Rate-distortion Theory for Secrecy Systems
Curt Schieler and Paul Cuff

Abstract

In this work, secrecy in communication systems is measured by the distortion that an adversary incurs. The transmitter and receiver share secret key, which they use to encrypt communication and ensure distortion at an adversary. A model is considered in which an adversary not only intercepts the communication from the transmitter to the receiver, but also potentially has access to noisy observations of the system. In particular, the adversary may have causal or noncausal access to a signal correlated with the source sequence or the receiver’s reconstruction sequence. The main contribution is the solution of the optimal tradeoff among communication rate, secret key rate, distortion at the adversary, and distortion at the legitimate receiver. It is demonstrated that causal side information at the adversary plays a pivotal role in this tradeoff. It is also shown that measures of secrecy based on normalized equivocation are a special case of the framework.

Index Terms
Rate-distortion theory, information-theoretic secrecy, shared secret key, causal disclosure, soft-covering lemma, equivocation.

I. INTRODUCTION

In “Communication Theory of Secrecy Systems” [6], Shannon regarded a communication system as perfectly secret if the source and the eavesdropped message are statistically independent. The secrecy system studied in [6] is referred to as the “Shannon cipher system” and is depicted in Figure 1. A necessary and sufficient condition for perfect secrecy is that the number

C. Schieler (schieler@princeton.edu) and P. Cuff (cuff@princeton.edu) are with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544.
This work was supported in part by the National Science Foundation under Grants CCF-1116013 and CCF-1017431, and also by the Air Force Office of Scientific Research under Grant FA9550-12-1-0196. Parts of this work were presented in [1], [2], [3], [4], [5].
of secret key bits per source symbol exceeds the entropy of the source. When the amount of key is insufficient, one must relax the requirement of statistical independence and invite new measures of secrecy.

One common way of measuring sub-perfect secrecy is with equivocation, the conditional entropy $H(X|M)$ of the source given the public message. The use of equivocation as a measure of secrecy was considered in the original work on the wiretap channel in [7] and [8], and it continues today. Although a distortion-based approach to secrecy might appear incomparable at first glance, it turns out that equivocation (when normalized by blocklength) becomes a special case of the framework developed here, under the proper choice of distortion measure.

In this work, we study an information-theoretic measure of secrecy that is directly inspired by rate-distortion theory. Whereas the objective in classical rate-distortion theory is to minimize a receiver’s distortion for a given rate of communication, our goal is to maximize an eavesdropper’s distortion for a given rate of secret key. If we relax the requirement of lossless communication in Shannon’s cipher system, then our goal is to maximize an eavesdropper’s distortion for a given secret key rate, communication rate, and distortion tolerance at the receiver. Although there are a variety of secrecy systems other than Shannon’s cipher system (such as a wiretap channel [7] or distributed correlated sources [9], [10]), this paper is concerned exclusively with settings involving shared secret key, a single discrete memoryless source, and a noiseless channel.

When distortion is used as a measure of secrecy, we are implicitly viewing an eavesdropper in the same way that one views a receiver in a standard rate-distortion setting – as an active participant whose goal is to produce a sequence that is statistically correlated with the source sequence. Because he plays an active role, the eavesdropper is thought of as an adversarial entity. To ensure robustness, we will design the communication and encryption schemes against the
worst-case adversarial strategy; that is, we wish to maximize the minimum distortion attainable by an adversary.

The study of information-theoretic secrecy via rate-distortion theory was initiated by Yamamoto in [11], in which the rate-distortion region was characterized for the special setting in which no secret key is available. Later, in “Rate-distortion Theory for the Shannon Cipher System” [12], Yamamoto considered the exact problem we have heretofore described, but only obtained an inner and outer bound on the achievable rate-key-distortion region. In this paper, we characterize the region; however, it is not our main focus. The following example serves to illustrate the care that should be exercised in a distortion-based approach to secrecy and motivates our primary investigation, which is centered around a salient feature of our model referred to as causal disclosure.

A. One-bit secrecy and causal disclosure

Consider an $n$-bit i.i.d. source sequence $X^n$ with $X_i \sim \text{Bern}(1/2)$. Suppose common randomness $K \sim \text{Bern}(1/2)$ is available to the transmitter and receiver; that is, there is one bit of shared secret key. Now suppose the transmitter uses $K$ to encrypt $X^n$ by transmitting the $n$-bit message $Y^n$, where $Y_i = X_i \oplus K$. In other words, he flips all of the bits of $X^n$ if $K = 1$, otherwise he simply sends $X^n$. Upon intercepting the public message $Y^n$, the adversary produces a reconstruction $Z^n$ and incurs expected distortion $\mathbb{E} \frac{1}{n} \sum_{i=1}^{n} d(X_i, Z_i)$, where $d(x, z)$ is a per-letter distortion measure. If $d(x, z) = 1\{x \neq z\}$, then an optimal strategy for the adversary is to simply set $Z^n = Y^n$, yielding an expected distortion of $1/2$. Observe that $1/2$ is also the maximum possible expected distortion that we could ever force on the adversary, regardless of the amount of secret key available! It appears as though we have maximized secrecy by only using one bit of secret key for an arbitrarily long $n$-bit source. However, this view is severely misleading because the adversary actually knows a great deal about $X^n$, namely that it is one of only two candidate sequences.

This example demonstrates the potential fragility of using distortion to measure secrecy without recognizing the ramifications. For, although maximum secrecy (in the distortion sense) is attained, it vanishes altogether if the adversary views just one true bit of the source sequence (the bit allows him to determine whether or not to flip the $Y^n$ sequence). In general, the consequences of the example apply to the setting that Yamamoto considered in [12]. An arbitrarily small rate
of secret key is enough to guarantee maximum distortion, but such secrecy is weak in the sense that even a small amount of additional knowledge (for example, observation of a few source symbols) is enough for the adversary to completely identify the source sequence.

The way that we strengthen a distortion-based approach to secrecy is through an assumption of causal disclosure, in which we design codes under the supposition that the adversary has noisy (or noiseless) access to the past behavior of the system. For example, in the one-bit secrecy example we might assume that the adversary produces the \( i \)th reconstruction symbol \( Z_i \) based not only on the public message \( M \), but also on the past source symbols \( X_{i-1} \). Incidentally, such a modification to the standard rate-distortion theory setting does not change the theory, though it has a dramatic effect in this secrecy setting. Regardless of whether or not an adversary actually has access to such information, designing our encryption under the assumption that he does leads to a much more robust notion of secrecy. In particular, it is resistant to disruptions in secrecy like those exhibited in the example. Despite the “pessimistic” nature of the causal disclosure assumption, we find that the optimal tradeoff between secret key and distortion in this regime is reasonable and not degenerate.

The assumption of causal disclosure is relevant not only for the sake of robustness, but also for its natural interpretations. In [13], an alternative view of rate-distortion theory was introduced in which source and reconstruction sequences are regarded as sequences of actions in a distributed system. Communication is used to coordinate the receiver’s actions with the transmitter’s actions (which are given by nature). In this context, an adversary can be viewed as an active participant in the system who produces a sequence of actions. With this interpretation, it is not unrealistic to assume that the adversary could have causal access to the system behavior. Depending on where the adversary is intercepting communication, he might be able to view the past actions of the transmitter or receiver (or both) and produce his current action accordingly.

To our delight, we find that optimal communication in this setting is not only fundamentally different than that of other source coding problems (often requiring a stochastic decoder), but in fact lends itself to a simple interpretation of injecting artificial memoryless noise into the adversary’s received signal.
The content of this paper is as follows. In Section II, we describe the problem setup. In Section III, we present a generalized version of the one-bit secrecy example in which there is no assumption of causal disclosure. In Section IV, we state our main result, Theorem I, in which causal disclosure is a primary assumption. Theorem I describes the optimal relationship among the communication rate, secret key rate, and distortion at the legitimate receiver and adversary. Section IV also establishes a number of relevant corollaries to Theorem I and provides several concrete examples. In Section V, we demonstrate how normalized equivocation arises as a special case of the causal disclosure framework. In Section VI, we give the achievability proof of Theorem I. Afterward, we discuss several important properties and implications of the optimal communication scheme used in the proof. Section VII provides the converse proof of Theorem I. In Section VIII, we consider some settings with noncausal disclosure that are not subsumed by Theorem I but that can be proved similarly. Lastly, Section IX gives results for settings involving causal disclosure with delay greater than one.
II. Preliminaries

The communication system model used throughout is shown in Figure 2. The transmitting node, Node A, observes an i.i.d. source sequence \( X^n \triangleq (X_1, \ldots, X_n) \), where \( X_i \) is distributed according to \( P_X \). Nodes A and B share a source of common randomness \( K \in \{1, \ldots, 2^{nR_0}\} \), referred to as secret key, which is uniformly distributed and independent of \( X^n \). Based on the source block \( X^n \) and the secret key \( K \), Node A transmits a message \( M \in \{1, \ldots, 2^{nR}\} \) that is received without loss by Nodes B and C. Once \( M \) is delivered, all three nodes sequentially produce actions: in the \( i \)th step, Nodes A, B and C produce \( X_i, Y_i, \) and \( Z_i \), respectively. Note that Node A has no control over his actions; they are simply given by \( X^n \). At the other end, Node B produces \( Y_i \) based on the pair \((M, K)\) and the adversarial Node C produces \( Z_i \) based on \( M \) and his observation of the past behavior of the system, \((W_{i-1}^x, W_{i-1}^y)\). At each step, the joint actions of the players incur a value \( \pi(x, y, z) \), which represents symbol-wise payoff; the block-average payoff is given by

\[
\frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, Z_i). \tag{1}
\]

Nodes A and B want to cooperatively maximize payoff, while Node C wants to minimize payoff through his actions \( Z^n \). Note that instead of evaluating secrecy and coordination separately, which could be done with two payoff functions \( \pi_1(x, y) \) and \( \pi_2(x, z) \), we have unified them in a single function \( \pi(x, y, z) \). This approach emphasizes the game-theoretic nature of the model, but the use of multiple payoff functions does have its own merits, and the results extend readily.

In Figure 2 we depict noisy causal disclosure by \((W_x^{i-1}, W_y^{i-1})\), where \( W_x^n \) is the output of a memoryless channel \( \prod_{i=1}^{n} P_{W_x|X} \) with input \( X^n \), and \( W_y^n \) is the output of a memoryless channel \( \prod_{i=1}^{n} P_{W_y|Y} \) with input \( Y^n \). Modeling the side information in this way covers a variety of scenarios. For example, if \( P_{W_x|X} \) and \( P_{W_y|Y} \) are identity channels, then the adversary has causal access \((X^{i-1}, Y^{i-1})\). If \((W_x, W_y) = (\emptyset, \emptyset)\), then the adversary is completely blind to the past and only views the public message \( M \).

Throughout, we assume that the alphabets \( \mathcal{X}, \mathcal{Y}, \) and \( \mathcal{Z} \) are finite. We denote the set \( \{1, \ldots, m\} \) by \([m]\) and use \( \Delta_A \) to denote the probability simplex of distributions with alphabet \( A \). The notation \( X \perp Y \) indicates that the random variables \( X \) and \( Y \) are independent, and \( X - Y - Z \) indicates a markov chain relationship.
**Definition 1.** An \((n, R, R_0)\) code consists of an encoder \(f : \mathcal{X}^n \times [2^{nR_0}] \to [2^{nR}]\) and a decoder \(g : [2^{nR}] \times [2^{nR_0}] \to \mathcal{Y}^n\). More generally, we allow a stochastic encoder \(P_{\mathcal{M}|\mathcal{X}^n,K}\) and a stochastic decoder \(P_{\mathcal{Y}^n|\mathcal{M},K}\). An \((n, R, R_0)\) code is said to have blocklength \(n\), communication rate \(R\), and secret key rate \(R_0\).

Permitting stochastic decoders that use local randomization is crucial (in contrast to Wyner’s wiretap channel, in which a stochastic encoder is needed). On the other hand, it is likely that the optimal encoder can be a deterministic function of the message and key, but this has not been shown. The proof of our main result uses a stochastic encoder and stochastic decoder.

Nodes A and B use an \((n, R, R_0)\) code to coordinate against Node C. To ensure robustness, we consider the payoff that can be assured against the worst-case adversary, i.e., the max-min payoff. There are several ways to define the payoff criterion for a block, and we consider three: expected payoff, probability of assured payoff, and symbol-wise minimum payoff.

**Definition 2.** Fix a source distribution \(P_X\), a symbol-wise payoff function \(\pi : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}\), and causal disclosure channels \(P_{\mathcal{W}_i|X}\) and \(P_{\mathcal{W}_i|Y}\). Denote the pair \((W^n_x, W^n_y)\) by \(W^n\). The triple \((R, R_0, \Pi)\) is achievable if there exists a sequence of \((n, R, R_0)\) codes such that

- **Under payoff criterion \(P_1\) (expected payoff):**
  
  \[
  \liminf_{n \to \infty} \min \left\{ \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, Z_i) \right\} \geq \Pi. \tag{2}
  \]

- **Under payoff criterion \(P_2\) (probability of assured payoff):**
  
  \[
  \forall \varepsilon > 0, \lim_{n \to \infty} \min \{ P_{\mathcal{Z}_i|M,W_{i-1}} \} \mathbb{P} \left[ \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, Z_i) \geq \Pi - \varepsilon \right] = 1. \tag{3}
  \]

- **Under payoff criterion \(P_3\) (symbol-wise minimum payoff):**
  
  \[
  \liminf_{n \to \infty} \min \left\{ \mathbb{E} \pi(X_i, Y_i, Z) \right\} \geq \Pi. \tag{4}
  \]

Under \(P_2\), the range of \(\pi(x, y, z)\) is extended to include \(-\infty\).

Several remarks concerning the preceding definitions are in order.

1) Although \(P_2\) and \(P_3\) are incomparable, they are both stronger than \(P_1\).

2) In each of the criteria, we allow the adversary to employ his best set of probabilistic strategies \(\{P_{\mathcal{Z}_i|M,W_{i-1}}\}_{i=1}^{n}\) that minimize payoff. However, since expectation is linear...
in $P_{Z_i|M,W^{i-1}}$ for all $i$, the expectation is minimized by extreme points of the probability simplex; thus, we can assume that Node C uses a set of deterministic strategies, \( \{z_i(m, w^{i-1})\}_{i=1}^n \).

3) It is assumed (although not explicit in the notation) that the adversary has full knowledge of the source statistics and the code that Nodes A and B use.

4) The optimal payoff does not increase if Node B is given direct causal access to Nodes A and C (i.e., if the decoder is given by \( \{P_{Y_i|M,K,X^{i-1},Z^{i-1}}\}_{i=1}^n \) instead of simply $P_{Y^n|M,K}$). This is shown in Section VII in the converse proof of the main result.

**Definition 3.** The rate-payoff region $\mathcal{R}(P_1)$ is the closure of achievable triples $(R, R_0, \Pi)$ under payoff criterion $P_1$. Regions $\mathcal{R}(P_2)$ and $\mathcal{R}(P_3)$ are defined in the same way.

### III. One-bit Secrecy, Generalized

In this section, we expand on the scenario in which lossless communication is required between Nodes A and B and there is no causal disclosure of the system behavior to Node C. Since $X^n$ must equal $Y^n$ with high probability, the payoff function is of the form $\pi(x, z)$. Thus, the achievability criteria for $(R, R_0, \Pi)$ under $P_3$ are that

$$\lim_{n \to \infty} \mathbb{P}[X^n \neq Y^n] = 0 \quad (5)$$

and

$$\lim_{n \to \infty} \inf \min_{i \in [n]} \min_{z(m)} \mathbb{E} \pi(X_i, z(M)) \geq \Pi. \quad (6)$$

**Proposition 1.** Fix $P_X$ and $\pi(x, z)$. If lossless communication is required and there is no causal disclosure, then $\mathcal{R}(P_3)$, the rate-payoff region under payoff criterion $P_3$, is equal to

$$\left\{ (R, R_0, \Pi) : R \geq H(X) \atop R_0 \geq 0 \atop \Pi \leq \min_{z} \mathbb{E} \pi(X, z) \right\}. \quad (7)$$

Thus, any positive rate of secret key\(^1\) guarantees maximum secrecy (in the distortion sense), as Node C can achieve $\min_z \mathbb{E} \pi(X, z)$ by only knowing the source statistics. In fact, we now prove that each point in (7) can be achieved with key size $\mathcal{K} = [n]$ instead of $\mathcal{K} = [2^{nR_0}]$. This

\(^1\)Note that $R_0 = 0$ is only included in Proposition 1 because we defined the region as the closure of achievable triples.
shows that even if the number of secret key bits is sublinear in the blocklength (in this case, \( \log n \)), one can achieve maximum secrecy when there is no assumption of causal disclosure. As in the example of one-bit secrecy, such guarantees are shattered if even a small number of source bits are available to the adversary.

The following lemma is useful for the payoff analysis.

**Lemma 1.** Let \( P_{XYZ} \) be a markov chain \( X - Y - Z \), and \( f \) an arbitrary function. Then

\[
\min_{g(x,y)} \mathbb{E} f(g(X,Y), Z) = \min_{g(y)} \mathbb{E} f(g(Y), Z). \tag{8}
\]

**Proof:** We have

\[
\min_{g(x,y)} \mathbb{E} f(g(X,Y), Z) = \min_{g(x,y)} \sum_{x,y} P_{X,Y}(x,y) \mathbb{E} f(g(X,Y), Z) | (X,Y) = (x,y) \tag{9}
\]

\[
= \sum_{x,y} P_{X,Y}(x,y) \min_{g} \mathbb{E} f(g, Z) | (X,Y) = (x,y) \tag{10}
\]

\[
= \sum_{x,y} P_{X,Y}(x,y) \min_{g} \mathbb{E} f(g, Z) | Y = y \tag{11}
\]

\[
= \min_{g(y)} \mathbb{E} f(g(Y), Z), \tag{12}
\]

where (a) follows from the markovity assumption.

Now we prove Proposition 1.

**Proof of Proposition 1** \[\square\] **Converse.** By the converse to the lossless source coding theorem, if (5) holds then we must have \( R > H(X) \). To see that the payoff never exceeds \( \min_z \mathbb{E} \pi(X,z) \), observe that the adversary can always let \( Z^n \) equal \((z^*, \ldots, z^*)\), where

\[
z^* = \arg\min_z \mathbb{E} \pi(X,z). \tag{13}
\]

**Achievability.** Let \( \varepsilon > 0 \). If Nodes A and B use the set of \( \varepsilon \)-typical sequences as their codebook, i.e. the set

\[
T^n_\varepsilon = \{x^n : |P_{x^n}(x) - P_X(x)| < \varepsilon P_X(x), \forall x \in \mathcal{X} \}, \tag{14}
\]

then the rate of communication is \((1 + \varepsilon)H(X)\) and the probability of error is

\[
\mathbb{P}[X^n \neq Y^n] < \varepsilon. \tag{15}
\]

In (14), \( P_{x^n} \in \Delta \mathcal{X} \) denotes the empirical distribution of the sequence \( x^n \), which is defined by

\[
P_{x^n}(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{x_i = x\}. \tag{16}
\]
The message is encrypted using common randomness $K \sim \text{Unif}(n)$. In order to encrypt, we first partition $T^n_\varepsilon$ into bins of size $n$ (in a manner specified shortly), and use $K$ to apply a one-time pad to the location of source sequence within the bin. More precisely, the encoder operates as follows: if $X^n$ is typical and is the $L$th sequence in the $J$th bin, then transmit the message $M = (J, L \oplus K)$, where $\oplus$ indicates addition modulo $n$. Thus, the adversary knows which bin $X^n$ lies in, but does not know which of the $n$ sequences it is. Node B can recover both $J$ and $L$, and produces the corresponding sequence.

We partition $T^n_\varepsilon$ according to the following equivalence relation:

$$x^n \sim y^n \text{ if } x^n \text{ is a cyclic permutation of } y^n.$$ \hspace{1cm} (17)

Although the resulting partition can contain bins of size less than $n$, the number of such bins is small enough that we can ignore them without affecting the communication rate or (15). Thus, we assume that partitioning $T^n_\varepsilon$ yields only bins of size $n$. Due to (17), each bin of size $n$ has the following property.

**Property 1.** If bin $b_j$ is written as an $n \times n$ matrix, then every row and column of the matrix has the same empirical distribution (denoted by $P_j$) and hence every row has the same probability (denoted by $\alpha_j$) under the source distribution $\prod_{i=1}^n P_X(x_i)$.

We now analyze the payoff. For sufficiently large $n$, we have for all $i \in [n]$ that

$$\min_{z(m)} \mathbb{E} \pi(X_i, z(M)) = \min_{z(j,l)} \mathbb{E} \pi(X_i, z(J, L \oplus K))$$ \hspace{1cm} (18)

$$\overset{(a)}{=} \min_{z(j)} \mathbb{E} \pi(X_i, z(J))$$ \hspace{1cm} (19)

$$= \min_{z(j)} \sum_j \sum_{x^n \in b_j} p(x^n) \pi(x_i, z(j))$$ \hspace{1cm} (20)

$$\overset{(b)}{=} \sum_j \alpha_j \min_z \sum_{x^n \in b_j} \pi(x_i, z)$$ \hspace{1cm} (21)

$$\overset{(c)}{=} \sum_j \alpha_j \min_z \sum_{x \in \mathcal{X}} nP_j(x) \pi(x, z)$$ \hspace{1cm} (22)

$$\overset{(d)}{\geq} \sum_j n\alpha_j \min_z \sum_{x \in \mathcal{X}} (1 - \varepsilon) P_X(x) \pi(x, z)$$ \hspace{1cm} (23)

$$= \mathbb{P}[X^n \in T^n_\varepsilon](1 - \varepsilon) \min_z \mathbb{E} \pi(X, z)$$ \hspace{1cm} (24)

$$\geq (1 - \varepsilon)^2 \min_z \mathbb{E} \pi(X, z), \hspace{1cm} (25)$$
where (a) is due to \((X_i, J) \perp (L \oplus K)\) and Lemma 1. (b) and (c) are due to Property 1 and (d) follows from the definition of \(T^n_\varepsilon\). Thus, we have (6).

**Discussion**

Suppose Nodes A and B use the binning scheme just described in the proof of Proposition 1 to achieve maximum secrecy. What if, instead of eavesdropping only the public message, the adversary is also able to view the past behavior of the system, namely \(X_{i-1}\)? Because of the structure of each bin (i.e., Property 1), knowledge of just the first symbol, \(X_1 = x_1\), is enough for the adversary to narrow down the size of the list of candidate source sequences from \(n\) to approximately \(nP_X(x_1)\). One can see that the adversary will be able to determine the true sequence quickly, well before the end of the block. In this manner, the adversary can take advantage of the causal disclosure to force the payoff to take on its minimum value instead of its maximum value. In general, causal disclosure benefits an adversary and gives rise to a nontrivial tradeoff between secret key and payoff. We remark that one of the key elements in the proof of the main result is that the benefits of causal disclosure can be voided if the right amount of secret key is available. In fact, it will become evident in Section VI that using secret key to sterilize the causal disclosure gives rise to the optimal tradeoff of secret key and payoff.

**IV. **Main Result

Our main result is the following.

**Theorem 1.** Fix \(P_X, \pi(x, y, z)\), and causal disclosure channels \(P_{W_x|X}\) and \(P_{W_y|Y}\). Then \(\mathcal{R}(P_1)\), the closure of achievable \((R, R_0, \Pi)\) under payoff criterion \(P_1\), is equal to

\[
\bigcup_{W_x - X - (U, V) - Y - W_y} \left\{ \left. \begin{array}{l}
(R, R_0, \Pi) : R \geq I(X; U, V) \\
R_0 \geq I(W_xW_y; V|U) \\
\Pi \leq \min_{z(u)} \mathbb{E} \pi(X, Y, z(U))
\end{array} \right\},
\right. \tag{26}
\]

where \(|U| \leq |\mathcal{X}| + 2\) and \(|V| \leq |\mathcal{X}||\mathcal{Y}|(|\mathcal{X}| + 2) + 1\). Furthermore,

\[
\mathcal{R}(P_1) = \mathcal{R}(P_2) = \mathcal{R}(P_3). \tag{27}
\]

We remark that the convexity of \(\mathcal{R}(P_1)\) and \(\mathcal{R}(P_2)\) can be shown from Definitions 2 and 3 by using a standard time-sharing argument. By (27), \(\mathcal{R}(P_3)\) is also a convex set.
We now elaborate on several corollaries to Theorem 1 that are obtained through different choices of the causal disclosure channels $P_{W_x|X}$ and $P_{W_y|Y}$. To begin, we consider scenarios in which lossless communication is required between Nodes A and B.

**A. Lossless communication**

In the following, we require $X^n$ to equal $Y^n$ with high probability. That is, we introduce into Definition 2 the additional constraint

$$\lim_{n \to \infty} \mathbb{P}[X^n \neq Y^n] = 0.$$ (28)

Conveniently, (28) can be ensured by considering payoff criterion $P_2$ with a payoff function $\pi(x, y, z)$ that evaluates to $-\infty$ when $x \neq y$.

**Corollary 1.** Fix $P_X$, $\pi(x, z)$, and causal disclosure channel $P_{W_x|X}$. If lossless communication is required (i.e., (28) is imposed), then the rate-payoff region $\mathcal{R}(P_2)$ is equal to

$$\bigcup_{U \subseteq X \setminus W_x} \left\{ (R, R_0, \Pi) : R \geq H(X) \right\}$$

$$R_0 \geq I(W_x; X|U)$$

$$\Pi \leq \min_{z(u)} \mathbb{E} \pi(X, z(U)).$$ (29)

**Proof:** Define a payoff function

$$\bar{\pi}(x, y, z) \triangleq \begin{cases} 
\pi(x, z) & \text{if } x = y \\
-\infty & \text{if } x \neq y.
\end{cases}$$ (30)

When $\Pi > -\infty$, it is easily verified that

$$\mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, Z_i) \geq \Pi - \varepsilon \right] \to 1 \iff \mathbb{P}[X^n \neq Y^n] \to 0$$

$$\mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Z_i) \geq \Pi - \varepsilon \right] \to 1.$$ (31)

Thus, $\mathcal{R}(P_2)$ is given by Theorem 1 with $W_y = \emptyset$. Denoting the region in (29) by $\mathcal{S}$, we now show that $\mathcal{R}(P_2) = \mathcal{S}$. Note that when $\Pi > -\infty$, we have

$$-\infty < \Pi \leq \min_{z(u)} \mathbb{E} \pi(X, Y, z(U)).$$ (32)
which implies \( X = Y \). When combined with \( X - (U,V) - Y \), this gives \( H(X|UV) = 0 \).

Therefore, \( \mathcal{R}(P_2) \subseteq \mathcal{S} \) follows from

\[
R \geq I(X;U,V) = H(X) \\
R_0 \geq I(W_x;V|U) = I(W_x;X,V|U) \geq I(W_x;X|U).
\]

To see \( \mathcal{S} \subseteq \mathcal{R}(P_2) \), let \( V = Y = X \).

Corollary \([1]\) in turn, spawns two important results. By invoking Corollary \([1]\) with \( W_x = \emptyset \), we recover Proposition \([1]\) under \( P_2 \).

**Corollary 2.** Fix \( P_X \) and \( \pi(x,z) \). If lossless communication is required and there is no causal disclosure, then the rate-payoff region \( \mathcal{R}(P_2) \) is equal to

\[
\{(R, R_0, \Pi) : R \geq H(X) \} \\
R_0 \geq 0 \\
\Pi \leq \min_z \mathbb{E} \pi(X,z).
\]

If we instead consider the disclosure channel \( W_x = X \), we have the following.

**Corollary 3.** Fix \( P_X \) and \( \pi(x,z) \). If lossless communication is required and \( X_{i-1} \) is disclosed, then the rate-payoff region \( \mathcal{R}(P_2) \) is equal to

\[
\bigcup_{P_{U|X}} \left\{ (R, R_0, \Pi) : R \geq H(X) \right\} \\
\bigcup_{P_{U|X}} \left\{ R_0 \geq H(X|U) \right\} \\
\bigcup_{P_{U|X}} \left\{ \Pi \leq \min_{z(u)} \mathbb{E} \pi(X,z(U)) \right\}.
\]

**B. Lossless communication example**

In this section, we present a concrete example of the region in Corollary \([3]\) (causal disclosure of Node A) and compare it to the region in Corollary \([2]\) (no causal disclosure).

We first show that \((34)\) can be written as a linear program. Since the constraint on \( R \) is fixed by the source distribution, we focus our attention on the boundary of the \((R_0, \Pi)\) tradeoff, namely

\[
\Pi(R_0) \triangleq \max_{P_U|X : H(X|U) \geq R_0} \min_{z(u)} \mathbb{E} \pi(X,z(U)).
\]
Notice that this can be rewritten as

\[ \Pi(R_0) = \max_{P_U: \sum_u P_{U|U=u} P_X = P_X, H(X|U) \geq R_0} \sum_u P_U(u) \min_z \mathbb{E}[\pi(X, z)|U = u]. \]  

(36)

If we are able to restrict the set \( \{P_{X|U=u}\}_{u \in \mathcal{U}} \) in the maximization to a finite set \( \mathcal{P} \subseteq \Delta_X \), then \( \Pi(R_0) \) can be expressed as a linear program. Indeed, viewing the distribution \( P_U \) as a vector \( p \in \mathbb{R}^{\mathcal{P}} \), (36) becomes

\[
\begin{align*}
\text{maximize} & \quad d^\top p \\
\text{subject to} & \quad p \geq 0 \\
& \quad 1^\top p = 1 \\
& \quad Tp = p_x \\
& \quad h^\top p \leq R_0
\end{align*} 
\]

where

- \( T \in \mathbb{R}^{|X| \times |\mathcal{P}|} \) is the transition matrix whose columns are the elements of \( \mathcal{P} \).
- The vector \( d \in \mathbb{R}^{|\mathcal{P}|} \) has entries

\[ d_u = \min_z \mathbb{E}[\pi(x, z)|U = u], \quad u \in \mathcal{U}. \]  

(37)

- The vector \( h \in \mathbb{R}^{|\mathcal{P}|} \) has entries

\[ h_u = H(X|U = u), \quad u \in \mathcal{U}. \]  

(38)

To see why there is always a choice of finite \( \mathcal{P} \) such that the rate-payoff boundary is unaffected, consider the function \( d: \Delta_X \to \mathbb{R} \) defined by

\[ d(p) = \min_z \mathbb{E}[\pi(X, z)], \quad \text{where } X \sim p. \]  

(39)

Observe that \( d(\cdot) \) is the boundary a convex polytope because it is the minimum of \( |Z| \) linear functions (and \( Z \) is finite). Define the set

\[ \mathcal{P} = \{p \in \Delta_X : d(p) \text{ is an extreme point of } d\} \]  

(40)

Given a set of distributions \( \{P_{X|U=u}\}_u \) that optimize (36), we can write each element \( P_{X|U=u} \) as a convex combination of the distributions in \( \mathcal{P} \) while maintaining the value of the objective.
Furthermore, due to the concavity of the entropy function, the constraint on $R_0$ is still satisfied. Thus, $\mathcal{P}$ is sufficient for the optimization.

In the particular case that the payoff function is hamming distance (i.e., $\pi(x, z) = 1\{x \neq z\}$), the set $\mathcal{P}$ has a particularly convenient form:

$$\mathcal{P} = \{p \in \Delta_X : p = \text{Unif}(\mathcal{A}) \text{ for some } \mathcal{A} \subseteq \mathcal{X}\}. \tag{41}$$

This allows us to give the following simple analytical expression for $\Pi(R_0)$. The proof is given in Appendix A.

**Theorem 2.** Fix $P_X$ and let $\pi(x, z) = 1\{x \neq z\}$. Define the function $f(\cdot)$ as the linear interpolation of the points $(\log n, \frac{n-1}{n}), n \in \mathbb{N}$. Also, define

$$\pi_{\text{max}} = 1 - \max_x P_X(x). \tag{42}$$

Then, the boundary of the rate-payoff region when lossless communication is required and $X^{i-1}$ is disclosed can be written as

$$\Pi(R_0) = \min\{f(R_0), \pi_{\text{max}}\}. \tag{43}$$

In Figure 3, we illustrate Theorem 2 for an arbitrary source distribution. Note that when there is no causal disclosure and $\pi(x, z)$ is hamming distance, the payoff is given by Corollary 2 as

$$\min_z \mathbb{E} \pi(X, z) = 1 - \max_x P_X(x) = \pi_{\text{max}}, \tag{44}$$

regardless of the rate of secret key. Comparing (44) with $\min\{f(R_0), \pi_{\text{max}}\}$ demonstrates the effect of causal disclosure (see Figure 3). In particular, we see that the assumption that the adversary does not view any of the true source bits can lead to a rather fragile guarantee of maximum secrecy. Indeed, at low rates of secret key, the gap that results from revealing the source causally is the difference between maximum secrecy and zero secrecy. This reduction in payoff is the price that is paid for increased robustness against an adversary (e.g., preventing pitfalls like those that we saw in the example of one-bit secrecy).

From Theorem 2, we also readily see that the payoff can saturate when $R_0 < H(X)$, which shows that maximum secrecy in the sense of distortion is not the same as Shannon’s perfect secrecy. For example, if $P_X = \{1/4, 1/4, 1/2\}$, then the maximum payoff of $1/2$ occurs at $R_0 = 1$, but $H(X) = 1.5$.

---

Here $n$ does not refer to blocklength.
Fig. 3: Illustration of Theorem 2 for a generic source $P_X$ with $1 - \max_x P_X(x) = \pi_{\max}$. The solid curve, $\Pi(R_0) = \min\{f(R_0), \pi_{\max}\}$, is the tradeoff between rate of secret key and payoff under the assumption of causal disclosure (Corollary 3). The loosely dashed line is $\pi_{\max}$, which also corresponds to the payoff when there is no causal disclosure (Corollary 2). The densely dashed curve is $f(R_0)$.

C. Lossy communication

In the previous section, the communication rate lay above $H(X)$ and did not affect the $(R_0, \Pi)$ tradeoff. However, when the requirement of lossless communication is relaxed, all three quantities interact. There are four natural special cases that are obtained by setting $W_x$ equal to $\emptyset$ or $X$ and setting $W_y$ equal to $\emptyset$ or $Y$. We denote the corresponding rate-payoff regions as $R_\emptyset, R_A, R_B, \text{ and } R_{AB}$ to distinguish which nodes’ actions are causally revealed.

Corollary 4. Fix $P_X$ and $\pi(x, y, z)$. In each of the following, the region holds under all three payoff criteria.

If there is no causal disclosure, then the rate-payoff region, $R_\emptyset$, is equal to

$$\bigcup_{P_{Y|X}} \left\{ (R, R_0, \Pi) : R \geq I(X; Y) \right\} \bigcup \left\{ R_0 \geq 0 \right\} \bigcup \left\{ \Pi \leq \min_z \mathbb{E} \pi(X, Y, z) \right\}. \quad (45)$$

May 16, 2013 DRAFT
If $X^{i-1}$ is disclosed, then the rate-payoff region, $\mathcal{R}_A$, is equal to
\[
\bigcup_{P_{Y,U|X}} \left\{ (R, R_0, \Pi) : \begin{array}{l}
R \geq I(X;Y,U) \\
R_0 \geq I(X;Y|U) \\
\Pi \leq \min_{Z(u)} \mathbb{E}\pi(X,Y,z(u))
\end{array} \right\}.
\] (46)

If $Y^{i-1}$ is disclosed, then $\mathcal{R}_B$ is given by directly substituting $W_x = \emptyset$ and $W_y = Y$ in (26). Similarly, if $(X^{i-1}, Y^{i-1})$ is disclosed, then $\mathcal{R}_{AB}$ is given by directly substituting $W_x = X$ and $W_y = \emptyset$ in (26).

**Proof:** Setting $(W_x, W_y) = (\emptyset, \emptyset)$ in Theorem 1 gives $\mathcal{R}_\emptyset$. Denote the region in (45) by $S$. If $(R, R_0, \Pi) \in \mathcal{R}_\emptyset$, then
\[
\begin{align*}
R &\geq I(X;U,V) = I(X;U,V,Y) \geq I(X;Y) \\
\Pi &\leq \min_{Z(u)} \mathbb{E}\pi(X,Y,z(U)) \leq \min_z \mathbb{E}\pi(X,Y,z),
\end{align*}
\]
which gives $\mathcal{R}_\emptyset \subseteq S$. To see $S \subseteq \mathcal{R}_\emptyset$, let $U = \emptyset$ and $V = Y$.

Setting $(W_x, W_y) = (X, \emptyset)$ in Theorem 1 gives $\mathcal{R}_A$. Denote the region in (46) by $T$. If $(R, R_0, \Pi) \in \mathcal{R}_A$, then
\[
\begin{align*}
R &\geq I(X;U,V) = I(I;U,V,Y) \geq I(X;U,Y) \\
R_0 &\geq I(X;V|U) = I(X;V,Y|U) \geq I(X;Y|U),
\end{align*}
\]
which gives $\mathcal{R}_A \subseteq T$. To see $T \subseteq \mathcal{R}_A$, let $V = Y$.

**D. Lossy communication examples**

In this section, we investigate concrete examples of Corollary 4 by considering the payoff function
\[
\pi(x, y, z) = \mathbb{1}\{x = y, x \neq z\}.
\] (51)

For this choice, the block-average payoff is the fraction of symbols in a block that Nodes A and B are able to agree on and keep hidden from Node C.

We now present achievable regions for the cases of Corollary 4 when $P_X \sim \text{Bern}(1/2)$ and $\pi(x, y, z)$ is given by (51). The region that we give for $R_\emptyset$ is optimal, and numerical computation...
suggests that the other regions are optimal as well. Setting $P_{Y|X} = \text{BSC}(\alpha)$, we have

$$\mathcal{R}_\emptyset = \bigcup_{\alpha \in [0, \frac{1}{2}]} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq 1 - h(\alpha) \\ R_0 \geq 0 \\ \Pi \leq \frac{1}{2}(1 - \alpha) \end{array} \right\}. \tag{52}$$

If we let $U = \emptyset$ and $P_{Y|X} = \text{BSC}(\alpha)$, then we have

$$\mathcal{R}_A \supseteq \bigcup_{\alpha \in [0, \frac{1}{2}]} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq 1 - h(\alpha) \\ R_0 \geq 1 - h(\alpha) \\ \Pi \leq \frac{1}{2}(1 - \alpha) \end{array} \right\}. \tag{53}$$

Letting $U = \emptyset$, $P_{Y|X} = \text{BSC}(\alpha)$, and $P_{V|Y} = \text{BSC}(\beta)$ gives

$$\mathcal{R}_B \supseteq \bigcup_{\alpha, \beta \in [0, \frac{1}{2}]} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq 1 - h(\alpha) \\ R_0 \geq 1 - h(\beta) \\ \Pi \leq \frac{1}{2}(1 - \alpha \ast \beta) \end{array} \right\}. \tag{54}$$

and also

$$\mathcal{R}_{AB} \supseteq \text{conv} \left\{ \bigcup_{\alpha, \beta \in [0, \frac{1}{2}]} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq 1 - h(\alpha) \\ R_0 \geq 1 + h(\alpha \ast \beta) - h(\alpha) - h(\beta) \\ \Pi \leq \frac{1}{2}(1 - \alpha \ast \beta) \end{array} \right\} \right\}. \tag{55}$$

where $\alpha \ast \beta = \alpha(1 - \beta) + \beta(1 - \alpha)$ and $\text{conv}(\cdot)$ denotes the convex hull operation. Regions (53) and (54) are convex as given.

Several observations concerning the regions in Figure 4 are in order. First, the minimum payoff is $1/4$, which occurs when there is no communication or secret key. This is achieved if Node B generates an i.i.d. sequence according to $\text{Bern}(1/2)$ and Node C produces an arbitrary sequence. Second, note the strict containment from top to bottom: causal access to Node A (Fig. 4b) is better for the adversary than access to Node B (Fig. 4c), and the combination (Fig. 4d) is strictly better for him than Node A alone. Finally, observe the effect of having a higher secret key rate than communication rate, and vice versa. When Node A is causally revealed, the payoff is a function of $\min(R, R_0)$ and there is no advantage in having excess of either rate. However, when Node B is revealed, both $R_0 > R$ and $R > R_0$ result in higher payoff than $R = R_0$. When both nodes are revealed, an excess of secret key rate increases payoff.$^3$

$^3$These relationships are not known to be true in general.
Fig. 4: Achievable regions of Corollary 4 for $P_X \sim \text{Bern}(1/2)$ and $\pi(x, y, z) = 1\{x = y, x \neq z\}$. Numerical computation suggests that these regions are optimal.
V. EQUIVCATION

In this section, we show that (normalized) equivocation-based measures of secrecy become a special case of the causal disclosure framework if we choose the payoff function to be a log-loss function. Relating distortion to conditional entropy via a log-loss function was done recently in the context of certain multiterminal source coding problems [14].

First, we remark that Theorem 1 can be readily extended to include multiple distortion functions. For example, if we wanted to separately evaluate coordination and secrecy, we could use two payoff functions \( \pi_1(x, y) \) and \( \pi_2(x, y, z) \). In this setting, it might be more natural to refer to distortion functions than payoff functions, with the goal of minimizing the distortion between Nodes A and B while maximizing the distortion between Nodes (A,B) and Node C. Then, the rate-distortion region becomes

\[
\bigcup_{W_x - X - (U,V) - Y - W_y} \left\{ (R, R_0, D_1, D_2) : \begin{array}{l}
R \geq I(X; U, V) \\
R_0 \geq I(W_x W_y; V | U) \\
D_1 \geq \mathbb{E} d_1(X, Y) \\
D_2 \leq \min_{z \in \Delta_X} \mathbb{E} d_2(X, Y, z(U))
\end{array} \right\}.
\]

(56)

Now consider \( W_x = X \) and a distortion function \( d_2 : \mathcal{X} \times \mathcal{Y} \times \Delta_X \rightarrow \mathbb{R} \) defined by

\[
d_2(x, y, z) = \log \frac{1}{z(x)}
\]

(57)

where \( z \) is a probability distribution on \( \mathcal{X} \), and \( z(x) \) denotes the probability of \( x \in \mathcal{X} \) according to \( z \in \Delta_X \). With this choice, the distortion in criterion \( P_1 \) can be written as

\[
\min_{\{P_{Z_i|M,X^{i-1}}\}_{i=1}^n} \mathbb{E} \frac{1}{n} \sum_{i=1}^n d_2(X_i, Y_i, Z_i) = \frac{1}{n} \sum_{i=1}^n \min_{P_{Z|M,X^{i-1}}} \mathbb{E} d_2(X_i, Y_i, Z) = \frac{1}{n} \sum_{i=1}^n \min_{P_{Z|M,X^{i-1}}} \mathbb{E} \log \frac{1}{Z(X_i)} = \min_{\{P_{Z_i|M,X^{i-1}}\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n H(X_i|Z^{i-1}) = \frac{1}{n} H(X^n|M),
\]

(58)

(59)

(60)

(61)

where (a) is due to the Lemma 2 (below). Thus, for the log-loss distortion function in (57), expected adversarial distortion under an assumption of causal disclosure simply becomes normalized equivocation.
Lemma 2. Fix a pair of random variables \((X, Y)\) and let \(Z = \Delta_X\). Then

\[
H(X|Y) = \min_{z: X \sim Y - z} \mathbb{E} \left[ \log \frac{1}{Z(X)} \right]
\]

where \(z(x)\) is the probability of \(x\) according to \(z\).

Proof: If \(X - Y - Z\), then

\[
\mathbb{E} \left[ \log \frac{1}{Z(X)} \right] = \mathbb{E} \left[ \log \frac{1}{P_{X|Y}(X|Y)} \right] + \mathbb{E} \left[ \log \frac{P_{X|Y}(X|Y)}{Z(X)} \right]
\]

\[
= H(X|Y) + \sum_{y,z} P_{YZ}(y, z) D(P_{X|Y=y}|z)
\]

\[
\geq H(X|Y),
\]

with equality if \(z = P_{X|Y=y}\) for all \((y, z)\).

So far, we have focused on the equivocation of \(X^n\); however, one might be interested in \(\frac{1}{n} H(Y^n|M)\) or \(\frac{1}{n} H(X^n, Y^n)\), instead. In these cases, the rate-distortion-equivocation regions can again be recovered from Theorem 1 (via the form in (56)) by considering \((W_x, W_y) = (\emptyset, Y)\), \(Z = \Delta_Y\) and

\[
d_2(x, y, z) = \log \frac{1}{z(y)}
\]

or \((W_x, W_y) = (X, Y)\), \(Z = \Delta_X \times Y\) and

\[
d_2(x, y, z) = \log \frac{1}{z(x, y)}.
\]

In all three cases, the regions can be simplified (in particular, the auxiliary random variable \(V\) can be eliminated). The results are given in the following theorem, part 1 of which was given by Yamamoto in [12].

Corollary 5. Fix \(P_X\) and \(d(x,y)\). Let \(\mathcal{R}\) denote the closure of achievable pairs \((R, R_0, D, E)\).

1) If the equivocation criterion is

\[
\liminf_{n \to \infty} \frac{1}{n} H(X^n|M) \geq E,
\]

then

\[
\mathcal{R} = \bigcup_{P_{Y|X}} \left\{ (R, R_0, D, E) : R \geq I(X; Y), \quad D \geq \mathbb{E} d(X, Y), \quad E \leq H(X) - [I(X; Y) - R_0]_+ \right\},
\]
where \([x]_+ = \max\{0, x\}\).

2) If the equivocation criterion is

\[
\liminf_{n \to \infty} \frac{1}{n} H(Y^n|M) \geq E, \tag{70}
\]

then

\[
\mathcal{R} = \bigcup_{X-U-Y} \left\{ (R, R_0, D, E) : R \geq I(X;U) \right\} \bigcup \left\{ D \geq \mathbb{E} d(X,Y) \right\} \bigcup \left\{ E \leq H(Y) - [I(Y;U) - R_0]_+ \right\}. \tag{71}
\]

3) If the equivocation criterion is

\[
\liminf_{n \to \infty} \frac{1}{n} H(X^n,Y^n|M) \geq E, \tag{72}
\]

then

\[
\mathcal{R} = \bigcup_{X-U-Y} \left\{ (R, R_0, D, E) : R \geq I(X;U) \right\} \bigcup \left\{ D \geq \mathbb{E} d(X,Y) \right\} \bigcup \left\{ E \leq H(X,Y) - [I(X,Y;U) - R_0]_+ \right\}. \tag{73}
\]

**Proof:** We only prove part 2, as parts 1 and 3 follow similar arguments. First, fix \(d_2(x,y,z)\) according to (66). Then, by Lemma 2,

\[
\min_{z(u)} \mathbb{E} d_2(X,Y,z(U)) = H(Y|U). \tag{74}
\]

From the discussion above, it is clear that setting \((W_x, W_y) = (\emptyset, Y)\) in (56) gives

\[
\mathcal{R} = \bigcup_{X-(U,V)-Y} \left\{ (R, R_0, D_1, D_2) : R \geq I(X;U,V) \right\} \bigcup \left\{ R_0 \geq I(Y;V|U) \right\} \bigcup \left\{ D \geq \mathbb{E} d(X,Y) \right\} \bigcup \left\{ E \leq H(Y|U) \right\}. \tag{75}
\]

Denote the region in (71) by \(\mathcal{S}\). To see \(\mathcal{R} \subseteq \mathcal{S}\), first consider \((R, R_0, D, E) \in \mathcal{R}\). Letting \(U' = (U,V)\), we have

\[
R \geq I(X;U,V) = I(X;U') \tag{76}
\]
\[
E \leq H(Y|U) = H(Y|U,V) + I(Y;V|U) \leq H(Y|U') + R_0 \tag{77}
\]
\[
E \leq H(Y), \tag{78}
\]
which implies \((R, R_0, D, E) \in S\). To see \(S \subseteq \mathcal{R}\), let \((R, R_0, D, E) \in S\). Set \(V' = U\) and find a random variable \(U'\) such that \(U' - U - (X, Y)\) form a markov chain and \(E = H(Y|U')\). This is always possible because \(E \in (H(Y|U), H(Y))\). Then, we can write

\[
R \geq I(X; U) = I(X; U', V') \tag{79}
\]

\[
R_0 \geq E - H(Y|U) = H(Y|U') - H(Y|U) = I(Y; V'|U') \tag{80}
\]

\[
E = H(Y|U'), \tag{81}
\]

which implies \((R, R_0, D, E) \in \mathcal{R}\). Thus, \(\mathcal{R} = S\).

VI. Achievability Proof

A. Soft covering lemma

The primary tool used in the achievability proof of Theorem 1 is a so-called “soft covering lemma”, a known result concerning the approximation of the output statistics of a channel\(^4\). Various forms of the lemma can be found in [16], [17], and [15], and related notions from the perspective of random binning can be found in [18]. In brief, the most basic version of the lemma is as follows. First, generate a random codebook of \(2^{nR}\) independent codewords, each drawn according to \(\prod_{i=1}^{n} P_U(u_i)\). Select a codeword, uniformly at random, as the input to a memoryless channel \(\prod_{i=1}^{n} P_{X|U}(x_i|u_i)\). The lemma states that if \(R > I(X; U)\), then the distribution of the channel output \(X^n\) converges to \(\prod_{i=1}^{n} P_X(x_i)\) in expected total variation, where the expectation is with respect to the random codebook.

Total variation, a measure of the distance between two distributions \(P\) and \(Q\) with common alphabet \(\mathcal{X}\), is defined by

\[
\|P - Q\| = \sup_{A} |P(A) - Q(A)|. \tag{82}
\]

Although the soft covering lemma holds for other measures (e.g., normalized divergence [16],[17]), we use total variation version found in [17] and [15] because of the following properties that total variation enjoys.

Property 2. Total variation distance satisfies the following.

\(^4\)The name “soft covering lemma” was given in [15]. The same lemma has also been referred to as the “resolvability lemma” and “cloud-mixing lemma”.

May 16, 2013 DRAFT
(a) If $X$ is countable, then total variation can be rewritten as

$$\|P - Q\| = \frac{1}{2} \sum_{x \in X} |P(x) - Q(x)|. \quad (83)$$

(b) Let $\varepsilon > 0$ and let $f(x)$ be a function bounded by $b \in \mathbb{R}$. Then

$$\|P - Q\| < \varepsilon \implies |\mathbb{E}_P f(X) - \mathbb{E}_Q f(X)| < \varepsilon b, \quad (84)$$

where $\mathbb{E}_P$ indicates that the expectation is taken with respect to the distribution $P$.

(c) Total variation satisfies the triangle inequality. For any $R \in \Delta_X$,

$$\|P - Q\| \leq \|P - R\| + \|R - Q\|. \quad (85)$$

(d) Let $P_XP_{Y|X}$ and $Q_XP_{Y|X}$ be two joint distributions on $\Delta_{X \times Y}$. Then

$$\|P_XP_{Y|X} - Q_XP_{Y|X}\| = \|P_X - Q_X\|. \quad (86)$$

That is, the distance between two joint distributions with the same channel $P_{Y|X}$ is equal to the distance between the input marginal distributions.

(e) For any $P, Q \in \Delta_{X \times Y}$,

$$\|P_X - Q_X\| \leq \|P_{XY} - Q_{XY}\|. \quad (87)$$

That is, the distance between marginal distributions is always less than the distance between the joint distributions.

We require the following generalization of the basic soft covering lemma to nonmemoryless channels.

**Definition 4.** Let $\{P_{X^n,Y^n}\}_{n=1}^\infty$ be a sequence of joint distributions. The sup-information rate of this sequence is defined as

$$\overline{T}(X; Y) \triangleq \limsup_{n \to \infty} \frac{1}{n} i_{P_{X^n,Y^n}}(X^n; Y^n), \quad (88)$$

where

$$\limsup_{n \to \infty} W_n \triangleq \inf \{ \tau : \mathbb{P}[W_n > \tau] \to 0 \} \quad (89)$$

and

$$i_{P_X,Y}(a; b) \triangleq \log \frac{P_{X,Y}(a, b)}{P_X(a)P_Y(b)}. \quad (90)$$

The function $i_{P_X,Y}(a; b)$ is called the information density.
Lemma 3 ([17], [15]). Let \( \{P_{X^n,Y^n}\}_{n=1}^\infty \) be a sequence of joint distributions. Let \( \mathcal{C}^{(n)} \) be a random codebook of \( 2^{nR} \) sequences in \( X^n \), each drawn independently according to \( P_{X^n} \) and indexed by \( m \in [2^{nR}] \). Let \( Q_{Y^n} \) denote the output distribution of the channel when the input is selected from \( \mathcal{C}^{(n)} \) uniformly at random; that is,

\[
Q_{Y^n}(y^n) = 2^{-nR} \sum_{m \in [2^{nR}]} P_{Y^n | X^n}(y^n | X^n(m)). \tag{91}
\]

If \( R > I(X;Y) \), then

\[
\lim_{n \to \infty} \mathbb{E}_{\mathcal{C}^{(n)}} \| Q_{Y^n} - P_{Y^n} \| = 0, \tag{92}
\]

where \( \mathbb{E}_{\mathcal{C}^{(n)}} \) indicates that the expectation is with respect to the random codebook.\(^5\)

![Superposition soft covering](image)

**Fig. 5:** Superposition soft covering: The pair \( (M_1, M_2) \) specify a codeword pair \( (U^n, V^n) \), which is passed through a memoryless channel to produce \( X^n \). This system induces a joint distribution \( P_{M_1,M_2,X^n} \).

We now turn to the particular variation of the soft covering that we will invoke multiple times in the achievability proof of Theorem [1]. As depicted in Figure 5, consider a setting involving superposition coding with two codebooks of rates \( R_1 \) and \( R_2 \). The codewords in codebooks \( \mathcal{C}^{(n)}_U \) and \( \mathcal{C}^{(n)}_V \) are generated according to \( \prod_{i=1}^n P_U \) and \( \prod_{i=1}^n P_{V|U} \), respectively. Consider \( P_{M_1,M_2,X^n} \), the joint input-output distribution corresponding to Figure 5, this distribution is defined by

\[
P_{M_1,M_2,X^n}(m_1, m_2, x^n) = 2^{-n(R_1+R_2)} P_{X^n|U^nV^n}(x^n | U^n(m_1), V^n(m_2)), \tag{93}
\]

where \( P_{X^n|U^nV^n} \) is a memoryless channel given by

\[
P_{X^n|U^nV^n}(x^n | u^n, v^n) = \prod_{i=1}^n P_{X|U,V}(x_i | u_i, v_i). \tag{94}
\]

\(^5\)Because the codebook is random, the output distribution \( Q_{Y^n} \) is a random variable taking values on \( \Delta Y^n \). One way to notate this is through the use of conditional distributions (i.e., write \( Q_{Y^n|\mathcal{C}^{(n)}} \)), but we choose to suppress such notation in order to simplify the presentation.
We want to know when the marginal distribution $P_{M_1,X^n}$ approximates the input-output distribution of the memoryless channel $\prod_{i=1}^nP_{X_i|U_i}$, where $P_{X_i|U} = \sum_v P_{X_i|U,V_i}P_{V_i|U}$. More precisely, we seek a sufficient condition on $R_2$ such that $P_{M_1,X^n}$ converges to a distribution $Q_{M_1,X^n}$ in expected total variation, where

$$Q_{M_1,X^n}(m_1, x^n) = 2^{-nR_1} \prod_{i=1}^nP_{X_i|U_i}(x_i|U_i(m_1)).$$

(95)

The following lemma establishes the rate condition as $R_2 > I(X; V|U)$. Furthermore, the lemma shows that the rate condition is unchanged if the channel $P_{X^n|U^n,V^n}$ is replaced by

$$P_{X^nZ^n|U^nV^n}(x^n, z^n|u^n, v^n) \triangleq \prod_{i=1}^nP_{X_i|U_i}(x_i|u_i, v_i)P_{Z_i|XUV}(z_i|x_i, u_i, v_i)^2\{i \in [k]\}$$

(96)

and $Q_{M_1,X^n}$ is replaced by

$$Q_{M_1,X^nZ^n}(m_1, x^n, z^n) = 2^{-nR_1} \prod_{i=1}^nP_{X_i|U_i}(x_i|U_i(m_1))P_{Z_i|XU}(z_i|x_i, U_i(m_1))^{2\{i \in [k]\}}$$

(97)

provided $k \in \mathbb{N}$ is not too large relative to the blocklength $n$.

**Lemma 4.** Fix $P_{U,V,X,Z}$. Let $C_U^{(n)}$ be a random codebook of $2^{nR_1}$ sequences in $U^n$, each drawn independently according to $\prod_{i=1}^nP_U(u_i)$ and indexed by $m_1 \in [2^{nR_1}]$. For each $m_1 \in [2^{nR_1}]$, let $C_V^{(n)}(m_1)$ be a random codebook of $2^{nR_2}$ sequences in $V^n$, each drawn independently according to $\prod_{i=1}^nP_{V_i|U_i}(v_i|u_i(m_1))$ and indexed by $(m_1, m_2)$, $m_2 \in [2^{nR_2}]$.

Let $P_{M_1M_2X^nZ^n}$ and $Q_{M_1X^nZ^n}$ be defined according to (93), (96), and (97).

If $R_2 > I(X; V|U)$, then there exists $\alpha \in (0, 1]$, depending only on the gap $R_2 - I(X; V|U)$, such that if $k \leq [\alpha n]$, then

$$\lim_{n \to \infty} \mathbb{E}_{C^{(n)}}\|P_{M_1X^nZ^k} - Q_{M_1X^nZ^k}\| \leq e^{-\gamma n}.$$  

(98)

for some $\gamma > 0$. Furthermore, (98) holds when $Z^k$ is replaced by $Z_B$ for any $B \subseteq [n]$ that satisfies $|B| \leq k$.

The proof, which uses Lemma 3, is relegated to Appendix B. We now turn to the proof of Theorem 1.
B. Achievability proof: design of encoder and decoder

Given \( \varepsilon > 0 \) and a joint distribution

\[
P_{UVXYW} = P_{UV} P_{X|U} P_{Y|U} P_{W_{x}|X} P_{W_{y}|Y},
\]

(99)
denote \( W = (W_x, W_y) \) and set \( R = I(X; U, V) + 2\varepsilon \) and \( R_0 = I(W; V|U) + \varepsilon \). We will represent the message \( M \) by the pair \( (M_p, M_s) \); the subscripts indicate that part of the message is intended to be “public” and part is intended to be “secure”. Although \( M = (M_p, M_s) \) is fully viewed by the adversary, we will use the secret key in such a way that only \( M_p \) is of any use to him. The two message components have corresponding rates \( R_p = I(X; U) + \varepsilon \) and \( R_s = I(X; V|U) + \varepsilon \), so that \( R = R_p + R_s \).

We first generate a random codebook of \( (U^n, V^n) \) pairs in the same manner as Lemma 4. Let \( C_U \) be a codebook of \( 2^{nR_p} \) sequences in \( U^n \), each drawn independently according to \( \prod_{i=1}^{n} P_U(u_i) \) and indexed by \( m_p \in [2^{nR_p}] \). For each \( m_p \), let \( C_V(m_p) \) be a codebook of \( 2^{n(R_s+R_0)} \) sequences in \( V^n \), each drawn independently according to \( \prod_{i=1}^{n} P_{V|U}(v_i|u_i(m_p)) \) and indexed by \( (m_p, m_s, k) \in [2^{nR_p}] \times [2^{nR_s}] \times [2^{nR_0}] \).

(100)

In order to establish the encoder and decoder, we first consider an auxiliary joint distribution \( \overline{P} \), defined by

\[
\overline{P}_{MKX^nY^nW^n} = \overline{P}_{MK} \overline{P}_{X^nY^n|MK} \overline{P}_{W^n|X^nY^n},
\]

(101)

where

- \( \overline{P}_{MK}(m, k) = 2^{-n(R+R_0)} \).
- \( \overline{P}_{X^nY^n|MK}(x^n, y^n, m, k) = \prod_{i=1}^{n} P_{XY|UV}(x_i, y_i|u_i(m_p), v_i(m_p, m_s, k)) \).
- \( \overline{P}_{W^n|X^nY^n}(w^n|x^n, y^n) = \prod_{i=1}^{n} P_{W|XY}(w_i|x_i, y_i) \).

Note that, due to the markov relation \( X - (U, V) - Y \) in (99), the distribution \( \overline{P} \) satisfies the markov chain

\[
X^n - (M, K) - Y^n.
\]

(102)

Contrast \( \overline{P} \) with the true distribution of the system,

\[
P_{MKX^nY^nW^n} = P_{X^n} P_K P_{M|X^nK} P_{Y^n|MK} P_{W^n|X^nY^n},
\]

(103)

where
\[ PX^n(x^n) = \prod_{i=1}^{n} P_X(x_i). \]
\[ PK(k) = 2^{-nR_0}. \]
\[ PM_{X^nK} \text{ is the encoder (to be defined).} \]
\[ PY^n|_{MK} \text{ is the decoder (to be defined).} \]
\[ PW^n|_{X^nY^n}(w^n|x^n,y^n) = \prod_{i=1}^{n} P_{W|XY}(w_i|x_i,y_i). \]

We will demonstrate that \( \hat{P} \) and \( P \) are close in expected total variation\(^6\) when the encoder and decoder are defined correctly. This being true, \( P \) will effectively inherit any properties that \( \hat{P} \) enjoys. In particular, the following property of \( P \), which holds when the rate of secret key \( R_0 \) is large enough, is the key to the payoff analysis:
\[
\lim_{n \to \infty} \mathbb{E} \| \hat{P}_{MX^nB^nW^n} - Q_{MX^nB^nW^n} \| = 0, \tag{104}
\]
where
\[
Q_{MX^nB^nW^n} \triangleq 2^{-n(R_R + R_s)} \prod_{i=1}^{n} P_{W|U}(w_i|u_i(M_p)) P_{XY|WU}(x_i,y_i|w_i,u_i(M_p)) 1_{\{i \in B\}}, \tag{105}
\]
and \( B \subseteq [n] \) is such that \( |B| \leq \lceil \alpha n \rceil \) for the appropriate choice of \( \alpha \in (0,1] \).

To see the significance of \( Q_{MX^nB^nW^n} \), consider \( B = \emptyset \) and \( W = (X,Y) \), so that
\[
Q_{MX^nY^n} = 2^{-n(R_R + R_s)} \prod_{i=1}^{n} P_{XY|U}(x_i,y_i|w_i(M_p)). \tag{106}
\]
Recall that \( W = (X,Y) \) implies direct causal disclosure of Nodes A and B; that is, the adversary has access to \((M,M_{X^{i-1}}Y_{i-1})\) at step \( i \). Now suppose that we could somehow define an encoder and decoder so that \( Q_{MX^nY^n} \) is actually the resulting distribution of the triple \((M,X^n,Y^n)\). Because \( Q \) factors as
\[
Q_{MX^nY^n} = Q_M Q_{M_p} Q_{X^nY^n|M_p}, \tag{107}
\]
we see that \( M_s \) is independent of \((M_p,X^n,Y^n)\). Therefore, the adversary’s estimate of \((X_i,Y_i)\) is not improved by his knowledge of \( M_s \) (it is effectively the secure part of the message). Furthermore, \( Q_{X^nY^n|M_p} \) is a memoryless channel from the codeword \( u^n(M_p) \) to the sequence pair \((X^n,Y^n)\); in particular, this implies
\[
(X_i,Y_i) - u_i(M_p) - (X_{i-1}^{i-1},Y_{i-1}^{i-1}), \forall i \in [n]. \tag{108}
\]

\(^6\)Note that \( P \) and \( \hat{P} \) are random variables because the codebook is random.
Therefore, the adversary’s best estimate of \((X_i, Y_i)\) only depends on \(u_i(M_p)\) and is not improved by the causal disclosure. In summary, if we can design the code so that \(P_{M^n X^n Y^n} \approx Q_{M^n X^n Y^n}\), then we will have essentially created an artificial noisy channel from the intercepted codeword \(u^n(M_p)\) to the pair \((X^n, Y^n)\), a property which greatly simplifies the payoff analysis and is interesting independent of the causal disclosure problem. We discuss this effect and some of its implications at the end of this section after completing the achievability proof.

Continuing with the proof, we want to show that (104) holds. Since \(R_0 > I(W; V|U)\), this property follows from invoking Lemma 4 with the assignment

\[
(R_1, R_2, U, V, X, Z) = (R, R_0, U, V, W, (X, Y)).
\]

(109)

We now show that \(\bar{P}\) and \(P\) are close in expected total variation as long as the communication rate \(R\) is large enough. First, define the encoder as \(P_{M\mid X^n K} \triangleq P_{M\mid X^n} P_{X^n\mid M K} P_{Y^n\mid X^n Y^n}\)

\[
P_{Y^n W^n M\mid X^n K} = P_{M\mid X^n K} P_{Y^n\mid M K} P_{W^n\mid X^n Y^n}
\]

\[
= \bar{P}_{M\mid X^n K} \bar{P}_{Y^n\mid M K} \bar{P}_{W^n\mid X^n Y^n}
\]

\[
= \bar{P}_{Y^n W^n M\mid X^n K},
\]

(112)

where (a) follows because \(\bar{P}\) satisfies (102). This allows us to write

\[
\mathbb{E}\|P_{M K X^n Y^n W^n} - \bar{P}_{M K X^n Y^n W^n}\| = \mathbb{E}\|P_{K X^n} \bar{P}_{Y^n W^n M\mid X^n K} - \bar{P}_{K X^n} P_{Y^n W^n M\mid X^n K}\|
\]

\[
= \mathbb{E}\|P_{K X^n} - \bar{P}_{K X^n}\|
\]

\[
\leq \frac{1}{2} \sum_{k \in K} 2^{-n R_0} \mathbb{E}\|P_{X^n} - \bar{P}_{X^n\mid K = k}\|,
\]

(115)

where (a) uses Property 2d of total variation and (b) uses Property 2a. Continuing, Properties 2c and 2e of total variation give

\[
\mathbb{E}\|P_{X^n} - \bar{P}_{X^n\mid K = k}\| \leq \mathbb{E}\|P_{X^n} - Q_{X^n}\| + \mathbb{E}\|\bar{P}_{X^n\mid K = k} - Q_{X^n}\|
\]

\[
\leq \mathbb{E}\|P_{X^n} - Q_{X^n}\| + \mathbb{E}\|\bar{P}_{M_p X^n\mid K = k} - Q_{M_p X^n}\|,\]

(117)

where \(Q_{M_p X^n}\) is defined by

\[
Q_{M_p X^n}(m_p, x^n) = 2^{-n R_p} \prod_{i=1}^{n} P_{X\mid U}(x_i \mid u_i(m_p))
\]

(118)
and $Q_{X^n}$ is the marginal of $Q_{M_p X^n}$.

Because $R_p > I(X; U)$, the first term in (117) vanishes by invoking Lemma 4 with the assignment

$$(R_1, R_2, U, V, X, Z) = (0, R_p, \emptyset, U, X, \emptyset). \quad (119)$$

Note that this is simply the basic soft covering lemma for memoryless channels that was described at the beginning of the section. Next, because $R_s > I(X; V | U)$, we can again invoke Lemma 4 to show that the second term in (117) vanishes, this time with the assignment

$$(R_1, R_2, U, V, X, Z) = (R_p, R_s, U, V, X, \emptyset). \quad (120)$$

Thus, we have

$$\lim_{n \to \infty} \mathbb{E} \| P_{MKX^n Y^n W^n} - \overline{P}_{MKX^n Y^n W^n} \| = 0. \quad (121)$$

It follows that $P$ inherits the property in (104) by writing

$$\lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - Q_{MXB Y^n W^n} \| \leq e^{-n\gamma}, \quad (126)$$

for some $\gamma > 0$, where $Q$ is defined in (105) with respect to a random codebook and $B \subseteq [n]$ is any subset of indices with $|B| \leq \lfloor \alpha n \rfloor$.

Lemma 5 encapsulates everything we need in order to analyze payoff criteria $P_2$ and $P_3$, both of which are stronger than $P_1$. We begin with $P_3$. 

$$\begin{align*}
\lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - Q_{MXB Y^n W^n} \| &\leq \lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - \overline{P}_{MXB Y^n W^n} \| \\
&\leq \lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - Q_{MXB Y^n W^n} \| + \lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - \overline{P}_{MXB Y^n W^n} \| \\
&\leq \lim_{n \to \infty} \mathbb{E} \| P_{MKX^n Y^n W^n} - \overline{P}_{MKX^n Y^n W^n} \| + \lim_{n \to \infty} \mathbb{E} \| P_{MKX^n Y^n W^n} - \overline{P}_{MKX^n Y^n W^n} \| \\
&= 0,
\end{align*}$$

where (a) and (b) come from Properties 2c and 2e of total variation, and (c) is due to (104) and (121).

In the following lemma, we summarize what has been shown so far.

**Lemma 5.** There exists an encoder, decoder and $\alpha \in (0, 1]$ such that $P_{MKX^n Y^n W^n}$ satisfies

$$\lim_{n \to \infty} \mathbb{E} \| P_{MXB Y^n W^n} - Q_{MXB Y^n W^n} \| \leq e^{-n\gamma}, \quad (126)$$

for some $\gamma > 0$, where $Q$ is defined in (105) with respect to a random codebook and $B \subseteq [n]$ is any subset of indices with $|B| \leq \lfloor \alpha n \rfloor$.

Lemma 5 encapsulates everything we need in order to analyze payoff criteria $P_2$ and $P_3$, both of which are stronger than $P_1$. We begin with $P_3$.
C. Achievability proof: analysis of payoff criterion $P_3$

We first argue the existence of a sequence of codebooks such that

$$\lim_{n \to \infty} \max_{i \in [n]} \| P_{M,X,Y,W^n} - Q_{M,X,Y,W^n} \| = 0$$

(127)

and

$$\lim_{n \to \infty} \max_{i \in [n]} \| Q_{u_i(M_p)} - P_U \| = 0.$$  (128)

To that end, since $R_p > 0$ we can invoke Lemma 4 with $k = 1$ and the assignment

$$(R_1, R_2, U, V, X, Z) = (0, R_p, \emptyset, U, \emptyset, U)$$

(129)

to yield

$$E \| Q_{u_i(M_p)} - P_U \| \leq e^{-\beta n}$$

(130)

for some $\beta > 0$. Then, by (130) and Lemma 5 with $B = \{i\}$, we can write

$$E_{C(n)} \left[ \sum_{i=1}^{n} \| P_{M,X,Y,W^n} - Q_{M,X,Y,W^n} \| + \sum_{i=1}^{n} \| Q_{u_i(M_p)} - P_U \| \right]$$

(131)

$$= \sum_{i=1}^{n} E_{C(n)} \| P_{M,X,Y,W^n} - Q_{M,X,Y,W^n} \| + \sum_{i=1}^{n} E_{C(n)} \| Q_{u_i(M_p)} - P_U \|$$

(132)

$$\leq ne^{-\gamma n} + ne^{-\beta n}.$$  (133)

Thus, there must exist a sequence of codebooks such that

$$\lim_{n \to \infty} \sum_{i=1}^{n} \| P_{M,X,Y,W^n} - Q_{M,X,Y,W^n} \| + \sum_{i=1}^{n} \| Q_{u_i(M_p)} - P_U \| = 0$$

(134)

because (134) holds when averaged over random codebooks. Finally, (134) implies (127) and (128).

With (127) and (128) in hand, we can use Property 2b of total variation to take advantage of the structure of $Q_{M,X,Y,W^n}$ in analyzing the payoff criterion. Define

$$b = \max_{(x,y,z) \in X \times Y \times Z} \pi(x, y, z).$$

(135)

For large enough $n$, we have for all $i \in [n]$ that

$$\min_{z(m,w^{i-1})} E_P \pi(X_i, Y_i, z(M, W^{i-1})) \overset{(a)}{=} \min_{z(m,w^{i-1})} E_Q \pi(X_i, Y_i, z(M, W^{i-1})) - \varepsilon b$$

(136)

$$\overset{(b)}{=} \min_{z(U)} E_Q \pi(X_i, Y_i, z(U_i(M_p))) - \varepsilon b$$

(137)

$$\overset{(c)}{=} \min_{z(U)} E \pi(X, Y, z(U)) - 2\varepsilon b$$

(138)
Step (a) is due to Property 2b of total variation and (127). Step (b) follows from Lemma 1 after noting that under $Q$ we have

$$ (X_i, Y_i) - U_i(M_p) - (M, W^{i-1}). \tag{139} $$

Lastly, step (c) follows from Property 2b of total variation and (128). This completes the proof for payoff criterion $P_3$.

**D. Achievability proof: analysis of payoff criterion $P_2$**

Analyzing criterion $P_2$ is more involved. Without loss of generality, we first restrict attention to those distributions $P_{UVXYW}$ that satisfy

$$ P_{XY}(x,y) > 0 \implies \pi(x,y,z) > -\infty, \forall x,y,z. \tag{140} $$

Otherwise, $\min_z \mathbb{E} \pi(X,Y,z) = -\infty$ and the region in Theorem 1 is trivial.

The analysis will take place over sub-blocks of length $k = [\alpha n]$ rather than over the full block. For ease of presentation, we assume that $[\alpha n] = \alpha n$ and that $k$ divides $n$ evenly; the analysis is readily adjusted when this is not the case. Define the $j$th sub-block of size $k$ by

$$ B(j) = \{(j-1)k, \ldots, jk - 1\}, j \in [1/\alpha]. \tag{141} $$

and denote the first $\ell$ indices of $B(j)$ by $B'_{(j)}$. Define $c = \min_z \mathbb{E} \pi(X,Y,z(U))$. To prove achievability for $P_2$, we claim that it is enough to show that

$$ \lim_{k \to \infty} \mathbb{E}_{C(\alpha)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B'_{(j)}} \pi(X_i, Y_i, Z^*_i) < c - \varepsilon \right] = 0, \forall j \in [1/\alpha], \tag{142} $$

where $Q$ is given by (105) and the $\{Z^*_i\}_{i=1}^n$ denote the adversary’s optimal strategy (i.e., the
strategy that minimizes the probability of assured payoff. To verify the claim, first write

\[ E_{C(n)} P \left[ \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, Z_i^*) \geq c - \varepsilon \right] \]  

\[ \geq E_{C(n)} P \left[ \cap_j \left\{ \frac{1}{k} \sum_{i \in B_j^T} \pi(X_i, Y_i, Z_i^*) \geq c - \varepsilon \right\} \right] \]  

\[ = 1 - E_{C(n)} P \left[ \cup_j \left\{ \frac{1}{k} \sum_{i \in B_j^T} \pi(X_i, Y_i, Z_i^*) < c - \varepsilon \right\} \right] \]  

\[ \overset{(a)}{\geq} 1 - \frac{1}{\alpha} \sum_j E_{C(n)} P \left[ \frac{1}{k} \sum_{i \in B_j^T} \pi(X_i, Y_i, Z_i^*) < c - \varepsilon \right] \]  

\[ \overset{(b)}{\rightarrow} 1 - \frac{1}{\alpha} \sum_j E_{C(n)} P_Q \left[ \frac{1}{k} \sum_{i \in B_j^T} \pi(X_i, Y_i, Z_i^*) < c - \varepsilon \right] \]  

\[ \overset{(c)}{\rightarrow} 1, \]  

where (a) uses the union bound, (b) follows from Lemma 5 and the definition of total variation, and (c) follows from (142). The analysis of criterion \( P_2 \) will be complete upon invoking Shannon’s random coding argument to guarantee the codebook existence.

We cannot use the standard law of large numbers to show (142) because the dependence of \( Z_i^* \) on \( W_{i-1} \) implies that the \( \{ \pi(X_i, Y_i, Z_i^*) \}_{i=1}^{n} \) are not mutually independent. Instead, we use a martingale argument.

Denote the random variable \( U^n(M_p) \) by \( \overline{U}^n \). Fix \( j \in [1/\alpha] \), and let \( S_t \) be the random variable defined by

\[ S_t = \sum_{i \in B_j^T} (\pi(X_i, Y_i, Z_i^*) - c(\overline{U}_i)), \]

where

\[ c(\overline{U}_i) \triangleq \min_{z(\overline{U})} E_{Q}[\pi(X_i, Y_i, z(\overline{U}_i))|\overline{U}_i], i \in B_j^T. \]

We claim that, conditioned on \( \overline{U}_t \), \( S_t \) is a submartingale, i.e.,

\[ E_Q[S_t|S_{t-1}, \overline{U}_t] \geq S_{t-1}, \forall t \in B(j). \]

To verify the claim, first observe that the definition of \( S_t \) gives

\[ E_Q[S_t|S_{t-1}, \overline{U}_t] = S_{t-1} + E_Q[\pi(X_t, Y_t, Z_t^*)|S_{t-1}, \overline{U}_t] - c(\overline{U}_t). \]
Then, write
\[
\mathbb{E}_Q[\pi(X_t, Y_t, Z_t^*)|S_{t-1}, U_t] \geq \min_{z(m,w_{t-1})} \mathbb{E}_Q[\pi(X_t, Y_t, z(M,W_{t-1}))|S_{t-1}, U_t] \tag{153}
\]
\[
\overset{(a)}{=} \min_{z(u)} \mathbb{E}_Q[\pi(X_t, Y_t, z(U_t))|S_{t-1}, U_t] \tag{154}
\]
\[
\overset{(b)}{=} \min_{z(u)} \mathbb{E}_Q[\pi(X_t, Y_t, z(U_t))|U_t] \tag{155}
\]
\[
= c(U_t). \tag{156}
\]

Step (a) follows by invoking Lemma [1] after noting that under $Q$ we have the Markov chain
\[
(X_t,Y_t) - (U_t, S_{t-1}) - (M,W_{t-1}). \tag{157}
\]

Step (b) follows from the Markov chain
\[
(X_t,Y_t) - U_t - S_{t-1}. \tag{158}
\]

Thus, conditioned on $U_t$, we have that $S_t$ is a submartingale. By Doob’s decomposition theorem, we can write $S_t = M_t + A_t$, where $M_t$ is a martingale and $A_t$ is an increasing process with $A_1 = 0$. Then, conditioned on $U^n$, we have
\[
\mathbb{P}_Q\left[\frac{1}{k} \sum_{i \in B_{(j)}^i} \pi(X_i,Y_i,Z_i^*) < \frac{1}{k} \sum_{i \in B_{(j)}^i} c(U_i) - \varepsilon \right] \tag{160}
\]
\[
= \mathbb{P}_Q[S_k < -n \varepsilon] \tag{161}
\]
\[
\leq \mathbb{P}_Q[M_k < -n \varepsilon] \tag{162}
\]
\[
= \mathbb{P}_Q[M_k - \mathbb{E}_Q[M_k] < -n \varepsilon - \mathbb{E}_Q[M_k]] \tag{163}
\]
\[
\overset{(a)}{\leq} \frac{\text{Var}_Q(M_k)}{(n \varepsilon + \mathbb{E}_Q[S_1])^2}, \tag{164}
\]

where (a) follows from Chebyshev’s inequality. Finally, we recursively bound the variance of $M_k$ (conditioned on $U^n$) by writing
\[
\text{Var}_Q(M_k) \overset{(a)}{=} \text{Var}_Q(\mathbb{E}_Q[M_k|M_{k-1}]) + \mathbb{E}_Q[\text{Var}_Q(M_k|M_{k-1})] \tag{165}
\]
\[
\leq \text{Var}_Q(\mathbb{E}_Q[M_k|M_{k-1}]) + \max_{(x,y): Q(x,y) > 0} \pi(x,y)^2 \tag{166}
\]
\[
\leq \text{Var}_Q(M_{k-1}) + \max_{(x,y): Q(x,y) > 0} \pi(x,y)^2. \tag{167}
\]
Step (a) uses the law of total variance. The recursion implies \( \operatorname{Var}(M_k) \in O(n) \), which, together with (164), shows

\[
\lim_{k \to \infty} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} \pi(X_i, Y_i, Z^*_i) < \frac{1}{k} \sum_{i \in B^t(j)} c(U_i) - \varepsilon \mid \bar{U}^n \right] = 0. \tag{168}
\]

Since this convergence is uniform for all \( \bar{U}^n \), we can take the expectation over random codebooks to get

\[
\lim_{k \to \infty} \mathbb{E}_{C(n)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} \pi(X_i, Y_i, Z^*_i) < \frac{1}{k} \sum_{i \in B^t(j)} c(U_i) - \varepsilon \right] = 0. \tag{169}
\]

Now observe that the law of large numbers guarantees

\[
\lim_{k \to \infty} \mathbb{E}_{C(n)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} c(U_i) > c - \varepsilon \right] = \lim_{k \to \infty} \mathbb{E}_{C(n)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} c(U_i(1)) > c - \varepsilon \right] \tag{170}
\]

\[
= \lim_{k \to \infty} \mathbb{E}_{C(n)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} c(U_i(M_p)) > c - \varepsilon \right] \tag{171}
\]

\[
= 1. \tag{172}
\]

where (a) uses the symmetry of the codebook construction. This, together with (169), yields

\[
\lim_{k \to \infty} \mathbb{E}_{C(n)} \mathbb{P}_Q \left[ \frac{1}{k} \sum_{i \in B^t(j)} \pi(X_i, Y_i, Z^*_i) < c - 2\varepsilon \right] = 0, \tag{173}
\]

completing the proof of (142). This concludes the achievability proof for Theorem 1.

E. Discussion: Optimal encoding produces artificial noise

The optimal encoding and decoding scheme designed in this section produces an effect that is worth investigating outside of this particular context of rate-distortion theory for secrecy systems. In particular, consider the most pessimistic disclosure assumption, that \( W = (X, Y) \). In this case, the communication system effectively corrupts the i.i.d. information signal \( X^n \) with noise by synthesizing a memoryless broadcast channel, with the information source \( X^n \) as input, actions at the intended receiver \( Y^n \) as one output, and a sequence \( U^n \) as the other output observed by the adversary. The synthesis is accurate in a particular sense relevant to secrecy. That is, the communication system, which uses public message \( M \) and secret key \( K \) to facilitate coordination, synthesizes memoryless noise characterized by \( P_{YU|X} \) by producing a distribution on \( (X^n, Y^n, M) \) such that \( P_{X^nY^n|M} \) closely approximates \( \prod_{i=1}^{n} P_{XY|U}(x_i, y_i | u_i(M)) \) for a set of
statistically typical $u^n(M)$ sequences. This behavior is revealed by $Q_{M,X^n,Y^n}$ in (106), which the
proof shows to converge to the induced joint distribution of the system in the limit of large $n$.

Let us now consider why this might be an operationally meaningful criterion for synthesizing
noise in a secrecy setting. For comparison, consider the work of Winter in [19]. There, he
considers a distribution on a triple of variables $(X,Y,U)$ and a communication system that
generates correlated random variables $X^n$ and $Y^n$ at two different nodes using communication
and secret key in the presence of an adversary. For the sake of comparison, imagine $Y^n$ as
a noisy version of $X^n$. The secrecy criterion in that work is very strong, requiring that the
public message reveal no more about the sequences $X^n$ and $Y^n$ then the correlated sequence
$U^n$ would, in the sense that $M$ is stochastically degraded from $U^n$ with respect to $(X^n,Y^n)$. This
is stronger than the secrecy criterion we gave in the previous paragraph, requiring more
communication resources as a consequence. Although the communication setting and results in
[19] are quite different from ours in that the setting does not have an information source provided
by nature, our proof and methods for achievability bear resemblance.

The noise synthesis achieved by the communication system of this section, even with the
weaker secrecy performance implied by (106), has some compelling operational significance.
Consider an adversary who actually does observe a noise-corrupted version of the information
signal, such as one of the outputs of a broadcast channel. As in any probabilistic situation,
rational behavior is based on the posterior distribution of the state of the universe given what
is known to the individual. In this situation that means $P_{X^n,Y^n|U^n}$ will dictate the adversary’s
optimal behavior, regardless of the objective that the adversary is trying to accomplish. Therefore,
a communication system that mimics $P_{X^n,Y^n|U^n}$ will elicit the same behavior by an adversary for
the same observed $U^n$ sequence as would occur if the noisy channel was genuine. Furthermore,
if the observed $U^n$ sequence is statistically representative of true noisy observations, then the
communication system performance in the presence of an adversary will be equivalent to the
memoryless broadcast channel that it mimics.

VII. Converse Proof

It is enough to prove the converse to Theorem 1 for payoff criterion $P_1$ alone, since it is the
weakest of the criteria. We further weaken the conditions by allowing Node B causal access to

---

7Exact characterization of this depends on the specific objectives of the communication system.
Nodes A and C (i.e., we permit decoders of the form \( \{P_{Y_i|MX_i\cdots Z_{i-1}}\}_{i=1}^n \)). We will see that this allowance does not increase the payoff.

Fix a source distribution \( P_X \), a payoff function \( \pi(x,y,z) \), and causal disclosure channels \( P_{W_x|X} \) and \( P_{W_y|Y} \). For ease of presentation, denote the pair \((W^n_x, W^n_y)\) by \( W^n \). Next, let \( Q \) be an auxiliary random variable drawn uniformly from \([n]\), independently of \((X^n, Y^n, W^n, M, K)\).

Identify the following random variables:

\[
X = X_Q \\
Y = Y_Q \\
Z = Z_Q \\
(W_x, W_y) = W_Q \\
U = (M, W^{Q-1}, Q) \\
V = K.
\] (174-179)

With these choices, it can be verified that

\[
W_x - X - (U,V) - Y - W_y \sim P_X \\
W_x|X \sim P_{W_x|X} \\
W_y|Y \sim P_{W_y|Y} \sim P_{W_y|Y}
\] (180-183)

The following properties of \( P_{MKX^nY^nW^n} \) can also be verified:

\[
X^n \perp K \sim P_X \\
X_i - (M, K, X_i^{i-1}) - W^{i-1}, \forall i \\
X_Q \perp Q.
\] (184-186)
Let \((R, R_0, \Pi)\) be an achievable triple. We first have

\[
nR \geq H(M) \quad (187)
\]

\[
\geq H(M|K) \quad (188)
\]

\[
\geq I(X^n; M|K) \quad (189)
\]

\[
^{(a)} I(X^n; M, K) \quad (190)
\]

\[
= \sum_{i=1}^{n} I(X_i; M, K|X^{i-1}) \quad (191)
\]

\[
= \sum_{i=1}^{n} I(X_i; M, K, X^{i-1}) \quad (192)
\]

\[
^{(b)} \sum_{i=1}^{n} I(X_i; M, K, X^{i-1}, W^{i-1}) \quad (193)
\]

\[
\geq \sum_{i=1}^{n} I(X_i; M, K, W^{i-1}) \quad (194)
\]

\[
^{(c)} nI(X_Q; M, K, W^{Q-1}, Q) \quad (195)
\]

\[
= nI(X; U, V), \quad (196)
\]

where (a), (b), and (c) follow from (184), (185), and (186). Next, we have

\[
nR_0 \geq H(K) \quad (197)
\]

\[
\geq H(K|M) \quad (198)
\]

\[
\geq I(W^n; K|M) \quad (199)
\]

\[
\geq \sum_{i=1}^{n} I(W_i; K|M, W^{i-1}) \quad (200)
\]

\[
^{(a)} nI(W_Q; K|M, W^{Q-1}, Q) \quad (201)
\]

\[
= nI(W; V|U), \quad (202)
\]
where (a) follows from (186). Finally, we have

\[ \Pi \leq \min_{z(m,w^{i-1},i)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \pi(X_i, Y_i, z(M, W^{i-1}, i)) \right] \]  

(203)

\[ = \min_{z(m,w^{i-1},i)} \mathbb{E} \mathbb{E}[\pi(X_Q, Y_Q, z(M, W^{Q-1}, Q))|Q] \]  

(204)

\[ = \min_{z(m,w^{i-1},i)} \mathbb{E} \pi(X_Q, Y_Q, z(M, W^{Q-1}, Q)) \]  

(205)

\[ = \min_{z(u)} \mathbb{E} \pi(X, Y, z(U)). \]  

(206)

It remains to bound the cardinality of \( \mathcal{U} \) and \( \mathcal{V} \), which is straightforward from the standard support lemma (e.g., \cite{20}). Note that the set of markov distributions forms a compact, connected set. To bound \( \mathcal{U} \), it suffices to have \(|\mathcal{X}| - 1\) elements to preserve \( P_X \) and 3 more elements to preserve \( H(X|U,V) \), \( I(W;V|U) \), and \( \min_{z(u)} \mathbb{E} \pi(X, Y, z(U)) \). To bound \( \mathcal{V} \), it suffices to have \(|\mathcal{X}||\mathcal{Y}||\mathcal{U}| - 1\) elements to preserve \( P_{XYU} \) and 2 more elements to preserve \( H(X|U,V) \) and \( H(W|U,V) \).

**VIII. OTHER FORMS OF DISCLOSURE**

Throughout this section, we denote \((W^n_x, W^n_y)\) by \( W^n \). We consider several relevant scenarios that are not subsumed by Theorem 1, but that can be solved by modifying the proof slightly. Instead of viewing \( W^{i-1} \), the adversary might have access to \( W_i \), \( W^i \), or even the full sequence \( W^n \). It turns out that the regions corresponding to \( W^i \) and \( W^n \) are the same.

**Theorem 3.** Fix \( P_X, \pi(x, y, z) \) and disclosure channels \( P_{W_x|x} \) and \( P_{W_y|y} \). If \( W_i \) is disclosed instead of \( W^{i-1} \), then the rate-payoff region for all three payoff criteria is equal to

\[ \bigcup_{P_{Y|X}} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq I(X;Y) \\ R_0 \geq 0 \\ \Pi \leq \min_{z(w_x,w_y)} \mathbb{E} \pi(X, Y, z(W_x, W_y)) \end{array} \right\}. \]  

(207)

**Proof:** For the proof of achievability, first invoke Lemma 4 with

\[ (R_1, R_2, U, V, X, Z) = (R_p, R_s, \emptyset, Y, \emptyset, (X, Y, W)) \]  

(208)

to show that \( R_0 > 0 \) suffices for the property

\[ \lim_{n \to \infty} \mathbb{E} \left\| P_{MXBYW} - Q_{MXBYW} \right\| = 0. \]  

(209)
to hold. Note that under $Q$,
\[ M \perp (X_B, Y_B, W_B). \]

The remainder of the achievability proof follows Section [VI]. To prove the converse, it is first straightforward to bound $R$ and $R_0$. To bound $\Pi$, identify $(W_x, W_y) = W_Q$, where $Q \sim \text{Unif}(n)$, and write
\[ \Pi \leq \min_{z(u, w)} \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, z(M, W_i)) \]  
\[ \leq \min_{z(u)} \mathbb{E} \frac{1}{n} \sum_{i=1}^{n} \pi(X_i, Y_i, z(W_i)) \]  
\[ = \min_{z(w)} \mathbb{E} \pi(X_Q, Y_Q, z(W_Q)) \]  
\[ = \min_{z(w)} \mathbb{E} \pi(X, Y, z(W_x, W_y)). \]

**Theorem 4.** Fix $P_X$, $\pi(x, y, z)$ and disclosure channels $P_{W_x | X}$ and $P_{W_y | Y}$. If $W^n$ or $W^i$ is disclosed instead of $W^{i-1}$, then the rate-payoff region for all three payoff criteria is equal to
\[ \bigcup_{W_x - X - (U; V) - Y - W_y} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq I(X; U, V) \\ R_0 \geq I(W_x, W_y; V | U) \\ \Pi \leq \min_{z(u, w)} \mathbb{E} \pi(X, z(U, W_x, W_y)) \end{array} \right\}. \]  

**Proof:** For the proof of achievability, suppose $W^n$ is disclosed. The only modification needed is to replace the markov chain in (139) with
\[ (X_i, Y_i) - (U_i(M_p), W_i) - (M, W^n) \]  
and the markov chains in (158) and (159) with
\[ (X_t, Y_t) - (\mathcal{U}_t, S^{t-1}, W_t) - (M, W^n) \]  
\[ (X_t, Y_t) - (\mathcal{U}_t, W_t) - S^{t-1}. \]

To show the converse, suppose that only $W^i$ is disclosed. The proof follows arguments similar to those in Section [VII] with the same identification of random variables. 

May 16, 2013 DRAFT
IX. CAUSAL DISCLOSURE WITH DELAY

In this section, we consider the effects of assuming that adversary has delayed causal access to the system behavior; that is, we replace causal disclosure $W^{i-1}$ with $W^{i-d}$, $d > 1$. We establish an inner and outer bound on the corresponding rate-payoff region and give an example in which the bounds meet. Using the bounds, we further show that if lossless communication is required, the minimum rate of secret key needed to ensure a given level of payoff is on the order of $1/d$.

A. Inner and outer bound

**Theorem 5** (Inner bound, causal disclosure with delay $d$). Fix $P_X$, $\pi(x, y, z)$, and causal disclosure channels $P_{W_x|x}$ and $P_{W_y|y}$. Let $R_d$ denote the closure of achievable $(R, R_0, \Pi)$ when the causal disclosure has delay $d$, $d \geq 1$.

$$R_d \supseteq \bigcup_{W^d_x = X^d - (U, V) - Y^d - W^d_y} \left\{ (R, R_0, \Pi) : \begin{array}{l} R \geq \frac{1}{d} I(X^d; U, V) \\ R_0 \geq \frac{1}{d} I(W^d_x W^d_y, V|U) \\ \Pi \leq \min_{z(u)} \mathbb{E} \left[ \frac{1}{d} \sum_{j=1}^d \pi(X_j, Y_j, z(U)) \right] \end{array} \right\}$$

(219)

where

$$P_{X^d W^d_x} = \prod_{j=1}^d P_X P_{W_x|x}$$

(220)

and

$$P_{W^d_y|Y^d} = \prod_{j=1}^d P_{W_y|y}.$$  

(221)

**Proof:** For simplicity, we present the proof for $d = 2$. The idea is to transform the problem into one involving delay $d = 1$ so that we can apply Theorem 1. To that end, we first treat the source $X^n$ as an i.i.d. sequence $\widetilde{X}$ of super-symbols of length 2 by defining

$$\widetilde{X}_i = (X_{2i-1}, X_{2i}), \quad i = 1, 2, \ldots, n/2.$$  

(222)

Similarly, treat $Y^n$, $W^n_x$, and $W^n_y$ as sequences of super-symbols by appropriately defining $\widetilde{Y}$, $\widetilde{W}_x$, and $\widetilde{W}_y$. At steps $i = 2, 4, \ldots, n$, disclose additional information $(W_{x,i-1}, W_{y,i-1})$ to the

---

8Numerical investigation reveals that the bounds are not tight in general.
adversary so that his estimate of the super-symbol pair \((\tilde{X}_i, \tilde{Y}_i)\) is based on the public message and \((\tilde{W}_{x}^{-1}, \tilde{W}_{y}^{-1})\). Note that supplying extra information to the adversary can only reduce the achievable region. To complete the transformation, define a payoff function \(\bar{\pi} : \mathcal{X}^2 \times \mathcal{Y}^2 \times \mathcal{Z}^2 \rightarrow \mathbb{R}\) by

\[
\bar{\pi}(x^2, y^2, z^2) = \sum_{j=1}^{2} \pi(x_j, y_j, z_j). \tag{223}
\]

If \((\bar{R}, \bar{R}_0, \bar{\Pi})\) is an achievable triple for this transformed problem, then \((\bar{R}/2, \bar{R}_0/2, \bar{\Pi}/2)\) is an achievable triple for the delayed causal disclosure problem with \(d = 2\). By applying Theorem 4, we obtain the region in (219) for \(d = 2\).

**Theorem 6** (Outer bound, causal disclosure with delay \(d\)). Fix \(P_X, \pi(x, y, z)\), and causal disclosure channels \(P_{W_x|x}\) and \(P_{W_y|y}\). Let \(\mathcal{R}_d\) denote the closure of achievable \((R, R_0, \Pi)\) when the causal disclosure has delay \(d\), \(d \geq 1\).

\[
\mathcal{R}_d \subseteq \bigcup_{W_x = X - (U, V) - Y - W_y} \left\{ (R, R_0, \Pi) : \begin{array}{l}
R \geq \frac{1}{d} I(X; U, V) \\
R_0 \geq \frac{1}{d} I(W_x W_y; V|U) \\
\Pi \leq \min_{z(u)} \mathbb{E} \pi(X, Y, z(U))
\end{array} \right\}. \tag{224}
\]

**Proof:** The key to the proof is the following lemma.

**Lemma 6.** For arbitrary random variables \((X^n, Y)\), it holds that

\[
d \cdot I(X^n; Y) \geq \sum_{i=1}^{n} I(X_i; Y|X_i^{i-d}) \tag{225}
\]

**Proof of Lemma 6**

\[
d \cdot I(X^n; Y) = \sum_{j=1}^{d} I(X^n; Y) \tag{226}
\]

\[
\geq \sum_{j=1}^{d} \sum_{i \in [n], i \equiv j \mod d} I(X_{i-d+1}^{i-d}; Y|X_i^{i-d}) \tag{227}
\]

\[
\geq \sum_{j=1}^{d} \sum_{i \in [n], i \equiv j \mod d} I(X_i; Y|X_i^{i-d}) \tag{228}
\]

\[
= \sum_{i=1}^{n} I(X_i; Y|X_i^{i-d}) \tag{229}
\]
where (a) uses the chain rule for mutual information.

The converse steps of Section VII can now be modified by identifying $U = (M, W^{Q-d}, Q)$. Bound $R$ by writing

$$d \cdot nR \geq d \cdot H(M)$$

$$\vdots$$

$$\geq d \cdot I(X^n; M, K)$$

$$\geq (a) \sum_{i=1}^{n} I(X_i; M, K|X^{i-d})$$

$$\vdots$$

$$\geq nI(X_Q; M, K, W^{Q-d}, Q)$$

$$= nI(X; U, V),$$

where (a) uses Lemma 6. The bound on $R_0$ follows a similar argument, and $\Pi$ can be bounded in the manner of Section VII.

B. Lossless communication

We now specialize the inner and outer bound to the setting in which lossless communication is required and $X^{i-d}$ is disclosed. In this regime, we are able to show explicitly how delay affects the tradeoff between rate of secret key and payoff.

**Theorem 7.** Fix $P_X$ and $\pi(x, z)$. Let $\mathcal{R}_d$ denote the closure of achievable $(R, R_0, \Pi)$ for the case of lossless communication and causal disclosure $X^{i-d}$, $d \geq 1$. Let $R_d(\Pi)$ denote the key-payoff boundary of $\mathcal{R}_d$. First, we have

$$\mathcal{R}_d \supseteq \bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

$$\bigcup_{X^d \sim \prod_{j=1}^{d} P_X} \left\{(R, R_0, \Pi) : R \geq H(X) \right\}$$

where (a) uses Lemma 6. The bound on $R_0$ follows a similar argument, and $\Pi$ can be bounded in the manner of Section VII.
Furthermore, for all $\Pi$, 

$$R_d(\Pi) = \Theta\left(\frac{1}{d}\right).$$ (239)

Proof: The bounds on $R_d$ follow directly from Theorems 5 and 6. From (238), we obtain $R_d(\Pi) \geq \frac{1}{d} R_1(\Pi)$. It remains to show that $R_d(\Pi) \leq c \cdot \frac{1}{d}$ for some constant $c$; we do this via (237). To that end, let $X^d \sim \prod_{j=1}^{d} P_X$. Let $K \sim \text{Unif}(\mathcal{X})$ be independent of $X^d$ and define

$$U = (X_1 \oplus K, X_2 \oplus K, \ldots, X_d \oplus K),$$ (240)

where $\oplus$ indicates addition modulo $\mathcal{X}$. With this choice of $U$, we have

$$H(X_i|X_j, U) = 0, \forall i, j \in [d]$$ (241)

and

$$X_j \perp U, \forall j \in [d].$$ (242)

Therefore, we can write

$$\frac{1}{d} H(X^d|U) = \frac{1}{d} \sum_{i=1}^{d} H(X_j|X^{j-1}, U)$$ (243)

$$\overset{(a)}{=} \frac{1}{d} H(X_1|U)$$ (244)

$$\overset{(b)}{=} \frac{1}{d} H(X),$$ (245)

where (a) and (b) follow from (241) and (242), respectively. Also, we have

$$\min_{z(u)} \mathbb{E}\left[\frac{1}{d} \sum_{j=1}^{d} \pi(X_j, z(U))\right] = \frac{1}{d} \sum_{j=1}^{d} \min_{z(u)} \mathbb{E} \pi(X_j, z(U))$$ (246)

$$\overset{(a)}{=} \min_{z} \mathbb{E} \pi(X, z)$$ (247)

$$= \pi_{\text{max}},$$ (248)

where (a) follows from Lemma 1 and (242). Thus, we have from the inner bound in (237) that this choice of $U$ achieves the point $(R_0, \Pi) = (\frac{1}{d} H(X), \pi_{\text{max}})$. Since $\pi_{\text{max}}$ is the maximum possible payoff, we have $R_d(\Pi) \leq \frac{1}{d} H(X)$, completing the proof of (239).
C. Example in which the bounds meet

In the preceding proof, we demonstrated that the point \((R_0, \Pi) = (\frac{1}{d}H(X), \pi_{\max})\) is in the region \((237)\) and is therefore achievable. If we choose the source distribution to be \(P_X \sim \text{Bern}(1/2)\), then from Theorem \([2]\) (which gives us \(R_1(\Pi)\)) and the convexity of the rate-payoff region, it is clear that \(R_d(\Pi) \leq \frac{1}{d}R_1(\Pi)\). Conversely, the outer bound in \((238)\) directly gives \(R_d \geq \frac{1}{d}R_1(\Pi)\).

APPENDIX A

PROOF OF THEOREM [2]

A. Supporting lemma

For each \(x \in \mathcal{X}\), define \(\mathcal{F}_n(x) \subseteq \Delta_X\) by

\[
\mathcal{F}_n(x) \triangleq \{ p \in \Delta_X : p = \text{Unif}(\mathcal{A}) \text{ for some } \mathcal{A} \subseteq \mathcal{X}, |\mathcal{A}| = n, \text{ and } p(x) = \max_{x'} p(x') \},
\]

and define \(\mathcal{A}_n(x)\) by

\[
\mathcal{A}_n(x) \triangleq \{ p \in \Delta_X : p(x) = \max_{x'} p(x') \text{ and } p(x) \in \left[\frac{1}{n+1}, \frac{1}{n}\right] \}.
\]

The key to the proof of Theorem [2] is the following technical lemma.

**Lemma 7.** For a random variable \(X\) with distribution \(P_X\), let \(\pi\) and \(N\) be such that \(P_X \in \mathcal{A}_N(\pi)\).

1) There exists a random variable \(V\), correlated with \(X\), such that for all \(v \in \mathcal{V}\),

\[
P_{X|V=v} \in \mathcal{F}_N(\pi) \cup \mathcal{F}_{N+1}(\pi).
\]

In other words, \(P_X\) can be written as a convex combination of distributions in \(\mathcal{F}_N(\pi) \cup \mathcal{F}_{N+1}(\pi)\).

2) Let \(n \in [N]\). There exists a random variable \(V\) such that for all \(v \in \mathcal{V}\),

\[
P_{X|V=v} \in \bigcup_{x \in \mathcal{X}} \mathcal{F}_n(x).
\]

In other words, for any \(n \in [N]\), \(P_X\) can be written as a convex combination of distributions in \(\bigcup_{x} \mathcal{F}_n(x)\).

**Proof:** Fix \(\pi \in \mathcal{X}\) and \(n \in \mathbb{N}\), and define

\[
\mathcal{F} \triangleq \mathcal{F}_n(\pi) \cup \mathcal{F}_{n+1}(\pi).
\]
First, one can verify that $A_n(x)$ is a convex set. Furthermore, it is well-known that every compact convex set is the convex hull of its extreme points. Thus, to prove part 1, it is enough to show that the set of extreme points of $A_n(x)$ is equal to $F$. Then $p \in A_n(x)$ can be written as a convex combination of the elements of $F$.

The set of extreme points of a convex set $C$ is defined by

$$\text{extr}(C) \triangleq \{ p \in C : \text{if } p = \theta q + (1 - \theta) r, \ q, r \in C, \ \theta \in (0,1) \text{ then } p = q = r \}. \quad (254)$$

We first show that $F \subseteq \text{extr}(A_n(x))$. Let $p \in F$, and let $q, r \in A_n(x)$, $\theta \in (0,1)$ be such that $q \neq p, r \neq p$, and

$$p = \theta q + (1 - \theta) r \quad (255)$$

If $p \in F_n(x)$, then $p = q = r$ is clear because $q(x) \in [0, \frac{1}{n}]$ and $r(x) \in [0, \frac{1}{n}]$ for all $x \in X$. On the other hand, suppose $p \in F_{n+1}(x)$. Because $q, r \in A_n(x)$ and $p(x) = \frac{1}{n+1}$, we have $q(x) = r(x) = \frac{1}{n+1}$. Thus, $q(x) \in [0, \frac{1}{n+1}]$ and $r(x) \in [0, \frac{1}{n+1}]$ for all $x \in X$, and again $p = q = r$.

To show $\text{extr}(A_n(x)) \subseteq F$, we proceed by way of contradiction and suppose that $p \in \text{extr}(A_n(x))$ and $p \notin F$. From $p \notin F$, it holds that $p(x') \in (0, \frac{1}{n+1}) \cup (\frac{1}{n+1}, \frac{1}{n})$ for some $x' \in X$. There are now three separate cases to consider depending on whether $p(x) = \frac{1}{n+1}$, $p(x) \in (\frac{1}{n+1}, \frac{1}{n})$, or $p(x) \in \frac{1}{n}$. For ease of exposition, we only consider $p(x) = \frac{1}{n+1}$; the other two cases use a similar argument. Since $p(x') \leq p(x)$, we have $p(x') \in (0, \frac{1}{n+1})$. It follows that there must exist $x'' \neq x'$ such that $p(x'') \in (0, \frac{1}{n+1})$; otherwise, we would have

$$\sum_{x \in X} p(x) = \frac{n}{n+1} + p(x') < 1 \quad (256)$$

Now we can write $p = \frac{1}{2} q + \frac{1}{2} r$, where

$$q(x) = \begin{cases} p(x), & x \neq x', x \neq x'' \\ p(x) + \varepsilon, & x = x' \\ p(x) - \varepsilon, & x = x'' \end{cases} \quad (257)$$

$$r(x) = \begin{cases} p(x), & x \neq x', x \neq x'' \\ p(x) - \varepsilon, & x = x' \\ p(x) + \varepsilon, & x = x'' \end{cases} \quad (258)$$
and
\[ \varepsilon = \frac{1}{2} \min \{ p(x'), p(x''), \frac{1}{n+1} - p(x'), \frac{1}{n+1} - p(x'') \}. \] (259)

Thus, \( p \notin \text{extr}(A_n(x)) \), giving the contradiction. We have shown \( F = \text{extr}(A_n(x)) \) and part 1.

To prove part 2 of the lemma, first define
\[ B_n \triangleq \bigcup_{x \in X} F_n(x). \] (260)

For any \( n \), it holds that
\[ B_{n+1} \subseteq \text{conv}(B_n). \] (261)

This follows from writing \( p \in B_{n+1} \) as
\[ p = \sum_{q \in B_n, \text{supp}(q) \subseteq \text{supp}(p)} \frac{1}{n+1} q. \] (262)

One can establish part 2 by using part 1 and (261).

\[ \blacksquare \]

\textbf{B. Proof of Theorem 2}

With Lemma 7 in hand, we are equipped to prove Theorem 2. Fix \( R_0 \) and let \( U^* \) be the maximizer of \( \Pi(R_0) \). When the payoff function is \( \pi(x, z) = 1\{x \neq z\} \), we can rewrite \( \Pi(R_0) \) as
\[ \Pi(R_0) = \min_{z(u)} \mathbb{E} \pi(X, z(U^*)) \] (263)

\[ = \min_{z(u)} \sum_u P_U(u) \sum_x P_{X|U^*}(x|u) 1\{x \neq z(u)\} \] (264)

\[ = \sum_u P_{U^*}(u) \min_z \sum_x P_{X|U^*}(x|u) 1\{x \neq z\} \] (265)

\[ = \sum_u P_{U^*}(u) \min_z (1 - P_{X|U^*}(z|u)) \] (266)

\[ = \sum_u P_{U^*}(u) (1 - \max_x P_{X|U^*}(x|u)). \] (267)

We now show that the set \( \{P_{X|U^* = u}\}_u \) in (36) can be restricted the finite set \( \mathcal{P}\text{unif} \), where
\[ \mathcal{P}\text{unif} \triangleq \{ p \in \Delta_X : p = \text{Unif}(\mathcal{A}) \text{ for some } \mathcal{A} \subseteq X' \}. \] (268)
By applying part 2 of Lemma 7 to each distribution in \( \{P_{X|U^* = u}\}_u \), we have that there exists a random variable \( V \) such that

\[
\forall u, v, P_{X|U^* = u, V = v} \in \mathcal{P}_{\text{unif}}
\]

\[
\forall u, v, v', \arg \max_x P_{X|U^*, V}(x|u, v) = \arg \max_x P_{X|U^*, V}(x|u, v').
\]  

(269) 

(270)

We now write

\[
\Pi(R_0) = \sum_u P_{U^*}(u)(1 - \max_x P_{X|U^*}(x|u))
\]

\[
= \sum_u P_{U^*}(u)(1 - \max_x \sum_v P_{X|U^*, V}(x|u, v) P_{V|U^*}(v|u))
\]

\[
= \min_u \max_v \mathbb{E} \pi(X, z(U^*, V)).
\]  

(271) 

(272) 

(273) 

(274)

where (a) is due to (267) and (b) follows from (270). By noting that \( R_0 \geq H(X|U^*) \geq H(X|U^*, V) \) and letting \( U = (U^*, V) \), we have

\[
\Pi(R_0) \leq \max_{U: P_{X|U^* = u} \in \mathcal{P}_{\text{unif}}} \min_{z(u,v)} \mathbb{E} \pi(X, z(U)).
\]  

(275)

This shows that we can restrict attention to \( \mathcal{P}_{\text{unif}} \) without hurting the payoff. Now, observe that \( p \in \mathcal{P}_{\text{unif}} \) satisfies

\[
(H(p), 1 - \max_x p(x)) = (\log n, \frac{n-1}{n})
\]

(276)

for some \( n \in \mathbb{N} \). Referring to (267) and noting that \( H(X|U) = \sum_u P_U(u)H(X|U = u) \), we see that \( \Pi(R_0) \) cannot lie outside of the convex hull of the pairs \( (\log n, \frac{n-1}{n}), n \in \mathbb{N} \). That is,

\[
\Pi(R_0) \leq f(R_0).
\]  

(277)

To see \( \Pi(R_0) \leq \pi_{\text{max}} \), simply write

\[
\Pi(R_0) = \sum_u P_{U^*}(u)(1 - \max_x P_{X|U^*}(x|u))
\]

\[
\leq 1 - \max_x \sum_u P_{U^*}(u) P_{X|U^*}(x|u)
\]

\[
= \pi_{\text{max}}.
\]  

(278) 

(279) 

(280)

It remains to show that \( \min\{f(R_0), \pi_{\text{max}}\} \) can be achieved through the proper choice of \( U \). To that end, let \( \pi \) and \( N \) be such that \( P_X \in \mathcal{A}_N(\pi) \). By the convexity of \( \mathcal{R} \), we will be done once
we show that we can achieve not only the points \((\log n, \frac{n-1}{n})\), \(n \in [N]\), but also the intersection of the \(f\) with \(\pi_{\text{max}}\). To achieve the point \((\log n, \frac{n-1}{n})\), invoke part 2 of Lemma\(^7\) produce \(U\). Denote the corresponding rate-payoff pair by \((R'_0, \Pi')\). Since the \(\{P_{X|U=u}\}_u\) all satisfy
\[
(H(X|U = u), 1 - \max_x P_{X|U=u}(x|u)) = (\log n, \frac{n-1}{n})
\] (281)
so must \((R'_0, \Pi')\) as well. To achieve the intersection of \(f\) with \(\pi_{\text{max}}\), first invoke part 1 of Lemma\(^7\) to produce \(U\). Denote the corresponding rate-payoff pair by \((R''_0, \Pi'')\). The \(\{P_{X|U=u}\}_u\) correspond to either \((\log n, \frac{n-1}{n})\) or \((\log(n + 1), \frac{n}{n+1})\). Thus, \((R''_0, \Pi'')\) lies on \(f\) because it is a convex combination of those two points. We also have that \((R''_0, \Pi'')\) satisfies \(\Pi'' = \pi_{\text{max}}\) because
\[
\arg\max_x P_{X|U=u}(x|u) = \pi, \ \forall u \in \mathcal{U}.
\] (282)
This completes the proof of Theorem\(^2\)

APPENDIX B

PROOF OF LEMMA\(^4\)

Let \(R_2 > I(X; V|U)\). Define the set
\[
\mathcal{A}_{\varepsilon', \varepsilon''} \triangleq \{u^n \in \mathcal{U}^n : u^n \in \mathcal{T}_{\varepsilon'}^{(n)} \text{ and } u^k \in \mathcal{T}_{\varepsilon''}^{(k)}\}.
\] (283)
First, write
\[
\|P_{M_1, X^n, Z^k} - Q_{M_1, X^n, Z^k}\| = \sum_{m_1: U^n(m_1) \in \mathcal{A}_{\varepsilon', \varepsilon''}} P_{M_1}(m_1) \|P_{X^n, Z^k|M_1=m_1} - Q_{X^n, Z^k|M_1=m_1}\|
\] (284)
\[
+ \sum_{m_1: U^n(m_1) \notin \mathcal{A}_{\varepsilon', \varepsilon''}} P_{M_1}(m_1) \|P_{X^n, Z^k|M_1=m_1} - Q_{X^n, Z^k|M_1=m_1}\|. \quad (285)
\]
The expected value of the second term in (285) can be bounded easily. We have
\[
\frac{1}{2} \mathbb{E} \sum_{m_1: U^n(m_1) \notin \mathcal{A}_{\varepsilon', \varepsilon''}} P_{M_1}(m_1) \|P_{X^n, Z^k|M_1=m_1} - Q_{X^n, Z^k|M_1=m_1}\| \quad (286)
\]
\[
\leq \mathbb{E} \sum_{m_1: U^n(m_1) \notin \mathcal{A}_{\varepsilon', \varepsilon''}} P_{M_1}(m_1) \quad (287)
\]
\[
= \mathbb{P}[U^n(M_1) \notin \mathcal{A}_{\varepsilon', \varepsilon''}] \quad (288)
\]
\[
= \mathbb{P}[U^n(1) \notin \mathcal{A}_{\varepsilon', \varepsilon''}] \quad (289)
\]
\[
\leq \mathbb{P}[U^n \notin \mathcal{T}_{\varepsilon'}^{(n)}] + \mathbb{P}[U^k \notin \mathcal{T}_{\varepsilon''}^{(k)}] \quad (290)
\]
\[
\leq \varepsilon' + \varepsilon'' \quad (291)
\]
for sufficiently large $n$. The expected value of the first term in (285) can be rewritten by moving
the expectation with respect to $C^{(n)}$ inside the sum.

\[
\mathbb{E} \sum_{m_1: U^n(m_1) \in A^{(n)}_{\epsilon', \epsilon''}} P_{M_1}(m_1) \left\| P_{X^n, Z^n | M_1 = m_1} - Q_{X^n, Z^n | M_1 = m_1} \right\| \tag{292}
\]

\[
= \mathbb{E}_{C_U} \sum_{m_1: U^n(m_1) \in A^{(n)}_{\epsilon', \epsilon''}} P_{M_1}(m_1) \mathbb{E}_{C_U(m_1)} \left\| P_{X^n, Z^n | M_1 = m_1} - Q_{X^n, Z^n | M_1 = m_1} \right\|. \tag{293}
\]

It remains to show that the inner expectation vanishes for each $m_1$.  We can invoke Lemma 3
once we show that the sup-information rate of $P_{X^n, Z^n, V^n | U^n = u^n}$ (with $u^n \in A^{(n)}_{\epsilon', \epsilon''}$) is strictly less
than $R_2$ for small enough $\epsilon' > 0$, $\epsilon'' > 0$, and $\alpha > 0$. To that end, observe that $P_{X^n, Z^n, V^n | U^n = u^n}$
is memoryless and the second moments of $\{i(X_i, Z_i; V_i)\}$ are uniformly bounded. Thus, we can
apply the law of large numbers and bound the sup-information rate by the expected information
density. We bound the expected information density as follows.

\[
\mathbb{E} \left[ \frac{1}{n} i_{P_{X^n, Z^n, V^n | U^n = u^n}} (X^n, Z^n; V^n) \right] \tag{294}
\]

\[
= \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{k} i_{P_{X^n, Z^n | U^n = u^n}} (X_i, Z_i; V_i) \right] + \mathbb{E} \left[ \frac{1}{n} \sum_{i=k+1}^{n} i_{P_{X^n, Z^n | U^n = u^n}} (X_i; V_i) \right] \tag{295}
\]

\[
= \frac{1}{n} \sum_{i=1}^{k} I(X, Z; V | U = u_i) + \frac{1}{n} \sum_{i=k+1}^{n} I(X; V | U = u_i) \tag{296}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} I(X; V | U = u_i) + \frac{1}{n} \sum_{i=1}^{k} I(Z; V | X, U = u_i) \tag{297}
\]

\[
= \sum_{u \in \mathcal{U}} \frac{1}{n} \left| \{i \in [n] : u_i = u\} \right| I(X; V | U = u) + \frac{1}{n} \left| \{i \in [k] : u_i = u\} \right| I(Z; V | X, U = u) \tag{298}
\]

\[
\leq \sum_{u \in \mathcal{U}} (1 + \epsilon') P_U(u) I(X; V | U = u) + (1 + \epsilon'') P_U(u) I(Z; V | X, U = u) \tag{299}
\]

\[
= I(X; V | U) + \epsilon' I(X; V | U) + \alpha (1 + \epsilon'') I(Z; V | X, U). \tag{300}
\]

Step (a) follows from $\frac{k}{n} \leq \alpha$ and the definition of $T_{\epsilon'}^{(n)}$ and $T_{\epsilon''}^{(k)}$. The expression in (300)
is strictly less than $R$ for the proper choice of positive $\epsilon'$, $\epsilon''$ and $\alpha$. Thus, when $u^n \in A^{(n)}_{\epsilon', \epsilon''}$,

\[
R > \limsup_{n \to \infty} \frac{1}{n} i_{P_{X^n, Z^n, V^n | U^n = u^n}} (X^n, Z^n, V^n). \tag{301}
\]

\footnote{Due to the symmetry of codebook construction, the behavior of the inner expectation is uniform for all $m_1$. Thus, the rate of convergence does not play a role in claiming that (293) vanishes.}
Invoking Lemma 3, we have

\[
\lim_{n \to \infty} E_{c^{(n)}}(m_1) \left\| P_{X^n,Z^k|M_1=m_1} - Q_{X^n,Z^k|M_1=m_1} \right\| = 0.
\]  

(302)

The same arguments hold when \( Z^k \) is replaced by \( Z_B \) for any \( B \subseteq [n] \) that satisfies \( |B| \leq k \). Furthermore, one can use arguments similar to those found in [15] to show that the expression in (302) converges exponentially fast with \( n \). This concludes the proof of Lemma 4.

REFERENCES

[1] P. Cuff, “A framework for partial secrecy,” in Proc. Global Telecomm. Conf. (GLOBECOM), Dec. 2010.
[2] ——, “Using a secret key to foil an eavesdropper,” in 48th Allerton Conf. on Communication, Control, and Computing, Sept. 2010, pp. 1405–1411.
[3] C. Schieler and P. Cuff, “Secrecy is cheap if the adversary must reconstruct,” in Proc. IEEE International Symposium on Information Theory (ISIT), Jul. 2012, pp. 66–70.
[4] P. Cuff, “Optimal equivocation in secrecy systems a special case of distortion-based characterization,” in Information Theory and Applications Workshop (ITA), Feb. 2013, pp. 1–3.
[5] C. Schieler and P. Cuff, “Rate-distortion theory for secrecy systems,” to be presented at ISIT 2013.
[6] C. Shannon, “Communication theory of secrecy systems,” Bell Syst. Tech. J., vol. 28, no. 4, pp. 656–715, Oct. 1949.
[7] A. Wyner, “The wire-tap channel,” Bell Syst. Tech. J., vol. 54, no. 8, pp. 1334–1387, 1975.
[8] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” IEEE Trans. Inf. Theory, vol. 24, no. 3, pp. 339–348, May 1978.
[9] U. Maurer, “Secret key agreement by public discussion from common information,” IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 733–742, May 1993.
[10] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography. i. secret sharing,” IEEE Trans. Inf. Theory, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.
[11] H. Yamamoto, “A rate-distortion problem for a communication system with a secondary decoder to be hindered,” IEEE Trans. Inf. Theory, vol. 34, no. 4, pp. 835–842, July 1988.
[12] ——, “Rate-distortion theory for the shannon cipher system,” IEEE Trans. Inf. Theory, vol. 43, no. 3, pp. 827–835, May 1997.
[13] P. Cuff, H. Permuter, and T. Cover, “Coordination capacity,” IEEE Trans. Inf. Theory, vol. 56, no. 9, pp. 4181–4206, Sept. 2010.
[14] T. Courtade and T. Weissman, “Multiterminal source coding under logarithmic loss,” IEEE Trans. Inf. Theory, 2013, to appear.
[15] P. Cuff, “Distributed channel synthesis,” Aug. 2012. [Online]. Available: http://arxiv.org/abs/1208.4415
[16] A. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, Mar. 1975.
[17] T. S. Han and S. Verdu, “Approximation theory of output statistics,” IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 752–772, May 1993.
[18] M. H. Yassaee, M. R. Aref, and A. Gohari, “Achievability proof via output statistics of random binning,” Mar. 2012. [Online]. Available: http://arxiv.org/abs/1203.0730
[19] A. Winter, “Secret, public and quantum correlation cost of triples of random variables,” in Proc. IEEE International Symposium on Information Theory (ISIT), Sept. 2005, pp. 2270–2274.

[20] I. Csiszár and J. Körner, Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.