\mathcal{Q}^n_{\mathcal{XY}}(Q_Y, D)$ and $\mathcal{Q}^n_{\mathcal{XY}}(Q_X, D)$ by their first argument. A summary of the defined sets is given in Table 4.1.*

The eavesdropper then proceeds assuming $g_1(y^n)$ is the correct joint type. S/he randomly chooses a sequence from a set that covers $T_{Q_{X|Y}}(y^n)$. To this end, we associate with each joint type Q_{XY} a joint type Q_{XYV} from $\mathcal{Q}^n(Q_{XY}, D_e)$ (cf. Table 4.1), and generate a sequence uniformly at random from $T_{Q_{V|Y}}(y^n)$, where $Q_{V|Y}$ is the conditional probability induced by Q_{XYV}. We denote the function corresponding to this stage by $g_2 : \mathcal{Y}^n \times \mathcal{Q}^n_{XY} \to \mathcal{V}^n$. Thus, $g(y^n) = g_2(y^n, g_1(y^n))$.

Remark 14 *The above strategy does not depend on the specifics of the function f implemented by the primary user, i.e., it only uses the fact that $d(X^n, f(X^n)) \leq D$. It is also independent of the source statistics.*

The following lemma lower-bounds the probability that $g_2(y^n, Q_{XY})$ generates a sequence V^n satisfying $d_e(x^n, V^n) \leq D_e$, for a given pair $(x^n, y^n) \in T_{Q_{XY}}$, i.e., assuming the eavesdropper guesses the joint type correctly.

Lemma 10 *Given joint type $Q_{XYV} \in \mathcal{Q}^n_{XYV}$ and $(x^n, y^n) \in T_{Q_{XY}}$, if V^n is chosen uniformly at random from $T_{Q_{V|Y}}(y^n)$, then $\mathbf{Pr}\left(V^n \in T_{Q_{V|X}}(x^n)\right) \geq c_n 2^{-nI_{Q_{XYV}}(X;V|Y)}$, where $c_n = (n+1)^{-|\mathcal{X}||\mathcal{Y}||\mathcal{V}|}$.*

Proof:

$$\mathbf{Pr}\left(V^n \in T_{Q_{V|X}}(x^n)\right) \geq \mathbf{Pr}\left(V^n \in T_{Q_{V|X,Y}}(x^n, y^n)\right) = \frac{\left|T_{Q_{V|X,Y}}(x^n, y^n)\right|}{\left|T_{Q_{V|Y}}(y^n)\right|} \geq \frac{c_n 2^{nH(V|X,Y)}}{2^{nH(V|Y)}}$$
$$= c_n 2^{-nI(X;V|Y)}.$$

where the second inequality follows from Lemma 1.2.5 in [11]. ■

Since the eavesdropper is interested in maximizing this probability, s/he will associate, with each Q_{XY}, the joint type achieving the maximum:

$$P_n^\star(Q_{XY}) = \underset{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e)}{\operatorname{argmin}} I(X; V|Y). \tag{4.30}$$

Note that this is the same joint type achieving the maximum in (4.23).

We can now lower-bound the probability that $x^n \in B_{D_e}(g(y^n))$, for a given pair (x^n, y^n) satisfying $d(x^n, y^n) \leq D$.

Lemma 11 *Given $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ satisfying $d(x^n, y^n) \leq D$, $\mathbf{Pr}\left(x^n \in B_{D_e}(g(y^n))\right) \geq c'_n 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)}$, where $c'_n = (n+1)^{-|\mathcal{X}||\mathcal{Y}|(|\mathcal{V}|+1)}$, $Q_{XY} = Q_{x^n y^n}$, and g is as described above.*

Proof:

$$\begin{aligned}
\mathbf{Pr}(x^n \in B_{D_e}(g(y^n))) &= \sum_{Q'_{XY} \in \mathcal{Q}^n_{XY}(Q_{y^n}, D)} p(g_1(y^n) = Q'_{XY}) p\left(x^n \in B_{D_e}\left(g_2(y^n, Q'_{XY})\right)\right) \\
&\geq p(g_1(y^n) = Q_{x^n y^n}) p\left(x^n \in B_{D_e}\left(g_2(y^n, Q_{x^n y^n})\right)\right) \\
&\geq (n+1)^{-|\mathcal{X}||\mathcal{Y}|} p\left(x^n \in B_{D_e}\left(g_2(y^n, Q_{XY})\right)\right) \\
&\geq (n+1)^{-|\mathcal{X}||\mathcal{Y}|(|\mathcal{V}|+1)} 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)},
\end{aligned}$$

where the last inequality follows from Lemma 10. ■

We now show that the above described scheme indeed achieves the exponent in (4.17). Consider any possibly random function f implemented by the

primary user (and satisfying the distortion constraint), and denote by P_f the induced joint probability on (X^n, Y^n). Now, consider the following chain of inequalities.

$$
\begin{aligned}
&\mathbf{Pr}\Big(d_e\Big(X^n, g(f(X^n))\Big) \le D_e\Big) \\
&= \sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} P(x^n) P_f(y^n|x^n) p\left(x^n \in B_{D_e}(g(y^n))\right) \\
&\overset{\text{(a)}}{\ge} c_n' \sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} P(x^n) P_f(y^n|x^n) 2^{-nI_{P_n^\star(Q_{x^n y^n})}(X;V|Y)} \\
&\ge c_n' \sum_{x^n \in \mathcal{X}^n} P(x^n) \sum_{y^n \in \mathcal{Y}^n} P_f(y^n|x^n) \min_{Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_{x^n}, D)} 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)} \\
&= c_n' \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{x^n \in T_{Q_X}} P(x^n) \min_{Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_X, D)} 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)} \\
&\overset{\text{(b)}}{=} c_n' \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{x^n \in T_{Q_X}} 2^{-n(D(Q_X\|P) + H_{Q_X}(X))} 2^{-nI_{P_n^\star(Q_X)}(X;V|Y)} \\
&\overset{\text{(c)}}{\ge} c_n' (n+1)^{-|\mathcal{X}|} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{-n\left(D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y)\right)} \\
&\ge c_n' (n+1)^{-|\mathcal{X}|} \max_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{-n\left(D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y)\right)} \\
&= c_n' (n+1)^{-|\mathcal{X}|} \exp\left\{-n\left(\min_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} [D(Q_X\|P) + I_{P_n^\star(Q_X)}(X;V|Y)]\right)\right\}, \qquad (4.31)
\end{aligned}
$$

where

(a) follows from Lemma 11.

(b) follows from (4.30) and (4.25).

(c) follows from Lemma 1.2.3 in [11] $\left(|T_{Q_X}| \ge (n+1)^{-|\mathcal{X}|} 2^{nH_{Q_X}(X)}\right)$.

Taking the limit as n goes to infinity, we get

$$
\begin{aligned}
E^+(P, D, D_e) &= \limsup_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\Big(d_e\Big(X^n, g_n(f_n(X^n))\Big) \le D_e\Big) \\
&\le \min_{Q} D(Q\|P) + R(Q, D, D_e), \qquad (4.32)
\end{aligned}
$$

where the last inequality follows from Proposition 9.

Combining (4.32) and (4.27) yields that the limit in (4.14) exists and is equal to the expression given in (4.17), thus establishing Theorem 15.

4.6.3 The Shannon Cipher System

We now consider the setup of the Shannon cipher system with lossy communication, as in Section 4.2. We assume again that the distortion constraint for the primary user is imposed almost surely.

The message M is overheard by the eavesdropper who knows the statistics of the source and the encoding and decoding functions f and h. However, s/he does not have access to the common randomness K.

As before, the relevant secrecy metric is the probability of a successful guess, i.e., a guess $V^n = g(M)$ satisfying $d_e(X^n, V^n) \leq D_e$. The optimal guess is determined, again, by the MAP rule g_o.

Let $\overrightarrow{D} = (D, D_e)$ and $\overrightarrow{R} = (R, r)$. For a given DMS P, distortion vector $\overrightarrow{D}$, rate vector $\overrightarrow{R}$, and reliability exponent α, we denote the optimal exponent by $E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha)$, i.e.,

$$E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) = \lim_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\Big(d_e\Big(X^n, g_n(f_n(X^n, K))\Big) \leq D_e\Big), \tag{4.33}$$

where $\{f_n\}$ is restricted to the class of functions ensuring the feasibility of the primary user's problem. Similarly to (4.33), we define $E^-(P, \overrightarrow{D}, \overrightarrow{R}, \alpha)$ and $E^+(P, \overrightarrow{D}, \overrightarrow{R}, \alpha)$ using the lim inf and lim sup, respectively.

We extend the definition of $R(P_X, D, D_e)$ to account for the rate constraint as

follows. For a given distribution P_X satisfying $R(P_X, D) \leq R$,

$$R(P_X, R, D, D_e) = \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \leq D \\ I(X;Y) \leq R}} R(P_{XY}, D_e). \tag{4.34}$$

Extending the properties of $R(P_X, D, D_e)$, we prove the following properties of $R(P_X, R, D, D_e)$ in Appendix C.3.

Proposition 12 *In the following statements, $D \geq D_{\min}$, $D_e \geq D_{e,\min}$, and a given pair (P_X, R) satisfy $R \geq R(P_X, D)$.*

(P4) For fixed P_X, $R(P_X, R, D, D_e)$ is a finite-valued function of (R, D, D_e). Moreover, for fixed D_e, $R(P_X, R, D, D_e)$ is continuous in the triple (P_X, R, D) over the set $\mathcal{S} = \{(P_X, R, D) : P_X \in \mathcal{P}_X, D \geq D_{\min}, R > R(P_X, D)\}$.

(P5) $R_e(P_X, D_e) - R(P_X, D) \leq R(P_X, R, D, D_e) \leq R(P_X, D, D_e) \leq R_e(P_X, D_e)$.

The main result is given by the following theorem.

Theorem 16 *Under assumptions (A1)-(A4), for any DMS P, distortion functions d and d_e with associated distortion levels $D \geq D_{\min}$ and $D_e \geq D_{e,\min}$, corresponding respectively to the primary user and the eavesdropper, and reliability exponent α:*

$$E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) = \min \left\{ E_0(P, D_e),\ r + \min_{Q: D(Q\|P) \leq \alpha} D(Q\|P) + R(Q, R, D, D_e) \right\}. \tag{4.35}$$

Remark 15 *The minimization over Q is due to the imposition of an exponentially decaying probability of violating the distortion constraint. If we replace it instead by $\mathbf{Pr}\,(d(X^n, Y^n) > D) \leq \delta$, for some small δ, then the second term of* (4.35) *would collapse to $r + R(P, R, D, D_e)$.*

Remark 16 *We can recover Theorem 15 by setting $\alpha = +\infty$, $r = 0$, and $R = \log|\mathcal{Y}|$. Weinberger and Merhav's result [54, Theorem 1] can also be recovered by noting that the leniency assumption implies $R(Q, R, D, D_e) = 0$ for all Q. Moreover, for any $D_e > \min_{v\in\mathcal{V}} \mathbf{E}_P[d(X, v)]$ and $r > 0$, $E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) > 0$. Indeed, the first condition implies $R_e(P, D_e) > 0$, hence $E_0(P, D_e) > 0$. This refines Schieler and Cuff's observation [40] that any positive key rate drives the distortion at the eavesdropper to its maximal expected value with high probability.*

A straightforward but useful corollary of Theorem 16 is a necessary and sufficient condition on the key rate for the achievability of the perfect secrecy exponent. In particular,

$$E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) = E_0(P, D_e) \text{ if and only if } r \geq E_0(P, D_e) - \min_{Q:D(Q\|P)\leq\alpha} D(Q\|P) + R(Q, R, D, D_e). \tag{4.36}$$

Let r_0 be the minimum rate needed to achieve $E_0(P, D_e)$. The condition in (4.36) is interesting in that it allows r_0 to be *strictly* less than $E_0(P, D_e)$, which itself satisfies $E_0(P, D_e) \leq R_e(P, D_e)$.

Remark 17 *One might suspect that $r \geq \max_{Q:D(Q\|P)\leq\alpha} R(Q, D)$ is sufficient to achieve $E_0(P, D_e)$, since we can use good rate-distortion codes for each type and the number of available keys is large enough to completely "hide" the source sequence within a type class. This is, indeed, true as it implies the condition in* (4.36)*:*

$$D(Q\|P) + R_e(Q, D_e) - D(Q\|P) - R_e(Q, D_e) + R(Q, D) = R(Q, D),$$

$$\Rightarrow \min_{Q'}[D(Q'\|P) + R_e(Q', D_e)] - D(Q\|P) - R_e(Q, D_e) + R(Q, D) \leq R(Q, D),$$

By Property (P5): $\Rightarrow \min_{Q'}[D(Q'\|P) + R_e(Q', D_e)] - D(Q\|P) - R(Q, R, D, D_e) \leq R(Q, D),$

$$\Rightarrow E_0(P, D_e) + \max_{Q:D(Q\|P)\leq\alpha}[-D(Q\|P) - R(Q, R, D, D_e)] \leq \max_{Q:D(Q\|P)\leq\alpha} R(Q, D).$$

The converse of Theorem 16 is based on the following analysis. To achieve the second exponent in (4.35), the eavesdropper tries to guess the value of the key and then applies the scheme suggested in the previous section. Taking into consideration the rate constraint, the term $R(Q, D, D_e)$ which appears in (4.17) is replaced by $R(Q, R, D, D_e)$. Also, taking into account the modified distortion constraint, the minimization over all Q's which appears in (4.17) is replaced by a minimization over Q's satisfying $D(Q||P) \leq \alpha$. The first exponent is the perfect secrecy exponent (given in (4.18)), which the eavesdropper can achieve even in the absence of any observation. The fact that one of these two schemes achieves the optimal exponent implies that the eavesdropper does not benefit from guessing only *part* of the key. Either s/he guesses the entire key correctly and proceeds, or s/he makes a completely blind guess. Interestingly, a similar observation has been made by Schieler and Cuff [41] in the context of minimum expected distortion over a list.

To describe the achievability result, it is helpful to rewrite (4.35) as:

$$E(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) = \min \Bigg\{ \min_{Q: D(Q||P) \leq \alpha} D(Q||P) + \min\{r + R(Q, R, D, D_e),\ R_e(Q, D_e)\}, \quad (4.37)$$
$$\min_{Q: D(Q||P) \geq \alpha} D(Q||P) + R_e(Q, D_e) \Bigg\}.$$

The primary user will operate as follows. For low-probability types Q, particularly Q's with $D(Q||P) > \alpha$, the transmitter will send a dummy message. This is feasible because we allowed some probability of violating the distortion constraint. For such Q's, the eavesdropper receives no information. Therefore, the guessing exponent conditioned on T_Q is given by $R_e(Q, D_e)$, yielding the second term of (4.37). For Q's satisfying $D(Q||P) \leq \alpha$, let $E(\overrightarrow{D}, \overrightarrow{R}, Q) = \min\{r + R(Q, R, D, D_e),\ R_e(Q, D)\}$. This can be understood as the exponent conditioned on $X^n \in T_Q$. For each such Q, we associate a joint type induced by a

$P_{Y|X}$ that achieves the maximum in (4.34). Similarly to Section 4.6.2, we use this joint type to generate a rate-distortion code. This roughly corresponds to the term $R(Q, R, D, D_e)$. To take advantage of the secret key, we in fact produce 2^{nr} such codes, and use the key to randomize the choice of the code, yielding the additional r term. Since the eavesdropper can always guess blindly and achieve the exponent $R_e(Q, D)$, we get $\min\{r + R(Q, R, D, D_e),\ R_e(Q, D)\}$. We will show in Lemma 15 that such random construction fails to achieve the desired exponent with only *doubly exponentially* small probability.

As mentioned earlier, the code construction for each type depends on the conditional $P_{Y|X}$ achieving the maximum in (4.34). A natural question arises: under what conditions does the optimal test channel (which we will denote by $P^{\star}_{Y|X}$) achieve that max? One can readily verify that this holds when $R(Q, R, D, D_e) = 0$ (e.g., the eavesdropper's constraint is more lenient than that of the legitimate receiver). We further investigate this question by considering special cases of Theorem 16. In the following, assume $\alpha = +\infty$. Hence, $R > \max_Q R(Q, D)$.

Perfect Reconstruction at the Eavesdropper

Suppose $\mathcal{V} = \mathcal{X}$, and the eavesdropper is required to reconstruct the source sequence perfectly, i.e., the secrecy metric is $\mathbf{Pr}(V^n = X^n)$. In our formulation, this is equivalent to setting d_e to be the Hamming distance and D_e to 0. Then, for each Q, we get

$$R(Q, R, D, 0) = \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \le D \\ I(X;Y) \le R}} R(P_{XY}, 0) = \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \le D \\ I(X;Y) \le R}} H(X|Y) = H_Q(X) - R(Q, D).$$

Note that the maximum is achieved by the optimal test channel, and the exponent is given by

$$E(P, \overrightarrow{D}, \overrightarrow{R}) = \min_{Q} D(Q||P) + \min\{r + H_Q(X) - R(Q, D),\ H_Q(X)\},$$

where we have used the equivalent form (4.37). Note that, in contrast to $R(Q, R, D, D_e) = 0$, this case corresponds to a more lenient constraint at the legitimate receiver, which leads us to our next example.

Binary Source with Hamming Distortion and $D_e \leq D$

Suppose $\mathcal{X} = \mathcal{Y} = \mathcal{V} = \{0, 1\}$, d and d_e are both the Hamming distance, and $D_e \leq D < 1/2$. We prove the following lemma in Appendix C.4.

Lemma 13 *If $D_e \leq D < 1/2$,*

$$R(Q, D, D_e) = R_e(Q, D_e) - R(Q, D) = \begin{cases} 0, & H(Q) \leq H(D_e), \\ H(Q) - H(D_e), & H(D_e) \leq H(Q) \leq H(D), \\ H(D) - H(D_e), & H(Q) \geq H(D). \end{cases}$$

It follows from property (P5) of Proposition 12 that

$$R(Q, D, D_e) = R_e(Q, D_e) - R(Q, D) \Rightarrow R(Q, R, D, D_e) = R_e(Q, D_e) - R(Q, D).$$

Therefore, the exponent is given by:

$$E(P, \overrightarrow{D}, \overrightarrow{R}) = \min \begin{cases} \min\{D(Q||P) : \quad H(Q) \leq H(D_e)\}, \\ \min\{D(Q||P) + H(Q) - H(D_e) : \quad H(D_e) \leq H(Q) \leq H(D)\}, \\ \min\{D(Q||P) + \min\{r + H(D) - H(D_e), H(Q) - H(D_e)\} : \quad H(Q) \geq H(D)\}. \end{cases}$$

If $X \sim \text{Ber}(1/2)$, then $D(Q\|P) = 1 - H(Q)$, and the minima corresponding to the first two cases reduce to $1 - H(D_e)$. The third minimum can be computed as follows:

$$\min_{\substack{Q: \\ H(Q)\geq H(D)}} \quad 1 - H(D_e) + \min\{r + H(D) - H(Q), 0\} = 1 - H(D_e) + \min\{r + H(D) - 1, 0\}.$$

Therefore,

$$E(P, \overrightarrow{D}, \overrightarrow{R}) = \begin{cases} 1 - H(D_e), & r \geq 1 - H(D), \\ r + H(D) - H(D_e), & r < 1 - H(D). \end{cases}$$

The resulting expression when $r < 1 - H(D)$ admits a simple geometric explanation, shown in Figure 4.3 below. Upon observing the public message, the

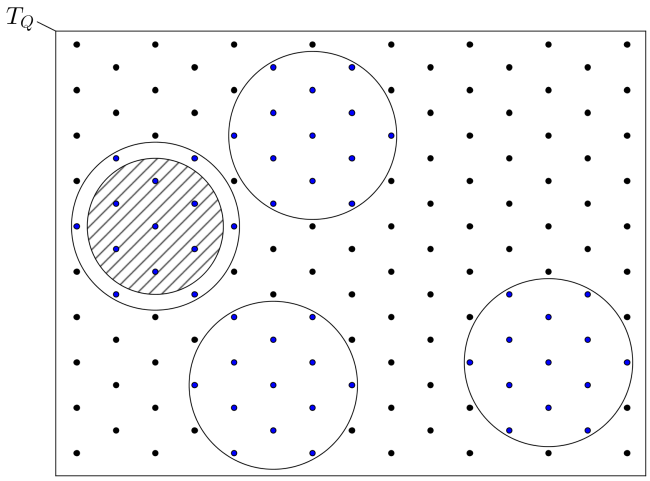

Figure 4.3: The dots represent sequences in a type class T_Q. Each of the 2^{nr} non-dashed circles represents a Hamming-distortion ball of radius D, corresponding to a possible reconstruction at the legitimate receiver. Thus, dots within the circle (in blue) represent candidate source sequences. The dashed circle represents the distortion ball of radius D_e around the eavesdropper's reconstruction, and it fits entirely in a non-dashed circle.

candidate source sequences are clustered into 2^{nr} balls. Each ball corresponds to a possible value of the key K, and has volume $2^{nH(D)}$ since it is the pre-image

of a possible reconstruction at the legitimate receiver. For the eavesdropper, the maximum volume of the ball that s/he can generate to "engulf" candidate sequences is $2^{nH(D_e)}$. Due to the structure of Hamming distortion, this maximally-sized ball can fit entirely into any one of the clusters, so that the probability of a successful guess is $2^{nH(D_e)}2^{-n(r+H(D))}$. Note that the geometric interpretation assumed that we are using good rate-distortion codes (to get pre-images of volume $2^{nH(D)}$). The described structure is also reminiscent of successive refinement [15]. These can be explained by the following lemma.

Lemma 14 *If $R(Q,D,D_e) = R_e(Q,D_e) - R(Q,D)$, then the optimal test channel $P^{\star}_{Y|X}$ achieves the maximum in* (4.16)*. Moreover, Q is successively refinable from D to D_e.*

Proof: Consider the proof of the lower bound in (P3) of Proposition 7. $R(Q,D,D_e) = R_e(Q,D_e) - R(Q,D)$ implies that (C.1) is an equality. Hence, $P^{\star}_{Y|X}$ achieves the maximum. Moreover, (C.2) becomes an equality. Let $P^{(1)}_{V|XY}$ be the minimizer in (C.1), and $P^{(2)}_{V|XY}$ the minimizer in (C.2). Then, $H_{P^{\star}_{XY}P^{(1)}_{V|XY}}(X|V) \leq H_{P^{\star}_{XY}P^{(2)}_{V|XY}}(X|V) = H_{P^{\star}_{XY}P^{(1)}_{V|XY}}(X|V,Y) \leq H_{P^{\star}_{XY}P^{(1)}_{V|XY}}(X|V)$. Therefore, $P^{(1)}_{V|XY}$ satisfies $\mathbf{E}[d_e(X,V)] \leq D_e$ and $H_{P^{\star}_{XY}P^{(1)}_{V|XY}}(X|V,Y) = H_{QP^{(1)}_{V|X}}(X|V)$, i.e. the Markov chain $X-V-Y$ holds. Finally, note that $I(X;V) = I(X;(V,Y)) = I(X;Y) + I(X;V|Y) = R_e(Q,D_e)$, implying that Q is successively refinable from D to D_e. ■

Proof of the Lower Bound

We show that $E^{-}(P,\overrightarrow{D},\overrightarrow{R},\alpha) \geq \min\{E_0(P,D), \; r + \min_{Q:D(Q\|P)\leq\alpha} D(Q\|P) + R(Q,R,D,D_e)\}$ by demonstrating an encoding-decoding strategy for the primary user that achieves the given exponent.

As before, the primary user will operate on the source sequences on a type-by-type basis. The result is driven by the following lemma, which is based on the analysis of Schieler and Cuff [41] and the proof of which is given in Appendix C.5.

Lemma 15 *Let $\epsilon > 0, n \in \mathbb{N}, Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}$ be given. Let N be an integer such that $2^{n(I_{Q_{XY}}(X;Y)+2\epsilon/3)} \leq N \leq 2^{n(I_{Q_{XY}}(X;Y)+\epsilon)}$. Generate a code $C^n = (Y_1^n, Y_2^n, \ldots, Y_N^n)$ by choosing N elements independently and uniformly at random from T_{Q_Y}.*

1. **Covering:** *For $x^n \in T_{Q_X}$, define*

$$C(x^n) = \{m \in [N] : (x^n, Y_m^n) \in T_{Q_{XY}}\}, \tag{4.38}$$

$$N_{x^n} = |C(x^n)|, \tag{4.39}$$

and the event

$$\mathcal{E} = \{C^n : \text{there exists } x^n \in T_{Q_X} \text{ such that } N_{x^n} = 0 \text{ or } N_{x^n} > 2^{2n\epsilon}\}. \tag{4.40}$$

Then, there exists $n_1(\epsilon, |\mathcal{X}|, |\mathcal{Y}|)$ (independent of Q_{XY}) such that, for all $n \geq n_1$,

$$\mathbf{Pr}(\mathcal{E}) \leq e^{-2^{n\epsilon/7}}. \tag{4.41}$$

2. **Guessing—single code:** *Suppose $X^n \sim \mathrm{Unif}(T_{Q_X})$ and $C^n \notin \mathcal{E}$. Let $P^C_{M|X^n}$ be as follows. Given x^n, M is chosen uniformly at random from $C(x^n)$. Then, for all $n \geq n_1$, for all $v^n \in \mathcal{V}^n$ and all $m \in [N]$,*

$$\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) \leq 2^{-n(R(Q_{XY}, D_e) - 4\epsilon)}, \tag{4.42}$$

and

$$\mathbf{E}[\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) | \mathcal{E}^c] \leq 2^{-n(R_e(Q_X, D_e) - 4\epsilon)}, \tag{4.43}$$

where the probabilities are computed with respected to the randomness in $P^{C}_{M|X^n}$, and the expectation with respect to the distribution of the code C^n.

3. **Guessing—multiple codes:** *Let K be uniform over $\mathcal{K} = [2^{nr}]$, $r > 0$, and independent of X^n (which is uniform over T_{Q_X}). For each $k \in [2^{nr}]$, generate C_k^n as described above. Define $P^{\{C_k^n\}_k}_{M|X^n,K}$ as follows. Given k, if $C_k \in \mathcal{E}$, then M is chosen uniformly at random from $[N]$ independently of X^n. If $C_k^n \notin \mathcal{E}$, then M is chosen uniformly at random from $C_k(X^n)$. Let,*

$$\tilde{\mathcal{E}} = \Big\{ \{C_k\}_{k=1}^{2^{nr}} : \text{ there exists } m \in [N] \text{ such that } \max_{v^n \in \mathcal{V}^n} \mathbf{Pr}\left(d_e(X^n, v^n) \le D_e | M = m, \{C_k^n\}_{k=1}^{2^{nr}} \right) > 2^{-n(\min\{R_e(Q_X, D_e), r + R(Q_{XY}, D_e)\} - 8\epsilon)} \Big\}, \tag{4.44}$$

where the probability is computed with respect to $P^{\{C_k^n\}_k}_{M|X^n,K}$. Then, for all $n \ge n_1$,

$$\mathbf{Pr}(\tilde{\mathcal{E}}) \le e^{-2^{n\epsilon/9}}, \tag{4.45}$$

where the probability is computed with respected to the distributions of the codes $\{C_k^n\}_k$. □

The first part of the Lemma asserts that if we generate the codebook randomly, then each $x^n \in T_{Q_X}$ will be covered by a small number of codewords (the probability that this event does not occur is doubly exponentially small). Therefore, if we encode x^n by choosing a codeword uniformly at random from its cover, the induced $P_{X^n|Y^n}(.|y^n)$ will be roughly uniform over the set $T_{Q_{X|Y}}(y^n) = \{x^n : (x^n, y^n) \in T_{Q_{XY}}\}$. Consequently, given a codeword index m and $v^n \in \mathcal{V}^n$, the second part bounds the probability that v^n covers X^n, and also bounds the expectation (over the choice of the codebook) of that probability.

Finally, the third part considers generating 2^{nr} codebooks and the induced distribution $P_{X^n|M}$, where M is the index of a chosen codeword. This distribution roughly corresponds to generating 2^{nr} elements uniformly at random from T_{Q_Y}, revealing the chosen elements to the adversary, then choosing one of them uniformly at random and generating X^n uniformly at random from $T_{Q_{X|Y}}(Y^n)$. This setup is similar to the one studied by Schieler and Cuff [41, Theorem 4]. Equation (4.45) states that, for most realizations of the codebooks, the probability that the adversary generates a successful guess, given a codeword index, is upper-bounded by $2^{-n(\min\{R_e(Q_X,D_e),r+R(Q_{XY},D_e)\}}$. The implication is that the best the adversary could do is 1) either ignore the index and guess X^n blindly, 2) or guess which codebook is being used (i.e., guess the value of the key K) and use the scheme suggested in the previous section.

Now, fix $\delta > 0$ such that $R_{\alpha+\delta} = \max_{Q:D(Q\|P)\leq\alpha+\delta} R(Q,D) < R$. Note that such δ exists since $\lim_{\delta\to 0} R_{\alpha+\delta} = R_\alpha$ (which follows from Proposition 23 in Appendix C.1 and the fact that $D(Q\|P)$ is convex). Fix R' such that $R_{\alpha+\delta} < R' < R$, and $\epsilon > 0$ such that $\epsilon < R - R'$. Let

$$\mathcal{Q}_{\mathcal{X}}^n(\alpha,\delta) = \{Q \in \mathcal{Q}_{\mathcal{X}}^n : D(Q\|P) \leq \alpha + \delta\}. \tag{4.46}$$

Let n be large as given by Lemma 15. For each type $Q_X \in \mathcal{Q}_{\mathcal{X}}^n(\alpha,\delta)$, we associate a joint type Q_{XY} and generate 2^{nr} codebooks $\{C_k\}_{k=1}^{2^{nr}} \in \tilde{\mathcal{E}}^c$ where the size of each codebook is upper-bounded by $2^{n(I_{Q_{XY}}(X;Y)+\epsilon)}$ (the existence of such codes follows from (4.45)). Since the primary user wants to minimize the probability of a successful guess by the eavesdropper, but must also satisfy a rate constraint, the associated type is chosen as follows:

$$Q_{R'}^\star(Q_X) \in \underset{\substack{Q_{XY}\in\mathcal{Q}_{\mathcal{XY}}^n(Q_X,D):\\ I_{Q_{XY}}(X;Y)\leq R'}}{\operatorname{argmax}} R(Q_{XY},D_e). \tag{4.47}$$

The encoding function f is as follows. Given a source sequence x^n satisfying $Q_{x^n} \in \mathcal{Q}_{\mathcal{X}}^n(\alpha, \delta)$, and a realization of the key k, a reconstruction sequence is chosen uniformly at random from $C_k(x^n)$ (cf. (4.38)). The associated message is then given by:

- $\lceil \log |\mathcal{Q}_{\mathcal{X}}^n| \rceil$ bits to describe Q_X.
- $\lceil \log |C_k^n| \rceil$ bits to describe the index of the reconstruction.

The legitimate receiver uses the first part of the message and the key to determine which codebook is being used, and then uses the second part of the message to recover the reconstruction Y^n. Finally, all sequences x^n such that $Q_{x^n} \notin \mathcal{Q}_{\mathcal{X}}^n(\alpha, \delta)$ are mapped to an arbitrary message m_0.

Remark 18 *One can check that this encoding is feasible by noting that the required number of bits satisfy:*

$$\begin{aligned}
\lceil \log |\mathcal{Q}_{\mathcal{X}}^n| \rceil + \lceil \log |C_k^n| \rceil &\leq |\mathcal{X}| \log(n+1) + 1 + n\left(I_{Q^\star_{R'}(Q_X)}(X;Y) + \epsilon\right) + 2 \\
&\leq n\left(R' + \epsilon + \frac{|\mathcal{X}|}{n}\log(n+1) + \frac{3}{n}\right) \\
&< nR,
\end{aligned}$$

for n large enough. Moreover, it satisfies the excess distortion probability constraint, since

$$\begin{aligned}
\mathbf{Pr}(d(X^n, Y^n) > D) &\leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} P(Q_X) \\
&\leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} 2^{-nD(Q_X \| P)} \\
&\leq (n+1)^{|\mathcal{X}|} 2^{-n(\alpha+\delta)} < 2^{-n\alpha},
\end{aligned}$$

where the last inequality holds for large enough n.

To analyze the performance of the eavesdropper, note that when s/he observes a message $m \neq m_0$, then the induced distribution $P_{X^n|M=m}$ is exactly the setup studied in part three of Lemma 15. Indeed, the message m indicates the type of the transmitted sequence and the index of the reconstruction (among 2^{nr} possible codebooks). For $m = m_0$, i.e., for sequences of type outside $\mathcal{Q}_n(\alpha, \delta)$, the performance can still be analyzed in light of Lemma 15 by considering the associated Q_{XY} to be of the form $Q_X Q_Y$ (i.e., X and Y are independent), in which case $\min\{R_e(Q_X, D_e), r + R(Q_{XY}, D_e)\} = \min\{R_e(Q_X, D_e), r + R_e(Q_X, D_e)\} = R_e(Q_X, D_e)$. Now, consider the following chain of inequalities.

$$
\begin{aligned}
&\mathbf{Pr}\left(d_e\left(X^n, g_0(f(X^n, K))\right) \leq D_e\right) \\
&= \sum_{Q_X \in \mathcal{Q}_X^n} \sum_{x^n \in T_{Q_X}} \sum_{m \in \mathcal{M}} P(x^n) P_f(m|x^n) \mathbf{1}\left\{x^n \in B_{D_e}\left(g_0(m)\right)\right\} \\
&= \sum_{Q_X \in \mathcal{Q}_X^n} P(Q_X) \sum_{m \in \mathcal{M}} \sum_{x^n \in T_{Q_X}} P_f(m|T_{Q_X}) P_f(x^n|m, T_{Q_X}) \cdot \mathbf{1}\left\{x^n \in B_{D_e}\left(g_0(m)\right)\right\} \\
&= \sum_{Q_X \in \mathcal{Q}_X^n} P(Q_X) \sum_{m \in \mathcal{M}} P_f(m|T_{Q_X}) \max_{v^n \in \mathcal{V}^n} P_f\left(d_e(X^n, v^n) \leq D_e | m, T_{Q_X}\right) \\
&\overset{\text{(a)}}{\leq} \sum_{Q_X \in \mathcal{Q}_X^n} P(Q_X) \sum_{m \in \mathcal{M}} P_f(m|T_{Q_X}) 2^{-n(\min\{R_e(Q_X, D_e), r + R(Q_{R'}^\star(Q_X), D_e)\} - 8\epsilon)} \\
&\leq \sum_{Q_X \in \mathcal{Q}_X^n} 2^{-n(D(Q_X\|P) + \min\{R_e(Q_X, D_e), r + R(Q_{R'}^\star(Q_X), D_e)\} - 8\epsilon)} \\
&\leq (n+1)^{|\mathcal{X}|} \max_{Q_X \in \mathcal{Q}_X^n} \exp\Big\{ - n \left(D(Q_X\|P) + \min\{R_e(Q_X, D_e), r + R(Q_{R'}^\star(Q_X), D_e)\} - 8\epsilon\right) \Big\} \\
&= (n+1)^{|\mathcal{X}|} \exp\Bigg(- n \min\Bigg\{ \min_{Q_X \in \mathcal{Q}_X^n(\alpha, \delta)} D(Q_X\|P) + \min\{R_e(Q_X, D_e), r + R(Q_{R'}^\star(Q_X), D_e)\}, \\
&\qquad\qquad \min_{Q_X \notin \mathcal{Q}_X^n(\alpha, \delta)} D(Q_X\|P) + R_e(Q_X, D_e) \Bigg\} + 8\epsilon n \Bigg) \\
&= (n+1)^{|\mathcal{X}|} \exp\Bigg(-n \min\Bigg\{ \min_{Q_X \in \mathcal{Q}_X^n(\alpha, \delta)} D(Q_X\|P) + r + R(Q_{R'}^\star(Q_X), D_e), \\
&\qquad\qquad \min_{Q_X \in \mathcal{Q}_X^n} D(Q_X\|P) + R_e(Q_X, D_e) \Bigg\} + 8\epsilon n \Bigg),
\end{aligned}
\tag{4.48}
$$

where (a) follows from (4.44) of Lemma 15 and the fact that the codebooks

$\{C_k\}_{k=1}^{2^{nr}} \notin \tilde{\mathcal{E}}$ by construction. Therefore,

$$E^-(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) = \liminf_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\left(d_e\Big(X^n, g_n(f_n(X^n, K))\Big) \le D_e\right)$$
$$\ge \min\left\{E_0(P, D), r + \min_{Q: D(Q\|P)\le\alpha+\delta} D(Q\|P) + R(Q, R', D, D_e)\right\} - 8\epsilon,$$

where the inequality follows from the following proposition, the proof of which is given in Appendix C.6.

Proposition 16

$$\lim_{n\to\infty} \min_{Q_X \in \mathcal{Q}_X^n(\alpha,\delta)} D(Q_X\|P) + R(Q^\star_{R'}(Q_X), D_e) = \min_{Q: D(Q\|P)\le\alpha+\delta} D(Q\|P) + R(Q, R', D, D_e).$$

Now, note that ϵ is arbitrary, and

$$\lim_{R'\to R} \min_{Q: D(Q\|P)\le\alpha+\delta} D(Q\|P) + R(Q, R', D, D_e) = \min_{Q: D(Q\|P)\le\alpha+\delta} D(Q\|P) + R(Q, R, D, D_e), \tag{4.49}$$

since $R(Q, \tilde{R}, D, D_e)$ is uniformly continuous in $(Q, \tilde{R})$ over the set $\{(Q, \tilde{R}) : D(Q\|P) \le \alpha + \delta, R' \le \tilde{R} \le R\}$ by Proposition 12. Finally, it follows from Proposition 12 and Proposition 23 (to follow in Appendix C.1) that

$$\lim_{\delta\to 0} \min_{Q: D(Q\|P)\le\alpha+\delta} D(Q\|P) + R(Q, R, D, D_e) = \min_{Q: D(Q\|P)\le\alpha} D(Q\|P) + R(Q, R, D, D_e) \tag{4.50}$$

As such,

$$E^-(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) \ge \min\left\{E_0(P, D), r + \min_{Q: D(Q\|P)\le\alpha} D(Q\|P) + R(Q, R, D, D_e)\right\}.$$

Proof of the Upper Bound

We now prove that $E^+(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) \le \min\{r + \min_{Q: D(Q\|P)\le\alpha} D(Q\|P) + R(Q, R, D, D_E), E_0(P, D_e)\}$. We have already shown, following Theorem 15, that

the perfect secrecy exponent $E_0(P, D_e)$ is achievable by the eavesdropper even in the absence of any observation. It follows immediately that

$$E^+(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) \leq E_0(P, D_e). \tag{4.51}$$

So we only need to demonstrate a strategy that achieves the first exponent. The strategy is based on the one suggested in Section 4.6.2. We will add an initial stage in which the eavesdropper tries to guess the value of K, by choosing an element uniformly at random from $\{1, 2, \cdots, 2^{nr}\}$. The eavesdropper's guess, denoted by $\tilde{K}$, is equal to K with probability 2^{-nr} (This will correspond to the r term in (4.35)). Then, s/he generates $\tilde{Y}^n = h(M, \tilde{K})$. Next, the eavesdropper implements the same stages suggested in Section 4.6.2, where $\tilde{Y}^n$ plays the role of Y^n. We denote the strategy by g'.

Remark 19 *If h is stochastic, it can be replaced by a deterministic h that still satisfies the reliability constraint. Since this does not change the conditional $P_{X^n|M}$, we can assume, without loss of generality, that h is deterministic.*

Now, consider any functions f and h implemented by the primary user (and satisfying the distortion constraint). Let P_f denote the induced joint probability of (X^n, M, K), and P_K denote the distribution of K. To analyze the performance of g', note that, unlike Section 4.6.2, not every realization of $\tilde{Y}^n$ necessarily satisfies the distortion constraint. To that end, define

$$\mathcal{M}_D(x^n, k) = \{m \in \mathcal{M} : d(x^n, h(m, k)) \leq D\}, x^n \in \mathcal{X}^n,\ k \in \mathcal{K}, \tag{4.52}$$

$$\text{and } \mathcal{A} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : d(x^n, y^n) > D\}. \tag{4.53}$$

The distortion constraint implies that

$$P_f(\mathcal{A}) \leq 2^{-n\alpha}. \tag{4.54}$$

Moreover, the analysis of g' should take into account the rate constraint R. The following Lemma by Weissman and Ordentlich [55] will be instrumental.

Lemma 17 ([55, Lemma 3]) *Let $Y^n(.)$ be an n-block code of rate $\leq R$. Then, for every $Q \in \mathcal{Q}_X^n$ and $\eta > 0$, if X^n is uniformly distributed over T_Q,*

$$\mathbf{Pr}\left(\left\{x^n \in T_Q : I_{Q_{x^n Y^n(x^n)}}(X;Y) > R + \eta\right\}\right) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|+|\mathcal{X}|} 2^{-n\eta}. \tag{4.55}$$

Remark 20 *This is not the exact statement found in [55], but it is a straightforward modification.*

So, define for every $\eta > 0$, $x^n \in \mathcal{X}^n$, and $k \in \mathcal{K}$,

$$\mathcal{M}_R(x^n, k, \eta) = \{m \in \mathcal{M} : I_{Q_{x^n y^n}}(X;Y) \leq R + \eta, \text{where } y^n = h(m,k)\}, \tag{4.56}$$

$$\mathcal{M}_{D,R}(x^n, k, \eta) = \mathcal{M}_D(x^n, k) \cap \mathcal{M}_R(x^n, k, \eta), \tag{4.57}$$

$$\text{and } \mathcal{B}(\eta) = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : I_{Q_{x^n y^n}}(X;Y) > R + \eta\}. \tag{4.58}$$

Finally, fix $\epsilon > 0$, $\delta > 0$ and $\eta > 0$, and consider the following chain of inequali-

ties.

$$
\begin{aligned}
&\mathbf{Pr}\Big(d_e\Big(X^n, g'(f(X^n,K))\Big) \le D_e\Big) \\
&= \sum_{x^n\in\mathcal{X}^n}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}} P(x^n)P_K(k)P_f(m|x^n,k)P_f\left(x^n \in B_{D_e}\Big(g'(m)\Big)\right) \\
&\ge \sum_{x^n\in\mathcal{X}^n}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}_{D,R}(x^n,k,\eta)} P(x^n)P_K(k)P_f(m|x^n,k)P_f\left(x^n \in B_{D_e}\Big(g'(m)\Big)\right) \\
&= \sum_{x^n\in\mathcal{X}^n}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}_{D,R}(x^n,k,\eta)} P(x^n)P_K(k)P_f(m|x^n,k)\sum_{\tilde{k}\in\mathcal{K}} P_K\left(\tilde{K}=\tilde{k}\right)P_f\left(x^n \in B_{D_e}\left(g(h(m,\tilde{k}))\right)\right) \\
&\ge 2^{-nr}\sum_{x^n\in\mathcal{X}^n}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}_{D,R}(x^n,k,\eta)} P(x^n)P_K(k)P_f(m|x^n,k)P_f\left(x^n \in B_{D_e}\left(g(h(m,k))\right)\right) \\
&\overset{(a)}{\ge} c_n' 2^{-nr}\sum_{x^n\in\mathcal{X}^n}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}_{D,R}(x^n,k,\eta)} P(x^n)P_K(k)P_f(m|x^n,k)2^{-nI_{P_n^\star(Q_{x^n y^n})}(X;V|Y)} \\
&\overset{(b)}{\ge} c_n' 2^{-nr}\sum_{Q_X\in\mathcal{Q}_{\mathcal{X}}^n}\sum_{x^n\in T_Q}\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{M}_{D,R}(x^n,k,\eta)} P(x^n)P_K(k)\cdot P_f(m|x^n,k)\min_{\substack{Q_{XY}\in\mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X,D):\\ I_{Q_{XY}}(X;Y)\le R+\eta}} 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)} \\
&= c_n' 2^{-nr}\sum_{Q_X\in\mathcal{Q}_{\mathcal{X}}^n}\min_{\substack{Q_{XY}\in\mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X,D):\\ I_{Q_{XY}}(X;Y)\le R+\eta}} 2^{-nI_{P_n^\star(Q_{XY})}(X;V|Y)}P_f(\mathcal{A}^c\cap\mathcal{B}^c(\eta)\cap T_Q) \\
&\overset{(c)}{\ge} c_n' 2^{-nr}\sum_{Q_X\in\mathcal{Q}_{\mathcal{X}}^n(\alpha,-\delta)}\min_{\substack{Q_{XY}\in\mathcal{Q}_{\mathcal{X}\mathcal{Y}}^n(Q_X,D):\\ I_{Q_{XY}}(X;Y)\le R+\eta}} 2^{-n(R(Q_{XY},D_e)+\epsilon)}P_f(\mathcal{A}^c\cap\mathcal{B}^c(\eta)\cap T_Q) \\
&\ge c_n' 2^{-nr}\sum_{Q_X\in\mathcal{Q}_{\mathcal{X}}^n(\alpha,-\delta)} 2^{-n(R(Q_X,R+\eta,D,D_e)+\epsilon)}P_f(\mathcal{A}^c\cap\mathcal{B}^c(\eta)\cap T_Q), \qquad (4.59)
\end{aligned}
$$

where

(a) follows from Lemma 11 and (4.52): $m \in \mathcal{M}_D(x^n,k)$ guarantees that $d(x^n,h(m,k)) \le D$ so that Lemma 11 is applicable.

(b) follows from (4.57): $m \in \mathcal{M}_{D,R}(x^n,k,\eta)$ allows us to restrict the minimum to $Q_{XY} \in \mathcal{Q}_{\mathcal{X}\mathcal{Y}}(Q_X,D)$ satisfying $I_{Q_{XY}}(X;Y) \le R+\eta$.

(c) follows from Proposition 24 in Appendix C.2.

Now, note that

$$P_f(\mathcal{A}|T_Q) \leq \frac{P_f(\mathcal{A})}{P(T_Q)} \leq (n+1)^{|\mathcal{X}|}2^{-n(\alpha - D(Q||P))}. \tag{4.60}$$

Moreover, by Lemma 17,

$$P_f(\mathcal{B}(\eta)|T_Q) \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|+|\mathcal{X}|}2^{-n\eta}. \tag{4.61}$$

Combining (4.60) and (4.61) yields, for every $Q_X \in \mathcal{Q}_X^n(\alpha, -\delta)$,

$$\begin{aligned} P_f(\mathcal{A}^c \cap \mathcal{B}^c(\eta) \cap T_Q) &= P(T_Q)P_f(\mathcal{A}^c \cap \mathcal{B}^c(\eta)|T_Q) \\ &\geq P(T_Q)(1 - P_f(\mathcal{A}|T_Q) - P_f(\mathcal{B}(\eta)|T_Q)) \\ &\geq P(T_Q)(1 - (n+1)^{|\mathcal{X}|}2^{-n\delta} - (n+1)^{|\mathcal{X}||\mathcal{Y}|+|\mathcal{X}|}2^{-n\eta}) \\ &\geq P(T_Q)/2, \end{aligned} \tag{4.62}$$

where the last inequality holds for large enough n. Continuing from (4.59), we get

$$\begin{aligned} &\mathbf{Pr}\Big(d_e\Big(X^n, g'(f(X^n, K))\Big) \leq D_e\Big) \\ &\geq (c_n'/2)2^{-nr}(n+1)^{-|\mathcal{X}|} \sum_{Q_X \in \mathcal{Q}_X^n(\alpha,-\delta)} 2^{-n(D(Q_X||P)+R(Q_X,R+\eta,D,D_e)+\epsilon)} \\ &\geq (c_n'/2)(n+1)^{-|\mathcal{X}|} \exp\left\{-n\left(\epsilon + r + \min_{Q_X \in \mathcal{Q}_X^n(\alpha,-\delta)} D(Q_X||P) + R(Q_X, R+\eta, D, D_e)\right)\right\}. \end{aligned} \tag{4.63}$$

Therefore, taking the limit as n goes to infinity, and noting that ϵ, δ, and η are arbitrary, we get

$$\begin{aligned} E^+(P, \overrightarrow{D}, \overrightarrow{R}, \alpha) &= \limsup_{n\to\infty} \max_{\{f_n\}} \min_{\{g_n\}} -\frac{1}{n} \log \mathbf{Pr}\Big(d_e\Big(X^n, g_n(f_n(X^n, K))\Big) \leq D_e\Big) \\ &\leq \lim_{\delta\to 0} \lim_{\eta\to 0} r + \min_{Q: D(Q||P)\leq \alpha-\delta} D(Q||P) + R(Q, R+\eta, D, D_e) \\ &= r + \min_{Q: D(Q||P)\leq\alpha} D(Q||P) + R(Q, R, D, D_e), \end{aligned} \tag{4.64}$$

where the last equality follows similarly to equations (4.49) and (4.50). Combining (4.51) and (4.64) yields our result.

CHAPTER 5

LEARNING COMPLEXITY

This chapter investigates the complexity of estimating $\mathcal{L}(X \to Y)$, i.e., the number of samples needed to estimate $\mathcal{L}(X \to Y)$, which we equivalently denote by $\mathcal{L}(P_X; P_{Y|X})$. To this end, an estimator is defined as a function $f : (\mathcal{X} \times \mathcal{Y})^\star \to \mathbb{R}$, which maps a sequence of samples drawn from a joint distribution to an estimate of its maximal leakage. Given a desired level of accuracy δ, and a probability of error ϵ, the sample complexity of an estimator f is defined as:

$$S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|)[f] = \min\{n : P_{XY}\left(\left|\mathcal{L}(P_X; P_{Y|X}) - f(X^n, Y^n)\right| > \delta\right) < \epsilon, \text{ for all } P_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}\}, \tag{5.1}$$

where $P_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}}$ is the set of all probability distributions on $\mathcal{X} \times \mathcal{Y}$, and (X^n, Y^n) are drawn independently from P_{XY}. Then, the sample complexity of maximal leakage is defined as:

$$S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|) = \inf_f S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|)[f]. \tag{5.2}$$

We show that $S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|)$ turns out to be infinity. This is mainly due to the discontinuity of maximal leakage in the support of X. Therefore, we assume we have a known lower bound θ on the minimum strictly positive probability of an element in $\mathcal{X}$, and we define

$$\mathcal{P}^\theta_{\mathcal{X}\times\mathcal{Y}} = \{P_{XY} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Y}} : \min_{x\in\mathcal{X}:P_X(x)>0} P_X(x) \geq \theta\}, \tag{5.3}$$

$$S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|, \theta) = \inf_f \min\{n : P_{XY}\left(\left|\mathcal{L}(P_X; P_{Y|X}) - f(X^n, Y^n)\right| > \delta\right) < \epsilon, \text{ for all } P_{XY} \in \mathcal{P}^\theta_{\mathcal{X}\times\mathcal{Y}}\}. \tag{5.4}$$

We prove the following upper and lower bounds in Sections 5.1 and 5.2 respectively.

Theorem 17 *For all $\theta \in (0,1)$, discrete alphabets $\mathcal{X}$ and $\mathcal{Y}$, $\delta > 0$, and $\epsilon \in (0,1)$,*

$$S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|, \theta) \leq \frac{8(\log(5/\epsilon) + |\mathcal{Y}| \log |\mathcal{X}|)}{\theta\left((2 - e^{-\delta}) \log(2 - e^{-\delta}) + e^{-\delta} - 1\right)}. \tag{5.5}$$

Note that, for small δ, the denominator behaves as δ^2. If θ is of the order of $1/|\mathcal{X}|$, we get $S(\theta, \delta, \epsilon) \leq O\left(|\mathcal{X}|(|\mathcal{Y}| \log |\mathcal{X}| + \log(1/\epsilon))/\delta^2\right)$.

Theorem 18 *Given $\epsilon \in (0, 1/3)$ and $\eta > 0$, there exists $c_{\epsilon,\eta} > 0$ (depending only on ϵ and η) and $\delta_\eta > 0$ (depending only on η) such that for all $\theta \in (0,1)$, for all discrete alphabets $\mathcal{X}$, and for infinitely many choices of $|\mathcal{Y}|$ with $\mathcal{Y}$ discrete,*

$$S_{\delta_\eta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|, \theta) \geq c_{\epsilon,\eta} \frac{|\mathcal{Y}|^{1-\eta}}{\theta}. \tag{5.6}$$

If $\theta \to 0$, the bound goes to infinity, which justifies our earlier claim that $S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|)$ is $+\infty$.

Remark 21 *In terms of the dependence on the alphabets, the upper and lower bounds are within sub-polynomial factors of each other.*

5.1 Proof of Theorem 17

Let

$$M(P_X; P_{Y|X}) := \exp\{\mathcal{L}(P_X; P_{Y|X})\} = \sum_{y \in \mathcal{Y}} \max_{\substack{x \in \mathcal{X}: \\ P_X(x) > 0}} P_{Y|X}(y|x). \tag{5.7}$$

It is straightforward to verify that a $(1 - e^{-\delta})$-multiplicative estimator for $M(P_X; P_{Y|X})$ translates to a δ-additive estimator for $\mathcal{L}(P_X; P_{Y|X})$, where a $\hat{\delta}$-multiplicative estimator means that $|M - \hat{M}| \leq \hat{\delta} M$. Therefore, in the remainder,

we will analyze multiplicative estimators of M. Now, consider $n \in \mathbb{N}$, and let $N \sim \mathrm{Poi}(n)$. (We consider Poisson sampling because it simplifies the analysis, and we connect it later to fixed-length sampling.) Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ be N independent samples drawn from a distribution P_{XY}. For each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let N_x denote the number of times x appears, N_y the number of times y appears, and $N_{x,y}$ the number of times (x, y) appears in the sequence. Then, $N_x \sim \mathrm{Poi}(nP_X(x))$, $N_y \sim \mathrm{Poi}(nP_Y(y))$, and $N_{x,y} \sim \mathrm{Poi}(nP_{XY}(x, y))$. Now, let $\theta' = \theta/4$. The estimator works as follows:

1. For each $x \in \mathcal{X}$ with $N_x > 0$, generate a random variable $\tilde{N}_x \sim \mathrm{Poi}(n\theta')$. If $N_x = 0$, set $\tilde{N}_x = 0$.
2. For each $x \in \mathcal{X}$ with $N_x > 0$, keep only the first $\tilde{N}_x$ samples containing x and disregard the rest.
   (a) If there are not enough samples for some x (i.e., $\tilde{N}_x > N_x$), then let $\hat{M} = 1$.
   (b) Otherwise, let
   $$\hat{M} = \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \frac{\tilde{N}_{x,y}}{n\theta'}, \tag{5.8}$$
   where $\tilde{N}_{x,y}$ is the number of times (x, y) appears in the truncated sequence.

To analyze the above estimator, we consider a slightly modified setting. In particular, suppose the estimator has access to an infinite sequence $(X_1, Y_1), (X_2, Y_2), \ldots$ Then, $N_x = +\infty$ with probability 1 for each $x \in \mathrm{supp}(X)$. In this case, for each (x, y) with $P_X(x) > 0$, $\tilde{N}_{x,y} \sim \mathrm{Poi}(n\theta' P_{Y|X}(y|x))$. Now, consider the following lemma, which is an application of the Chernoff bound to Poisson random variables. The proof is straightforward and omitted.

Lemma 18 *Consider $\delta \in (0,1)$, $\lambda > 0$, and let $N \sim \mathrm{Poi}(\lambda)$.*

$$\mathbf{Pr}(N \geq (1+\delta)\lambda) \leq \exp\{\lambda(\delta - (1+\delta)\log(1+\delta))\}, \tag{5.9}$$

and

$$\mathbf{Pr}(N \leq (1-\delta)\lambda) \leq \exp\{\lambda(-\delta - (1-\delta)\log(1-\delta))\}. \tag{5.10}$$

Remark 22 *It is a simple exercise to check that the exponents are negative for all $\delta \in (0,1)$.*

For each $y \in \mathcal{Y}$, let $x(y) \in \mathrm{argmax}_{x:P_X(x)>0} P_{Y|X}(y|x)$. Let $\hat{\delta} = 1 - e^{-\delta}$, and consider the following:

$$\begin{aligned}
\mathbf{Pr}\left(\hat{M} - M \leq -\hat{\delta}M\right) &= \mathbf{Pr}\left(\sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} \tilde{N}_{x,y}/n\theta' \leq (1-\hat{\delta})M\right) \\
&= \mathbf{Pr}\left(\sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} \tilde{N}_{x,y} \leq (1-\hat{\delta})Mn\theta'\right) \\
&\leq \mathbf{Pr}\left(\sum_{y\in\mathcal{Y}} \tilde{N}_{x(y),y} \leq (1-\hat{\delta})Mn\theta'\right) \\
&\overset{(a)}{=} \mathbf{Pr}\left(\mathrm{Poi}\left(n\theta' M\right) \leq (1-\hat{\delta})Mn\theta'\right) \\
&\overset{(b)}{\leq} \exp\left\{Mn\theta'\left(-\hat{\delta} - (1-\hat{\delta})\log(1-\hat{\delta})\right)\right\} \\
&\overset{(c)}{\leq} \exp\left\{n\theta'\left(-\hat{\delta} - (1-\hat{\delta})\log(1-\hat{\delta})\right)\right\},
\end{aligned} \tag{5.11}$$

where (a) follows from the fact that $\tilde{N}_{x,y}$'s are independent $\mathrm{Poi}\left(n\theta' P_{Y|X}(y|x(y))\right)$, (b) follows from Lemma 18, and (c) follows from the fact that $M \geq 1$. Now consider

the probability that $\hat{M}$ exceeds M by a factor of at least $\hat{\delta}M$:

$$\begin{aligned}
\mathbf{Pr}\left(\hat{M} - M \geq \hat{\delta}M\right) &= \mathbf{Pr}\left(\sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} \tilde{N}_{x,y} \geq (1+\hat{\delta})Mn\theta'\right) \\
&= \mathbf{Pr}\left(\bigcup_{(x_1,\ldots,x_{|\mathcal{Y}|})\in\mathcal{X}^{|\mathcal{Y}|}} \left(\sum_{y\in\mathcal{Y}} \tilde{N}_{x_y,y} \geq (1+\hat{\delta})Mn\theta'\right)\right) \\
&= \mathbf{Pr}\left(\bigcup_{(x_1,\ldots,x_{|\mathcal{Y}|})\in\mathcal{X}^{|\mathcal{Y}|}} \left(\mathrm{Poi}\left(n\theta' \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x_y)\right) \geq (1+\hat{\delta})Mn\theta'\right)\right) \\
&\overset{(a)}{\leq} |\mathcal{X}|^{|\mathcal{Y}|}\mathbf{Pr}\left(\mathrm{Poi}(n\theta' M) \geq (1+\hat{\delta})Mn\theta'\right) \\
&\overset{(b)}{\leq} |\mathcal{X}|^{|\mathcal{Y}|} \exp\left\{Mn\theta'\left(\hat{\delta} - (1+\hat{\delta})\log(1+\hat{\delta})\right)\right\} \\
&\overset{(c)}{\leq} |\mathcal{X}|^{|\mathcal{Y}|} \exp\left\{n\theta'\left(\hat{\delta} - (1+\hat{\delta})\log(1+\hat{\delta})\right)\right\}, \qquad (5.12)
\end{aligned}$$

where (a) follows from Lemma 19 below and the fact that for any $(x_1, \ldots, x_{|\mathcal{Y}|}) \in \mathcal{X}^{|\mathcal{Y}|}$, $\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x_y) \leq M$, (b) follows from Lemma 18, and (c) follows from the fact that $M \geq 1$.

Lemma 19 *Consider $\lambda_1 > 0$, $\lambda_2 > 0$ such that $\lambda_1 \geq \lambda_2$, and let $N_1 \sim \mathrm{Poi}(\lambda_1)$, and $N_2 \sim \mathrm{Poi}(\lambda_2)$. Then, for all k,*

$$\mathbf{Pr}(N_1 \geq k) \geq \mathbf{Pr}(N_2 \geq k).$$

Proof: Let $\lambda_3 = \lambda_1 - \lambda_2$, and $N_3 \sim \mathrm{Poi}(\lambda_3)$ independent of N_2. Then $N_2 + N_3 \sim \mathrm{Poi}(\lambda_1)$. Hence $\mathbf{Pr}(N_1 \geq k) = \mathbf{Pr}(N_2 + N_3 \geq k) \geq \mathbf{Pr}(N_2 \geq k)$. □

Let

$$n^\star = \frac{\log(5/\epsilon) + |\mathcal{Y}|\log|\mathcal{X}|}{\theta'\left((1+\hat{\delta})\log(1+\hat{\delta}) - \hat{\delta}\right)}. \qquad (5.13)$$

For such a choice, we get by (5.11) and (5.12),

$$\mathbf{Pr}\left(|\hat{M} - M| \geq \hat{\delta}M\right) \leq 2\epsilon/5. \qquad (5.14)$$

Remark 23 *For all* $\hat{\delta} \in (0,1)$, $\hat{\delta} + (1-\hat{\delta})\log(1-\hat{\delta}) \geq (1+\hat{\delta})\log(1+\hat{\delta}) - \hat{\delta}$.

Note that the Poisson estimator behaves identically to the infinite-sequence estimator unless there exists $x \in \text{supp}(X)$ for which $N_x = 0$ or $\tilde{N}_x > N_x$. Therefore, we need to compute the probability of that event.

$$\begin{aligned}
\mathbf{Pr}\left(\text{there exists } x \in \text{supp}(X) : \tilde{N}_x > N_x\right) &\leq \sum_{x \in \mathcal{X}} \mathbf{Pr}(\tilde{N}_x > N_x) \\
&\leq \sum_{x \in \in \text{supp}(X)} \mathbf{Pr}\left(\text{Poi}(n^\star \theta') \geq \text{Poi}\left(n^\star P_X(x)\right)\right) \\
&\overset{(a)}{\leq} \sum_{x \in \in \text{supp}(X)} \exp\left\{-\left(\sqrt{n^\star P_X(x)} - \sqrt{n^\star \theta'}\right)^2\right\} \\
&\overset{(b)}{\leq} \sum_{x \in \in \text{supp}(X)} \exp\left\{-\left(\sqrt{4n^\star \theta'} - \sqrt{n^\star \theta'}\right)^2\right\} \\
&= |\mathcal{X}| e^{-n^\star \theta'} \\
&\overset{(c)}{\leq} \epsilon/5, \qquad (5.15)
\end{aligned}$$

where (a) follows from the Chernoff bound, (b) follows from the fact that for all $x \in \text{supp}(X)$, $P_X(x) \geq \theta = 4\theta'$, and (c) follows from the fact that $(1+\hat{\delta})\log(1+\hat{\delta}) - \hat{\delta} < 2\log 2 - 1 < 1$ for $\hat{\delta} \in (0,1)$. Similarly,

$$\begin{aligned}
\mathbf{Pr}\left(\text{there exists } x \in \text{supp}(X) : N_x = 0\right) \leq \sum_{x \in \text{supp}(X)} \mathbf{Pr}(N_x = 0) = \sum_{x \in \text{supp}(X)} e^{-n^\star P_X(x)} \leq |\mathcal{X}| e^{-n^\star \theta'} \\
\leq \epsilon/5. \quad (5.16)
\end{aligned}$$

Finally, we compare fixed-length sampling with Poisson sampling. Consider an optimal fixed-length estimator with $2n^\star$ samples. Then, a $\text{Poi}(n^\star)$ estimator can outperform it only if $N > 2n^\star$. However, by Lemma 18,

$$\mathbf{Pr}(\text{Poi}(n^\star) > 2n^\star) \leq e^{-n^\star(2\log 2 - 1)} \leq \epsilon/5. \qquad (5.17)$$

By equations (5.14), (5.15), (5.16), and (5.17), there exists a fixed-length estimator with $\hat{\delta}$ multiplicative accuracy (i.e., δ additive accuracy) and ϵ probability of

error that uses $2n^\star$ samples only. Hence,

$$S_{\delta,\epsilon}(|\mathcal{X}|, |\mathcal{Y}|, \theta) \leq 2\frac{\log(5/\epsilon) + |\mathcal{Y}| \log |\mathcal{X}|}{\theta' \left((1+\hat{\delta}) \log(1+\hat{\delta}) - \hat{\delta}\right)}.$$

Plugging in $\hat{\delta} = 1 - e^{-\delta}$ and $\theta' = \theta/4$ yields Theorem 17.

5.2 Proof of Theorem 18

Let $|\mathcal{Y}| = k$. We will derive a lower-bound on complexity by considering a subproblem, i.e., we will restrict our attention to a subset of $\mathcal{P}^{\theta}_{\mathcal{X}\times\mathcal{Y}}$ (cf. (5.3)). In particular, consider $P_{XY} \in \mathcal{P}^{\theta}_{\mathcal{X}\times\mathcal{Y}}$ which satisfy: $P_X(x_1) = \theta \in (0,1)$, and $P_{Y|X}$ has the following form:

$$P_{Y|X} = \begin{bmatrix} p_1 & p_2 & \cdots & p_k \\ 1/k & 1/k & \cdots & 1/k \\ \vdots & \vdots & & \vdots \\ 1/k & 1/k & \cdots & 1/k \end{bmatrix}, \tag{5.18}$$

where $p_Y = (p_1, p_2, \cdots, p_k)$ is some distribution over $\mathcal{Y}$. Now, for any distribution p_Y over $\mathcal{Y}$, define

$$h(p_Y) = \log\left(\sum_{y\in\mathcal{Y}} \max\left\{\frac{1}{k},\ p_y\right\}\right). \tag{5.19}$$

Therefore,

$$\mathcal{L}(P_X; P_{Y|X}) = h(P_{Y|X}(.|x_1)). \tag{5.20}$$

Hence, estimating maximal leakage for this subproblem is the same as estimating a property of $P_{Y|X}(.|x_1)$. We further simplify the analysis by considering Pois-

son sampling. In particular, let

$$\tilde{S}_{\delta,\epsilon}(|\mathcal{X}|,|\mathcal{Y}|,\theta) = \inf_f \min\{n : N \sim \mathrm{Poi}(n), \ \mathbf{Pr}\left(\left|\mathcal{L}(P_X; P_{Y|X}) - f(X^N, Y^N)\right| > \delta\right) < \epsilon, \\ \text{for all } P_{XY} \in \mathcal{P}^{\theta}_{\mathcal{X}\times\mathcal{Y}}\}, \tag{5.21}$$

and

$$\tilde{S}^h_{\delta,\epsilon}(|\mathcal{Y}|) = \inf_f \min\{n : N \sim \mathrm{Poi}(n), \mathbf{Pr}\left(\left|h(P_Y) - f(Y^N)\right| > \delta\right) < \epsilon, \ \text{for all } P_Y \in \mathcal{P}_{\mathcal{Y}}\}, \tag{5.22}$$

where $\mathcal{P}_{\mathcal{Y}}$ is the set of all probability distributions on $\mathcal{Y}$, and Y^n is drawn independently according to P_Y. Since sampling $\mathrm{Poi}(n)$ from P_{XY} gives $\mathrm{Poi}(n\theta)$ samples from $P_{Y|X}(.|x_1)$, we get

$$\tilde{S}_{\delta,\epsilon}(|\mathcal{X}|,|\mathcal{Y}|,\theta)) \geq \tilde{S}^h_{\delta,\epsilon}(|\mathcal{Y}|)/\theta. \tag{5.23}$$

On the other hand, given $N \sim \mathrm{Poi}(n)$,

$$\begin{aligned} &\inf_f \mathbf{Pr}\left(\left|\mathcal{L}(P_X; P_{Y|X}) - f(X^N, Y^N)\right| > \delta\right) > \epsilon \\ \Rightarrow &\inf_f \mathbf{Pr}\left(\left|\mathcal{L}(P_X; P_{Y|X}) - f(X^{n/2}, Y^{n/2})\right| > \delta\right) > \epsilon/2, \end{aligned} \tag{5.24}$$

when $n > \log(2/\epsilon)/\log(e/2)$. It remains to show that, for each $\eta > 0$,

$$\tilde{S}^h_{\delta,\epsilon}(k) \geq \Omega(k^{1-\eta}). \tag{5.25}$$

To that end, let $Y_1, Y_2, \cdots, Y_N$ be N independent samples drawn from p_Y, where $N \sim \mathrm{Poi}(n)$. Since $h(\cdot)$ is symmetric, and the sampling is done i.i.d, the *profile* $\Phi(Y^N)$ is a sufficient statistic, i.e., $\Phi = (\Phi_1, \Phi_2, \cdots)$ where Φ_l is the number of elements y that appeared l times in the sequence Y^N. To get a lower bound on $\tilde{S}^h_{\delta,\epsilon}(k)$, we will exhibit two distribution p_Y and q_Y such that

$$|h(p_Y) - h(q_Y)| > 2\delta, \tag{5.26}$$

and when $n \leq O(k^{1-\eta})$,

$$\|p_\Phi - q_\Phi\|_{TV} < \epsilon, \tag{5.27}$$

where p_Φ and q_Φ are the induced distributions on the profiles under Poisson sampling by distributions p_Y and q_Y, respectively, and $\|.\|_{TV}$ is the total variation distance. If we can exhibit such distributions, then the probability of error is at least $1/2 - \epsilon/2 > \epsilon$ for $\epsilon < 1/3$. To that end, the following lemma, shown by Valiant [48], will be useful:

Lemma 20 ([48] [1, Theorem 19]) *Given distributions p and q such that*

$$\max_y \max\{p_y, q_y\} \leq \frac{\epsilon}{40n}, \tag{5.28}$$

for Poisson sampling with $N \sim \text{Poi}(n)$, it holds that

$$\|p_\Phi - q_\Phi\|_{TV} \leq \epsilon/2 + 5\sum_{m=1}^{\infty} n^m |P_m(p) - P_m(q)|, \tag{5.29}$$

where $P_m(p) = \sum_y p_y^m$.

So we will construct distributions such that they satisfy (5.26) and their moments are equal up to an arbitrarily high degree. The following construction is based on the work of Acharya *et al.* [1] on Renyi entropy estimation. Fix $\eta > 0$. Let d be a fixed even integer satisfying $d < 1/\eta$, and let $k = d\ell$ for some $\ell \in \mathbb{N}$. Let $v = (v_1, v_2, \cdots, v_d) \in \mathbb{R}^d$, and associate with v the following distribution:

$$p_{ij}^v = \frac{|v_i|}{\ell \|v\|_1}, \;\; 1 \leq i \leq d, \; 1 \leq j \leq \ell. \tag{5.30}$$

Note that (by a simple calculation),

$$P_m(p^v) = \frac{1}{\ell^{m-1}} \left(\frac{\|v\|_m}{\|v\|_1} \right)^m. \tag{5.31}$$

Now, let $v = (1, 2, \cdots, d)$. Consider the polynomial

$$p(z) = (z-1)(z-2)\cdots(z-d), \tag{5.32}$$

and let $q(z) = p(z) - \Delta$, where Δ is small enough so that $q(z)$ has d positive roots. Let $w = (w_1, w_2, \cdots, w_d)$ represent the roots of $q(z)$. Since the sum of the first $d-1$ powers of a root of a polynomial of dth order do not depend on its constant (by the Newton-Girard identities [32]), then

$$\|v\|_m = \|w\|_m, \quad 1 \le m \le d-1. \tag{5.33}$$

By letting q^w be the distribution associated with w (similarly to (5.30)), we get by (5.31) and (5.33)

$$\sum_{m=1}^{\infty} n^m |P_m(p^v) - P_m(q^w)| = \sum_{m=d}^{\infty} \frac{n^m}{\ell^{m-1}} \left| \left(\frac{\|v\|_m}{\|v\|_1}\right)^m - \left(\frac{\|w\|_m}{\|v\|_1}\right)^m \right|. \tag{5.34}$$

Suppose $n \le C_{\epsilon,d}\ell^{1-1/d}$ for some constant $C_{\epsilon,d}$. Then,

$$\begin{aligned}
\sum_{m=1}^{\infty} n^m |P_m(p^v) - P_m(q^w)| &= \sum_{m=d}^{\infty} \frac{n^m}{\ell^{m-1}} \left| \left(\frac{\|v\|_m}{\|v\|_1}\right)^m - \left(\frac{\|w\|_m}{\|v\|_1}\right)^m \right| \\
&\le \sum_{m=d}^{\infty} \frac{(C_{\epsilon,d}\ell^{1-1/d})^m}{\ell^{m-1}} \left| \left(\frac{\|v\|_m}{\|v\|_1}\right)^m - \left(\frac{\|w\|_m}{\|v\|_1}\right)^m \right| \\
&= \sum_{m=d}^{\infty} \frac{C_{\epsilon,d}^m}{\ell^{m/d-1}} \left| \left(\frac{\|v\|_m}{\|v\|_1}\right)^m - \left(\frac{\|w\|_m}{\|v\|_1}\right)^m \right| \\
&\le \ell \sum_{m=d}^{\infty} \left(\frac{C_{\epsilon,d}}{\ell^{1/d}}\right)^m \\
&= \frac{C_{\epsilon,d}^d}{1 - C_{\epsilon,d}/\ell^{1/d}} \\
&\le \frac{C_{\epsilon,d}^d}{1 - C_{\epsilon,d}} \\
&\le \epsilon/10,
\end{aligned} \tag{5.35}$$

where the last inequality holds for the proper choice of $C_{\epsilon,d}$, which we will denote by $\hat{C}_{\epsilon,d}$. Therefore, when (5.28) holds, $\|p_\Phi - q_\Phi\|_{TV} \le \epsilon$. For (5.26), we will

show that $|h(p_Y) - h(q_Y)|$ is a non-zero constant independent of θ and ℓ. To that end, let $p_i = \frac{v_i}{\|v\|_1}$, $q_i = \frac{w_i}{\|v\|_1}$. Recall

$$h(p_Y) = \log\left(\sum_{i,j} \max\left\{\frac{1}{k},\ p_{ij}\right\}\right),$$

and note that

$$p_{ij} > \frac{1}{k} \Leftrightarrow \frac{|v_i|}{\ell\|v\|_1} > \frac{1}{\ell d} \Leftrightarrow p_i > \frac{1}{d} \Leftrightarrow \frac{i}{d(d+1)/2} > \frac{1}{d} \Leftrightarrow i > \frac{d+1}{2}.$$

Then,

$$\begin{aligned} e^{h(p)} &= \sum_{i=d/2+1}^{d}\sum_{j=1}^{\ell} p_{ij} + \sum_{i=1}^{d/2}\sum_{j=1}^{\ell} 1/k = \sum_{i=d/2+1}^{d} p_i + \sum_{i=1}^{d/2} 1/d = \frac{1}{2} + \frac{1}{d(d+1)/2}\sum_{i=d/2+1}^{d} i \\ &= \frac{5}{4} - \frac{1}{4(d+1)}. \end{aligned} \tag{5.36}$$

To evaluate $h(q)$, we will approximate $w_i - v_i$. First, note that

$$\int_{v_i}^{w_i} p'(z)dz = p(w_i) - p(v_i) = q(w_i) + \Delta = \Delta. \tag{5.37}$$

Then, for small Δ, we get

$$(w_i - v_i)p'(v_i) = \Delta + o(\Delta) \Rightarrow w_i - v_i = \frac{\Delta}{p'(v_i)} + o(\Delta). \tag{5.38}$$

Now, note that

$$p'(z) = \sum_{i=1}^{d}\prod_{j\neq i}(z-j).$$

Then, for $i \in \{1, 2, \cdots, d\}$,

$$p'(i) = (i-1)...(i-(i-1))\cdot(i-(i+1))...(i-d) = (-1)^{d-i}\frac{d!}{i\binom{d}{i}}. \tag{5.39}$$

It follows from (5.38) and (5.39) that for Δ small enough, $p_i > 1/d$ implies $q_i > 1/d$, and similarly $p_i < 1/d$ implies $q_i < 1/d$. (d was chosen to be even so that

$p_i \neq 1/d$ for all i.) Then,

$$\begin{aligned}
e^{h(q)} &= \sum_{i=d/2+1}^{d} \sum_{j=1}^{\ell} q_{ij} + \sum_{i=1}^{d/2} \sum_{j=1}^{\ell} 1/k \\
&= \frac{1}{2} + \sum_{i=d/2+1}^{d} q_i \\
&= \frac{1}{2} + \frac{1}{d(d+1)/2} \sum_{i=d/2+1}^{d} [v_i + (w_i - v_i)] \\
&\overset{(a)}{=} \frac{5}{4} - \frac{1}{4(d+1)} + \frac{2\Delta}{d(d+1)} \sum_{i=d/2+1}^{d} \left(\frac{1}{p'(i)} + o(1) \right) \\
&= \frac{5}{4} - \frac{1}{4(d+1)} + \frac{2\Delta}{d(d+1)} \sum_{i=d/2+1}^{d} \left((-1)^{d-i} \frac{i\binom{d}{i}}{d!} + o(1) \right) \\
&\overset{(b)}{=} \frac{5}{4} - \frac{1}{4(d+1)} - \frac{2\Delta}{d(d+1)} (-1)^{d/2} \frac{d(d+2)\binom{d}{d/2+1}}{4(d-1)(d!)} + o(\Delta), \qquad (5.40)
\end{aligned}$$

where (a) follows from equations (5.36) and (5.38), and (b) follows from the Lemma below (the proof of which is given in Appendix D). Note that (5.36) and (5.40) imply $h(p) \neq h(q)$, and the difference is independent of ℓ and ϵ i.e., (5.26) holds for small enough (constant) δ.

To sum up, if $n \leq \hat{C}_{\epsilon,d} \ell^{1-1/d}$ and (5.28) holds, then (5.27) holds and the probability of error subsequently exceeds ϵ. Since $\max_y \max\{p_y, q_y\} \approx 2/(d\ell)$, (5.28) holds when $n \leq \epsilon d\ell/80$. That is, we need $n \geq \min\{\epsilon d\ell/80, \hat{C}_{\epsilon,d} \ell^{1-1/d}\}$. Hence,

$$n \geq \Omega(\tilde{C}_{\epsilon,d} \ell^{1-1/d}) = \Omega(\tilde{C}_{\epsilon,d} d^{1/d-1} k^{1-1/d}). \qquad (5.41)$$

Since d was arbitrary, we have established Theorem 18. ■

Lemma 21 *For d even, and $1 \leq j \leq d/2$,*

$$\sum_{i=d/2+j}^{d} (-1)^{d-i} \binom{d}{i} i = (-1)^{d/2-j} \frac{(2j+d-2)(2j+d)\binom{d}{d/2+j}}{4(d-1)}.$$

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

In this thesis, we introduced maximal leakage as an operationally-defined measure to quantify information leakage, particularly in side-channel problems. The measure satisfies axiomatic properties of an information measure, is simple to compute, and is robust to variations in its definition. Moreover, it is sandwiched between mutual information and local differential privacy (which is known to be pessimistic).

Furthermore, we studied the Shannon cipher system using maximal leakage as a performance metric. Counter-intuitively, we showed that memoryless schemes are strictly suboptimal, whereas quantization-based schemes achieve optimality. A limitation of this study is that it allows for non-causal encoding. Such an assumption is in many cases unrealistic. Even block coding can introduce unacceptable delays, which is particularly true for the SSH example. We propose then to consider *causal* encoding for the IBS (with a memoryless source). We conjecture that the following causal scheme is optimal (and in fact matches the non-causal performance): Let $P_{Y|X}$ be the optimal test channel. Then for each $x \in \mathcal{X}$, choose a sequence $Y_x^{nP_X(x)}$ from the typical set corresponding to $P_{Y|X=x}$. As the source sequence arrives, for the ith appearance of symbol x, we output $Y_x(i)$. Concurrently, we monitor the empirical type of the incoming source sequence: this is an $|\mathcal{X}|$-dimensional random walk. The distribution of P_X implies that there is a certain (small enough) ball in which the trajectory will reside with high probability (e.g., the concentration at the endpoint is dictated by the law of large numbers). If the empirical trajectory ever leaves that ball, we will output junk for the remainder of the block. The intuition is that this will only happen

for atypical sequences. Moreover, we propose to study a timing version of the IBS, which most resembles the SSH side-channel. In this case, the input and output are point processes, which can be handled by the general formula given in Theorem 4, and the fidelity constraint is replaced with a maximum and an average delay constraint. Additionally, maximal leakage can be used to revisit other design problems for leakage mitigation, which have been studied under a different leakage metric, such as the wiretap channel.

Our study of the sample complexity of estimating maximal leakage from data suggests that, for mechanism design problems, it is better to restrict the space of schemes to those that are amenable to mathematical analysis. Indeed, the task of learning maximal leakage from data is infeasible unless we make a certain assumption about the source distribution. Therefore, in the absence of this assumption, it is infeasible to analyze a given scheme solely from simulations.

Finally, we explored connections between maximal leakage and existing information metrics. In particular, we used the guessing framework to provide operational definitions for commonly used leakage measures (such as mutual information and local differential privacy). Such connections provide an interesting area for further study. For instance, we noted that maximal leakage is equal to Sibson mutual information of order infinity, which is a generalization of Renyi entropy and Renyi divergence. However, there is no definition in the literature for conditional Sibson mutual information, and as such, conditional maximal leakage might inspire such a definition. It is also worth noting that $I_\infty(X;Y)$ has been recently proposed as a complexity measure in the communication complexity literature, and has appeared in the compression literature as

the Shtarkov sum. The connection between the maximal leakage formula and hypothesis testing suggests that it could be useful in the study of classification and learning problems.

APPENDIX A

PROOFS FOR SECTION 3.5

A.1 Proof of Theorem 11

To show ($\leq$) direction, fix U, $\hat{\mathcal{U}}$ and d, and consider:

$$\begin{aligned}
\inf_{\hat{U}:X-Y-\hat{U}} \mathbf{E}[d(U,\hat{U})] &= \sum_{y\in\mathcal{Y}} \min_{\hat{u}} \sum_{u\in\mathcal{U}} P_{UY}(u,y)d(u,\hat{u}) \\
&= \sum_{y\in\mathcal{Y}} \min_{\hat{u}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathrm{supp}(X)} P_X(x)P_{U|X}(u|x)P_{Y|X}(y|x)d(u,\hat{u}) \\
&\geq \sum_{y\in\mathcal{Y}} \left(\min_{\hat{u}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathrm{supp}(X)} P_X(x)P_{U|X}(u|x)d(u,\hat{u}) \right) \min_{\tilde{x}\in\mathrm{supp}(X)} P_{Y|X}(y|\tilde{x}) \\
&= \sum_{y\in\mathcal{Y}} \left(\min_{\hat{u}} \mathbf{E}[d(U,\hat{u})] \right) \min_{\tilde{x}\in\mathrm{supp}(X)} P_{Y|X}(y|\tilde{x}).
\end{aligned}$$

For the reverse direction, let $U = X$, $\hat{\mathcal{X}} = \mathrm{supp}(X)$, and

$$d(x,\hat{x}) = \begin{cases} \frac{1}{P_X(x)}, & x = \hat{x}, \\ 0, & x \neq \hat{x}. \end{cases} \tag{A.1}$$

Then,

$$\min_{\hat{x}\in\mathrm{supp}(X)} \mathbf{E}[d(X,\hat{x})] = \sum_{x\in\mathrm{supp}(X)} P_X(x)d(x,\hat{x}) = \min_{\hat{x}\in\mathrm{supp}(X)} P_X(\hat{x})d(\hat{x},\hat{x}) = 1, \tag{A.2}$$

and for a given $y \in \mathcal{Y}$,

$$\min_{\hat{x}\in\mathrm{supp}(X)} \sum_{x\in\mathrm{supp}(X)} P_X(x)P_{Y|X}(y|x)d(x,\hat{x}) = \min_{\hat{x}\in\mathrm{supp}(X)} P_{Y|X}(y|\hat{x}), \tag{A.3}$$

which concludes the proof. ■

A.2 Proof of Corollary 6

In the following, assume X has full support.

1) The data processing inequality follows directly from the definition.

2) If $P_{Y|X}$ is non-trivial and deterministic, then for each y such that $P_Y(y) > 0$, there exists $x \in \text{supp}(X)$ such that $P_{Y|X}(y|x) = 0$.

3) The "if" direction is straightforward. The "only if" direction follows from the fact that, for each y, $\min_x P_{Y|X}(y|x) \leq P_Y(y)$. Thus, $\sum_y \min_x P_{Y|X}(y|x) = 1 \Rightarrow \forall y, \min_x P_{Y|X}(y|x) = P_Y(y) \Rightarrow X$ and Y are independent.

4) Local-differential privacy upper-bounds maximal cost leakage since:

$$\frac{1}{\sum_y \min_x P_{Y|X}(y|x)} = \frac{\sum_y P_Y(y)}{\sum_y \min_x P_{Y|X}(y|x)} \leq \max_y \frac{P_Y(y)}{\min_x P_{Y|X}(y|x)} \leq \max_{x,x',y} \frac{P_{Y|X}(y|x')}{P_{Y|X}(y|x)}.$$

Example 9 is an example of 5), and 6) is straightforward.

7) Convexity follows from the fact that $\min_x P_{Y|X}(y|x)$ is concave in $P_{Y|X}$, and $(-\log)$ is a non-increasing convex function.

A.3 Proof of Theorem 12

Without loss of generality, assume X and Y have full support. To show LHS $\leq$ RHS, fix any $\hat{\mathcal{X}}$, d and $y \in \mathcal{Y}$, and consider:

$$\begin{aligned}
\min_{\hat{u}\in\hat{\mathcal{U}}} \mathbf{E}[d(U,\hat{u})|Y=y] &= \min_{\hat{u}\in\hat{\mathcal{U}}} \sum_{u\in\mathcal{U}} P_{U|Y}(u|y) d(u,\hat{u}) \\
&= \min_{\hat{u}\in\hat{\mathcal{U}}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_{X|Y}(x|y) P_{U|X}(u|x) d(u,\hat{u}) \\
&= \min_{\hat{u}\in\hat{\mathcal{U}}} \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} \frac{P_{Y|X}(y|x)}{P_Y(y)} P_X(x) P_{U|X}(u|x) d(u,\hat{u}) \\
&\geq \min_{\hat{u}\in\hat{\mathcal{U}}} \frac{\min_{x'} P_{Y|X}(y|x')}{P_Y(y)} \sum_{u\in\mathcal{U}} \sum_{x\in\mathcal{X}} P_X(x) P_{U|X}(u|x) d(u,\hat{u}) \\
&= \frac{\min_{x'} P_{Y|X}(y|x')}{P_Y(y)} \min_{\hat{u}\in\hat{\mathcal{U}}} \mathbf{E}[d(U,\hat{u})].
\end{aligned}$$

The reverse direction follows by using the same d as in (A.1).

A.4 Proof of Corollary 7

To show that the left-hand side is upper-bounded by the right-hand side, note that $P_Y(y) \leq \max_x P_{Y|X}(y|x)$. For the reverse direction, consider the following. Let $y^\star$ be an element achieving the max of L^{dp}. Let $x_0 \in \operatorname{argmin}_x P_{Y|X}(y^\star|x)$ and $x_1 \in \operatorname{argmax}_x P_{Y|X}(y^\star|x)$. Finally, for a given $\alpha > 0$, let $P_X(x_0) = 1-\alpha$ and $P_X(x_1) = \alpha$. Then,

$$\begin{aligned}
\max_{P_X} \mathcal{L}^{rc}(X\to Y) &\geq \log \frac{P_Y(y^\star)}{P_{Y|X}(y^\star|x_0)} \\
&= \log \frac{(1-\alpha)P_{Y|X}(y^\star|x_0) + \alpha P_{Y|X}(y^\star|x_1)}{P_{Y|X}(y^\star|x_0)} \\
&\xrightarrow{\alpha\to 1} \log \frac{P_{Y|X}(y^\star|x_1)}{P_{Y|X}(y^\star|x_0)} = L^{dp}(X\to Y).
\end{aligned}$$

■

APPENDIX B

PROOF OF EQUATION (4.12)

Let $P_{Y|X} = \begin{bmatrix} 1 - W_{10} & W_{10} \\ W_{01} & 1 - W_{01} \end{bmatrix}$ (where the first column corresponds to $y = 0$, the second to $y = 1$). Dropping the log, we can rewrite the problem as:

$$\begin{aligned} &\text{minimize } \max\{1 - W_{10}, W_{01}\} + \max\{W_{10}, 1 - W_{01}\} \\ &\text{subject to } (1 - p)W_{10} + pW_{01} \leq D, \;\; 0 \leq W_{10}, W_{01} \leq 1. \end{aligned} \tag{B.1}$$

Now note that

$$W_{10} + W_{01} \overset{(a)}{\leq} \frac{1 - p}{p} W_{10} + W_{01} = \frac{1}{p}((1 - p)W_{10} + pW_{01}) \overset{(b)}{\leq} \frac{D}{p} \overset{(c)}{\leq} 1, \tag{B.2}$$

where (a) follows because $p \leq 1/2$, (b) follows from the constraint in (B.1), and (c) follows because $D \leq p$. Using (B.2), we can rewrite (B.1) as

$$\begin{aligned} &\text{minimize } 2 - (W_{10} + W_{01}) \\ &\text{subject to } (1 - p)W_{10} + pW_{01} \leq D, \;\; 0 \leq W_{10}, W_{01} \leq 1. \end{aligned} \tag{B.3}$$

Therefore, we need to maximize $(W_{10} + W_{01})$. By (3), the sum is upper-bounded by D/p. The upper bound can be achieved by setting

$$W_{10}^{\star} = 0 \text{ and } W_{01}^{\star} = D/p, \tag{B.4}$$

which clearly satisfies the constraint in (B.1). Therefore, for the optimal $P_{Y|X}^{\star}$, we get

$$\mathcal{L}(X \to Y^{\star}) = \log(2 - D/p). \tag{B.5}$$

APPENDIX C

PROOFS FOR SECTION 4.6

C.1 Proof of Proposition 7

C.1.1 Proof of Property (P1)

(P1): For fixed P_{XY}, $R(P_{XY}, D_e)$ is a finite valued, non-increasing convex function of D_e. Furthermore, $R(P_{XY}, D_e)$ is a uniformly continuous function of the pair (P_{XY}, D_e).

Fix P_{XY}. The minimization in (4.15) is over a compact set, which is non-empty due to assumption (A3). Since $I(X;V|Y)$ is a continuous function of $P_{V|X,Y}$, the minimum is achieved. The monotonicity in D_e follows directly from the definition. It is easy to check that $I(X;V|Y)$ is convex in $P_{V|X,Y}$ for fixed P_{XY}. Then, the proof of the convexity of $R(P_{XY}, D_e)$ in D_e follows similarly to the case of the rate-distortion function with no side information (see Lemma 2.2.2 in [11]).

To show the uniform continuity in the pair (P_{XY}, D_e), consider the following proposition, the proof of which is given in Appendix C.7.1.

Proposition 22 *Let N_1 and N_2 be in $\mathbb{N}$, and let $\mathcal{S}$ and $\mathcal{U}$ be compact subsets of $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, respectively. Let ν be a non-negative continuous function defined on $\mathcal{S} \times \mathcal{U}$, and let ϑ be a real-valued continuous function defined on $\mathcal{S} \times \mathcal{U}$. Suppose they satisfy the following condition:*

(PA) If $(s, u_1) \in \mathcal{S} \times \mathcal{U}$ satisfies $\nu(s, u_1) = \min_{u' \in \mathcal{U}} \nu(s, u')$, then there exists u_2 such that $\vartheta(s, u_2) = \vartheta(s, u_1)$, and for all $s' \in \mathcal{S}$, $\nu(s', u_2) = \min_{u' \in \mathcal{U}} \nu(s', u')$.

Let $t_0 = \max_{s \in \mathcal{S}} \min_{u \in \mathcal{U}} \nu(s, u)$, and let φ be a function on $\mathcal{S} \times [t_0, +\infty)$ defined as follows:

$$\varphi(s, t) = \min_{u : \nu(s,u) \leq t} \vartheta(s, u).$$

If for fixed $s \in \mathcal{S}$, $\varphi(s, t)$ is continuous in t, then $\varphi(s, t)$ is continuous in the pair (s, t).

Remark 24 *The proposition generalizes Lemma 2.2.2 in [11], which shows the continuity of the regular rate-distortion function, and the proof follows along similar lines.*

The proposition yields immediately the continuity of $R(P_{XY}, D_e)$ by identifying $\mathcal{S}$ with $\mathcal{P}_{\mathcal{X}\mathcal{Y}}$, $\mathcal{U}$ with the set of conditional probability distributions $P_{V|XY}$, t_0 with $D_{e,\min}$, and the functions ν, ϑ, and φ with $\mathbf{E}[d_e(X, V)]$, $I(X; V|Y)$, and $R(P_{XY}, D_e)$ respectively. It is easy to check that $D_{e,\min} = \max_{P_{XY}} \min_{P_{V|XY}} \mathbf{E}[d_e(X, V)]$ so that we can identify it with t_0. To see why $\mathbf{E}[d_e(X, V)]$ and $I(X; V|Y)$ satisfy (PA), note the following. For notational convenience, we write $\mathbf{E}[d_e(X, V)]$ as $d_e(P_{XY}, P_{V|XY})$, and $I(X; V|Y)$ as $I(P_{XY}, P_{V|XY})$. Suppose $d_e(P_{XY}, P_{V|XY}) = \min_{\hat{P}_{V|XY}} d_e(P_{XY}, \hat{P}_{V|XY})$ and let $D_e(x) = \min_{v \in \mathcal{V}} d_e(x, v)$ for $x \in \mathcal{X}$. Then for all (x, v) such that $d_e(x, v) > D_e(x)$, $P_{XV}(x, v) = 0$. Expanding $P_{XV}(x, v)$:
$P_{XV}(x, v) = \sum_y P_{XY}(x, y) P_{V|XY}(v|x, y) = 0 \Rightarrow$ for all $y \in \mathcal{Y}$, $P_{XY}(x, y) = 0$ or $P_{V|XY}(v|x, y) = 0$. Then, define $P'_{V|XY}$ as follows:

- If $P_{XY}(x, y) > 0$, let $P'_{V|(X=x,Y=y)} = P_{V|(X=x,Y=y)}$.
- If $P_{XY}(x, y) = 0$, let $P'_{V|(X=x,Y=y)}$ satisfy $P'_{V|(X=x,Y=y)}(v|x, y) = 0$ if $d_e(x, v) > D_e(x)$.

Then $P_{XY}P_{V|X,Y} = P_{XY}P'_{V|X,Y}$, thus $I(P_{XY}, P_{V|XY}) = I(P_{XY}, P'_{V|XY})$. Moreover, the definition of $P'_{V|XY}$ guarantees that $d_e(x,v) > D_e(x) \Rightarrow P'_{V|XY}(v|x,y) = 0$ for all y. Therefore, for any joint distribution P'_{XY}, $d_e(P'_{XY}, P'_{V|XY}) = \min_{\hat{P}_{V|XY}} d_e(P_{XY}, \hat{P}_{V|XY})$.

Finally, to prove uniform continuity, note that $R(P_{XY}, D_e) = R(P_{XY}, D_{e,\max})$ for all $D_e \geq D_{e,\max}$. Therefore, $R(P_{XY}, D_e)$ is uniformly continuous on the set $\mathcal{P}_{\mathcal{X}\mathcal{Y}} \times [D_{e,\max}, \infty)$. Since it is also uniformly continuous on $\mathcal{P}_{\mathcal{X}\mathcal{Y}} \times [D_{e,\min}, D_{e,\max}]$, the result is established. ■

C.1.2 Proof of Property (P2)

(P2): For fixed P_X, $R(P_X, D, D_e)$ is a finite-valued function of (D, D_e). Moreover, for fixed D_e, $R(P_X, D, D_e)$ is a uniformly continuous function of the pair (P_X, D).

Fix P_X. The maximization in (4.16) is over a compact set, which is non-empty due to assumption (A3). Since $R(P_{XY}, D_e)$ is a continuous function of P_{XY}, it is also continuous in $P_{Y|X}$ for fixed P_X. Therefore, the maximum is achieved.

As for the continuity in (P_X, D) for fixed D_e, we view $R(P_X, D, D_e)$ as a function of (P_X, D), and $R(P_{XY}, D_e)$ as function of $(P_X, P_{Y|X})$. In the terminology of Proposition 22, we identify $\mathcal{S}$ with $\mathcal{P}_X$, $\mathcal{U}$ with the set of conditional probability distributions $P_{Y|X}$, t_0 with $D_{\min}$, and the functions ν, ϑ, and φ with $\mathbf{E}[d(X,Y)]$, $-R(P_{XY}, D_e)$, and $-R(P_X, D, D_e)$ respectively. Proving that $\mathbf{E}[d(X,Y)]$ and $R(P_{XY}, D_e)$ satisfy (PA) follows along the same lines as proving $\mathbf{E}[d_e(X,V)]$ and $I(X;V|Y)$ satisfy (PA) . Moreover, if continuity holds, uniform continuity follows from the fact that $R(P_X, D, D_e)$ is constant for all $D \geq D_{\max}$.

It remains to show that $R(P_X, D, D_e)$ is a continuous function of D for fixed P_X

and D_e. The result of Proposition 22 then applies immediately. To this end, consider the following proposition, the proof of which is given in Appendix C.7.2

Proposition 23 *Let N be in $\mathbb{N}$, and let $\mathcal{T}$ be a non-empty compact subset of $\mathbb{R}^N$. Let L be a real-valued continuous function defined on $\mathcal{T}$. Let $T_1 \supseteq T_2 \supseteq \cdots$ be a decreasing sequence of non-empty compact subsets of $\mathcal{T}$. Let $T = \bigcap_{i\geq 1} T_i$. Then,*

$$\lim_{k\to\infty} \max_{t\in T_k} L(t) = \max_{t\in T} L(t).$$

Moreover, let $S_1 \subseteq S_2 \subseteq \cdots$ be an increasing sequence of non-empty compact subsets of $\mathcal{T}$. Let $S = \overline{\bigcup_{i\geq 1} S_i}$ (where the bar denotes closure of the set). Then

$$\lim_{k\to\infty} \max_{t\in S_k} L(t) = \max_{t\in S} L(t).$$

Consequently, if $\mathcal{T}$ is also convex, and L_c is a real-valued convex and continuous function defined on $\mathcal{T}$ with $s_0 = \min_{t\in\mathcal{T}} L_c(t)$, then

$$\hat{L}(s) := \max_{t:L_c(t)\leq s} L(t)$$

is continuous in $s \in [s_0, +\infty)$.

It follows immediately then that $R(P_X, D, D_e)$ is continuous in D for fixed P_X and D_e, since $\mathbf{E}[d(X,Y)]$ is convex and continuous in $P_{Y|X}$, and $R(P_{XY}, D_e)$ is continuous in $P_{Y|X}$ (for fixed P_X).

C.1.3 Proof of Property (P3)

(P3): $R_e(P_X, D_e) - R(P_X, D) \leq R(P_X, D, D_e) \leq R_e(P_X, D_e)$.

The upper bound is straightforward since $R(P_{XY}, D_e)$, the rate-distortion function with side information, is always upper-bounded by $R_e(P_X, D_e)$. The

lower bound is derived by considering a conditional $P^{\star}_{Y|X}$ that achieves the rate-distortion function.

$$
\begin{aligned}
R(P_X, D, D_e) = & \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \leq D}} \min_{\substack{P_{V|X,Y}: \\ \mathbf{E}[d_e(X,V)] \leq D_e}} I(X;V|Y) \\
& \geq \min_{\substack{P_{V|X,Y}: \\ \mathbf{E}[d_e(X,V)] \leq D_e}} H_{P^{\star}_{XY}}(X|Y) - H_{P^{\star}_{XY} P_{V|XY}}(X|V,Y) \qquad \text{(C.1)} \\
& \geq \min_{\substack{P_{V|X,Y}: \\ \mathbf{E}[d_e(X,V)] \leq D_e}} H_{P^{\star}_{XY}}(X|Y) - H_{P_X P_{V|X}}(X|V) \qquad \text{(C.2)} \\
& = -H_{P_X}(X) + H_{P^{\star}_{XY}}(X|Y) + \\
& \qquad \min_{\substack{P_{V|X,Y}: \\ \mathbf{E}[d_e(X,V)] \leq D_e}} H_{P_X}(X) - H_{P_X P_{V|X}}(X|V) \\
& = -R(P_X, D) + R_e(P_X, D_e).
\end{aligned}
$$

C.2 Proof of Proposition 9

First, consider the following proposition.

Proposition 24 *For all $\epsilon > 0$, there exists $n_2(\epsilon, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{V}|)$, such that for all $n \geq n_2$, for all $D_e \geq D_{e,\min}$, for each $Q_{XY} \in \mathcal{Q}^n_{XY}$,*

$$
\left| \min_{\substack{P_{XYV} \in \\ \mathcal{Q}^n_{XYV}(Q_{XY}, D_e)}} I_{P_{XYV}}(X;V|Y) - R(Q_{XY}, D_e) \right| \leq \epsilon.
$$

Proof: It follows directly from the definition that

$$
\min_{\substack{P_{XYV} \in \\ \mathcal{Q}^n(Q_{XY}, D_e)}} I_{P_{XYV}}(X;V|Y) \geq R(Q_{XY}, D_e).
$$

So, we only need to show the other direction. To that end, let $\delta > 0$ be small enough such that

$$
\|P_{XYV} - P'_{XYV}\| \leq \delta \Rightarrow \left| I_{P_{XYV}}(X;V|Y) - I_{P'_{XYV}}(X;V|Y) \right| \leq \epsilon, \qquad \text{(C.3)}
$$

where $\|\cdot\|$ is used to indicate the L_2-norm. Let $n \geq n_2 \geq |\mathcal{V}|\sqrt{|\mathcal{X}||\mathcal{Y}||\mathcal{V}|}/\delta$. Fix $Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}$, and let $P^\star_{V|XY}$ be the conditional distribution achieving the minimum in $R(Q_{XY}, D_e)$. We construct a conditional distribution $P'_{V|XY}$ as follows. For each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we will choose $P'_{V|X=x,Y=y}$ from $\mathcal{Q}^{nQ_{XY}(x,y)}_{\mathcal{V}}$, i.e., the set of rational PMFs over $\mathcal{V}$ with denominator $nQ_{XY}(x, y)$ (if $Q_{XY}(x, y) = 0$, then we can choose $P'_{V|X=x,Y=y}$ to be any distribution). This guarantees that $Q_{XY}P'_{V|XY}$ is in $\mathcal{Q}^n_{\mathcal{X}\mathcal{Y}\mathcal{V}}$. Let $v(x) = \operatorname{argmin}_{v\in\mathcal{V}} d_e(x, v)$ for $x \in \mathcal{X}$ (if more than one v achieves the minimum, choose one arbitrarily). We construct $P'_{V|XY}$ by rounding $P^\star_{V|XY}$ as follows. For each $(x, y) \in \mathcal{X} \times \mathcal{V}$, for $v \neq v(x)$, we set $P'_{V|XY}(v|x, y)$ to be the largest integer multiple of $1/(nQ_{XY}(x, y))$ that is smaller than $P^\star_{V|XY}(v|x, y)$, i.e., we round down with resolution $1/(nQ_{XY}(x, y))$ and denote this operation by $\lfloor . \rfloor_{nQ_{XY}(x,y)}$. Finally, we set $P'_{V|XY}(v(x)|x, y)$ appropriately to make $P'_{V|XY}(.|x, y)$ a valid probability distribution. It is easy to see that, for such a choice,

$$\left|P'_{V|XY}(v|x, y) - P^\star_{V|XY}(v|x, y)\right| \leq \frac{|\mathcal{V}|}{nQ_{XY}(x, y)}.$$

Moreover, this readily implies that

$$\left\|Q_{XY}P'_{V|XY} - Q_{XY}P^\star_{V|XY}\right\| \leq \frac{|\mathcal{V}|\sqrt{|\mathcal{X}||\mathcal{Y}||\mathcal{V}|}}{n} \leq \delta. \tag{C.4}$$

Let $P^\star_{XYV} = Q_{XY}P^\star_{V|XY}$, and $P'_{XYV} = Q_{XY}P'_{V|XY}$. Now, note that

$$\begin{aligned}
D_e &\geq \mathbf{E}_{P^\star_{XYV}}[d_e(X,V)] \\
&= \sum_{x,y}\sum_{v} Q_{XY}(x,y)P^\star_{V|XY}(v|x,y)d_e(x,v) \\
&= \sum_{x,y}\sum_{v} Q_{XY}(x,y)\lfloor P^\star_{V|XY}(v|x,y)\rfloor_{nQ_{XY}(x,y)}d_e(x,v) \\
&\quad + \sum_{x,y}\sum_{v} Q_{XY}(x,y)\left(P^\star_{V|XY}(v|x,y) - \lfloor P^\star_{V|XY}(v|x,y)\rfloor_{nQ_{XY}(x,y)}\right)d_e(x,v) \\
&\geq \sum_{x,y}\sum_{v} Q_{XY}(x,y)\lfloor P^\star_{V|XY}(v|x,y)\rfloor_{nQ_{XY}(x,y)}d_e(x,v) \\
&\quad + \sum_{x,y}\sum_{v} Q_{XY}(x,y)\left(P^\star_{V|XY}(v|x,y) - \lfloor P^\star_{V|XY}(v|x,y)\rfloor_{nQ_{XY}(x,y)}\right)d_e(x,v(x)) \\
&= \mathbf{E}_{P'_{XYV}}[d_e(X,V)].
\end{aligned}$$

Therefore,

$$\min_{\substack{P_{XYV}\in \\ \mathcal{Q}^n(Q_{XY},D_e)}} I_{P_{XYV}}(X;V|Y) \leq I_{P'_{XYV}}(X;V|Y) \leq I_{P^\star_{XYV}}(X;V|Y) + \epsilon = R(Q_{XY},D_e) + \epsilon,$$

where the second inequality follows from (C.3) and (C.4). ■

Similarly, we have the following proposition.

Proposition 25 *For all $\epsilon > 0$, there exists $n_3(\epsilon, |\mathcal{X}|, |\mathcal{Y}|, d_e)$, such that for all $n \geq n_3$, $D \geq D_{\min}$, $D_e \geq D_{e,\min}$, and for each $Q_X \in \mathcal{Q}^n_X$,*

$$\left| \max_{\substack{P_{XY}\in \\ \mathcal{Q}^n_{XY}(Q_X,D)}} R(Q_{XY},D_e) - R(Q_X,D,D_e)\right| \leq \epsilon.$$

The proof follows along the same lines as that of Proposition 24, and is thus omitted. ■

By the previous two propositions, for any given $\epsilon > 0$, we can set n large enough to satisfy

$$\left| I_{P_n^\star(Q_X)}(X;V|Y) - R(Q_X, D, D_e) \right| \leq \epsilon,$$

for all $Q_X \in \mathcal{Q}_X^n$. Therefore,

$$\begin{aligned} \min_{Q_X \in \mathcal{Q}_X^n} D(Q_X||P) + R(Q_X, D, D_e) - \epsilon &\leq \min_{Q_X \in \mathcal{Q}_X^n} D(Q_X||P) + I_{P_n^\star(Q_X)}(X;V|Y) \\ &\leq \min_{Q_X \in \mathcal{Q}_X^n} D(Q_X||P) + R(Q_X, D, D_e) + \epsilon \end{aligned}$$

By taking the limit as n goes to infinity, and noting that ϵ is arbitrary, the proof is concluded.

C.3 Proof of Proposition 12

C.3.1 Proof of Property (P4)

(P4): For fixed P_X, $R(P_X, R, D, D_e)$ is a finite-valued function of (R, D, D_e). Moreover, for fixed D_e, $R(P_X, R, D, D_e)$ is continuous in the triple (P_X, R, D) over the set $\mathcal{S} = \{(P_X, R, D) : P_X \in \mathcal{P}_X, D \geq D_{\min}, R > R(P_X, D)\}$.

Recall, for P_X satisfying $R(P_X, D) \leq R$,

$$R(P_X, R, D, D_e) = \max_{\substack{P_{Y|X}: \\ \mathbf{E}[d(X,Y)] \leq D \\ I(X;Y) \leq R}} R(P_{XY}, D_e).$$

For fixed P_X, let $S_{D,R} = \{P_{Y|X} : \mathbf{E}[d(X,Y)] \leq D, I(X;Y) \leq R\}$. Then $S_{D,R}$ is compact, and non-empty since $D \geq D_{\min}$ and $R \geq R(P_X, D)$. Since $R(P_{XY}, D_e)$ is a continuous function of P_{XY} (by Proposition 7), it is also continuous in $P_{Y|X}$ for fixed P_X. Therefore, the maximum is achieved.

To prove continuity of $R(P_X, R, D, D_e)$ in (P_X, R, D), first consider the following claims.

Claim 1: For fixed P_X, D_e, and D, $R(P_X, R, D, D_e)$ is continuous in R, where $R \in [R(P_X, D), +\infty)$.

This follows from the third part of Proposition 23 in Appendix C.1 by identifying $\mathcal{T}$ with $\{P_{Y|X} : \mathbf{E}[d(X,Y)] \leq D\}$ (which is compact, convex, and non-empty since $D \geq D_{\min}$), L_c with $I(X;Y)$ which is convex and continuous in $P_{Y|X}$, s_0 with $R(P_X, D)$, L with $R(P_X P_{Y|X}, D_e)$, and $\hat{L}$ with $R(P_X, R, D, D_e)$.

Claim 2: For fixed P_X, D_e, and R, $R(P_X, R, D, D_e)$ is continuous in D, where $D \in [D(P_X, R), +\infty)$ and $D(P_X, R) := \min_{P_{Y|X}: I(X;Y) \leq R} \mathbf{E}[d(X,Y)]$ is the distortion-rate function.

This follows from a similar argument.

We are now ready to prove continuity in the triple (P_X, R, D) over $\mathcal{S} = \{(P_X, R, D) : P_X \in \mathcal{P}_X, D \geq D_{\min}, R > R(P_X, D)\}$.

To that end, fix any $(P, R, D) \in \mathcal{S}$ and consider any sequence (P_k, R_k, D_k) converging to (P, R, D). First, we show that $\liminf_{k\to\infty} R(P_k, R_k, D_k, D_e) \geq R(P, R, D, D_e)$. Consider any $\epsilon > 0$. By continuity of $R(P, R, D, D_e)$ in R (for fixed P, D, and D_e), we can choose R' such that $R(P_X, D) < R' < R$ and $R(P, R', D, D_e) \geq R(P, R, D, D_e) - \epsilon/2$. We now consider two cases depending on the value of D. Let $D_0 = \min_{P_{Y|X}} \mathbf{E}[d(X,Y)]$.

If $D > D_0$: note that $D(P, R)$ is non-increasing in R, therefore $D'(P, R) \leq 0$. Moreover, it is convex in R and $R(P, D)$ does not achieve its minimum ($D(P, R(P, D)) = D > D_0$), hence $D'(P, R(P, D)) < 0$. Therefore, $R > R(P, D) \Rightarrow$

$D(P,R) < D$. Now choose D' such that $D(P,R) < D' < D$ and $R(P,R',D',D_e) \geq R(P,R',D,D_e) - \epsilon/2$. Let $P^\star_{Y|X}$ be a maximizer for $R(P,R',D',D_e)$.

If $D = D_0$: set $D' = D$ and $P'_{Y|X}$ be a maximizer for $R(P,R',D,D_e)$. Let $D(x) = \min_{y\in\mathcal{Y}} d(x,y)$ for $x \in \mathcal{X}$. Then $P'_{Y|X}$ must satisfy the following property: for all (x,y) such that $d(x,y) > D(x)$, $P(x) = 0$ or $P'_{Y|X}(y|x) = 0$. We can construct $P^\star_{Y|X}$ such that $d(x,y) > D(x) \Rightarrow P^\star_{Y|X}(y|x) = 0$, and $P(x) > 0 \Rightarrow P^\star_{Y|X=x} = P'_{Y|X=x}$. As such, $PP'_{Y|X} = PP^\star_{Y|X}$.

We claim that $P^\star_{Y|X}$ is feasible for the maximization in $R(P_k,R_k,D_k,D_e)$ for sufficiently large k. Indeed, $I(P_k;P^\star_{Y|X}) \to I(P;P^\star_{Y|X}) \leq R' < R$. Then for sufficiently large k, $I(P;P^\star_{Y|X}) \leq R_k$. Moreover, if $D > D_0$, then $\mathbf{E}[d(P_k,P^\star_{Y|X})] \to \mathbf{E}[d(P,P^\star_{Y|X})] \leq D' < D$. Then for sufficiently large k, $\mathbf{E}[d(P_k,P^\star_{Y|X})] \leq D_k$. Similarly, if $D = D_0$, then $\mathbf{E}[d(P_k,P^\star_{Y|X})] = \min_{P_{Y|X}} \mathbf{E}[d(P_k,P_{Y|X})] \leq D_{\min} \leq D_k$, where the first equality follows from the construction of $P^\star_{Y|X}$. So we get

$$\begin{aligned}
\liminf_{k\to\infty} R(P_k,R_k,D_k,D_e) &\geq \liminf_{k\to\infty} R(P_kP^\star_{Y|X},D_e) \\
&= R(PP^\star_{Y|X},D_e) \\
&= R(P,R',D',D_e) \\
&\geq R(P,R,D,D_e) - \epsilon,
\end{aligned}$$

where the first equality follows from the continuity of $R(P_{XY},D_e)$ in P_{XY}. Noting that ϵ is arbitrary, we get our first inequality.

On the other hand, let $P^{(k)}_{Y|X}$ be a maximizer for $R(P_k,R_k,D_k,D_e)$. Consider a sequence of integers $\{k_j\}$ such that

$$R(P_{k_j},R_{k_j},D_{k_j},D_e) \to \limsup_{k\to\infty} R(P_k,R_k,D_k,D_e).$$

Let $P^{(k_j)}_{Y|X}$ be the corresponding subsequence of maximizers. Since the set of conditional distributions $\{P_{Y|X}\}$ is bounded, $\{P^{(k_j)}_{Y|X}\}$ has a convergent subsequence $P^{(k_{j_\ell})}_{Y|X}$.

Let $P^{\star}_{Y|X}$ be its limit. We have, $I(P; P^{\star}_{Y|X}) = \lim_{\ell\to\infty} I(P; P^{(k_{j_\ell})}_{Y|X}) \leq \lim_{\ell\to\infty} R_{k_{j_\ell}} = R$. Similarly, $\mathbf{E}[d(P, P^{\star}_{Y|X})] = \lim_{\ell\to\infty} \mathbf{E}[d(P, P^{(k_{j_\ell})}_{Y|X})] \leq \lim_{\ell\to\infty} D_{k_{j_\ell}} = D$. Therefore,

$$\begin{aligned} R(P, R, D, D_e) &\geq R(PP^{\star}_{Y|X}, D_e) \\ &= \lim_{\ell\to\infty} R(P_{k_{j_\ell}} P^{k_{j_\ell}}_{Y|X}, D_e) \\ &= \lim_{\ell\to\infty} R(P_{k_{j_\ell}}, R_{k_{j_\ell}}, D_{k_{j_\ell}}, D_e) \\ &= \limsup_{k\to\infty} R(P_k, R_k, D_k, D_e). \end{aligned}$$

C.3.2 Proof of Property (P5)

(P5): $R_e(P_X, D_e) - R(P_X, D) \leq R(P_X, R, D, D_e) \leq R(P_X, D, D_e) \leq R_e(P_X, D_e)$.

The upper bound follows straightforwardly from the definition and (P3). The lower bound follows from the proof of (P3). Indeed, the bound in (P3) was derived by considering a conditional $P^{\star}_{Y|X}$ that achieves the rate-distortion function. As such, this choice is feasible since $I_{P^{\star}_{XY}}(X; Y) = R(P_X, D) \leq R$.

C.4 Proof of Lemma 13

Note that the second equality follows simply from the evaluation of $R_e(Q, D_e) - R(Q, D)$. So we only need to show the first equality.

Note that (P3) asserts that $R(Q, D, D_e) \geq R_e(Q, D_e) - R(Q, D)$, so we only need to show the reverse direction. Moreover, if $H(Q) \leq H(D)$, $R(Q, D) = 0$. It then follows from (P3) that $R(Q, D, D_e) = R_e(Q, D_e)$. It remains to show that, for Q

satisfying $H(Q) \geq H(D)$,

$$R(Q, D, D_e) \leq R_e(Q, D_e) - R(Q, D).$$

Remark 25 *The following proof was suggested by the reviewer, and it significantly simplifies our previous proof.*

To that end, let $P_{Y|X}$ satisfy $\mathbf{E}[d(X, Y)] \leq D$, $\tilde{X} = X \oplus Y$, $\tilde{V} = V \oplus Y$ and consider

$$\begin{aligned}
\min_{P_{V|XY}:\mathbf{E}[d(X,V)]\leq D_e} I(X; V|Y) &= \min_{P_{V|XY}:\mathbf{E}[d(X\oplus Y, V\oplus Y)]\leq D_e} I(X \oplus Y; V \oplus Y|Y) \\
&= \min_{P_{\tilde{V}|\tilde{X}Y}:\mathbf{E}[d(\tilde{X},\tilde{V})]\leq D_e} I(\tilde{X}; \tilde{V}|Y) \\
&\leq \min_{\substack{P_{\tilde{V}|\tilde{X}}:\mathbf{E}[d(\tilde{X},\tilde{V})]\leq D_e \\ V-X-Y}} I(\tilde{X}; \tilde{V}|Y) \\
&\leq \min_{P_{\tilde{V}|\tilde{X}}:\mathbf{E}[d(\tilde{X},\tilde{V})]\leq D_e} I(\tilde{X}; \tilde{V}) = [H(\tilde{X}) - H(D_e)]^+ \overset{\text{(a)}}{\leq} H(D) - H(D_e),
\end{aligned}$$

where (a) follows from the fact that $\mathbf{Pr}(\tilde{X} = 1) = \mathbf{E}[d(X, Y)] \leq D$. Therefore,

$$\begin{aligned}
R(Q, D, D_e) = \max_{P_{Y|X}:\mathbf{E}[d(X,Y)]\leq D} \min_{P_{V|XY}:\mathbf{E}[d(X,V)]\leq D_e} I(X; V|Y) &\leq H(D) - H(D_e) \\
&= R_e(Q, D_e) - R(Q, D),
\end{aligned}$$

as desired. ■

C.5 Proof of Lemma 15

Proof of 1): For $x^n \in T_{Q_X}$, and $m \in [N]$, let

$$N_{x^n,m} = \mathbb{I}\{(x^n, Y_m^n) \in T_{Q_{XY}}\}, \text{ so that } N_{x^n} = \sum_{m=1}^{N} N_{x^n,m}.$$

Note that, $N_{x^n,m} \sim Ber(\beta)$, where

$$\beta = \mathbf{Pr}((x^n, Y_m^n) \in T_{Q_{XY}}) = \frac{|T_{Q_{Y|X}}(x^n)|}{|T_{Q_Y}|},$$
$$\Rightarrow 2^{-n(I_{Q_{XY}}(X;Y)+\epsilon/2)} \leq \beta \leq 2^{-n(I_{Q_{XY}}(X;Y)-\epsilon/2)}. \tag{C.5}$$

Therefore,

$$\mathbf{Pr}(N_{x^n} = 0) = \mathbf{Pr}(N_{x^n,m} = 0, \forall m \in [N]) \overset{(a)}{=} \prod_{m=1}^{N}(1-\beta) \overset{(b)}{\leq} e^{-\beta N} \leq e^{-2^{n\epsilon/6}}, \tag{C.6}$$

where (a) follows from the independence of $N_{x^n,m}$ for different m's, and (b) follows from the fact that $(1-t)^N \leq e^{-tN}$. On the other hand,

$$\mathbf{Pr}(N_{x^n} > 2^{2n\epsilon}) = \mathbf{Pr}\left(\sum_{m=1}^{N} N_{x^n,m} > 2^{2n\epsilon}\right) \overset{(a)}{\leq} \left(\frac{eN\beta}{2^{2n\epsilon}}\right)^{2^{2n\epsilon}} \leq (e2^{-n\epsilon/2})^{2^{2n\epsilon}}, \tag{C.7}$$

where (a) follows from the Chernoff bound (cf. [41, Lemma 2]). Using equations (C.6) and (C.7) and the union bound, we get

$$\mathbf{Pr}(\mathcal{E}) \leq |\mathcal{X}|^n \left((e2^{-n\epsilon/2})^{2^{2n\epsilon}} + e^{-2^{n\epsilon/6}}\right) \leq e^{-2^{n\epsilon/7}},$$

establishing (4.41). ■

Proof of 2): To show that (4.42) holds, consider $C^n \notin \mathcal{E}$, and (x^n, m) where $m \in C(x^n)$,

$$P^C_{X^n|M}(x^n|m) = \frac{P^C_{M|X^n}(m|x^n)}{\sum_{x^n \in T_{Q_{X|Y}}(y_m^n)} P^C_{M|X^n}(m|x^n)} \leq \frac{1}{2^{n(H_{Q_{XY}}(X|Y)-\epsilon)}2^{-2n\epsilon}} = 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)}. \tag{C.8}$$

Then,

$$\begin{aligned}
\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) &= \sum_{x^n : d_e(x^n, v^n) \leq D_e} P^C_{X^n|M}(x^n|m) \\
&\leq \sum_{\substack{x^n : d_e(x^n, v^n) \leq D_e \\ (x^n, y_m^n) \in T_{Q_{XY}}}} 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&\overset{(a)}{\leq} \max_{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e)} 2^{n\left(H_{P_{XYV}}(X|V,Y)+\epsilon\right)} 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&= \max_{P_{XYV} \in \mathcal{Q}^n(Q_{XY}, D_e)} 2^{-n(I_{P_{XYV}}(X;V|Y)-4\epsilon)} \\
&\leq 2^{-n(R(Q_{XY}, D_e)-4\epsilon)},
\end{aligned} \tag{C.9}$$

where (a) follows from (4.23). It remains to show (4.43). To that end, note that, given $y^n \in \mathcal{Y}^n$ and $m \in [N]$,

$$\mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) \leq \frac{\mathbf{Pr}(Y_m^n = y^n)}{\mathbf{Pr}(\mathcal{E}^c)} \leq \frac{2^{-n(H_{Q_Y}(Y)-\epsilon/2)}}{1 - e^{-2^{n\epsilon/7}}} \leq 2^{-n(H_{Q_Y}(Y)-\epsilon)}.$$

Therefore,

$$\begin{aligned}
&\mathbf{E}[\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) | \mathcal{E}^c] \\
&= \sum_{C^n \in \mathcal{E}^c} \mathbf{Pr}(C^n | \mathcal{E}^c) \mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) \\
&= \sum_{y^n \in \mathcal{Y}^n} \sum_{C^n \in \mathcal{E}^c} \mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) \mathbf{Pr}(C^n | Y_m^n = y^n, \mathcal{E}^c) \mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C^n) \\
&= \sum_{y^n \in \mathcal{Y}^n} \sum_{C^n \in \mathcal{E}^c} \mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) \mathbf{Pr}(C^n | Y_m^n = y^n, \mathcal{E}^c) \sum_{\substack{x^n : d_e(x^n, v^n) \leq D_e \\ (x^n, y^n) \in T_{Q_{XY}}}} P_{X^n|M}^{\mathcal{C}}(x^n | m) \\
&\overset{(a)}{\leq} \sum_{y^n \in \mathcal{Y}^n} \sum_{C^n \in \mathcal{E}^c} \mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) \mathbf{Pr}(C^n | Y_m^n = y^n, \mathcal{E}^c) \sum_{\substack{x^n : d_e(x^n, v^n) \leq D_e \\ (x^n, y^n) \in T_{Q_{XY}}}} 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&= \sum_{y^n \in \mathcal{Y}^n} \mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) \sum_{\substack{x^n : d_e(x^n, v^n) \leq D_e \\ (x^n, y^n) \in T_{Q_{XY}}}} 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&= \sum_{x^n : d_e(x^n, v^n) \leq D_e} \sum_{y^n \in T_{Q_{Y|X}(x^n)}} \mathbf{Pr}(Y_m^n = y^n | \mathcal{E}^c) 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&\leq \sum_{x^n : d_e(x^n, v^n) \leq D_e} \sum_{y^n \in T_{Q_{Y|X}(x^n)}} 2^{-n(H_{Q_Y}(Y)-\epsilon/2)} 2^{-n(H_{Q_{XY}}(X|Y)-3\epsilon)} \\
&\leq \sum_{x^n : d_e(x^n, v^n) \leq D_e} 2^{nH_{Q_{XY}}(Y|X)} 2^{-n(H_{Q_{XY}}(X,Y)-7\epsilon/2)} \\
&= \sum_{x^n : d_e(x^n, v^n) \leq D_e} 2^{-n(H_{Q_X}(X)-7\epsilon/2)} \\
&\overset{(b)}{\leq} \max_{P_{XV} \in Q_{XV}^n} 2^{n(H_{P_{XV}}(X|V)+\epsilon/2)} 2^{-n(H_{Q_X}(X)-7\epsilon/2)} \\
&\leq 2^{-n(R_e(Q_X, D_e)-4\epsilon)},
\end{aligned}$$

where (a) follows from (C.8), and (b) can be shown analogously to (4.23). ■

Proof of 3): For notational convenience, let $E = \min\{R_e(Q_X, D_e), r + R(Q_{XY}, D_e)\}$.

Note that,

$$\mathbf{Pr}(\tilde{\mathcal{E}}) \leq \mathbf{Pr}\left(\bigcup_{k=1}^{2^{nr}} \mathcal{E}_k\right) + \mathbf{Pr}\left(\tilde{\mathcal{E}} \middle| \left(\bigcup_{k=1}^{2^{nr}} \mathcal{E}_k\right)^c\right) \leq e^{-2^{n\epsilon/8}} + \mathbf{Pr}\left(\tilde{\mathcal{E}} \middle| \bigcap_{k=1}^{2^{nr}} \mathcal{E}_k^c\right), \tag{C.10}$$

where the second inequality follows from the union bound and (4.41). Now, fix $\{C_k^n\}_{k=1}^{2^{nr}} \in \cap_{k=1}^{2^{nr}} \mathcal{E}_k^c$, $m \in [N]$, and $v^n \in \mathcal{V}^n$, and suppose $K = k_0$. Then,

$$\begin{aligned}\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, K = k_0, \{C_k^n\}_{k=1}^{2^{nr}}) &= \mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C_{k_0}^n) \\ &\leq 2^{-n(R(Q_{XY}, D_e) - 4\epsilon)}, \end{aligned} \tag{C.11}$$

where the inequality follows from (4.42). Furthermore,

$$\begin{aligned}&\mathbf{E}\left[\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, K = k_0, \{C_k^n\}_{k=1}^{2^{nr}}) \middle| \bigcap_{k=1}^{2^{nr}} \mathcal{E}_k^c\right] && \text{(C.12)} \\ &= \mathbf{E}\left[\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C_{k_0}^n) \middle| \bigcap_{k=1}^{2^{nr}} \mathcal{E}_k^c\right] \\ &= \mathbf{E}\left[\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C_{k_0}^n) | \mathcal{E}_{k_0}^c\right] \\ &\leq 2^{-n(R_e(Q_X, D_e) - 4\epsilon)}, && \text{(C.13)}\end{aligned}$$

where the last inequality follows from (4.43). Now, consider $\{C_k^n\}_{k=1}^{2^{nr}} \in \left(\cup_{k=1}^{2^{nr}} \mathcal{E}_k\right)^c$.

$$\begin{aligned}&\mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, \{C_k^n\}_{k=1}^{2^{nr}}) \\ &= \sum_{j=1}^{2^{nr}} \mathbf{Pr}(K = j | M = m, \{C_k^n\}_{k=1}^{2^{nr}}) \mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, K = j, \{C_k^n\}_{k=1}^{2^{nr}}) \\ &\leq \sum_{j=1}^{2^{nr}} 2^{-n(r-2\epsilon)} \mathbf{Pr}(d_e(X^n, v^n) \leq D_e | M = m, C_j^n), \end{aligned} \tag{C.14}$$

where the inequality follows from:

$$\begin{aligned}
\mathbf{Pr}(K=j|M=m,\{C_k^n\}_{k=1}^{2^{nr}}) &= \frac{\mathbf{Pr}(K=j)\mathbf{Pr}(M=m|K=j,\{C_k^n\}_{k=1}^{2^{nr}})}{\sum_{\ell=1}^{2^{nr}}\mathbf{Pr}(K=\ell)\mathbf{Pr}(M=m|K=\ell,\{C_k^n\}_{k=1}^{2^{nr}})} \\
&= \frac{\mathbf{Pr}(M=m|K=j,C_j^n)}{\sum_{\ell=1}^{2^{nr}}\mathbf{Pr}(M=m|K=\ell,C_\ell^n)} \\
&= \frac{\sum\limits_{\substack{x^n:\\ m\in C_j(x^n)}}\mathbf{Pr}(X^n=x^n)\mathbf{Pr}(M=m|X^n=x^n,K=j,C_j^n)}{\sum\limits_{\ell=1}^{2^{nr}}\sum\limits_{\substack{x^n:\\ m\in C_\ell(x^n)}}\mathbf{Pr}(X^n=x^n)\mathbf{Pr}(M=m|X^n=x^n,K=\ell,C_\ell^n)} \\
&\overset{(a)}{\leq} \frac{\sum_{x^n:m\in C_j(x^n)}1}{\sum_{\ell=1}^{2^{nr}}\sum_{x^n:m\in C_\ell(x^n)}2^{-2n\epsilon}} \\
&\overset{(b)}{=} 2^{-n(r-2\epsilon)}
\end{aligned}$$

where (a) follows from the fact that $1 \leq N_{x^n} \leq 2^{2n\epsilon}$, and (b) follows from the fact that, for any j, $\sum_{x^n:m\in C_j(x^n)} 1 = |\{x^n : \left(x^n, y_m^n(C_j)\right) \in T_{Q_{XY}}\}| = |T_{Q_{X|Y}(y_m^n(C_j))}| = |T_{Q_{X|Y}(y^n)}|$ for any $y^n \in T_{Q_Y}$.

Given $\left(\cup_{k=1}^{2^{nr}}\mathcal{E}_k\right)^c$, the terms in the summands of (C.14) are independent and identically distributed random variables, with an upper bound given by (C.11), and an expectation upper-bounded by (C.13). It follows from Chernoff's bound [41, Corollary 2] that

$$\begin{aligned}
&\mathbf{Pr}\left(\mathbf{Pr}(d_e(X^n,v^n)\leq D_e|M=m,\{C_k^n\}_{k=1}^{2^{nr}}) > 2^{-n(E-8\epsilon)}\middle|\bigcap_{k=1}^{2^{nr}}\mathcal{E}_k^c\right) \\
&= \mathbf{Pr}\left(\sum_{j=1}^{2^{nr}}2^{-n(r-2\epsilon)}\mathbf{Pr}(d_e(X^n,v^n)\leq D_e|M=m,C_j^n) > 2^{-n(E-8\epsilon)}\middle|\bigcap_{k=1}^{2^{nr}}\mathcal{E}_k^c\right) \\
&= \mathbf{Pr}\left(\sum_{j=1}^{2^{nr}}\mathbf{Pr}(d_e(X^n,v^n)\leq D_e|M=m,C_j^n) > 2^{-n(E-r-6\epsilon)}\middle|\bigcap_{k=1}^{2^{nr}}\mathcal{E}_k^c\right) \\
&\leq \left(\frac{e2^{nr}2^{-n(R_e(Q_X,D_e)-4\epsilon)}}{2^{-n(E-r-6\epsilon)}}\right)^{\frac{2^{-n(E-r-6\epsilon)}}{2^{-n(R(Q_{XY},D_e)-4\epsilon)}}} \\
&\leq \left(e2^{-n(R_e(Q_X,D_e)-E-4\epsilon+6\epsilon)}\right)^{2^{-n(E-r-R(Q_{XY},D_e)-6\epsilon+4\epsilon)}} \\
&\leq 2^{-\epsilon n 2^{2\epsilon n}}, \qquad \text{(C.15)}
\end{aligned}$$

where the last inequality follows from the fact that $R_e(Q_X, D_e) - E \geq 0$, and $E - r - R(Q_{XY}, D_e) \leq 0$. By the union bound,

$$\mathbf{Pr}\left(\tilde{\mathcal{E}} \middle| \bigcap_{k=1}^{2^{nr}} \mathcal{E}_k^c\right) \leq N 2^{-\epsilon n 2^{2\epsilon n}} \leq 2^{n(I_{Q_{XY}}(X;Y)+\epsilon-\epsilon 2^{2\epsilon n})} \leq 2^{n(\log|\mathcal{X}|+\epsilon-\epsilon 2^{2\epsilon n})} \leq 2^{-\frac{\epsilon}{2} n 2^{2\epsilon n}}. \tag{C.16}$$

Combining (C.10) and (C.16) yields

$$\mathbf{Pr}\left(\tilde{\mathcal{E}}\right) \leq e^{-2^{n\epsilon/8}} + 2^{-\frac{\epsilon}{2} n 2^{2\epsilon n}} \leq e^{-2^{n\epsilon/9}}, \tag{C.17}$$

as desired. ■

C.6 Proof of Proposition 16

Consider the following proposition.

Proposition 26 *Given $\epsilon > 0$, $\beta > 0$, and $R' > \max_{Q:D(Q\|P)\leq\beta} R(Q, D) =: R_\beta$, there exists $n_4(\epsilon, |\mathcal{X}|, |\mathcal{Y}|, R', d_e)$ such that for all $n \geq n_4$, $D \geq D_{\min}$, $D_e \geq D_{e,\min}$, and for each $Q_X \in Q_{\mathcal{X}}^n(\beta, 0)$ (cf. (4.46))* ,

$$\left| \max_{\substack{Q_{XY} \in Q_{\mathcal{X}\mathcal{Y}}^n(Q_X, D): \\ I_{Q_{XY}}(X;Y) \leq R'}} R(Q_{XY}, D_e) - R(Q_X, R', D, D_e) \right| \leq \epsilon.$$

Proof: Note that Proposition 25 is a special case in which $\beta = +\infty$ and $R \geq \max_Q R(Q, D)$. As such, the proof follows along similar lines as Propositions 25 and 24, but must account for the rate constraint R.

It follows directly from the definition that

$$\max_{\substack{Q_{XY} \in Q_{\mathcal{X}\mathcal{Y}}^n(Q_X, D): \\ I_{Q_{XY}}(X;Y) \leq R'}} R(Q_{XY}, D_e) \leq R(Q_X, R', D, D_e).$$

So, we only need to show the reverse direction. To that end, choose R'' such that $R_\beta < R'' < R'$. By Proposition 12, $R(Q_X, R, D, D_e)$ is uniformly continuous in (Q_X, R) over the set $\{(Q_X, R) : D(Q_X\|P) \le \beta, R'' \le R \le R'\}$. Then let $\delta_1 > 0$ be small enough such that, for all $Q_X \in \mathcal{Q}_{\mathcal{X}}^n(\beta, 0)$,

$$|R(Q_X, R' - \delta_1, D, D_e) - R(Q_X, R', D, D_e)| \le \epsilon/2. \tag{C.18}$$

Let $\delta_2 > 0$ be small enough such that

$$\begin{aligned} \|P_{XY} - P'_{XY}\| \le \delta_2 \Rightarrow |R(P_{XY}, D_e) - R(P'_{XY}, D_e)| &\le \epsilon/2 \\ \text{and } |I_{P_{XY}}(X;Y) - I_{P'_{XY}}(X;Y)| &\le \delta_1. \end{aligned} \tag{C.19}$$

Let $n \ge n_4 \ge |\mathcal{Y}|\sqrt{|\mathcal{X}|}/\delta_2$. Fix $Q_X \in \mathcal{Q}_{\mathcal{X}}^n(\beta, 0)$ and let $P^\star_{Y|X}$ be the conditional distribution achieving the maximum in $R(Q_X, R' - \delta_1, D, D_e)$. We construct $P'_{Y|X}$ by rounding the values of $P^\star_{Y|X}$, as done in Proposition 24. Similarly to Proposition 24, this guarantees that $Q_X P'_{Y|X} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}$,

$$\|Q_X P'_{Y|X} - Q_X P^\star_{Y|X}\| \le \delta_2, \text{ and } \mathbf{E}_{Q_X P'_{Y|X}}[d(X,Y)] \le D. \tag{C.20}$$

Moreover, it follows from (C.19) and (C.20) that $I_{Q_X P'_{XY}}(X;Y) \le I_{Q_X P^\star_{XY}}(X;Y) + \delta_1 \le R'$. Therefore,

$$\begin{aligned} \max_{\substack{Q_{XY} \in \mathcal{Q}^n_{\mathcal{X}\mathcal{Y}}(Q_X, D): \\ I_{Q_{XY}}(X;Y) \le R'}} R(Q_{XY}, D_e) &\ge R(Q_X P'_{Y|X}, D_e) \\ &\ge R(Q_X P^\star_{Y|X}, D_e) - \epsilon/2 \\ &\ge R(Q_X, R', D, D_e) - \epsilon, \end{aligned}$$

where the second inequality follows from (C.19) and (C.20), and the third inequality from (C.18). ■

The proposition yields

$$\begin{aligned}
&\min_{Q_X \in \mathcal{Q}_X^n(\beta,0)} D(Q_X\|P) + R(Q, R', D, D_e) - \epsilon \\
&\leq \min_{Q_X \in \mathcal{Q}_X^n(\beta,0)} D(Q_X\|P) + R(Q^\star_{R'}(Q_X), D_e) \\
&\leq \min_{Q_X \in \mathcal{Q}_X^n(\beta,0)} D(Q_X\|P) + R(Q, R', D, D_e).
\end{aligned}$$

By taking the limit as n goes to infinity, and noting that ϵ is arbitrary, the proof is concluded.

C.7 Proofs of Propositions 22 and 23

C.7.1 Proof of Proposition 22

We restate the proposition.

Proposition 22: Let N_1 and N_2 be in $\mathbb{N}$, and let $\mathcal{S}$ and $\mathcal{U}$ be compact subsets of $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, respectively. Let ν be a non-negative continuous function defined on $\mathcal{S} \times \mathcal{U}$, and let ϑ be a real-valued continuous function defined on $\mathcal{S} \times \mathcal{U}$. Suppose they satisfy the following condition:

(PA) If $(s, u_1) \in \mathcal{S} \times \mathcal{U}$ satisfies $\nu(s, u_1) = \min_{u' \in \mathcal{U}} \nu(s, u')$, then there exists u_2 such that $\vartheta(s, u_2) = \vartheta(s, u_1)$, and for all $s' \in \mathcal{S}$, $\nu(s', u_2) = \min_{u' \in \mathcal{U}} \nu(s', u')$.

Let $t_0 = \max_{s \in \mathcal{S}} \min_{u \in \mathcal{U}} \nu(s, u)$, and let φ be a function on $\mathcal{S} \times [t_0, +\infty)$ defined as follows:

$$\varphi(s, t) = \min_{u : \nu(s,u) \leq t} \vartheta(s, u).$$

If for fixed $s \in \mathcal{S}$, $\varphi(s,t)$ is continuous in t, then $\varphi(s,t)$ is continuous in the pair (s,t).

First, note that, for all $s \in \mathcal{S}$ and all $t \geq t_0$, $\nu^{-1}(s,[0,t]) \triangleq \{u : \nu(s,u) \leq t\}$ is closed by continuity of ν, so it is compact since it is also bounded. Moreover it is non-empty since $t \geq t_0$. Since ϑ is continuous and the minimization is over a compact set, φ is well defined.

Now fix $(s,t) \in \mathcal{S} \times [t_0, +\infty)$, and consider any sequence $(s_k, t_k) \to (s,t)$. Let $t_s = \min_{u \in \mathcal{U}} \nu(s,u)$ and consider any $\epsilon > 0$.

If $t > t_s$:

By continuity of $\varphi(s,t)$ as a function of t for fixed s, there exists $\delta > 0$ such that $|t - t'| \leq \delta \Rightarrow |\varphi(s,t) - \varphi(s,t')| \leq \epsilon$. Let $t' = t - \min\{\delta/2, (t - t_s)/2\}$, and let $u' \in \operatorname{argmin}_{u:\nu(s,u)\leq t'} \vartheta(s,u)$. Then, $\nu(s,u') < t$ and $\varphi(s,t') = \vartheta(s,u') \leq \varphi(s,t) + \epsilon$.

If $t = t_s$:

Let u' be a minimizer for $\varphi(s,t_s)$ satisfying $\nu(s',u') = t_{s'}$ for all $s' \in \mathcal{S}$. Such choice is possible by assumption (PA). Note that $\vartheta(s,u') = \varphi(s,t_s)$.

We claim that the choice of u' is feasible for the minimization in $\varphi(s_k,t_k)$, i.e., $\nu(s_k,u') \leq t_k$ for sufficiently large k. Indeed, if $t > t_s$, $\nu(s_k,u') \to \nu(s,u') = t' < t$, then for sufficiently large k, $\nu(s_k,u') \leq t_k$. If $t = t_s$, then $\nu(s_k,u') = t_{s_k} \leq t_0 \leq t_k$.

Moreover, by continuity of ϑ, $\vartheta(s_k,u') \to \vartheta(s,u')$. Then, for sufficiently large

k, $\vartheta(s_k, u') \leq \varphi(s,t) + \epsilon/2$. So, we get

$$\limsup_{k\to\infty} \varphi(s_k, t_k) \leq \limsup_{k\to\infty} \vartheta(s_k, u') \leq \varphi(s,t).$$

On the other hand, let u_k be a minimizer for $\varphi(s_k, t_k)$. Consider a sequence of integers $\{k_j\}$ such that

$$\varphi(s_{k_j}, t_{k_j}) \to \liminf_{k\to\infty} \varphi(s_k, t_k).$$

Let $\{u_{k_j}\}$ be the corresponding subsequence of minimizers. Since $\mathcal{U}$ is a bounded set, then $\{u_{k_j}\}$ has a convergent subsequence $\{u_{k_{j_\ell}}\}$. Let u' be its limit. By continuity of ν, we have $\nu(s,u') = \lim_{\ell\to\infty} \nu(s_{k_{j_\ell}}, u_{k_{j_\ell}}) \leq \lim_{\ell\to\infty} t_{k_{j_\ell}} = t$. Therefore,

$$\begin{aligned}
\varphi(s,t) \leq \vartheta(s,u') &= \lim_{\ell\to\infty} \vartheta(s_{k_{j_\ell}}, u_{k_{j_\ell}}) \\
&= \lim_{\ell\to\infty} \varphi(s_{k_{j_\ell}}, t_{k_{j_\ell}}) \\
&= \liminf_{k\to\infty} \varphi(s_k, t_k).
\end{aligned}$$

■

C.7.2 Proof of Proposition 23

We restate the proposition.

Proposition 23: Let N be in $\mathbb{N}$, and let $\mathcal{T}$ be a non-empty compact subset of $\mathbb{R}^N$. Let L be a real-valued continuous function defined on $\mathcal{T}$. Let $T_1 \supseteq T_2 \supseteq \cdots$ be a decreasing sequence of non-empty compact subsets of $\mathcal{T}$. Let $T = \bigcap_{i\geq 1} T_i$. Then,

$$\lim_{k\to\infty} \max_{t\in T_k} L(t) = \max_{t\in T} L(t).$$

Moreover, let $S_1 \subseteq S_2 \subseteq \cdots$ be an increasing sequence of non-empty compact

subsets of $\mathcal{T}$. Let $S = \overline{\bigcup_{i\geq 1} S_i}$ (where the bar denotes closure of the set). Then,

$$\lim_{k\to\infty} \max_{t\in S_k} L(t) = \max_{t\in S} L(t).$$

Consequently, if $\mathcal{T}$ is also convex, and L_c is a real-valued convex and continuous function defined on $\mathcal{T}$ with $s_0 = \min_{t\in\mathcal{T}} L_c(t)$, then

$$\hat{L}(s) := \max_{t:L_c(t)\leq s} L(t)$$

is continuous in $s \in [s_0, +\infty)$.

First, note that T is non-empty and compact since a countable intersection of non-empty decreasing compact sets is non-empty and compact. Let

$$t_k = \operatorname*{argmax}_{t\in T_k} L(t) \quad \text{and} \quad t^\star = \operatorname*{argmax}_{t\in T} L(t).$$

We need to show that $L(t_k) \to L(t^\star)$. Let $\mathcal{B}_\delta(t) = \{t' \in \mathcal{T} : \|t' - t\| < \delta\}$, and consider the following claim.

Claim 1: For all $\delta > 0$, there exists k_0 such that for all $k \geq k_0$, $T_k \subseteq \mathcal{B}_\delta(T)$, where

$$\mathcal{B}_\delta(T) = \bigcup_{t\in T} \mathcal{B}_\delta(t).$$

We show first how the claim yields our result. Let $\epsilon > 0$ be given. By the uniform continuity of L (continuity on a compact set), there exists $\delta > 0$ such that $\|t - t'\| \leq \delta \Rightarrow |L(t) - L(t')| \leq \epsilon$. Let k be large enough as guaranteed by the claim. Then, for all $t \in T_k$, there exists $t' \in T$ such that $\|t - t'\| \leq \delta$, and subsequently $|L(t) - L(t')| \leq \epsilon$. In particular, there exists $t' \in T$ such that $|L(t_k) - L(t')| \leq \epsilon$. Then, we get $L(t_k) \leq L(t') + \epsilon \leq L(t^\star) + \epsilon$. Since $L(t_k) \geq L(t^\star)$, we get $|L(t_k) - L(t^\star)| \leq \epsilon$. Therefore, $L(t_k) \to L(t^\star)$. It remains to prove the claim to establish the first part of the proposition.

Proof of Claim 1: Fix $\delta > 0$. $\mathcal{B}_\delta(T)$ is open in $\mathcal{T}$ by construction. Therefore, $T_k \backslash \mathcal{B}_\delta(T)$ is closed in $\mathcal{T}$. Since $\mathcal{T}$ is closed in $\mathbb{R}^N$, then $T_k \backslash \mathcal{B}_\delta(T)$ is also closed in $\mathbb{R}^N$. Moreover, it is bounded, so it is compact. Since

$$\bigcap_{i\geq 1} T_k \backslash \mathcal{B}_\delta(T) = \left(\bigcap_{i\geq 1} T_k\right) \Big\backslash \mathcal{B}_\delta(T) = T \backslash \mathcal{B}_\delta(T) = \varnothing,$$

and $T_k \backslash \mathcal{B}_\delta(T)$ is a decreasing sequence of compact sets, there exists k_0 such that for all $k \geq k_0$, $T_k \backslash \mathcal{B}_\delta(T)$ is empty.

Similarly, to prove the second part of the proposition, let

$$s_k = \operatorname*{argmax}_{t\in S_k} L(t) \quad \text{and} \quad s^\star = \operatorname*{argmax}_{t\in S} L(t).$$

We need to show that $L(s_k) \to L(s^\star)$. To this end, consider the following claim.

Claim 2: For all $\delta > 0$, there exists k_1 such that for all $k \geq k_1$, $S \subseteq \mathcal{B}_\delta(S_k)$.

We show first how the claim yields our result. Let $\epsilon > 0$ be given. By the uniform continuity of L, there exists $\delta > 0$ such that $\|t - t'\| \leq \delta \Rightarrow |L(t) - L(t')| \leq \epsilon$. Let k be large enough as guaranteed by the claim. Then, for all $t \in S$, there exists $t' \in S_k$ such that $\|t - t'\| \leq \delta$, and subsequently $|L(t) - L(t')| \leq \epsilon$. In particular, there exists $t' \in S_k$ such that $|L(s^\star) - L(t')| \leq \epsilon$. Then, we get $L(s_k) \geq L(t') \geq L(s^\star) - \epsilon$. Since $L(s_k) \leq L(s^\star)$, we get $|L(s_k) - L(s^\star)| \leq \epsilon$. Therefore, $L(s_k) \to L(s^\star)$. It remains to prove the claim.

Proof of Claim 2: Fix $\delta > 0$. $\mathcal{B}_\delta(S_k)$ is open in $\mathcal{T}$ by construction. Therefore, $S \backslash \mathcal{B}_\delta(S_k)$ is closed in $\mathcal{T}$. Then $S \backslash \mathcal{B}_\delta(S_k)$ is closed in $\mathbb{R}^N$. Moreover, it is bounded, so it is compact. Since

$$\bigcap_{i\geq 1} S \backslash \mathcal{B}_\delta(S_k) = S \Big\backslash \left(\bigcup_{i\geq 1} \mathcal{B}_\delta(S_k)\right) = \overline{\bigcup_{i\geq 1} S_i} \Big\backslash \mathcal{B}_\delta\left(\bigcup_{i\geq 1} S_i\right) = \varnothing,$$

and $S \setminus \mathcal{B}_\delta(S_k)$ is a decreasing sequence of compact sets, there exists k_1 such that for all $k \geq k_1$, $S \setminus \mathcal{B}_\delta(S_k)$ is empty.

Finally, consider $\hat{L}(s)$. If L_c is a constant function, then the statement is trivial. If not, consider $s \geq s_0$, and let s_k be a decreasing sequence converging to s. Then,

$$\lim_{k\to\infty} \hat{L}(s_k) = \lim_{k\to\infty} \max_{t:L_c(t)\leq s_k} L(t) = \max_{t:L_c(t)\leq s} L(t) = \hat{L}(s),$$

where the second equality follows from the first part of the proposition. Therefore, $\hat{L}(s)$ is right-continuous. Now, consider $s > s_0$, and let s_k be an increasing sequence converging to s. Note that,

$$\bigcup_{k\geq 1} \{t \in \mathcal{T} : L_c(t) \leq s_k\} = \{t \in \mathcal{T} : L_c(t) < s\}.$$

Denote the above set by S^- and let $S = \{t \in \mathcal{T} : L_c(t) \leq s\}$. The second part of the proposition implies that

$$\lim_{k\to\infty} \hat{L}(s_k) = \lim_{k\to\infty} \max_{t:L_c(t)\leq s_k} L(t) = \max_{t\in\overline{S^-}} L(t).$$

So it suffices to show that $\overline{S^-} = S$. Clearly, $\overline{S^-} \subseteq S$ since S is closed and $S^- \subseteq S$. It remains to show that any point $\tilde{t}$ satisfying $L_c(\tilde{t}) = s$ is a boundary point of S^-. To that end, note that $L_c(\tilde{t})$ is not a local minimum since $L_c(\tilde{t}) = s > s_0$ and L_c is convex by assumption. Therefore, any neighborhood of $\tilde{t}$ intersects S^-. As such $\overline{S^-} = S$, and $\hat{L}(s)$ is left-continuous, as desired. ■

APPENDIX D

PROOF OF LEMMA 21

We will proceed by induction on j. Let $j = d/2$. It is easy to check then that the left-hand side and the right-hand side are both equal to d. Next, the induction step shows that if the formula holds for $j+1$, then it holds for j:

$$\sum_{i=d/2+j}^{d} (-1)^{d-i} i \binom{d}{i} = (-1)^{d/2-j} (d/2+j) \binom{d}{\frac{d}{2}+j} + (-1)^{d/2-j-1} \frac{(2j+d)(2j+d+2)\binom{d}{d/2+j+1}}{4(d-1)}.$$

Note that

$$\binom{d}{d/2+j+1} = \binom{d}{d/2+j} \frac{d/2-j}{d/2+j+1}.$$

Then, continuing,

$$\begin{aligned}
\sum_{i=d/2+j}^{d} (-1)^{d/2-i} i \binom{d}{i} &= \frac{(-1)^{d/2-j}\binom{d}{d/2+j}}{4(d-1)} \left(4(d-1)(d/2+j) - \frac{(2j+d)(2j+d+2)(d/2-j)}{d/2+j+1}\right) \\
&= \frac{(-1)^{d/2-j}\binom{d}{d/2+j}}{4(d-1)} (2(d-1)(d+2j) - 2(2j+d)(d/2-j)) \\
&= \frac{(-1)^{d/2-j}\binom{d}{d/2+j}}{4(d-1)} (2j+d)(2(d-1) - 2(d/2-j)) \\
&= \frac{(-1)^{d/2-j}\binom{d}{d/2+j}}{4(d-1)} (2j+d)(d+2j-2),
\end{aligned}$$

as desired. ■

BIBLIOGRAPHY

[1] J. Acharya, A. Orlitsky, A. T. Suresh, and H. Tyagi. Estimating renyi entropy of discrete distributions. *IEEE Trans. Inf. Theory*, 63(1):38–56, Jan 2017.

[2] Mário S Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. Additive and multiplicative notions of leakage, and their capacities. In *IEEE 27th Computer Security Foundations Symposium*, pages 308–322. IEEE, 2014.

[3] M.S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. Measuring information leakage using generalized gain functions. In *Computer Security Foundations Symposium (CSF), 2012 IEEE 25th*, pages 265–279, June 2012.

[4] V. Anantharam, A. Gohari, S. Kamath, and C. Nair. On hypercontractivity and a data processing inequality. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 3022–3026, June 2014.

[5] Robert G Bartle. *The elements of integration and Lebesgue measure*. John Wiley & Sons, 2014.

[6] Mihir Bellare, Stefano Tessaro, and Alexander Vardy. Semantic security for the wiretap channel. In *Advances in Cryptology–CRYPTO 2012*, pages 294–311. Springer, 2012.

[7] Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Quantitative notions of leakage for one-try attacks. *Electronic Notes in Theoretical Computer Science*, 249:75–91, 2009.

[8] F.P. Calmon, M. Varia, M. Médard, M.M. Christiansen, K.R. Duffy, and S. Tessaro. Bounds on inference. In *51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 567–574, Oct 2013.

[9] William E Cobb. Exploitation of unintentional information leakage from integrated circuits. Technical report, DTIC Document, 2011.

[10] Imre Csiszár and Janos Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, 24(3):339–348, May 1978.

[11] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Budapest: Akadémiai Kiadó, 1997.

[12] Yevgeniy Dodis and Adam Smith. Entropic security and the encryption of high entropy messages. In Joe Kilian, editor, *Theory of Cryptography*, volume 3378 of *Lecture Notes in Computer Science*, pages 556–577. Springer Berlin Heidelberg, 2005.

[13] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 429–438, Oct 2013.

[14] Cynthia Dwork. Differential privacy: A survey of results. In *Theory and applications of models of computation*, pages 1–19. Springer, 2008.

[15] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. Inf. Theory*, 37(2):269–275, Mar 1991.

[16] Barbara Espinoza and Geoffrey Smith. Min-entropy as a resource. *Information and Computation*, 226:57 – 75, 2013. Special Issue: Information Security as a Resource.

[17] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270 – 299, 1984.

[18] Praveen Kumar Gopala, Lifeng Lai, and H. El Gamal. On the secrecy capacity of fading channels. *IEEE Trans. Inf. Theory*, 54(10):4687–4698, Oct. 2008.

[19] Michael Greenacre. *Correspondence analysis in practice*. CRC press, 2007.

[20] D. Gunduz, E. Erkip, and H. V. Poor. Lossless compression with security constraints. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 111–115, July 2008.

[21] Te Sun Han and Sergio Verdú. Generalizing the Fano inequality. *IEEE Trans. Inf. Theory*, 40(4):1247–1251, Jul 1994.

[22] I. Issa and A. B. Wagner. Measuring secrecy by the probability of a successful guess. In *53rd Annual Allerton Conference on Communication, Control, and Computing*, pages 980–987, Sept 2015.

[23] I. Issa and A. B. Wagner. Measuring secrecy by the probability of a successful guess. *IEEE Trans. Inf. Theory*, 63(6):3783–3803, June 2017.

[24] S. Kadloor, N. Kiyavash, and P. Venkitasubramaniam. Mitigating timing side channel in shared schedulers. *Networking, IEEE/ACM Transactions on*, PP(99):1–12, 2015.

[25] Paul Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In *Annual International Cryptology Conference*, pages 388–397. Springer, 1999.

[26] Paul C Kocher. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In *Annual International Cryptology Conference*, pages 104–113. Springer, 1996.

[27] Eric A Koziel. Effects of architecture on information leakage of a hardware advanced encryption standard implementation. Technical report, DTIC Document, 2012.

[28] S. Leung-Yan-Cheong and M.E. Hellman. The Gaussian wire-tap channel. *IEEE Trans. Inf. Theory*, 24(4):451–456, Jul. 1978.

[29] Cheuk Ting Li and Abbas El Gamal. Maximal correlation secrecy. *CoRR*, 2014.

[30] Ninghui Li, Tiancheng Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115, April 2007.

[31] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS '07. 48th Annual IEEE Symposium on*, pages 94–103, Oct 2007.

[32] D. G. Mead. Newton's identities. *The American Mathematical Monthly*, 99(8):749–751, 1992.

[33] Neri Merhav and Erdal Arıkan. The Shannon cipher system with a guessing wiretapper. *IEEE Trans. Inf. Theory*, 45(6):1860–1866, 1999.

[34] Manoj M. Prabhakaran and Vinod M. Prabhakaran. Rényi information complexity and an information theoretic characterization of the partition bound. *CoRR*, abs/1511.07949, 2015.

[35] Josyula R Rao, Pankaj Rohatgi, Helmut Scherzer, and Stephane Tinguely. Partitioning attacks: or how to rapidly clone some GSM cards. In *IEEE Proc. Symposium on Security and Privacy*, pages 31–41. IEEE, 2002.

[36] Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959.

[37] Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage. Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pages 199–212, New York, NY, USA, 2009. ACM.

[38] A. Russell and Hong Wang. How to fool an unbounded adversary with a short key. *IEEE Trans. Inf. Theory*, 52(3):1130–1140, March 2006.

[39] L. Sankar, S.R. Rajagopalan, and H.V. Poor. Utility-privacy tradeoffs in databases: An information-theoretic approach. *IEEE. Trans. Inf. Forensics Security*, 8(6):838–852, June 2013.

[40] C. Schieler and P. Cuff. Rate-distortion theory for secrecy systems. *IEEE Trans. Inf. Theory*, 60(12):7584–7605, Dec 2014.

[41] C. Schieler and P. Cuff. The henchman problem: Measuring secrecy by the minimum distortion in a list. *IEEE Trans. Inf. Theory*, 62(6):3436–3450, June 2016.

[42] Curt Schieler and Paul Cuff. The henchman problem: measuring secrecy by the minimum distortion in a list. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 596–600. IEEE, 2014.

[43] Claude E Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[44] Robin Sibson. Information radius. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(2):149–160, 1969.

[45] Geoffrey Smith. On the foundations of quantitative information flow. In Luca de Alfaro, editor, *Foundations of Software Science and Computational Structures*, volume 5504 of *Lecture Notes in Computer Science*, pages 288–302. Springer Berlin Heidelberg, 2009.

[46] Dawn Xiaodong Song, David Wagner, and Xuqing Tian. Timing analysis of keystrokes and timing attacks on SSH. In *Proceedings of the 10th USENIX Security Symposium - Volume 10*, Berkeley, CA, USA, 2001. USENIX Association.

[47] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[48] Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011.

[49] P. Venkitasubramaniam, Ting He, and Lang Tong. Anonymous networking amidst eavesdroppers. *IEEE Trans. Inf. Theory*, 54(6):2770–2784, June 2008.

[50] S. Verdú. α-mutual information. In *Information Theory and Applications Workshop (ITA), 2015*, pages 1–6, Feb 2015.

[51] Isabel Wagner and David Eckhoff. Technical privacy metrics: a systematic survey. *arXiv preprint arXiv:1512.00327*, 2015.

[52] Yao Wang and G. E. Suh. Efficient timing channel protection for on-chip networks. In *Networks on Chip (NoCS), 2012 Sixth IEEE/ACM International Symposium on*, pages 142–151, May 2012.

[53] N. Weinberger and N. Merhav. A large deviations approach to secure lossy compression. *IEEE Trans. Inf. Theory*, PP(99):1–1, 2017.

[54] Nir Weinberger and Neri Merhav. A large deviations approach to secure lossy compression. May 2015.

[55] T. Weissman and E. Ordentlich. The empirical distribution of rate-constrained source codes. *IEEE Trans. Inf. Theory*, 51(11):3718–3733, Nov 2005.

[56] David Williams. *Probability with martingales*. Cambridge university press, 1991.

[57] A. D. Wyner. The wire-tap channel. *Bell System Technical Journal*, 54(8):1355–1387, Oct. 1975.

[58] Hirosuke Yamamoto. Rate-distortion theory for the Shannon cipher system. *IEEE Trans. Inf. Theory*, 43(3):827–835, 1997.

[59] Kehuan Zhang and XiaoFeng Wang. Peeping tom in the neighborhood: Keystroke eavesdropping on multi-user systems. In *Proceedings of the*

18th USENIX Security Symposium (USENIX Security 09), Montreal, Canada, 2009. USENIX.